

RESEARCH ARTICLE

Open Access



# Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*

John F. Wolters<sup>1</sup>, Kenneth Chiu<sup>2</sup> and Heather L. Fiumera<sup>1\*</sup>

## Abstract

**Background:** Rigorous study of mitochondrial functions and cell biology in the budding yeast, *Saccharomyces cerevisiae* has advanced our understanding of mitochondrial genetics. This yeast is now a powerful model for population genetics, owing to large genetic diversity and highly structured populations among wild isolates. Comparative mitochondrial genomic analyses between yeast species have revealed broad evolutionary changes in genome organization and architecture. A fine-scale view of recent evolutionary changes within *S. cerevisiae* has not been possible due to low numbers of complete mitochondrial sequences.

**Results:** To address challenges of sequencing AT-rich and repetitive mitochondrial DNAs (mtDNAs), we sequenced two divergent *S. cerevisiae* mtDNAs using a single-molecule sequencing platform (PacBio RS). Using *de novo* assemblies, we generated highly accurate complete mtDNA sequences. These mtDNA sequences were compared with 98 additional mtDNA sequences gathered from various published collections. Phylogenies based on mitochondrial coding sequences and intron profiles revealed that intraspecific diversity in mitochondrial genomes generally recapitulated the population structure of nuclear genomes. Analysis of intergenic sequence indicated a recent expansion of mobile elements in certain populations. Additionally, our analyses revealed that certain populations lacked introns previously believed conserved throughout the species, as well as the presence of introns never before reported in *S. cerevisiae*.

**Conclusions:** Our results revealed that the extensive variation in *S. cerevisiae* mtDNAs is often population specific, thus offering a window into the recent evolutionary processes shaping these genomes. In addition, we offer an effective strategy for sequencing these challenging AT-rich mitochondrial genomes for small scale projects.

**Keywords:** *Saccharomyces cerevisiae* mitochondrial genome, mtDNA, Fungal genetics, Mitochondrial genetics, Intron, Mobile elements, Single-molecule sequencing

## Background

*Saccharomyces cerevisiae* has long been at the center of mitochondrial genetics, owing to a facultative anaerobic lifestyle and powerful genetic tools. Most mitochondrial research has focused on a limited number of laboratory strains, allowing for exacting functional studies of mitochondrial processes. Recently, this budding yeast has blossomed into a model for evolutionary biology [1–3]. Genome resequencing projects have revealed genetic diversity and natural population structures of *S. cerevisiae* [4–8]. The diversity in mitochondrial genomes has not been so thoroughly assessed.

Evolution of *S. cerevisiae* mitochondrial DNAs (mtDNAs) differs from nuclear genome evolution in multiple ways. Despite strong purifying selection on mtDNAs, intraspecific mitochondrial variation in *S. cerevisiae* is extensive, owing mainly to differences in intergenic sequences and mobile elements (reviewed in [9] and described below). Replication of mtDNA is not tied to the cell cycle [10], contributing to higher mutation rates in mtDNAs than in nuclear genomes [11]. In yeast, inheritance of mtDNAs is usually biparental [12], although the distribution of parental mitochondrial alleles in progeny is difficult to predict. This is due, in part, to different admixtures of parental mtDNAs in zygotes, mitochondrial recombination, and subsequent loss of heteroplasmy [13]. Additionally, mobile elements in mtDNA may move laterally within populations [14].

\* Correspondence: hfiumera@binghamton.edu

<sup>1</sup>Department of Biological Sciences, Binghamton University, Binghamton, NY, USA

Full list of author information is available at the end of the article

Together, these factors may cause mitochondrial sequences to diverge from nuclear population structure.

The mtDNAs of *S. cerevisiae* contain three subunits of the ATP synthase complex (*atp6*, *atp8* and *atp9*), apocytochrome *b* (*cob*), three subunits of the cytochrome *c* oxidase (*cox1*, *cox2*, and *cox3*), and one ribosomal protein (*rps3/VARI*). The mitochondrial genome also encodes large and small rRNAs (*rnl* and *rns*), an RNA component of the mitochondrial RNase P (*rnpB*) and 24 tRNAs [15]. In *S. cerevisiae*, these genes are separated by long AT-rich intergenic sequences and numerous introns. Repetitive GC-rich regions, known as GC clusters, intersperse the otherwise AT-rich mtDNAs [16, 17]. These clusters typically fall within non-coding sequences, but are also inserted into *rps3/VARI* and ribosomal RNA genes, where they alter the size of the resulting gene products [18–20]. Their palindromic nature likely influences mtDNA structure, which may explain associations with mtDNA instability [21] and mitochondrial recombination [22]. It has been proposed, but never formally tested, that GC cluster-induced structural changes may affect gene regulation [23].

Optional group I and group II introns (differentiated by characteristic RNA secondary structures) also contribute to intraspecific mtDNA variation. Self-encoded homing endonucleases and reverse transcriptases facilitate intron mobility [24] and acquired maturase activities aid in their preservation [25]. In *S. cerevisiae*, mitochondrial introns are found within *cox1* (group I: a13 $\alpha$ , a13 $\gamma$ , a14 $\alpha$ , a14 $\beta$ , a15 $\alpha$ , a15 $\beta$ , a15 $\gamma$ ; group II: a11, a12, and a15 $\gamma$ ), *cob* (group I: b12, b13, b14, b15; group II: b11 $\beta$ ) and *rnl* (group I:  $\Omega$ ) [15]. Additional introns observed in other *Saccharomyces* species include the group I introns a13 $\beta$  in *cox1* and b11 $\alpha$  in *cob* [26]. Incompatibilities between nuclear-encoded splicing factors and non-native introns provide credible support to theories that mitochondrial-nuclear coevolution have contributed to speciation of *Saccharomyces* yeasts through Dobzhansky-Muller-type incompatibilities [26–30]. However, some incompatibilities are strain-specific [26, 27] and highlight the importance of investigating mitochondrial diversity within, in addition to between, species.

The low number of available mtDNA sequences for *S. cerevisiae* yeasts has limited population genetic analyses. The mitochondrial genome of the reference strain was fully sequenced in 1998 [15], and until recently, very few additional mtDNAs were solved [31–33]. The lack of mitochondrial genomes produced by most high-throughput sequencing projects is most likely based on biases against the AT-rich and repetitive DNA during library preparation, sequencing and alignment [34, 35] and discussed in [27], but complete mtDNA sequence reconstruction is possible [32]. A particularly robust resequencing project recently released mtDNA sequences for 93 strains [6], thus providing substantial new resources for mtDNA population genetics. Despite these methodological advances in large-scale

projects, sequencing these AT-rich and complex mtDNAs remains challenging, especially for smaller scale studies.

In this study, we sequenced two mitochondrial genomes using PacBio-RS. This single-molecule sequencing platform was successfully used for both chloroplast and microbial genomes [36, 37], suggesting it may be useful for solving *Saccharomyces* genomes for a small number of strains. We then compared these two newly generated sequences with 98 additional mtDNA sequences to provide a comprehensive picture of intraspecific mtDNA sequence variation in *S. cerevisiae*. Our analyses revealed population-specific genic and intergenic sequence structure including novel intron variation.

## Results

### *De Novo* assembly of *S. cerevisiae* mtDNAs

To assess the feasibility of resolving AT-rich yeast mitochondrial genomes utilizing single-molecule real time sequencing (SMRT), we first generated a complete mitochondrial genome for *S. cerevisiae* strain NCYC3594 [38], a haploid derivative of YJM975 [39]. The nuclear genome of this wine/European isolate has been sequenced numerous times using Illumina sequencing [4, 6, 8, 40]. Only recently was a complete mitochondrial sequence for YJM975 solved [6]. We sequenced an 800 bp library created from a sample enriched for mtDNA. The long read lengths (average = 606 bp) facilitated a *de novo* assembly that produced a single contig with length and GC content consistent with the S288C mitochondrial genome. In addition to the mitochondrial contig, the sequencing reaction produced numerous reads that assembled into shorter (<5 kb) contigs with GC contents more representative of the nuclear genome (>30 %). Following assembly quality improvements, the resulting mtDNA sequence for NCYC3594 was 78,917 bp with a GC content of 16.1 %.

We also sequenced an additional mtDNA from strain NCYC3585 [38], a haploid derivative of 273614N [4]. To reduce nuclear DNA contamination in the sequencing sample, intact mitochondria were treated with DNase prior to isolation of mtDNA. This increased the mtDNA:nDNA ratio in the DNA samples from 0.3:1 (for NCYC3594) to 776:1 (for NCYC3585). We used a longer insert size (6 kbp) in the sequencing reaction to obtain average read lengths of 2055 bp. Following *de novo* assembly and quality refinements, the mtDNA sequence for NCYC3585 was 76,596 bp with a GC content of 15.1 %.

To assess the quality of these assemblies, we aligned these sequences with mitochondrial sequences from the respective parental strains [6]. The differences between the mtDNA sequences for NCYC3594 and YJM975 were limited to 7 small indel polymorphisms, ranging from 1 to 25 bp (summing to 41 bp in total, <0.0006 %

disagreement). The differences between NCYC3585 and YJM1450 (alias 273614N) were 3 indels, including 2 singletons and a 7 bp indel (<0.0002 % disagreement, Additional file 1: Table S1). All indels occurred within AT-rich intergenic regions. It is not known whether these small differences were due to strain specific polymorphisms or sequencing/assembly errors, but overall, the mtDNAs were nearly identical. Thus, single-molecule sequencing approaches generated highly accurate sequences of *Saccharomyces* AT-rich mtDNA.

#### Intraspecific diversity of mitochondrial protein coding sequences

To explore intraspecific mitochondrial evolution in *S. cerevisiae*, we first investigated phylogenetic relationships among protein coding sequences from 99 unique strains. We compared coding sequences from complete mtDNA sequences including the newly obtained sequences presented here, with those from the reference strain S288C [15], industrial isolate NCIM3107 [32], sake strain Kyokai No. 7 [33], clinical isolate YJM789 [31], and 93 additional strains from a recently released dataset [6]. In sum, these strains include those from distinct ancestral populations, as previously described [4] including wine/European (n = 30), North American (n = 2), West African (n = 4), Malaysian (n = 1), and Sake (n = 5) lineages, as well as a large number of strains with admixed genetic backgrounds (n = 57). A complete strain list and accession numbers provided in Additional file 2: Table S2.

A phylogenetic tree was built based on alignment of the concatenated coding sequences of *cox1-atp8-atp6-cob-atp9-rps3-cox2-cox3*, using sequences from *S. paradoxus* [27] as an outgroup (Fig. 1). Based on a total of 457 polymorphic positions across 6762 total aligned base pairs, mitochondrial sequences grouped into three broad clades. Mitochondrial genes from Asian strains (sake and Malaysian) and North American strains formed one large clade. Within this clade, sequences from the North American strains formed a distinct lineage from the Asian strains. A second distinct clade consisted mainly of wine/European strains. Sequences from three West African strains formed a third distinct clade. A single West African isolate grouped nearest to the sake lineage.

The strains with admixed (mosaic) ancestries had mtDNA sequences that mainly grouped within the larger Asian/North American clade, or within the European strains. We subdivided these mosaic strains into mosaic-A (Asian/North American), and mosaic-B (European) populations. Outliers in these groupings included two mosaic strains more closely related to the West African strains, and a single mosaic strain, YJM1399. Mitochondrial sequences from YJM1399 clustered more closely to

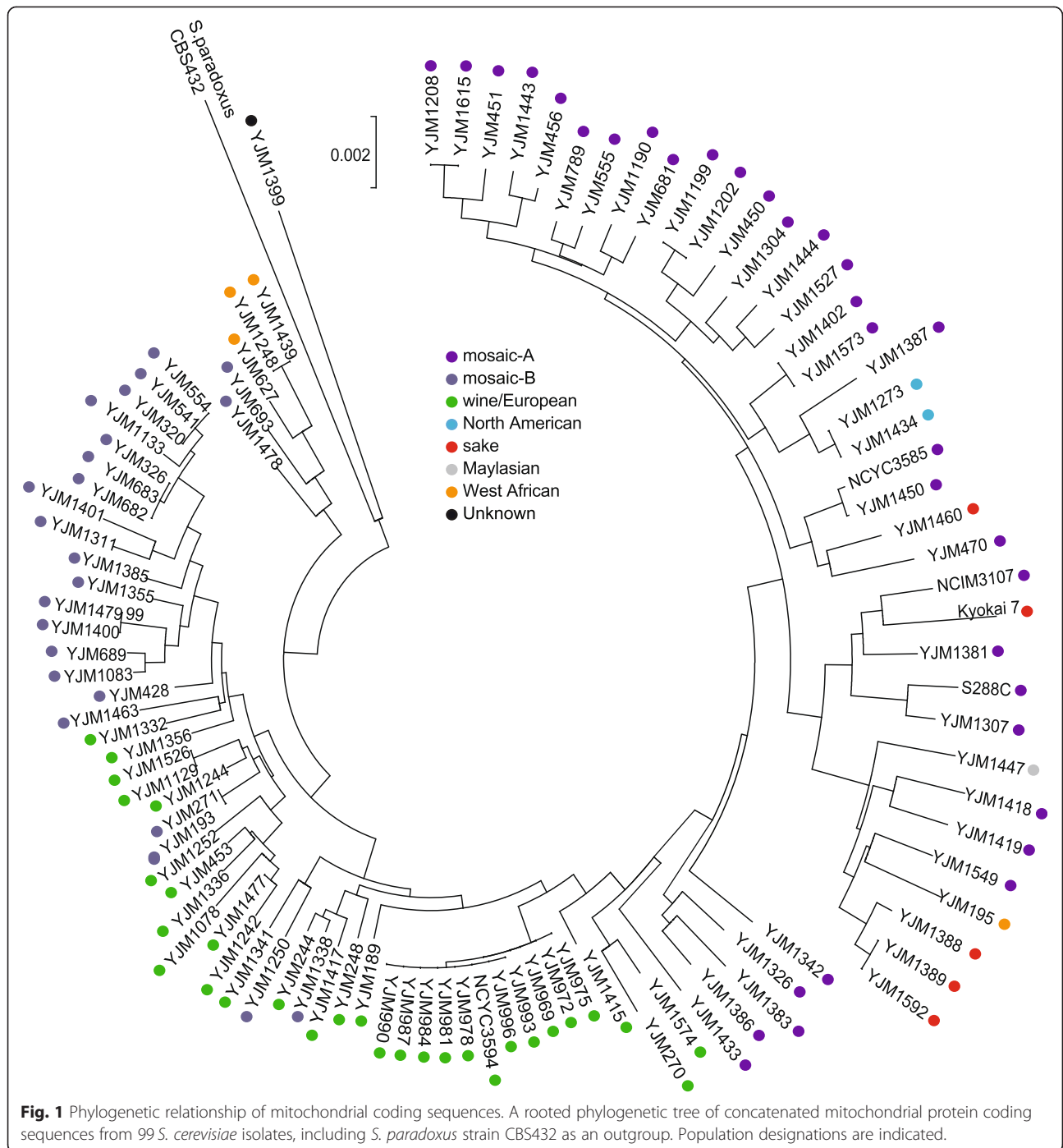
*S. paradoxus* than to other *S. cerevisiae* strains. The nuclear background of this mosaic strain was also significantly diverged (particularly in the number of insertions and deletions) [6], and thus, was treated independently for subsequent analyses. Overall, the phylogenetic analysis of mitochondrial sequences largely recapitulated the population structures obtained by previous analysis of their nuclear genomes; wine/European, West African, sake/Asian, and North American strains were phylogenetically distinct from each other and from most strains with mosaic ancestry.

We assessed polymorphisms within the species in each gene separately (Additional file 3: Table S3). The genes *atp8* and *atp9* each contained no nonsynonymous mutations and only one synonymous site, and the lowest nucleotide diversities of all coding sequences ( $\pi = 0.0033$  and 0.0017, respectively). Nonsynonymous variation was observed in all other genes, with the highest nonsynonymous/synonymous polymorphism ratio in *atp6* (pN/pS = 0.172). Consistent with the known intraspecific size variations in *rps3/VARI* [19], most of the coding sequence polymorphism occurred in this gene (161 polymorphic sites). We obtained similar phylogenetic groupings using alignments of concatenated coding sequences omitting *rps3* or of individual genes (not shown). Despite the different degrees of variation between the coding sequences, the phylogeny was not overly sensitive to variation within any one gene.

#### Divergent strains contain extensive indel variation across their mtDNAs

To assess whether patterns of mitochondrial variation are consistent between divergent populations, nine mtDNAs were chosen to reflect mitochondrial diversity across the species. These included the mtDNAs from NCYC3594 and YJM1078 (European), YJM1273 (North American), YJM1388 (sake), YJM1439 (West African), NCYC3585 and YJM789 (mosaic-A), YJM1401 (mosaic-B), and the reference strain, S288C. We performed a multiple genome alignment followed by extensive manual curation to properly align intron/exon boundaries and correct misalignments of large, repetitive intergenic sequences. A fully annotated and interactive alignment file is provided in Additional file 4: File S1.

Consistent with known size variation, extensive indel polymorphism was observed across the genomes, particularly in intergenic and intronic regions (Fig. 2). These mtDNAs ranged from 76,596 bp (NCYC3585) to 86,214 bp (YJM789), with size differences due to intron content, many small indels (<100 bp), and a small number of large indels that generally corresponded to known variable hypothetical ORFs. Coding sequences, representing less than 8.6 % of mitochondrial genome, were conserved and syntenous. Nucleotide diversity of intergenic

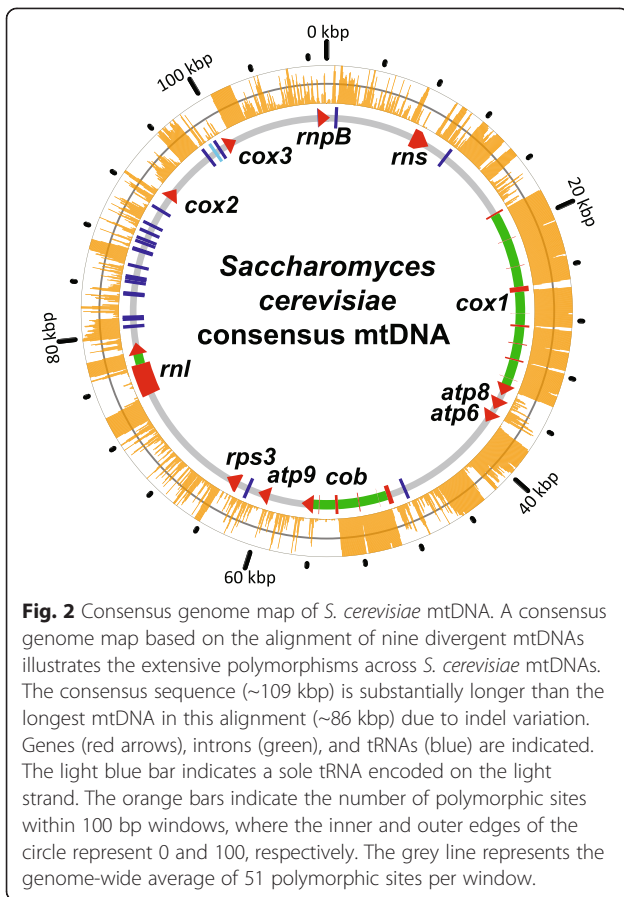


regions ( $\pi = 0.1782$ ) was significantly higher than for exons ( $\pi = 0.0138$ ), mainly due to indel variation. Excluding indels, nucleotide diversity of intergenic regions was significantly reduced ( $\pi = 0.0147$ ) but was still twice that of exons ( $\pi = 0.007$ ). Within the coding sequences, indels occurred in the hypervariable *rps3/VAR1* gene [19], and one instance of an in-frame 3 bp insertion in *cob*.

**Mobile GC clusters exhibit population specific patterns of variation**

Mobile GC-clusters are a known source of indel variation in *Saccharomyces* mtDNAs [21]. These clusters range between 30–80 bp and have been characterized into distinct classes based on consensus sequences [16, 17]. Most GC clusters fall into the M1 and M2 classes (following the classifications defined in [17]) in both *S. cerevisiae* and *S.*



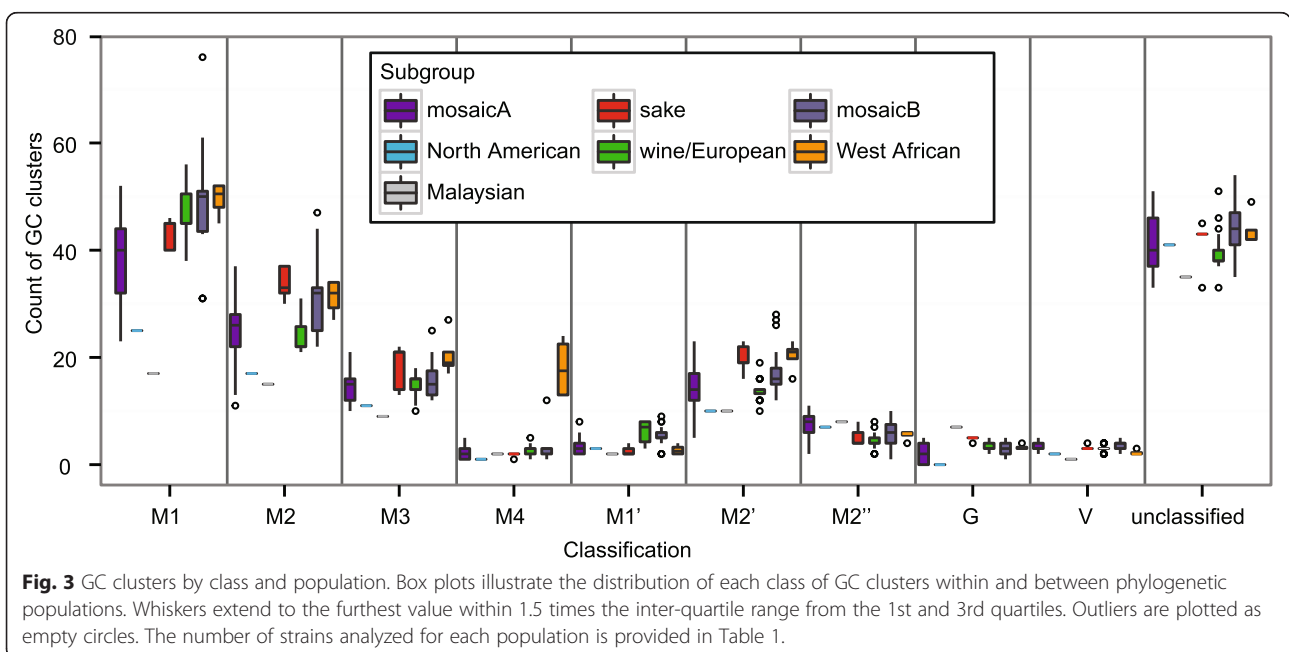


*paradoxus* [21, 27, 41]. Subclasses of M1 and M2 clusters (M1', M2', and M2'') are similar to their parent classes but contain specific insertions or deletions. M3 and M4 classes are found within tandem arrays of GC cluster repeats. The G and V classes are optionally found in *ori* sequences and the *rps3* gene, respectively.

We first determined the number and classes of GC clusters in all 99 complete mitochondrial genomes (Additional file 5: Table S4). On average, each strain had  $120 \pm 22$  classifiable GC-rich motifs. The majority of these GC clusters appeared as single elements, with an average of  $21 \pm 7$  tandem arrays of 2 or more clusters per strain. Consistent with previous descriptions [15–17], the numbers of M1 ( $43 \pm 10$ ) and M2 ( $26 \pm 6$ ) clusters were highest. On average, M3 and M2' clusters were equally represented ( $15 \pm 4$ ) and observed more frequently than the remaining classes (mean < 7). We also scanned each genome for GC-rich regions that did not fit a consensus sequence, identifying an average of  $42 \pm 5$  additional positions per strain. These unclassified GC-rich regions were often associated with tandem arrays of GC clusters and are likely degenerate variants of the main classes.

The variation in GC clusters demonstrated population specific patterns (Fig. 3 and Table 1). While the number of clusters varied between populations for each class (individual ANOVAs,  $P \leq 1.0 \times 10^{-3}$ ), the M4 clusters demonstrated the largest population specific effect ( $P \leq 1.0 \times 10^{-26}$ ). West African strains contained significantly more M4 clusters ( $18 \pm 6$ ) than any other population ( $2 \pm 1$ ).

To determine if GC clusters ever occurred in conserved positions, we examined the nine mtDNA multiple



**Table 1** Mitochondrial GC clusters in *S. cerevisiae*

Population	n	M1	M2	M3	M4	M1'	M2'	M2''	G	V	total	U	T
All strains	98	43.3 ± 9.5	26.4 ± 6.4	15.2 ± 3.5	3.1 ± 3.5	4.7 ± 2.1	15.4 ± 4.0	6.1 ± 2.4	3.0 ± 1.6	3.2 ± 0.8	120.3 ± 22.1	41.7 ± 5.1	20.7 ± 6.6
mosaic-A	33	38.0 ± 7.8	25.2 ± 6.1	14.4 ± 3.0	2.2 ± 1.0	3.5 ± 1.6	14.6 ± 4.0	7.8 ± 2.0	2.4 ± 1.9	3.3 ± 0.7	111.4 ± 20.2	41.5 ± 5.4	19.2 ± 4.4
N. American	2	25.0	17.0	11.0	1.0	3.0	10.0	7.0	0.0	2.0	76.0	41.0	12.0
Malaysian	1	17.0	15.0	9.0	2.0	2.0	10.0	8.0	7.0	1.0	71.0	35.0	12.0
sake	5	43.2 ± 3.0	33.8 ± 3.1	18.2 ± 4.3	1.8 ± 0.5	2.6 ± 0.9	19.8 ± 2.8	5.2 ± 1.8	4.8 ± 0.5	3.2 ± 0.5	132.6 ± 8.1	41.4 ± 4.8	24.6 ± 3.0
wine/Euro	30	47.2 ± 4.5	23.7 ± 2.7	15.3 ± 2.0	2.7 ± 0.9	6.2 ± 1.9	13.9 ± 1.6	4.5 ± 1.3	3.5 ± 0.9	3.0 ± 0.6	119.9 ± 8.3	39.6 ± 3.3	19.0 ± 2.6
mosaic-B	23	48.7 ± 9.3	31.2 ± 6.2	15.8 ± 3.2	2.9 ± 2.1	5.3 ± 1.7	17.5 ± 4.4	6.2 ± 2.5	2.8 ± 1.2	3.5 ± 0.9	133.9 ± 20.8	44.2 ± 4.8	22.9 ± 6.5
West African	4	49.5 ± 3.3	31.3 ± 3.4	20.5 ± 4.4	18.0 ± 5.8	2.75 ± .96	20.3 ± 3.0	5.5 ± 1.0	3.3 ± 0.5	2.3 ± 0.5	153.3 ± 11.8	43.8 ± 3.5	40.0 ± 9.3
<i>S.paradoxus</i> YJM1399	1	14.0	1.0	6.0	0	1.0	0	1.0	0	4.0	27.0	37.0	4.0

The number of GC cluster for each classification is provided for all strains, and for each population. The number of strains within each population is given (n). Values are means ± 1 standard deviation. GC cluster counts were identical for North American strains. U indicates unclassified GC-rich regions. T indicates the number of tandem arrays.

alignment. These mtDNAs contained a total of 1087 classifiable GC clusters that populated 282 unique positions. GC clusters were conserved at only 13 positions. These represented most classes (7 M1, 2 M2, and 1 each in M2', M3, M4, and V classes) and were evenly distributed across the genome.

We more closely examined the apparent expansion of M4 clusters in the West African lineage by examining the West African representative in the multiple alignment. All 22 of the M4 clusters in YJM1439 (West Africa) were found in short tandem arrays containing the M4 cluster and an M1 cluster, including 2 of the 7 conserved M1 clusters. In nearly all identified cases (21/22), the M4 cluster was located upstream of its associated M1 cluster, suggesting that the M4 cluster may target the upstream TAG motif in conjunction with the 5' region of the M1 cluster.

Additional repetitive characters across the genomes were identified in three mtDNAs using a k-mer counting method [42]. Highly repetitive short AT-rich sequences were observed only when using 15-mer scans (Additional file 6: File S2). These consisted of di- to penta-nucleotide repeats consistent with slippage during replication. Highly repetitive GC-rich sequences were only observed at k-mer scans under 50 bp.

**Novel mosaicism in intron patterns**

Fifteen distinct introns have been described in *S. cerevisiae* mitochondrial genomes [15, 31]. We compared the

mitochondrial intron profiles for 104 unique strains, including data from a deep sequencing project that did not provide complete mitochondrial genomes [40] but did provide read coverage for most coding sequences. The intron profiles generated were similar for strains appearing both in partial sequences from Bergstrom et al. [40] and complete sequences from Strobe et al. [6], thus providing a measure of confidence for deducing intron content from the partially sequenced mtDNAs.

Mitochondrial introns are remarkable variable, resulting in mosaic presence/absence patterns between individual strains. Seven of the nine introns in *cox1* are known as optional [15, 31]. In this analysis, the presence of these variable introns ranged from 25-91 % in the 104 mtDNAs analyzed (Table 2). We also observed variation in the *cox1*-ai5β and *cox1*-ai5γ intron. We could not find reports documenting absence of these introns within *S. cerevisiae*. However, inter- and intra-specific variation in other *Saccharomyces* species has been observed [26]; *cox1*-ai5β was prevalent throughout the genus but absent in *S. kudriavzevii* and the *cox1*-ai5γ was variable between *S. paradoxus* isolates. Five introns interrupt the *cob* gene in *S. cerevisiae* and “long” (containing all five introns) and “short” (containing just the last two introns) forms of this gene have been reported [15]. Consistent with these observations, the final two introns (*cob*-bi4 and -bi5) were invariably present in all *S. cerevisiae* strains. We also found strains lacking just *cob*-bi1β or both

**Table 2** Intron content within and between *Saccharomyces* mtDNAs

Intron	<i>cox1</i>									<i>cob</i>									<i>rnl</i>
	ai1	ai2	ai3α	ai3β	ai3γ	ai4α	ai4β	ai4γ	ai5α	ai5β	ai5γ	bi1α	bi1β	bi2	bi3	bi4	bi5	Ω	
	I1	I2	I3α	I3β	I3γ	I4α	I4β		I5α	I5β	I5α	I1α	I1β	I2	I3	I4	I5		
Group	(n)	II	II	I	I**	I	I	I	I**	I	I*	II*	I**	II	I	I	I*	I*	I
all <i>cerevisiae</i>	103	59	51	91	0	40	59	38	0	25	80	74	5	86	89	96	100	100	39
Mosaic-A	32	66	47	97	0	34	81	59	0	38	94	72	0	81	81	91	100	100	63
N. American	3	0	100	0	0	0	0	0	0	100	100	100	0	100	100	100	100	100	0
Malaysian	1	-	+	-	-	+	-	+	-	-	+	-	-	-	-	+	+	+	-
sake	5	100	80	100	0	40	40	0	0	20	100	100	0	20	20	80	100	100	0
wine/Euro	34	38	59	94	0	41	21	12	0	9	76	84	0	100	100	100	100	100	9
Mosaic-B	23	74	43	96	0	52	91	65	0	13	74	70	22	100	100	100	100	100	74
West African	5	100	20	80	0	20	100	0	0	80	0	0	0	40	100	100	100	100	0
Unknown	1	+	-	+	+	-	-	-	-	+	+	-	-	-	-	-	-	-	-
<i>paradoxus</i>	3	0	33	0	100	67	0	0	0	67	100	33	67	67	67	67	0	0	?
<i>mikatae</i>	1	+	-	+	+	-	+	-	-	+	+	+	+	+	+	+	+	+	?
<i>kudriavzevii</i>	1	-	-	+	+	-	+	-	-	+	-	+	+	+	+	+	-	-	?
<i>cariocanus</i>	1	-	-	-	+	-	-	-	-	-	+	+	+	+	+	-	-	-	?
<i>bayanus</i>	1	-	-	-	+	+	-	-	+	+	+	+	-	+	+	+	-	+	?
Insertion site		169	205	243	386	709	720	900	927	972	1107	1132	393	415	429	506	756	807	2716

The percentage of strains containing specific introns is provided for *S. cerevisiae* strains and related *Saccharomyces* species. The number of strains within each population is provided (n). When n = 1, intron presence/absence is indicated as +/- . Introns not known as variable (\*) or observed (\*\*) in *S. cerevisiae* are indicated next to their respective group heading. Insertion sites indicate the base pair in the CDS that precedes the intron insertion site.

*cob*-bi1 $\beta$  and *cob*-bi2 introns, indicating that additional intron mosaics exist in natural populations.

Though mitochondrial intron content is known to vary in a strain-dependent manner, comparative studies have revealed that the occurrence of certain introns follows species divisions [26, 43]. The *cox1*-ai3 $\beta$  intron has never been found in *S. cerevisiae* species, and our findings were true to this classification. Similarly, we did not find evidence of a rare *cox1* intron (ai4 $\gamma$ ) that has only been observed in *S. bayanus*. We did, however, find evidence of the *cob*-bi1 $\alpha$  intron in a small number of strains (5 %). To our knowledge, this intron has not previously been reported in *S. cerevisiae* but has been observed in much of the *Saccharomyces* genus [26]. Interestingly, strains with this intron also contained the five common *S. cerevisiae* introns, and showed no evidence of introgression of *cob* exon sequences.

One exception to the species-specific intron structure was the mosaic strain, YJM1399. This strain contained the *cox1*-ai3 $\beta$  intron (inserted at the same location as in other species), and lacked the otherwise conserved *cob*-bi4 and -bi5 introns. The sequence of the *cox1*-ai3 $\beta$  intron most closely matched a sequence from the distantly related yeast *Lachancea meyersii* (Genbank: HE983614.1, 53 % coverage, 92 % identity) and shared little, if any, homology with the *cox1*-ai3 $\beta$  intron from *S. paradoxus*. Other portions of the mitochondrial genome were highly homologous to those of other *S. cerevisiae* strains.

To determine if intron profiles were related to population structure, we compared intron content between each population (Table 2). The highly prevalent *cob*-bi1 $\beta$  and *cob*-bi2 introns were nearly absent in the sake population and Malaysian strain. No West African strain contained the *cox1*-ai5 $\beta$  or -ai5 $\gamma$  intron (previously thought to be conserved). The nearly identical mtDNA sequences of the three North American isolates lacked the frequently observed *cox1*-ai3 $\alpha$  intron. A phylogenetic tree based on character states of intron presence/absence (Additional file 7: Figure S3) was highly concordant with the phylogeny built from coding sequences (Figure 1); West African, sake/Malaysian, and North American strains were organized into distinct groups (with one exception within the sake strains). Intron patterns of wine/European and mosaic strains were more variable. Individual intron profiles for each strain are reported in Additional file 8: Table S5.

## Discussion and conclusions

Comparisons of mtDNAs between *Saccharomyces* species and other *Hemiascomycetes* yeasts have revealed broad evolutionary changes in mtDNA evolution, particularly in regards to genome organization [12, 44, 45]. Few population genetic investigations on intraspecific mtDNAs in yeasts exist [45–47], and none for *Saccharomyces* species.

To provide a window onto recent evolutionary changes in the mtDNA, we compared the intraspecific genetic variation in mitochondrial genomes from 100 strains of *S. cerevisiae*.

Through phylogenetic analysis of coding sequences, we found that these mtDNAs were organized into three broad clades that shared remarkable resemblance to clades constructed using the thousands of SNPs across nuclear sequences [4, 6], albeit at lower resolution. Populations according to nuclear divergences organize strains into 5 non-mosaic populations (wine/European, West African, sake, Malaysian, and North American). From mitochondrial coding sequences alone, discrete populations for West African and wine/European ancestry were observed. Resolution of North American and Asian mtDNAs clades were more obscured, although each maintained distinct lineages within a broader clade. Strains with known mosaic ancestry grouped predominantly as two subclades peripheral to the wine/European or Asian/North American clades. Many of these mosaics are strains domesticated for human activities (or perhaps the result of admixing between wild and domesticated strains [48]). High prevalence of mitochondrial mosaic strains were intermixed among the wine/European and Asian strains suggests that mitochondrial genomes, like their nuclear counterparts, also contain mixed ancestry. This implies that mitochondrial recombination occurs frequently during admixture.

Mobile elements, including introns, also had population specific profiles. Intron pattern were not fixed within each population (wine/Euro, West African, etc.), but substantial trends between populations emerged. One example is in the optional *cox1*-ai1 intron, which was omnipresent in all sake and West African strains but only found in 38 % of wine/European strains and none of the North American isolates. Newly discovered variation in introns previously thought fixed in *S. cerevisiae* was also related to population structure; the *cob*-bi1 $\beta$  intron was only missing in sake and West African strains while being highly prevalent in all other populations. Similarly, the *cob*-bi1 $\alpha$  intron, never previously reported within *S. cerevisiae* mtDNAs, was observed only in a small number of mosaic strains whose mtDNAs were most closely related to those of the wine/European strains (our “mosaic-B” group). While the ancestry of these mosaic strains are a subject of speculation, intron flanking sequences in their *cob* exons were more similar to the *cob* genes in *S. cerevisiae* than *S. paradoxus*, suggesting that this intron is not the result of introgression of a non-native *cob* allele. Intron analysis also revealed an interesting mitochondrial ancestry for mosaic strain, YJM1399. This strain contained an intron at the (non-*cerevisiae*) *cox1*-ai3 $\beta$  insertion site that most closely resembled a sequence from *Lachancea*, and contained an



intron-less *cob* despite complete conservation of several *cob* introns throughout the rest of the *S. cerevisiae* strains. Several fixed substitutions in the coding sequences of this mosaic relative to *S. cerevisiae* and *S. paradoxus* suggested this was a non-*cerevisiae* mtDNA. Other regions of the mtDNA, however, were homologous to *S. cerevisiae* mtDNA. This mosaic strain likely provides an example of mitochondrial introgression and not replacement.

Genetic diversity in *S. cerevisiae* extends past what has been measured here [7], and it is likely that as larger genetic space is sampled, additional mitochondrial mosaics and intron variants will be revealed. Analysis of allelic variation within intron sequences may provide deeper insight into mitochondrial evolution, however, we found that a simple binary presence/absence analysis was sufficient to reconstruct the populations described here. Insertion mechanisms that occur during intron homing are mutagenic to residues in flanking sequences [49, 50]. Thus, the phylogenies created by coding sequences are likely influenced by population-specific intron profiles and the accompanying co-conversion of exon sequences.

Patterns of GC clusters also demonstrated population structure. The total numbers of clusters ranged from an average of ~76 in North American strains to ~153 clusters in West African strains. Each population had significantly different patterns in the numbers and types of GC clusters, however the West African strains appear to have undergone a recent expansion of the rare M4 cluster. While M4 clusters can be at the first position of a tandem GC cluster array [17], the M4 clusters in the West African strains were almost exclusively found following an M1 cluster. M2 clusters generally paired with M3 clusters in tandem arrays.

Phenotypic variation in wild yeasts is believed to follow population history [51]. The population specific genetic structure in mtDNAs may play a role in phenetic groupings. In *S. cerevisiae*, strains with a common nuclear background but harboring different mtDNAs had slightly different growth rates [52], demonstrating that naturally variation in mtDNAs can affect phenotype. It is easy to imagine how allelic differences in oxidative phosphorylation genes (and corresponding gene networks) could contribute to different efficiencies in mitochondrial respiration. Intergenic mtDNA sequences may also contribute to phenotypic differences, directly or indirectly. Mobile GC clusters affecting recombination [22] may influence mtDNA replication and inheritance or interfere with mitochondrial translation (as in the case of translational bypass elements that are believed to have evolved from GC clusters [53]). GC clusters are also correlated with petite frequency in *Saccharomyces* yeasts [21]. Possibly, the elevated petite frequency observed in the West African-related laboratory strain, SK1 [54], is due to an elevated number of GC clusters

predicted in this phylogenetically distinct group. While *S. cerevisiae* introns are not essential for mitochondrial respiration [55], they may offer an underappreciated regulatory role [56–58]. Absence of introns in certain populations may relax selection on the nuclear encoded splicing factors, thus contributing to the creation of cytonuclear incompatibilities that play a role in post-zygotic speciation in these lower eukaryotes [26–30].

The whole-genome duplication preceding the evolution of the *Saccharomyces* genus is thought to have relaxed selection on mitochondrial functions, as evidenced by increased nonsynonymous mutations and relaxed codon bias in mitochondrially-targeted nuclear genes involved in respiration [59]. We observed more frequent nonsynonymous polymorphisms in *atp6*, consistent with relaxed purifying selection on this gene. The nonsynonymous to synonymous ratio of intraspecific polymorphisms for *atp6* (pN/pS = 0.172) was actually higher than that for the *rps3/VARI* excluding indel variations (pN/pS = 0.142), a mitochondrial gene known for intraspecific diversity [19]. Interestingly, this mirrors intraspecific polymorphisms observed for the *atp6* gene in *L. kluyveri* [47], a yeast that evolved before the whole genome duplication. This gene has also been implicated in tests of positive selection in other organisms [60–62]. There were almost no nonsynonymous substitutions between *S. cerevisiae* and *S. paradoxus*, precluding formal tests of selection. Even if a distinct evolutionary pattern for *atp6* exists, it is likely that these closely related species are under very similar selection patterns.

In *Saccharomyces*, mitochondrial population genetics studies have been limited by low numbers of available mtDNA sequences. High coverage sequencing projects routinely produce only partial mitochondrial sequences [8, 40]. The recovery of complete mtDNAs proves that robust methodologies can be used to reconstruct these challenging genomes from Illumina-based datasets [6, 47]. Here, we have shown that high quality mtDNA sequences can be produced from single molecule sequencing data. While Illumina sequencing proves to be useful for large-scale mitochondrial genome sequencing, the PacBio RS platform offers a cost-efficient method when only a small number of mitochondrial genomes are required.

## Methods

### Isolation of mtDNA

Prior to DNA isolation, crudely purified mitochondria were prepared as previously described [63]. Strain NCYC3594 (a haploid derivative of YJM975 [38]) was grown overnight at 30 °C in 1.5 L YPD media (1 % yeast extract, 2 % peptone, 2 % dextrose). Late-log phase cells were harvested, washed, and incubated for 10 min at 30 °C in approximately 30 ml of 100 mM Tris-SO<sub>4</sub> pH 9.4 containing 10 mM DTT. Cells were resuspended

in 1.2 M sorbitol, 20 mM  $\text{KH}_2\text{PO}_4$  pH 7.4 (7 ml/g wet weight cells) containing 0.5 mg/ml Zymolase 20T (Nacalai Tesque Inc.) and incubated while rocking at 30 °C until spheroplasts occurred (~45 min., determined by optical clearing of 50  $\mu\text{l}$  cell suspension added to 0.5 ml  $\text{H}_2\text{O}$ ), followed by physical shearing of cells using a 40 ml tissue grinder. Mitochondria were separated from unbroken cells and cell debris through alternating rounds of centrifugation of 5 min at 5000 rpm and 12 min 12000 rpm in a Sorvall F21S-8x50Y rotor.

Mitochondrial fractions for strain NCYC3585 (a haploid derivative of 273614N [38]) were collected as described above, except that cells were grown in YPEG media (1 % yeast extract, 2 % peptone, 3 % ethanol, 3 % glycerol). These mitochondrial enrichments were also subjected to a DNase treatment [64] by incubating in 1 ml 0.3 M sucrose, 5 mM  $\text{MgCl}_2$ , 50 mM Tris-HCl pH 8.0, 10 mM  $\text{CaCl}_2$  containing 100 units of DNase (New England BioLabs) for 30 min. at 37 °C. DNases were inactivated by the addition of 0.5 M EDTA (pH 8.0) to a final concentration of 0.2 M. The mitochondria were washed to remove DNases through 3 repeated cycles of centrifugation (15000 rpm at 4 °C, 10 min) and resuspended in sucrose buffer.

The resulting mitochondria-enriched cell fractions were lysed in ~500  $\mu\text{l}$  of 1 % Sarkosyl, 100 mM NaCl, 10 mM EDTA pH 8.0, Tris pH 8.0, and incubated at room temperature until optical clearing occurred (~30 min). DNA was isolated using phenol-chloroform extraction and ethanol precipitation. Purity of mtDNA (O.D. 260/280 = 1.8-2.0) was determined using a NanoDrop-1000 spectrophotometer.

#### Real Time PCR

Real-time PCR was used to determine the relative abundance of nuclear and mitochondrial DNA. Primer sequences for *ACT1* (5' GTATGTGTAAAGCCGGTTTTG and 5' CATGATACCTTGGTGTCTTGG) and *cox1* (5' C TACAGATACAGCATTTCCAAGA and 5' GTGCCTGATAGATGATAATGGT) were obtained from Taylor, et al. [65]. Each gene was amplified from dilutions of the purified mtDNAs using SYBR Green master mix (ABi Research) on an Abi 7300 Real Time PCR System. The ratio of mtDNA to nuclear DNA was determined as the logarithm of the difference in the cycle threshold values ( $\log_2 \Delta\text{CT}$ ), after correcting for genome size.

#### Sequencing

Approximately 2  $\mu\text{g}$  of purified mtDNAs were sequenced using the PacBio RS at the Yale Center for Genomic Analysis following circular consensus sequencing protocols. A library created from ~800 bp fragments of the mtDNA from NCYC3594 was sequenced using C1 chemistry,

producing 38,681 reads. A library created from ~6 kbp fragments of fractionated mtDNA from NCYC3585 was sequenced using C2 chemistry, producing 7,360 reads. A single SMRT Cell was used for each library. Reads less than 50 bp were removed.

#### Mitochondrial genome assembly and annotation

Circular consensus reads were assembled using MIRA v3.4.1 (NCYC3594) and MIRA 3.9.17 (NCYC3585) with parameters: denovo, genome, accurate [66]. The assembly for NCYC3585 produced two major contigs that overlapped (identified using MUMmer 3.0 [67]) by ~3000 bp and were manually joined to form a single scaffold. Consensus sequences and assembly qualities of the final scaffolds were improved by employing the Quiver consensus algorithm to map reads back to contigs and correct sequencing errors [37].

The average read quality for NCYC3594 was Q16.24, which improved to an average assembly quality of Q17.85 after Quiver. The average quality for the 273614N reads was Q15.95, which improved to Q45.83 after Quiver. Hand curation of each genome revealed a large duplicated sequence in an AT-rich intergenic region of NCYC3594 and was manually removed. Sequences were annotated using MFANNOT [24]. The annotations were verified with BLAST searches of features in the reference sequence [15]. Annotations included only tRNA sequences that were triply identified by MFANNOT, BLAST, and tRNAscan-SE [68]. Annotations were manually curated to ensure correct intron/exon boundaries.

The mitochondrial genome from *S. cerevisiae* is organized as a collection of linear concatemers that map to a circular genome [69, 70]. Consistent with circularity, we found reads that aligned to both ends of each linear scaffold ( $\geq 360$  bp on both ends with  $> 80$  % identity). The linear scaffolds were reorganized to match the S288C reference genome start position.

#### Alignments and phylogenetics

Mitochondrial protein coding sequences were extracted from 99 strains. Strain names and accession numbers are provided in Additional file 2: Table S2. A 6765 bp alignment of the concatenated CDS was generated utilizing Clustal Omega [71]. A neighbor-joining [72] tree was constructed utilizing MEGA6 [73]. The *S. paradoxus* strain CBS432 was used as an outgroup. The proportion of synonymous to non-synonymous polymorphisms within *S. cerevisiae* (pN/pS) was calculated using PAML [74, 75].

An initial multiple alignment was constructed using the LAGAN [76] algorithm in the mVISTA suite of programs [77] using the mtDNA sequences from S288C, YJM789, NCYC3594, and NCYC3585. Sequences from YJM1078, YJM1273, YJM1388, YJM1401, and YJM1439

were added to the multiple alignment utilizing MAFFT [78]. Alignments were manually curated to fix exon/intron boundaries and indels. Nucleotide diversity was calculated to allow for multiple minor alleles using the per site summation method. Genome polymorphism was assessed by counting individual polymorphic sites in sliding windows of 100 bp in 50 bp steps across the multiple alignment.

#### Additional analyses

GC clusters classifications were defined according to Weiller et al. [17], using consensus sequences (Additional file 9; Table S6). The GC clusters were identified using BLAST. GC-rich areas (>60 % GC) not matching consensus sequences were identified through 30 bp sliding window scans. Tandem GC cluster assays were defined as GC-rich regions that overlapped with at least two classified clusters. One-way ANOVAs were performed within each classification (count ~ population) and significance determined following a Bonferonni correction. Conserved GC clusters were identified as sites in the multiple alignment with overlapping GC-rich regions, followed by manual verification. Genome wide scans for repetitive elements in the mtDNAs from S288C, NCYC3594 and YJM789 were performed utilizing Jellyfish [79] with k-mer values of 15, 30, 50, and 100.

Intronic sequences were identified using BLAST with query sequences from S288C (*cox1*-a1 $\alpha$ , a12, a13 $\alpha$ , a15 $\alpha$ , a15 $\beta$ , a15 $\gamma$ , *rnl*-11), YJM789 (*cox1*-a13 $\gamma$ , a14 $\beta$ ) and *S. paradoxus* CBS432 (*cox1*-a13 $\beta$ ). Introns were classified as 0 or 1 based on clear presence of a homologous intron at that specific site and agreement with prior annotations (with the exception of a13 $\beta$  in YJM1399, which shares the insertion site but no homology with *S. paradoxus*). Additional intron information was obtained from unassembled contigs from <http://www.moseslab.csb.utoronto.ca/sgrp/> [40], based on identification of coding sequences in contigs and the presence of intervening sequence between exons. A distance matrix was calculated utilizing presence of an intron at the specific insertion sites as binary character trait, and a neighbor-joining tree was constructed with the ape package in R [80]. The strain NCIM3107 was omitted due to the presence of intervening sequences between exons atypical of the pattern observed in all other strains.

#### Availability of supporting data

The annotated mtDNA sequences are available as Genbank accession numbers KR260476 (strain NCYC3594) and KR260477 (NCYC3585). Phylogenetic data has been deposited at TreeBase (<http://purl.org/phylo/treebase/phyloids/study/TB2:S17639>). All scripts are available upon request.

#### Additional files

**Additional file 1: Table S1.** Concordance of sequence data. Polymorphic sites between PacBio and the Illumina-generated datasets from [6].

**Additional file 2: Table S2.** Strain list. Strain names and accession numbers are provided.

**Additional file 3: Table S3.** Nucleotide diversities. Polymorphic assessment, including nucleotide diversity ( $\pi$ ), for each mitochondrial gene in *S. cerevisiae* (excluding YJM1399). Percent difference was calculated as ( $\pi \times 100$ ). pN/pS indicates the ratio of nonsynonymous to synonymous polymorphism.

**Additional file 4: File S1.** An interactive multiple alignment of the nine complete mtDNAs from *S. cerevisiae*. To move to a particular feature, select the feature name from the searchable list on the left. Nucleotide positions for each strain represent the ungapped genomes, while positions in the consensus sequence represent the position in the gapped multiple alignment.

**Additional file 5: Table S4.** GC clusters. For each strain, the total counts of GC clusters in all classes are provided.

**Additional file 6: File S2.** Repetitive elements. A k-mer analysis of repetitive sequences in strains S2883 and NCYC3594. Each square represents a single, potentially overlapping, k-mer.

**Additional file 7: Figure S3.** Intron phylogenetics. A phylogenetic tree built from intron profiling, treated as binary character states. Any strain with a questionable intron profile was excluded.

**Additional file 8: Table S5.** Mitochondrial introns. Individual intron profiles for each strain are provided, where 1 indicates the presence of the intron and a blank cell, the absence.

**Additional file 9: Table S6.** GC cluster consensus sequences. Slightly modified from [17] to remove 5' TAG motif.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

This study was conceived and designed by JFW and HLF. JFW prepared DNA samples for sequencing, performed genome assemblies and conducted computational analyses. KC participated in the interpretation of raw sequencing reads and *de novo* assemblies. JFW and HLF wrote the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

We thank Anthony C. Fiumera for use of a real time thermocycler, and Gary D. Wilber and Joseph K. Palazzolo for assistance with certain scripts. This work was supported by National Institutes of Health award GM10132 (to HLF and KC).

#### Author details

<sup>1</sup>Department of Biological Sciences, Binghamton University, Binghamton, NY, USA. <sup>2</sup>Computer Science Department, Binghamton University, Binghamton, NY, USA.

Received: 13 December 2014 Accepted: 29 May 2015

Published online: 11 June 2015

#### References

- Borneman AR, Pretorius IS. Genomic insights into the *Saccharomyces sensu stricto* complex. *Genetics*. 2015;199(2):281–91.
- Zarin T, Moses AM. Insights into molecular evolution from yeast genomics. *Yeast*. 2014;31(7):233–41.
- Hittinger CT. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet*. 2013;29(5):309–17.
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. *Nature*. 2009;458(7236):337–41.
- Cromie GA, Hyma KE, Ludlow CL, Garmendia-Torres C, Gilbert TL, May P, et al. Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda)*. 2013;3(12):2163–71.

6. Strobe PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, Dietrich FS, McCusker JH. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res* 2015;25(5):762–74.
7. Wang QM, Liu WQ, Liti G, Wang SA, Bai FY. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol Ecol*. 2012;21(22):5404–17.
8. Skelly DA, Merrihew GE, Riffle M, Connelly CF, Kerr EO, Johansson M, Jaschob D, Graczyk B, Shulman NJ, Wakefield J et al. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* 2013;23(9):1496–504.
9. Groth C, Petersen RF, Piskur J. Diversity in organization and the origin of gene orders in the mitochondrial DNA molecules of the genus *Saccharomyces*. *Mol Biol Evol*. 2000;17(12):1833–41.
10. Lecrenier N, Foury F. New features of mitochondrial DNA replication system in yeast and man. *Gene*. 2000;246(1–2):37–48.
11. Foury F, Hu J, Vanderstraeten S. Mitochondrial DNA mutators. *Cell Mol Life Sci*. 2004;61(22):2799–811.
12. Solieri L. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends Microbiol*. 2010;18(11):521–30.
13. Westermann B. Mitochondrial inheritance in yeast. *Biochim Biophys Acta*. 2014;1837(7):1039–46.
14. Goddard MR, Burt A. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A*. 1999;96(24):13880–5.
15. Foury F, Roganti T, Lecrenier N, Purnelle B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett*. 1998;440(3):325–31.
16. de Zamaroczy M, Bernardi G. The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. *Gene*. 1986;41(1):1–22.
17. Weiller G, Schueller CM, Schweyen RJ. Putative target sites for mobile G + C rich clusters in yeast mitochondrial DNA: single elements and tandem arrays. *Molecular General Genet*. 1989;218(2):272–83.
18. Butow RA, Perlman PS, Grossman LI. The unusual var1 gene of yeast mitochondrial DNA. *Science*. 1985;228(4707):1496–501.
19. Hudspeth ME, Vincent RD, Perlman PS, Shumard DS, Treisman LO, Grossman LI. Expandable var1 gene of yeast mitochondrial DNA: in-frame insertions can explain the strain-specific protein size polymorphisms. *Proc Natl Acad Sci U S A*. 1984;81(10):3148–52.
20. Dujon B. Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell*. 1980;20(1):185–97.
21. Spirek M, Soltesova A, Horvath A, Slavikova E, Sulo P. GC clusters and the stability of mitochondrial genomes of *Saccharomyces cerevisiae* and related yeasts. *Folia Microbiol (Praha)*. 2002;47(3):263–70.
22. Dieckmann CL, Gandy B. Preferential recombination between GC clusters in yeast mitochondrial DNA. *EMBO J*. 1987;6(13):4197–203.
23. Hausner G. Fungal Mitochondrial Genomes, Plasmids and Introns. In: Arora DK, Khachatourians GG, editors. *Applied Mycology and Biotechnology: Fungal Genomics*, vol. III. New York: Elsevier Science; 2003. p. 101–31.
24. Lang BF, Laforest MJ, Burger G. Mitochondrial introns: a critical view. *Trends Genet*. 2007;23(3):119–25.
25. Belfort M. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev*. 2003;17(23):2860–3.
26. Spirek M, Polakova S, Jatzova K, Sulo P. Post-zygotic sterility and cytonuclear compatibility limits in *S. cerevisiae* xenomitochondrial cybrids. *Front Genet*. 2014;5:454.
27. Prochazka E, Franko F, Polakova S, Sulo P. A complete sequence of *Saccharomyces paradoxus* mitochondrial genome that restores the respiration in *S. cerevisiae*. *FEMS Yeast Res*. 2012;12(7):819–30.
28. Sulo P, Spirek M, Soltesova A, Marinoni G, Piskur J. The efficiency of functional mitochondrial replacement in *Saccharomyces* species has directional character. *FEMS Yeast Res*. 2003;4(1):97–104.
29. Chou JY, Hung YS, Lin KH, Lee HY, Leu JY. Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol*. 2010;8(7):e1000432.
30. Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*. 2008;135(6):1065–73.
31. Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, Cao Z, et al. Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc Natl Acad Sci U S A*. 2007;104(31):12825–30.
32. Ulaganathan K, Sravanthi Goud B, Reddy MM, Praveen Kumar V, Radhakrishna S, Balsingh J: Genome Sequence of *Saccharomyces cerevisiae* NCIM3107, Used in Bioethanol Production. *Genome Announce* 2015, 3(1):e01557-14. doi:10.1128/genomeA.01557-14.
33. Akao T, Yashiro I, Hosoyama A, Kitagaki H, Horikawa H, Watanabe D, et al. Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. *DNA Res*. 2011;18(6):423–34.
34. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. *Genome Biol*. 2013;14(5):R51.
35. Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One*. 2013;8(4):e62856.
36. Ferrarini M, Moretto M, Ward JA, Surbanovski N, Stevanovic V, Giongo L, et al. An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics*. 2013;14:670.
37. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10(6):563–9.
38. Cubillos FA, Louis EJ, Liti G. Generation of a large set of genetically tractable haploid and diploid *Saccharomyces* strains. *FEMS Yeast Res*. 2009;9(8):1217–25.
39. McCullough MJ, Clemons KV, Farina C, McCusker JH, Stevens DA. Epidemiological investigation of vaginal *Saccharomyces cerevisiae* isolates by a genotypic method. *J Clin Microbiol*. 1998;36(2):557–62.
40. Bergstrom A, Simpson JT, Salinas F, Barre B, Parts L, Zia A, et al. A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol*. 2014;31(4):872–88.
41. Piskur J, Smole S, Groth C, Petersen RF, Pedersen MB. Structure and genetic stability of mitochondrial genomes vary among yeasts of the genus *Saccharomyces*. *Int J Syst Bacteriol*. 1998;48:1015–24.
42. Kurtz S, Narechania A, Stein JC, Ware D. A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. *BMC Genomics*. 2008;9:517.
43. Chou JY, Leu JY. Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays*. 2010;32(5):401–11.
44. Gaillardin C, Neuvéglise C, Kerscher S, Nicaud JM. Mitochondrial genomes of yeasts of the *Yarrowia* clade. *FEMS Yeast Res*. 2012;12(3):317–31.
45. Friedrich A, Jung PP, Hou J, Neuvéglise C, Schacherer J. Comparative mitochondrial genomics within and among yeast species of the *Lachancea* genus. *PLoS One*. 2012;7(10):e47834.
46. Freel KC, Friedrich A, Hou J, Schacherer J. Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species *Lachancea thermotolerans*. *Genome Biol Evolution*. 2014;6(10):2586–94.
47. Jung PP, Friedrich A, Reisser C, Hou J, Schacherer J. Mitochondrial genome evolution in a single protoplast yeast species. *G3 (Bethesda)*. 2012;2(9):1103–11.
48. Hyma KE, Fay JC: Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. *Mol Ecol* 2013;22(11):2917–30.
49. Sanchez-Puerta MV, Abbona CC, Zhuo S, Tepe EJ, Bohs L, Olmstead RG, et al. Multiple recent horizontal transfers of the *cox1* intron in *Solanaceae* and extended co-conversion of flanking exons. *BMC Evol Biol*. 2011;11:277.
50. Mueller JE, Smith D, Belfort M. Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes Dev*. 1996;10(17):2158–66.
51. Warringer J, Zorgo E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait variation in yeast is defined by population history. *PLoS Genet*. 2011;7(6):e1002111.
52. Paliwal S, Fiumera AC, Fiumera HL. Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics*. 2014;198(3):1251–65.
53. Nosek J, Tomaska L, Burger G, Lang BF: Programmed translational bypassing elements in mitochondria: structure, mobility, and evolutionary origin. *Trends Genet* 2015;16(4):5926–31.
54. Dimitrov LN, Brem RB, Kruglyak L, Gottschling DE. Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains. *Genetics*. 2009;183(1):365–83.
55. Seraphin B, Boulet A, Simon M, Faye G. Construction of a yeast strain devoid of mitochondrial introns and its use to screen nuclear genes involved in mitochondrial splicing. *Proc Natl Acad Sci U S A*. 1987;84(19):6810–4.



56. Grand RS, Martienssen R, O'Sullivan JM. Potential roles for interactions between the mitochondrial and nuclear DNA throughout the cell cycle of *Schizosaccharomyces pombe*. *Mitochondrion*. 2014;17:141–9.
57. Turk EM, Das V, Seibert RD, Andrulis ED. The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. *PLoS One*. 2013;8(10):e78105.
58. Kaniak-Golik A, Skoneczna A. Mitochondria-nucleus network for genome stability. *Free radical biology & medicine* 2015;82:73–104.
59. Jiang H, Guan W, Pinney D, Wang W, Gu Z. Relaxation of yeast mitochondrial functions after whole-genome duplication. *Genome Res*. 2008;18(9):1466–71.
60. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*. 2004;303(5655):223–6.
61. Gagnaire PA, Normandeau E, Bernatchez L. Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Mol Biol Evol*. 2012;29(10):2909–19.
62. Bayona-Bafaluy MP, Muller S, Moraes CT. Fast adaptive coevolution of nuclear and mitochondrial subunits of ATP synthetase in orangutan. *Mol Biol Evol*. 2005;22(3):716–24.
63. Glick BS. Pathways and energetics of mitochondrial protein import in *Saccharomyces cerevisiae*. *Methods Enzymol*. 1995;260:224–31.
64. Storti RV, Sinclair JH. Sequence homology between mitochondrial DNA and nuclear DNA in the yeast, *Saccharomyces cerevisiae*. *Biochemistry*. 1974;13(21):4447–55.
65. Taylor SD, Zhang H, Eaton JS, Rodeheffer MS, Lebedeva MA, O'Rourke TW, et al. The conserved Mec1/Rad53 nuclear checkpoint pathway regulates mitochondrial DNA copy number in *Saccharomyces cerevisiae*. *Mol Biol Cell*. 2005;16(6):3010–8.
66. Chevreaux B, Wetter T, Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. In: *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*: 1999. 1999. p. 45–56.
67. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5(2):R12.
68. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 2005;33(Web Server issue):W686–9.
69. Maleszka R, Skelly PJ, Clark-Walker GD. Rolling circle replication of DNA in yeast mitochondria. *EMBO J*. 1991;10(12):3923–9.
70. Bendich AJ. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. *J Mol Biol*. 1996;255(4):564–88.
71. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
72. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4(4):406–25.
73. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.
74. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
75. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3(5):418–26.
76. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Program NCS, et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res*. 2003;13(4):721–31.
77. Dubchak I, Ryaboy DV. VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. *Methods Mol Biol*. 2006;338:69–89.
78. Katoh K, Frith MC. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics*. 2012;28(23):3144–6.
79. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764–70.
80. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 2004;20(2):289–90.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

