Hypotheses:

**Hypothesis**

**~~1) Does gene expression at the level of the whole chromosome scale with DNA levels for aneuploid chromosomes?~~**

**~~If not, is it a significant difference?~~**

**~~Do different lines with the same aneuploid chromosome(s) have similar expression levels on that chromosome?~~**

~~Analysis~~

~~t-test (or non-parametric equivalent) between expected gene expression level and average observed expression level~~

~~t-test on gene expression levels between different lines with the same aneuploid chromosome~~

~~\*also: use JMP to plot distribution of gene expression levels on aneuploid chromosome – should not be normally distributed~~

**2) Are there certain genes that are up or down regulated in all aneuploid lines? (ESR?)**

**Which ones are on the aneuploid chromosome and which ones are elsewhere?**

**Are there certain genes shared between ALL lines that are up- or down-regulated, regardless of ploidy? (MA effects?)**

**Are there certain genes on euploid lines that are commonly up- or down-regulated? Are any of these different than the aneuploid lines?**

Using cuffdiff, compare the output files listing DE genes to one another

Only aneuploid lines >50%, >75%, >90%

Sort resulting file by chromosome

then do again using all of the lines

Also separate by experiment and sequencing time

Cuffdiff and DESeq2 data match: the lines that seem to have little to no DE genes are the same between the two.

Going to primarily use DESeq2 but can use Cuffdiff to corroborate data.

Throwing out lines: 21, 66, 31, 69(?)

**3) What is the rate of aneuploidy? Is it different between GC and MA, and MA and old MA?**

Look up how to calculate this

Does the rate of aneuploidy correlate with wild/lab strain status? yes

Do our aneuploid lines have more gene conversion events? don’t know- question for another day?

What chromosomes are aneuploid? Are there any that are commonly aneuploid? Never aneuploid?

Look up the rate of aneuploidy in wild yeast versus lab yeast

Chromosomal instability and heterozygosity?

Mating between 2 haploids, so no lethals to uncover

Average rate of events per chromosome: 4 would expect this for every chromosome but don’t get that, also implies selection against certain aneuploid chromosomes

(monosomies and trisomies) and 2 events (tetrasomies)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # events | GC | MA | total |  |
| 0 | 54 | 116 | 186 |  |
| 1 | 38 | 27 | 66 |
| 2 | 1 | 2 | 1 |  |

Given rate, should be able to calculate the expected numbers

Rate: # trisomies + # monosomies OR rather, #trisomies\*2 since when you get a trisomy you also get a monsomy (but if you don’t end up seeing them, implies strong selection against a monosomy)

GC:

observed – 37

expected – 74

MA:

observed – 27

expected – 54

rate

Equation 1

Equation 3

Equation 4

**4) What are the fitness consequences of being aneuploid?**

Sam’s part

**5) Are histone genes dosage compensated? Are any ESR genes up- or down-regulated?**

**Other DE genes:**

**What are they? Are there any housekeeping genes?**

Cuffdiff

Do at same time as Hypothesis #2

**Dosage compensation/RNAseq analysis methods/notes:**

Reference genome from SGD – latest update (31-Jan-2015) (S288C)

1. Need to build an index using bowtie

use bowtie2-build (code #1 in text file)

2. Need to use Cufflinks to change .gff transcript file into a .gtf file

<http://cole-trapnell-lab.github.io/cufflinks/file_formats/#the-gffread-utility>

I used the Illumina Stranded RNAseq kit. The majority of the libraries got both an i5 and an i7 adapter. Some of them, the first batch you gave me, got single adapters. The second batch I used the i5 / i7 dual adapters. I hope this helps, I don't know what to tell you otherwise.

**Introduction**

Aneuploidy is when an organism contains an abnormal chromosome number, i.e. one not a multiple of the haploid state. There is some debate as to why aneuploidy is often maintained (or tolerated) in populations. Some hypothesize there is an intrinsic mechanism of dosage compensation to buffer the deleterious effects of imbalanced gene dosage [1, 2]. Others contest this argument, claiming there is no evidence for dosage compensation at the whole-chromosome level in S. cerevisiae [3]. An alternate hypothesis is that the accumulation or loss of chromosomes is an adaptive advantage to stressful environments, such as the case with yeast in an oxide-rich media that accumulate an extra copy of chromosome XI [4]. Here, we investigate whether there is an innate dosage compensation response in spontaneously aneuploid yeast that have been put through a 2000-generation mutation accumulation experiment with a single-cell bottleneck every 20 generations [5]. This allows us to examine the rate, type, and effects of spontaneous aneuploidies in an environment with little to no selection. The absence of selection allows us to analyze the transcriptome without bias from environmental factors, enabling us to determine the effects of aneuploidy on the genome in a stable condition and without selecting for any particular aneuploidies or mutations.

How often do aneuploidies arise in the absence of selection?

and what are they, what are their fitness effects?

-Get aneuploids in the wild, but that could be from selection in a potentially stressful environment (ie oxidative stress)

Looking at spontaneous aneuploidies in the absence of selection allows us to get a baseline for how often these spontaneously arise

Looking for if there is a common mechanism of DC in autosomes. Sex chromosomes evolved from autosomes, so they may have brought with them some inherent way of compensating for multiple gene copies.

**Methods**

*Strains, RNA extraction, and sequencing*

One of the strains used was previously described in [5]. Briefly, lines were obtained from previous mutation accumulation experiments, one from a highly heterozygous ancestor (hybrid of NCYC 3631, a Mat alpha derivative of YPS 606 (oak strain, PA, USA), and NCYC 3596, a Mat a derivative of DBPVG1106 (lici fruit, Indonesia, wine clade)). The other strain used was a lab strain obtained from one of the lab’s previous mutation accumulation experiments (described in [5]). This strain was homozygous at every locus except for the mating-type locus and was of the genotype *ade2, lys2-801, his3-∆D200, leu2-3.112, ura 3-*52. Both strains from the separate mutation accumulation experiments were diploid, and put through a bottleneck every ~20 generations for ~2000 generations.

30 lines from both hybrid and lab strains (12 aneuploid hybrid, 8 euploid hybrid, 6 aneuploid lab, 2 euploid lab), 3 replicates of each, were inoculated in 3ml liquid YPD (yeast extract 2%, peptone 2%, glucose 2%), and incubated on a rotator at 30° C for 24 hours. RNA was extracted using the MasterPure Yeast RNA Purification Kit (Epicentre). Integrity, concentration, and quality of RNA samples were assessed using the Qubit. Libraries were prepared using the Illumina Stranded RNAseq Kit. They were sequenced at the Georgia Genomics Facility on the Illumina NextSeq (75 cycles) SE75 High Output flow cell. Samples were multiplexed and split across two runs.

Some of the lines were sequenced on a different platform. 45 colonies of the lab strain (3 replicates each of 15 lines) were inoculated each into 3 ml liquid YPD medium, and incubated on a rotator at 30 for 24 hours. After 24 hours, mRNA was extracted using the MasterPure Yeast RNA Purification kit (Epicentre). RNA libraries were constructed using the Illumina Tru-seq mRNA Stranded Kit, amplified using 13 cycles of PCR and sequenced on an Illumina HiSeq 2500.

Because these datasets were extracted by different individuals, libraries were made with different kits, and the samples were run on different sequencing machines, analyses were carried out identically, but separately.

*RNAseq data analysis*

Raw sequence data was obtained from the Georgia Genomics Facility, with adapters removed. Quality control was done using FastQC version 1.8.0\_20 using default parameters (available at <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Low-quality bases were trimmed using Trimgalore v. 0.4.4 using -phred 33, -q 20 (available at https://www.bioinformatics.babraham.ac.uk/projects/trim\_galore/). RNA samples were aligned to the Ensembl R64-1-1 reference genome and annotated transcripts (obtained from <https://support.illumina.com/sequencing/sequencing_software/igenome.html>) using Tophat v. 2.1.1 with -i 10 -I 10000 [6]. Cufflinks v. 2.2.1 was used to assemble the transcriptomes using default parameters [6]. Normalization was done using Cuffnorm v. 2.2.1 with default parameters, differential expression was found using Cuffdiff v. 2.2.1 with default parameters, and read counts were found using HTseq v. 0.6.1pl (Python v. 2.7.8) [6]. Samtools (version 1.3.1) was used to view .sam files as .bam files and sort .bam files [7]. Scripts can be found at <https://github.com/hollygene/Dosage_Compensation/blob/master/assembly_script.sh>.

Cuffnorm was used to find FPKM (fragments per kilobase per million reads). A homebrew bash script was used to join the FPKM values for each line with the gene attributes file, turn the file into a .csv, remove mitochondrial sequences, and change the chromosome names from Roman numerals to numbers (script can be found at <https://github.com/hollygene/Dosage_Compensation/blob/master/DC_workflow_April2017.sh>). This data file was then read into the R statistical software [8]. Average FPKM between the replicates for each line at each chromosome was found, followed by the average FPKM ratio (average FPKM of descendent line/average FPKM of ancestral line). Boxplots and ANOVAs were done in R. Comparisons made between aneuploid lines and the expected log2 ratio of trisomic, monosomic, or tetrasomic chromosomes were made using a one-sample t-test with mu = log2(1.5) for trisomic, mu = log2(0.5) for monosomic, and mu = log2(2) for tetrasomic. Two separate, but identical, analyses were done for each of the datasets. R scripts are located at <https://github.com/hollygene/Dosage_Compensation/blob/master/DC_workflow_old_MA.Rmd> and <https://github.com/hollygene/Dosage_Compensation/blob/master/DC_workflow.Rmd>.

DESeq2 (CITE) was used to find differential expression at the gene-level and compare across samples. Read counts were filtered so that the sum of each row of counts was greater than 20 (\*\*I wonder if I change this so that I filter individual reads out that have read counts less than 10, would it significantly decrease the number of genes analyzed?). The function DESeq() was ran with default parameters, except that during the normalization, only chromosomes that were NOT aneuploid in any line were used to normalize to.

For individual gene analysis, we first wanted to know which genes were differentially expressed in multiple MA lines from the same experiment (hybrid and lab strains). To do so, we used DESeq2 to find the DE genes in all samples, then filtered the genes based on a Benjamini-Hochberg-adjusted p-value of <0.1. We then found the number of common genes found between samples in the filtered dataset.

We then asked which genes were differentially expressed on the aneuploid chromosomes of the aneuploid samples. To do this we performed a similar method as above, adding an extra filtering step to isolate the aneuploid chromosome(s) in each sample. We then compared samples with the same aneuploid chromosome and found what genes were commonly DE between samples aneuploid for the same chromosome.

Finally, we wanted to know if there were any genes on the aneuploid chromosome(s) that were dosage compensated. To do this, we first removed genes with a BH-adjusted p-value less than 0.1 from the aneuploid datasets. We then removed genes that were NOT on the aneuploid chromosome, resulting in a dataset containing genes located on the aneuploid chromosome that had a BH-adjusted p-value > 0.1. We wanted to know what percentage of the genes on the aneuploid chromosome were above this threshold value, so we found the number of genes interrogated on each aneuploid chromosome. We divided the number of non-DE genes by the total number of genes interrogated and multiplied by 100 to get a percentage of non-DE genes.

**Result***s*

*Rate of spontaneous aneuploidy is nearly twice as high in hybrid strain than lab strain*

Rate of aneuploidy in hybrid strain: 1.82E-04

Rate of aneuploidy in lab strain: 1.04E-04

Fisher’s exact test results: p-value = 0.0106 \*statistically significantly different numbers of aneuploidies between hybrid and lab strains

For the lab strain, 31 out of 145 sequenced samples were found to be aneuploid [5]. Of these, 29 had whole-chromosomal duplications, and 2 had whole-chromosomal deletions [5]. No segmental duplications were found in the lab strain lines. 4 lines contained multiple aneuploidies (i.e. two separate chromosomes had become aneuploid). Half of the chromosomes were aneuploid in multiple lines (Figure 2(?)).

proportion of lines with segmental duplications:

MA: 0

GC: 1

proportion of lines with multiple aneuploidies:

MA: 4 MA lines multiple aneuploidies

4 Lines had multiple events: 43: 2&3 trisomic, 108: 9 monosomic, 8 trisomic

152: 1, 8 10: 4, 16

GC: 7 lines GC with multiple aneuploidies

11: 1 (1n), 15 (3n)

18: 1(3n), 12(3n)

31: 7(3n), 16(3n)

33: 14(3n), 16(3n)

40: 5(3n), 16(3n)

47: 9(3n), 16(3n)

66: 7(3n),10(3n)

76: 9(3n), 10(n/2), 14(3n)



The rate of aneuploidy was calculated using Equation 1. For the lab strain, the rate of aneuploidy was found to be 1.06 x10-4. The probability of getting one aneuploid event (i.e. a single monosomy or trisomy in a single line) was calculated using equation 2. For the lab strain this probability was found to be 0.178. Multiplying this by the number of lines in the lab strain gives 25.8, indicating we expect 25 lines to have one aneuploidy event. Our data shows we found 27 lines with one event. The probability of two aneuploid events was calculated from equation 3. In the lab strain we found the expected number of two aneuploid events to be \*\*\*. The probability of more than two aneuploid events was found using Equation 4. This probability in the lab strain was found to be 0.022, indicating that we would expect to see 3.19 lines with more than two events. Zero lines with more than two events were found, indicating strong selection against this type of aneuploidy.

Hybrids of two yeast species have been shown to systematically lose all or part of one parent’s genome early in mitosis [9]. It is possible that our hybrid of distantly-related S. cerevisiae strains may show a milder version of genome incompatibility as exemplified by the higher rate of aneuploidy compared to the homozygous lab strain.

Chromosome IX was aneuploid most often in (at least MA, maybe GC lines?). A previous study found that chromosome IX was highly unstable in terms of chromosome loss in both diploid *S. cerevisiae* and diploid hybrid *S. cerevisiae-S. bayanus* strains [10] .

Chromosome V is lost spontaneously in S cerevisiae at a rate of 2-8 x 10-6 cell generations [11].

*Segmental Duplications Arose x% of the time*

Are the segmental duplications where you would expect them to be? (i.e. the ratios?)

\*\*How would you calculate the expected # of segmental duplications?

For line 76, where is the breakpoint? Presumably the section has to contain the centromere. Is the breakpoint in a gene?

*No evidence for whole-chromosome dosage compensation in either lab or hybrid strains*

Comparisons using Tukey’s Honestly Significantly Different Test (cite) were made between euploid and the aneuploid lines of interest (i.e. for chromosome 1, only those lines aneuploid for chromosome one were analyzed against each of the euploid lines). All aneuploid lines analyzed showed significant differential expression against each euploid line for the chromosome of interest. Most, but not all, aneuploid lines had nonsignificant p-values when comparing the gene expression on the aneuploid chromosome to the expected value of gene expression of a monosomic/trisomic/tetrasomic chromosome. Together, these observations support the conclusion that there is no whole-chromosome dosage compensation occurring in either the hybrid or lab strains. These findings are supported by previous work showing no dosage compensation in aneuploid yeast [12].

\*Maybe include a plot of the correlation between DNA copy number and RNA copy number in aneuploid chromosomes.

even though some lines have statistically significantly different gene expression levels, it could be because there is more RNA present to be degraded in trisomic lines, and less RNA present to be degraded in monosomic lines

Although autosomal dosage compensation has been observed in higher eukaryotes (cite some drosophila autosomal DC papers), it may be that these loci are more susceptible to dosage imbalances, or the phenomenon and mechanism of autosomal dosage compensation has evolved later in eukaryotes. Perhaps yeast do not require a mechanism of autosomal DC as they are so numerous and have a short generation time, so selection can act quickly to get rid of segmental or whole chromosome aneuploidies.

**Individual Genes**

Histone genes are known to be dosage compensated in S. cerevisiae (CITE). Only one of our aneuploid lines was aneuploid for a chromosome containing a histone gene (Chrom XV, histone gene is called \*\*\*). Analyzing the gene expression on this chromosome, we found that the histone gene \*\*\*\*\*\*\*\*, suggesting \*\*\*\*\*.

Yeast are known to undergo what is known as the environmental stress response (CITE), when conditions are unfavorable due to various factors, including \*\*\* (CITE). Torres(?) et al (CITE) found that similarly, aneuploid yeast undergo what they refer to as the “aneuploid stress response,” in which certain genes NOT located on the aneuploid chromosome are differentially expressed. Curious as to if we found the same genes and patterns in our spontaneously aneuploid samples, we investigated the same genes as Torres et al and found \*\*\*\*\*.

What about individual genes? Environmental stress response genes? Is the “aneuploid stress response” from Torres et al 2010 turned on?

Are there certain regions on the chromosome that are dosage compensated?

Comparisons across all aneuploid/euploid/GC/MA lines

ESR genes?

Histone gene compensation?

What genes are up/down regulated in aneuploid lines with higher fitness than the ancestor?

Discussion

Acknowledgements

Figures

References

1. James Hose, C.M.Y., Maria Sardi, Zhishi Wang, Michael A Newton, Audrey P Gasch, *Dosage compensation can buffer copy-number variation in yeast.* eLIFE, 2015. **4**: p. 1-27.

2. Audrey P Gasch, J.H., Michael A Newton, Maria Sardi, Mun Yong, Zhishi Wang, *Further support for aneuploidy tolerance in wild yeast and effects of dosage compensation on gene copy-number evolution.* eLIFE, 2016. **5**: p. 1-12.

3. Eduardo M Torres, M.S., Angelika Amon, *No current evidence for widespread dosage compensation in S. cerevisiae.* eLIFE, 2016. **5**: p. 1-19.

4. Kaya, A., et al., *Adaptive aneuploidy protects against thiol peroxidase deficiency by increasing respiration via key mitochondrial proteins.* Proc Natl Acad Sci U S A, 2015. **112**(34): p. 10685-90.

5. Zhu, Y.O., et al., *Precise estimates of mutation rate and spectrum in yeast.* Proceedings of the National Academy of Sciences, 2014. **111**(22): p. E2310-E2318.

6. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nature protocols, 2012. **7**(3): p. 562-578.

7. Li, H., et al., *The sequence alignment/map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.

8. Team, R.C., *R: A language and environment for statistical computing.* 2013.

9. Marinoni, G., et al., *Horizontal transfer of genetic material amongSaccharomyces yeasts.* Journal of Bacteriology, 1999. **181**(20): p. 6488-6496.

10. Kumaran, R., S.-Y. Yang, and J.-Y. Leu, *Characterization of chromosome stability in diploid, polyploid and hybrid yeast cells.* PLoS One, 2013. **8**(7): p. e68094.

11. Mulla, W., J. Zhu, and R. Li, *Yeast: a simple model system to study complex phenomena of aneuploidy.* FEMS microbiology reviews, 2014. **38**(2): p. 201-212.

12. Torres, E.M., et al., *Identification of aneuploidy-tolerating mutations.* Cell, 2010. **143**(1): p. 71-83.