

**Molecular characterization of the black yeast
Hortaea werneckii in saline environments**

by

Sean Philip Formby

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF**

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Pharmaceutical Sciences)

**THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)**

July 2017

© Sean Philip Formby, 2017

Abstract

As of 2007, over 30 million hectares are affected by salinization resulting in poor crop yield and a reduction of food production. Reversing salinization of soil is an expensive and long-term process. The bioengineering of plants to better cope with salinization of the soil is an ongoing research effort. *Hortaea werneckii* is an extremely halotolerant (salt tolerant) black yeast and can grow in the absence of salt or in almost saturating conditions (5M NaCl). Its natural ecological niche is the solar salterns of Slovenia which have range of environmental extremities such as the salt concentration, low oxygen, and high UV intensity. Recently it was discovered that this yeast has had recent genome duplication and 90% of the proteins exist in duplicate. The whole genome duplication and the extreme NaCl tolerance of *H. werneckii* provide an interesting model to investigate molecular mechanisms involved in salt stress.

In this study, *H. werneckii's* genome assembly is improved (increased contiguity) and used for subsequent molecular experiments such as MNase-seq and RNA-seq. These experiments were used to examine differences of gene expression and the corresponding chromatin architecture across a range of saline conditions to determine important molecular mechanism in salt tolerance. *H. werneckii* increases respiration in response to salt stress exemplified by the upregulation of mitochondrial associated genes and antioxidant defense genes. Additionally, *H. werneckii* genes encoding zinc transporters and genes involved in glycerol assimilation were increased in response to high salt. The chromatin landscape of some of these genes differs from other yeasts such as *S. cerevisiae*. Using next generation sequencing and third generation sequencing a more complete picture of *H. werneckii's* mechanisms of salt tolerance has been obtained while also creating an extensive data-base for future research.

Lay Abstract

Extremophiles are organisms that thrive in extreme environmental conditions normally detrimental to other organisms. Extremophiles are crucial for better understanding mechanisms of stress tolerance and in this case, salt stress. The increase of salinity in agricultural land is a huge global problem as plant growth is inhibited, resulting in lower or absent crop yield. Efforts using transgenes from extremophiles to bioengineer salt tolerant plants has proven somewhat effective, yet more research is needed to reach the desired level of salt tolerance to reclaim salinized land. Here, *Hortaea werneckii*, a salt tolerant black yeast, is molecularly characterized using the latest sequencing technology. This has resulted in an improved genome which can be used for future research and has revealed detailed insights into salt tolerance which can be applied in the creation of salt tolerant organisms.

Preface

This thesis is based on the experimental design developed by Dr. Corey Nislow and the experiments performed at the Nislow Giaever Laboratory, UBC.

Genome assembly, pipeline development, differential expression analysis, data-set curation, nucleosome occupancy analysis and all subsequent bioinformatic analysis was performed by the author of this thesis.

Culture cultivation, genomic DNA isolation, total RNA isolation, and nucleosome experiments were optimized and performed by the author of this thesis.

Sequencing and RNA library preparation was performed by Dr. Sunita Sinha, UBC.

Transcriptome assembly and annotation was done by Dr. Jason Stajich, University of California Riverside.

Genome assembly preparation from SMRT-Portal was conducted by Dr. Mauricio Neira, UBC.

Table of Contents

Abstract	ii
Lay Abstract	iii
Preface	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
List of Abbreviations	ix
Acknowledgments	x
Dedication	xi
1. Introduction	1
1.1. Overview	1
1.2. Salinization, a global problem, and its effect on plants	1
1.3. Halophiles and halotolerant organisms.....	4
1.4. <i>Hortaea werneckii</i> ecology and lifestyle.....	5
1.5. Adaptations to hypersaline stress and the role of duplicated genome	6
1.5.1. The role of glycerol, melanin, and cation homeostasis in salt tolerance	6
1.5.2. Evidence of a whole genome duplication and its implications	8
1.6. Combining next generation sequencing and third generation sequencing	10
1.7. Chromatin and nucleosomes	11
1.8. Study objectives	13
2. Methods and Materials	14
2.1. Culture and growth conditions	14
2.2. Genomic DNA isolation and PacBio sequencing	14
2.3. Nucleosome DNA isolation	14
2.4. NGS library preparation of mononucleosomal DNA fragments	16
2.5. RNA-Isolation and sequencing	17
2.6. Hybrid Genome Assembly	18
2.7. Transcriptome analysis	19
2.8. MNase-seq analysis	20

3. Using third and next generation sequencing to assemble of <i>H. werneckii</i>'s genome	21
3.1. Characterizing a halotolerant fungus	21
3.2. Combining next generation sequencing and third generation sequencing to assemble a contiguous genome	22
3.3. The Hybrid Assembly	31
4. The transcriptome of <i>H. werneckii</i>	33
4.1. Insights from a transcriptome assembly	34
4.2. Analyzing abundance of gene expression	35
4.3. Differential expression of <i>H. werneckii</i> grown in 0% NaCl versus 10% NaCl media	37
4.4. The gene expression profile of high salt relative to optimum growth conditions	40
4.5. Comparison of high salt (20% NaCl) and no salt (0% NaCl) gene expression in <i>H. werneckii</i>	45
5. Evaluating the chromatin landscape of <i>H. werneckii</i>	49
5.1. The influence of chromatin and nucleosomes	49
5.2. Inferring nucleosome positions with MNase-seq	49
5.3. <i>H. werneckii</i> and genome <i>cis</i> determinants of nucleosome positioning	50
5.4. Evaluating nucleosome occupancy and positioning at the transcription start site	52
6. The relationship of gene expression and chromatin states across salt conditions	59
6.1. Steady transcription and chromatin states, growing in high salt.	59
6.2. Gene expression and nucleosome dynamics	59
6.3. Changes of nucleosome positioning and occupancy in response to salt	60
7. The genome of <i>Hortaea acidophila</i>	69
7.1. <i>H. acidophila</i> genome assembly and comparison.....	69
8. Conclusion	74
References	77
Appendix A: Hybrid Assembly Pipeline	88
Appendix B: RNA-Seq Pipeline	92
Appendix C: MNase-seq Pipeline	95
Appendix D: Mitochondrial annotation	97

List of Tables

Table 1: Comparison of hybrid assemblies	27
Table 2: Assembly statistics of a hybrid assembly	30
Table 3: Transcriptome assembly statistics	35
Table 4: Summary of up and down regulated genes	37
Table 5: Gene expression abundance (TPM)	60
Table 6: Assembly statistics of <i>Hortaea acidophila</i>	70

List of Figures

Figure 1: Micrographs of <i>H. werneckii</i>	5
Figure 2: Mircococcal nuclease digestion gel	16
Figure 3: Hybrid Assembly Pipeline	19
Figure 4: Growth characterization of <i>H. werneckii</i>	22
Figure 5: Phenotype of <i>H. werneckii</i>	22
Figure 6: Phlyogentic tree of <i>H. werneckii</i> and other fungi	23
Figure 7: BUSCO assessment of hybrid assemblies	28
Figure 8: BUSCO assessment of final hybrid assembly	30
Figure 9: Whole genome duplication synteny plot of <i>H. werneckii</i>	31
Figure 10: 0% NaCl vs 10% NaCl differential expression	38
Figure 11: Expression of 1,3-beta-glucanosyltransferase transcripts	39
Figure 12: Differential expression of 20% NaCl versus 10% NaCl.....	41
Figure 13: Expression of zinc transporter associated genes.....	42
Figure 14: Carbon metabolism and oxidative stress associated genes	44
Figure 15: Differential gene expression between 0% NaCl versus 20% NaCl.....	47
Figure 16: Sequence (<i>cis</i>) composition of nucleosomes in <i>H. werneckii</i>	51
Figure 17: Nucleosome configuration of the transcription start site (TSS)	53
Figure 18: Nucleosome profile clusters	55
Figure 19: <i>De novo</i> motifs upstream of the TSS	56
Figure 20: Heatmap nucleosome profiles of 10% NaCl and 20% NaCl.....	62
Figure 21: Nucleosome and expression profiles of Int1, SodA and SodC.....	64
Figure 22: Dihydroxyacetone kinase like transcripts in <i>H. werneckii</i>	67
Figure 23: Glyceraldehyde-3-phosphate dehydrogenase like transcripts in <i>H. werneckii</i>	68
Figure 24: Comparison of <i>H. acidophila</i> versus <i>H. werneckii</i> and <i>Z. tritici</i>	71
Figure 25: Analysis of the histone variant gene loci	73

List of Abbreviations

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
ROS	Reactive oxygen species
HKT	High-Affinity Potassium Transporter
GPD	Glycerol-3-phosphate dehydrogenase
HOG	High osmolarity glycerol
NGS	Next generation sequencing
3GS	Third generation sequencing
MAPK	Mitogen-activated protein kinase
STL	Sugar Transporter-Like protein
WGD	Whole genome duplication
OLC	Overlap-layout-consensus
DBG	de Bruijn graph
NDR	Nucleosome depleted region
Mnase	Micrococcal nuclease
GO	Gene ontology
PMA	Plasma membrane ATPase
UTR	Untranslated region
TTS	Transcription termination site
TSS	Transcription start site
CDS	Coding DNA sequence
DAK	Dihydroxyacetone kinase
TPM	Transcripts per kilobase million
FC	Fold change
DHA	Dihydroxyacetone
BWA	Burrow-Wheelers aligner
ABS	Absolute value
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
DHN	1,8-dihydroxynaphthalene

Acknowledgements

I would like to thank the Dr. Corey Nislow for all the help he gave in writing this thesis and the opportunity to do so.

I offer my deepest gratitude to Dr. Thomas Chang, Dr. Theresa Rogers, Dr. Wayne Riggs, Dr. Abby Collier, Shirley Wong and Shirley Nakata who ensured that this thesis came to fruition and have been extremely supportive throughout my academic career.

I thank Dr. Amy Lee, Dr. Elisa Wong, and Dr. Kahlin Cheung-ong who taught me how to think critically.

I would like to also thank Dr. Zamir Punja and Andrew Wylie (MSc) for their discussions and support.

I would also like to thank my parents and friends for their support throughout a turbulent time.

To my parents

1. Introduction

1.1. Overview

There has been major progress in molecular biology research over the last decade, and many of these advances can be ascribed to innovative nucleic acid sequencing techniques[1]. Sequencing essentially determines the base constituents of DNA and/or RNA molecules and such sequencing methods are now able to provide this data in a timely and high throughput manner[2]. The two current, widely used sequencing methods broadly fall into two categories, Next Generation Sequencing (NGS) and Third Generation Sequencing (3GS), named according to when these innovations were developed and also the type of data produced i.e. short reads versus long reads[1, 3]. Both take advantage of DNA polymerase to analyze nucleic acid sequence (discussed below)[1]. Moreover, these methods have been refined and modified to examine a variety of biological processes on the molecular level. Thus, they have become essential tools for genomics, transcriptomics and epigenetics, especially for non-model organisms where previous molecular work is lacking. Such is the case for extremophiles, organisms that can thrive in otherwise detrimental environments. In particular, the salt tolerant fungus *Hortaea werneckii* is of interest due to its ability to survive a wide range of salt concentrations[4]. The research present here, molecularly characterizes *H. werneckii* in response to salt via combining these sequencing technologies. This work provides further insight into mechanisms involved in fungal salt tolerance while also laying the framework for future transgenic manipulation in salt sensitive eukaryotes.

1.2. Salinization, a global problem, and its effect on plants

Salinization is the harmful buildup of excess soluble salts in farmland mostly due to improper irrigation techniques such as poor drainage and the use of brackish water for irrigation [5]. As of 2016, it was estimated that over 45 million hectares (19.5%) of irrigated land was affected by salinization, the majority occurring in developing countries[6, 7]. Salts are normal constituents of all soils and are necessary because they contain essential elements required for normal plant growth. However, in excess, they become toxic severely inhibiting growth and eventually proving

fatal for the plant[8]. Therefore addressing this global problem is of utmost importance considering population growth and the corresponding demand for food production[9].

The predominant salt affecting soils is sodium chloride (NaCl), the most widespread natural occurring salt, and soils are considered saline if the concentration exceeds 40 mM NaCl[10]. NaCl accumulates in the soil from irrigation water which always contains trace amounts with the concentration dependent on the origin. If there is not adequate drainage, NaCl will accumulate to detrimental concentrations. Glycophytes are plants that are sensitive to excess NaCl and these include most agriculturally important crops such as wheat, rice, corn and the legumes[11-13]. Salt tolerance varies among species and strains yet the physiological response to salt remains largely similar[7].

Salt stress imparts a complex physiological response in plants cells and can broadly be subdivided into early onset osmotic stress, later occurring ionic stress and resultant oxidative stress[8-10]. The osmotic stress is due to the lowering of osmotic potential in the external environment caused by the low water activity imposed by excess salt. This hinders efficient water uptake, elicits stomata closure in photosynthetic tissues and reduces cell expansion rate in growing tissues[8, 12]. The ionic stress occurs by the buildup of ions (Na^+ and Cl^-) to toxic levels within the cell, that is, the plant cell cannot efficiently sequester or exclude these ions and they accumulate in photosynthetic tissues[12, 14]. The cation Na^+ in most plants reaches a toxic concentration before the anion Cl^- , although this is species dependent, as woody crops are more likely to be sensitive to Cl^- due to persistent woody stems and roots (soybean and citrus)[12]. Na^+ imparts toxicity by competing with K^+ for major binding sites of cytoplasmic enzymes, it upsets ion homeostasis resulting in K^+ deficiency and other nutrient imbalances, and impairs the function of chlorophylls, carotenoids and photosynthetic enzymes[8, 9, 14, 15]. Both stresses impair photosynthesis and as a result there is an increase in the production of reactive oxygen species (ROS)[8, 9, 12]. Normally these molecules are regulated and detoxified by antioxidant mechanisms but when they exceed the cells detoxification capacity they induce oxidative stress. It should be mentioned that ROS in plants are also important signalling molecules in abiotic stress, however

imbalance of ROS negatively impacts cellular processes as they directly damage membranes, proteins and inhibit cellular function[16, 17]. These stresses are cumulative, adversely affecting many processes including photosynthesis, germination, growth, and overall crop yield. Although glycophytes are sensitive to salt, their tolerance varies across species and genetic varieties as plants have several mechanisms to withstand salinity[7, 13].

There are three main categories of salt tolerance mechanisms as proposed by Munns and Tester [12]. These include ion exclusion involving the exclusion of Na^+ from leaf and shoot tissues, tissue ion tolerance involving the sequestering of Na^+ into cell compartments, and osmotic tolerance involving increased water uptake at the root which is ion independent [12, 13]. Moreover, other components have also been proposed to contribute to salinity tolerance such as the leaf area of the plant, production of antioxidants and transpiration use efficiency[7, 13]. Salt tolerance can thus be seen as a multigenic trait involving multiple physiological responses. Thus, there is great interest to augment salt tolerance mechanisms in plants via genetic engineering. For example, overexpression of *Arabidopsis thaliana* Na^+ transporter, AtHKT1;1, increased salt tolerance in transgenic rice lines by removing Na^+ from the xylem thereby preventing the transport to photosynthetic tissues[14, 18]. This coincides with maintaining a low Na^+/K^+ ratio within the cell, which has been shown to be a common phenotype of salt tolerant plant lines and species[11-13, 19]. Additionally, other abiotic influences can affect salinity tolerance in plants such as supplemental addition of essential elements including K^+ and Mn^{+2} , both of which have been shown to alleviate symptoms of salt stress[19-22].

Reversing soil salinization is expensive and a time consuming process and as a result there is great interest to bioengineer plants that can cope with salinized land without a substantial reduction of yield[23]. Due to the complex nature of plant genomes, multicellularity, long generation times and multifaceted physiological responses to salt, it may prove more effective to study fungal halophiles (salt loving fungi). These fungi can provide novel genes or shared mechanisms with the potential to confer greater salt tolerance in crops [4, 7, 24]. Efforts using fungal halophiles as novel gene sources have already proven effective, however more research is

still needed to reach the desired level of salt tolerance [25-28]. Moreover, insights from these fungi can prove fruitful in other industrial processes, as osmotolerance is an important trait in industries such as food fermentation and enzyme production[29].

1.3. Halophiles and halotolerant organisms

Halophilic and halotolerant fungi are examples of extremophiles which are organisms that thrive in environments usually detrimental to life. In this case they are able to flourish in high salinity[30]. An example of this natural abiotic stress occurs in the solar salterns of Slovenia[31]. The salterns are pools of trapped sea water which occur with tidal change; as the sea water in these pools evaporates, the concentration of dissolved salt (NaCl) increases. As water continues to evaporate, NaCl concentrations increase and water activity (a_w) decreases. When the NaCl concentration eventually reaches its saturation point (32% w/v) it will crystallize out of solution. The hypersaline salterns have additional abiotic stresses such as high temperature, high ultra violet (UV) radiation, low oxygen levels and display fluctuation of nutrient levels[31]. These conditions create a hostile environment to most microbes but halophilic and halotolerant fungi, like black yeasts (e.g. *Hortaea werneckii*), have adapted to exploit this niche. As in plants, the hypersaline environment imposes osmotic, ionic, and subsequent oxidative stress on the fungi, albeit these stresses and their adaptations occur more rapidly (within hours of insult).

Studying these eukaryotic life strategies may lead to understanding biological pathways which can be manipulated in higher order organisms such as plants. For example, the expression of a ribosomal subunit from *Aspergillus glaucus*, a halophilic fungus, in a transgenic tobacco plant resulted in an increase of salt tolerance when compared with the control [26]. Likely, this ribosomal subunit is more resistant to the excess Na^+ concentrations within cell thus complementing the existing mechanisms. Thus, there is a great need to study these organisms especially on the molecular level.

1.4. *Hortaea werneckii* ecology and lifestyle

Hortaea werneckii is an ascomycetous, halotolerant, melanized polymorphic black yeast belonging to the order *Capnodiales* (figure 1)[31]. Its halotolerance is reflected by its ability to grow in solutions lacking NaCl as well as solutions with near saturating concentrations (0 M - 5.2 M NaCl). It has broad growing optimum which lies between 0.8 M and 1.7 M NaCl[32]. This remarkable capability allows *H. werneckii* to thrive in the hypersaline solar eutrophic salterns of Slovenia, its primary ecological niche[33]. It is the most dominant and adapted fungal species in this environment, especially when salinity reaches 20% NaCl (w/v) or higher[34]. All of these contribute to *H. werneckii*'s polymorphic phenotype as the changes in its morphology correlate with its environment.

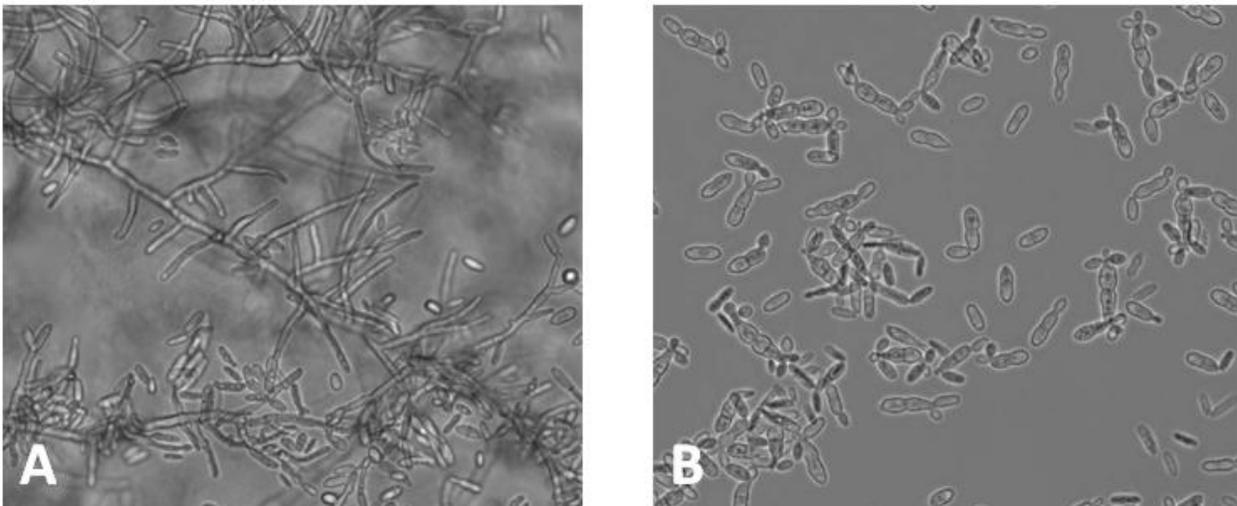


Figure 1. Micrographs of *Hortaea werneckii* at 400x magnification (using bright field) from different cultures. A) Hyphal stage of *H. werneckii* grown on minimal 5% NaCl yeast nitrogen base (YNB) agar. B) Yeast stage of *H. werneckii* grown in minimal 10% NaCl YNB broth.

Interestingly, *H. werneckii* been isolated from a variety of environments with low water activity such as salty food and wood immersed in brine[33]. During periods of drought, *H. werneckii* changes from its hydrophilic yeast form to a hydrophobic hyphal stage where conidia are produced. These structures are subsequently dispersed by air currents and able to germinate in hypersaline waters (Figure 1A & 1B)[32]. This, along with changes in cell number and nutrition also facilitate change between hyphal and yeast morphology, contributing to its complex asexual

lifestyle[31]. So far, no sexual cycle has been described for *H. werneckii* though a recent study has suggested the species is heterothallic (self-incompatible)[35]. Only one type of mating locus, the *MAT1-1* idiomorph, was found in the genome of the sequenced strain and fungi require the presence of the other idiomorph *MAT1-2* to initiate mating [33]. This, along with no evidence of a cryptic opposite mating type locus *MAT1-2*, indicate that *H. werneckii* is a heterothallic fungus in which two partners of opposite mating type are required for mating[35]. However, this may be due in part to a fragmented genome assembly and further work is required to determine this claim.

1.5. Adaptations to hypersaline stress and the role of duplicated genome

1.5.1. The role of glycerol, melanin, and cation homeostasis in salt tolerance

H. werneckii uses many strategies to combat the hyperosmotic and ionic stresses from salt stress, with the main one being the compatible solute strategy[32]. The compatible solute strategy involves the synthesis and accumulation of organic compatible solutes (mainly glycerol) in the cytoplasm which balances the osmotic gradient and protects cellular structures from an increase of toxic sodium ions[28]. The main enzyme involved in glycerol synthesis in fungi is glycerol-3-phosphate dehydrogenase (*GPD1*), which is under the control of the high-osmolality glycerol (HOG) signalling pathway. This osmoreponsive pathway is a well characterized mitogen activated protein kinase (MAPK) cascade and is crucial for adaptation to hyperosmolar environments[28]. The downstream kinase of the pathway, MAPK *HwHOG1*, when activated, is transported to the nucleus, and can down or up regulate osmoreponsive genes such as *GPD1*. Glycerol is a low molecular weight molecule that easily diffuses through the plasma membrane of the cell to the external environment [35]. *H. werneckii* melanises its cell wall and has glycerol uptake systems to maintain optimal intracellular concentrations, such as the *STL1*- like transporters, to counteract the passive glycerol leakage [4]. Furthermore, the compatible solute strategy is always coupled with the expulsion of toxic ions from the cell[4], accomplished via the multiple diverse cation transporters that maintain a low intracellular Na^+/K^+ ratio over a range of salinities[24, 35]. This low ion ratio is an observed phenotype of salt resistant plant cultivars and there is considerable similarity between the cation detoxification systems between plant and

fungi. This makes *H. werneckii* an excellent model to study eukaryotic salt tolerance however most studies of *H. werneckii* pertain to osmotic shock and do not address the state of acclimation [36-39]. Cation homeostasis is crucial in hypersaline environments and researchers found that there was an expansion of genes families encoding for alkali metal transporters in *H. werneckii* [35, 39]. The expansion of these metal cation transporter families presumably is due in part to a recent whole genome duplication (WGD), discussed later [31].

In response to high salinity, *H. werneckii* increases respiratory processes to meet the high energy demands of maintaining ion and glycerol homeostasis [40]. In previous studies some of the largest changes of gene expression in response to high salt were found to be associated with the metabolism and biogenesis of mitochondria[41, 42]. An increase in respiration results in the production of reactive oxygen species (ROS) which negatively interact with enzymes such as aconitase, a crucial component of the Krebs cycle, which is very sensitive to superoxide radicals [43]. To detoxify ROS, cells have evolved antioxidant defense pathways to scavenge these by-products of aerobic metabolism which include superoxide dismutase (SOD), peroxidase, catalase, and the glutathione and thioredoxin systems[44, 45]. These pathways may be potential candidates for genetic manipulation as transgenic popular plants expressing heterologous manganese SOD from *Tamarix androssowii* had increased salt tolerance[46, 47].

Micronutrients can also have a role in ROS detoxification. It is well known that zinc deficiency is associated with oxidative stress[48] and a recent study observed that supplementation of Zn⁺² to acetic acid exposed yeast cells (a method to induce oxidative stress), resulted in the reduction of ROS compared to controls [49]. This could reflect another approach for alleviating oxidative stress resultant from hypersalinity and highlights the importance of micronutrients to mitigate abiotic stress. Moreover, *H. werneckii* displays high ROS degradation ability (specifically H₂O₂) across a range of salinities and its capability to detoxify ROS has been proposed to be a determining factor of salt tolerance[41, 45]. However, the exact cellular mechanisms behind this, remain elusive.

In other fungi melanin plays a key role in oxidative stress protection, as observed in the pathogenic fungus *Fonsecaea monophora*[50]. In *H. werneckii*, cells exposed to H₂O₂, to induce oxidative stress, melanin had no effect on cell viability and thus is thought to have no role in H₂O₂ detoxification. Additionally, cells exposed to hypersalinity concentrations prior to the H₂O₂ insult, were able to survive higher H₂O₂ concentrations than cells with no exposure, suggesting different ROS detoxification mechanisms other than melanin must exist and pre-exposure to salt may prime antioxidant cellular machinery [45].

Although melanin may not be involved in ROS detoxification it still contributes to *H. werneckii*'s halotolerance as described by Kejzar [51]. When melanin biosynthesis was inhibited by tricyclazole, *H. werneckii* growth was arrested in high salinity media. Cell wall morphology was also altered. In ascomycetes, melanin is synthesized from the precursor 1,8-dihydroxynaphthalene (DHN) [52]. It has diverse roles in resistance to abiotic and biotic stresses while also contributing to virulence in melanised pathogenic fungi [53]. For example, increased expression of melanin synthesis genes (and a increase of melanin pigment) was associated with increased virulence of *Cryptococcus gattii* [54]. In *H. werneckii* melanin has been shown to limit the loss of glycerol from the cell but only at low NaCl concentrations (1M), at higher concentrations glycerol still diffused through the membrane [51]. Melanin likely contributes to halotolerance via changes in cell morphology and cell-wall integrity essentially limiting the exposure of *H. werneckii* to the environment[51]. Therefore, melanin biosynthentic pathways are pathways of research interest as they contribute to the halotolerant nature of *H. werneckii* and may provide additional clues into stress resistance of virulent pathogenic fungi.

1.5.2 Evidence of a whole genome duplication and its implications

In 2013, researchers found *H. werneckii* yeast had undergone a recent whole genome duplication when compared to phylogenetically related fungi (the order Capnodiales). Over 90% of predicted proteins existed in duplicate and contained similar amino acid identity [35]. These genes are referred to as ohnologs, which are genes that have been duplicated as a result of WGD

to distinguish them from small scale duplications [31, 55, 56] . The WGD means that although *H. werneckii* is a haploid, it can benefit from informal genetic redundancy [35].

The origins of the WGD remain unclear, however the amino acid similarity in the protein predicted genes strongly suggest an autopolyploidization event, which is the duplication of the genome within the species [35]. Thus, comparative genomics of related species may reveal insights into what drove this WGD and the expansion of metal cation transporters. For example, phylogenetic comparison of clades which diverged before or after an ancient WGD event in the *S. cerevisiae* lineage was used to deduce the nature of an ancient WGD. The researchers uncovered significant phylogenetic heterogeneity between the clades and hypothesized a model in which the ancient WGD was likely an allopolyploidization event, which is an interspecies hybridization resulting in a doubling of the genome[57, 58].

The somatic genome duplication in Sorghum root cells has been attributed to salt tolerance, which occurs through a controlled process called endoreduplication[59]. Endoreduplication refers to a variant cell cycle variant in which doubling of the genomic content occurs without subsequent segregation of chromosome and cytokinesis[60-64]. This is a common, essential, developmental process found in many multicellular organisms, such as plants and animals. However, this process is tightly controlled only occurring in specific tissues and does not pass through not the germ line [60, 64-66]. Endoreduplication has been reported in pathogenic fungi, but this is distinct from evolutionary polyploidy because the genome doubling is programmed, uniform and associated with the complex pathogenic life cycle[61, 67, 68]. Nonetheless, genome duplication is commonly associated with an increases of stress tolerance and likely an important component in determining *H. werneckii*'s salt tolerance[24, 65].

This leads to the question of when the WGD occurred, that is, before or after speciation events of related taxa. Understanding this may aid in understanding key evolutionary questions surrounding ohnolog gene fate[69]. *H. werneckii* has only two other species within its genus, and one, *Hortaea acidophila*, an acidophilic black yeast, has publicly available sequencing reads which

can be used to assemble its genome[70]. This would allow a comparative overview of the two genomes using synteny (conservation of chromosomal segments) between the two species to determine more information on the WGD and shared metabolic processes of these fungi [71, 72].

To address the underlying molecular mechanisms responsible for salt tolerance in *H. werneckii*, new DNA sequencing technologies are an attractive method. Here it is used delineate the genomic, transcriptional and epigenetic factors that contribute to salt tolerance.

1.6. Combining next generation sequencing and third generation sequencing

DNA and RNA comprise the cellular instructions for all living cells, understanding these molecules and their regulation is crucial for understanding biological systems. Within the last decade, the cost associated with DNA sequencing has been substantially reduced, giving many laboratories access to sequencing. These techniques allow a robust, high through put investigation into molecular processes involved in environmental stress responses. This has led to a surge in genome research using next generation sequencing. A major hurdle associated with DNA sequencing however, is data analysis.

Modern sequencing technologies produce vast amounts of data that require heavy computational processing. The two main sequencing technologies used are referred to as next generation sequencing (NGS) and third generation sequencing (3GS). Prior to being sequenced, DNA is fragmented into pieces as there is no technology at the moment that can sequence entire chromosomes. NGS technologies, such as the Illumina platforms, produce millions of short sequencing reads up to 500 bp in length depending on the platform and chemistry. Third generation sequencing technologies, such as the PacBio platform, produce reads greater than 1000 bases and average 5–20 kb per run[73]. NGS technologies are more accurate with approximately one sequence error per 1 kb, whereas 3GS produce 10-20 errors per 100 bp[1]. These error rates are important components in *de novo* assembly where the millions of reads are stitched back together to form the original continuous DNA molecules they originated from in the cell, akin to a jigsaw puzzle.

In *de novo* genome assembly, there are two major algorithms, overlap-layout-consensus (OLC) and de Bruijn graph (DBG). OLC is often used in long read assemblers such as Canu and Celera , whereas DBG assembly is used in short read assemblers such as AbySS, ALLPATHS-LG and SOAPdenovo[74]. As the name implies, OLC has three steps, it finds overlaps between reads, creates a layout of all reads and their overlap and from this a consensus of continuous DNA sequence is inferred[75]. This process works well with long reads but does not work well with short reads produced by NGS. Therefore, to handle this problem DBG assemblers chop up short reads into even shorter k-mers and construct links between the k-mers in a de Bruijn graph [75, 76]. Unbranched paths within the graph are inferred as continuous stretches of DNA and result in contigs. If a path in both cases cannot be transversed it results in fragmentation, as the program cannot confidently place a link[74]. Longer reads in general are more advantageous for *de novo* assembly as they transverse repeats allowing the assembler to overcome ambiguous branch points and thus resulting in a less fragmented genome but they are more error prone and must be corrected prior to OLC assembly[73]. The type of sequencing technology and subsequent software depends on the type of experiment, the computational power available and the biological questions asked.

1.7. Chromatin and nucleosomes

In eukaryotes, genomic DNA is packaged into chromatin. Nucleosomes are the basic repeating unit of chromatin and are composed of a histone octamer with 147 base pairs (bp) of DNA wrapped ~1.65 times around it in a left handed super helix. The histone octamer is made of four different types of histones, two copies of H2A, H2B, H3 and H4. The dyads (centre) of nucleosomes are spaced ~165-200 bp apart along the DNA sequence, and the amount of linker DNA (the DNA between nucleosomes) varies depending on the organism, genomic location and cell-type[77-79]. Their relative positions are dynamic and highly regulated along the DNA sequence and are influenced by other DNA associated proteins[80, 81]. DNA-histone interactions are composed of hydrogen bonds between the DNA minor groove negatively charged phosphate

backbone and the positively charged amino acid residues of the histones [82]. They provide stable packaging of DNA while allowing for epigenetic regulation by influencing DNA accessibility for other DNA binding proteins. Thus, they impact many DNA dependent processes such as DNA transcription, regulation, repair and recombination[81]. Moreover, nucleosomes are not randomly placed throughout the genomes, they form distinct repeating patterns and are highly organized around genic regions[83]. How these patterns are formed is still a matter of debate, but the general consensus is that a suite of molecular interactions are involved [77, 83]. These include the thermodynamic properties of DNA-sequence (*cis*-determinants), ATP-dependent chromatin remodelers and other DNA-binding proteins such as the RNA-pol II machinery and transcription factors[84, 85]. In terms of DNA sequence, certain conformations of nucleotides can influence the bending properties of DNA. Poly (dA:dT) and poly (dG:dC) tracts are intrinsically stiff and unfavorable to nucleosome formation whereas the 10 bp repeat of certain dinucleotides favors DNA bending and thus nucleosome formation[86]. Different yeast species have been shown to have different repeating and oscillating dinucleotide signatures [87].

For DNA to be transcribed, nucleosomes must be evicted or shifted so that the transcriptional machinery can pass through the nucleosome[79]. There are histone variants that make up nucleosome octamers and post translational modifications of histones which can also alter the dynamics of nucleosomes [81, 88]. Regions upstream of transcription start sites are typically depleted of nucleosomes, allowing transcription factors to establish the preinitiation complex (PIC) at the promoter[89]. These regions are known as “nucleosome depleted regions” (NDRs) and are important for proper gene function [82, 84, 85, 90, 91]. In *Saccharomyces cerevisiae*, more than 90% of promoters have this architecture and this is shared across eukaryotes[92-95]. NDRs can be altered to induce and repress transcription of genes in a controlled manner and are being explored as tools in genetic engineering [96, 97]. Therefore, determining the location and sequence properties within NDRs is important to understand transcriptional regulation[98]. Furthermore, the WGD of *H. werneckii*'s genome and its extremophilic nature provide an interesting model for looking at nucleosome dynamics. For *H. werneckii*, identification of nucleosome depleted regions could also be used to address gene divergence of ohnologs.

Early studies addressing the influence of the nucleosome on gene expression led to the general consensus that nucleosomes are inhibitory for gene expression, and *in vitro* experiments support the model[99]. High resolution techniques have revealed a more nuanced picture *in vivo* in which some promoters occupied by nucleosomes still display high levels of gene expression[100]. While the exact mechanisms of chromatin-level control of gene expression have not yet been fully elucidated, nucleosomes have a predominant regulatory function [101]. Moreover, the distance of the NDR, determined from dyad to dyad, correlates with expression levels and can be used to classify genes [77]. Wide NDRs are characteristic of constitutively expressed genes such as those required for mRNA processing and organelle organization. Whereas closed NDR conformation is associated with sporadically expressed genes such as those involved in stress responses [100]. As environmental changes influence transcription, they must also influence chromatin in terms of reconfiguration. Moreover, in plants epigenetic regulation of chromatin is a Thus, determining the nucleosome landscape of *H. werneckii* in response to chronic salt stress is of considerable interest.

1.8. Study objectives

Studying *H. werneckii*, a black yeast capable of surviving harsh abiotic systems may provide insight into mechanisms to alleviate salt stress in plants and fungi [50, 52, 54]. The large genome and extreme NaCl tolerance exhibited by *H. werneckii* make it an excellent model to study extremophile stress responses on the molecular level. Most studies pertaining to *H. werneckii* have studied hyperosmotic shock as opposed to acclimatized cells. This presents an opportunity to study *H. werneckii* under chronic salt exposure in the context of transcription and epigenetic changes using DNA sequencing. This data can then be compared to other eukaryote salt stress responses to delineate the genes and chromatin configuration associated with salt tolerance. The physiological responses expected are mechanisms involved in micronutrient acquisition, compatible solute production, ion homeostasis and antioxidant defense. The aim of this study is to provide the necessary data, using next generation and third generation sequencing technologies (NGS and 3GS respectively), to explore the molecular mechanisms on the gene and chromatin level that confer salt tolerance in *H. werneckii*.

2. Methods and Materials

2.1. Culture and growth conditions

H. werneckii strain, EXF-2000, was isolated from marine solar salterns on the Adriatic coast in Slovenia. It is maintained in the Ex Culture Collection of the Department of Biology, Biotechnical Faculty, University of Ljubljana (Infrastructural Centre Mycosmo, MRIC UL, Slovenia) and in the CBS culture collection (Centraalbureau voor Schimmelcultures, the Netherlands; strain number CBS 100457). *H. werneckii* cells were grown in supplemented synthetic defined yeast nitrogen base liquid medium (SC) (yeast nitrogen base 6.7 g/L, glucose 20 g/L, amino-acid complete supplement mix 2 g/L). The medium was made to pH 7.0 using 2N NaOH and supplemented with 0%, 10%, and 20% NaCl (w/v). Growth rate was assessed using a 48 well plate and measured on spectrophotometer at 28° C.

2.2. Genomic DNA isolation and PacBio sequencing

Two separate cultures of *H. werneckii* (grown in 50 mL of 10% NaCl SC) were grown to mid exponential phase (OD 0.8-1.0) at 28° C in 250 mL flasks at 180 rpm. Next, 1 mL of each culture was aliquoted into 1.5 mL microcentrifuge tubes five times and each tube spun at 11,000 rpm on a table top centrifuge. The supernatant was poured off and DNA was isolated from the cell pellet using Genra Purgene Yeast kit (Qiagen) according to manufacturer's instructions. Samples were pooled and then sequenced on five SMRT cells using PacBio RS II (Pacific Biosciences).

2.3. Nucleosome DNA isolation

Three separate *H. werneckii* cultures were grown in 200 mL of 0%, 10% or 20% NaCl SC media in 1 L baffled flasks until mid-exponential phase (OD 0.8-1.0). Nucleosomal DNA was isolated from these cultures according to Tsui *et al* (2012) [102]. Methanol-free formaldehyde (16%) was added to the cultures to a final concentration of 2%. Cultures were incubated for 30 minutes at 30 ° C on rotating shaker at 180 rpm to allow fixation. Glycine (2.0M) was added to quench the culture at a final concentration of 125 mM for 5 minutes. Each culture was spun down in 50 mL conical centrifuge tubes for 5 minutes at 4°C (7500 x g).

The supernatant was then poured off into a waste container and cell pellets were kept on ice. Each pellet was washed with 10 mL of 1X phosphate buffered saline (PBS) and combined back into a single 50 mL conical tube respective of the condition it came from. The cells were spun down for 5 minutes at 4°C (7500 x g) and the supernatant poured off. To digest the cell wall, 15 mL of zymolyase buffer (1 M sorbitol, 50 mM tris pH 7.4, 16.2 mM 2-mercaptoethanol) and 250 µL of zymolase (~312 U) were added to each sample. Each sample was put in the 37 °C shaker and incubated for ~60-90 minutes or until 90% of the cells lyse under pressure of a coverslip. Samples were then spun down at 8,000 x g ultracentrifuge for 15 minutes at 4°C. The supernatant was poured off and 6 mL of MNase buffer (1 M sorbitol, 10 mM tris pH 7.4, 5 mM MgCl₂, 1 mM CaCl₂, 0.075% igepal, 1 mM 2-mercaptoethanol, 500 µM spermidine) was added to each sample. Each sample was divided into six 2 mL microcentrifuge tubes (1 mL each).

Micrococcal nuclease (MNase) was added to each tube at concentrations of 0U, 2.0U, 12.5U, 22U, 30U and 50U respectively for 40 minutes at 37°C on a rotary shaker. The enzyme in each tube was deactivated by adding 75 µL of inactivation enzyme buffer (5% SDS, 50 mM EDTA) and 6 µL of proteinase K (20 mg/mL). Samples were incubated overnight in a 65°C water bath. Next, 1 mL of phenol pH 8.0 was added to each tube, vortexed for 1 minute, and spun in a centrifuge for 10 minutes (12,000 rpm) at 4°C. The top aqueous layer was transferred to a new 1.5 mL microcentrifuge tube and 900 µL of phenol:chloroform:isoamyl alcohol pH 8.0 was added. The samples were vortexed for 1 minute and spun in a centrifuge for 10 minutes (12,000 rpm) at 4°C. The top aqueous layer was transferred to a new 2 mL microcentrifuge tube. DNA was precipitated by addition of sodium-acetate (final concentration of 300 mM) and 2x the sample volume of 100% cold ethanol (tubes were split prior to accommodate the volume). The DNA pellet was suspended in 80 µL of water and RNA was depleted via addition of 2µL RNase A for 2 hours at 40°C. DNA was then run on a 2% agarose gel and samples that displayed 80% mononucleosomes were selected for downstream processing(Figure 2). The selected samples were used for NGS library preparation without gel extraction to minimize potential bias as described in Henikoff's alternate protocol [103]. The experiment was repeated three times for a total of three biological replicates for each condition.

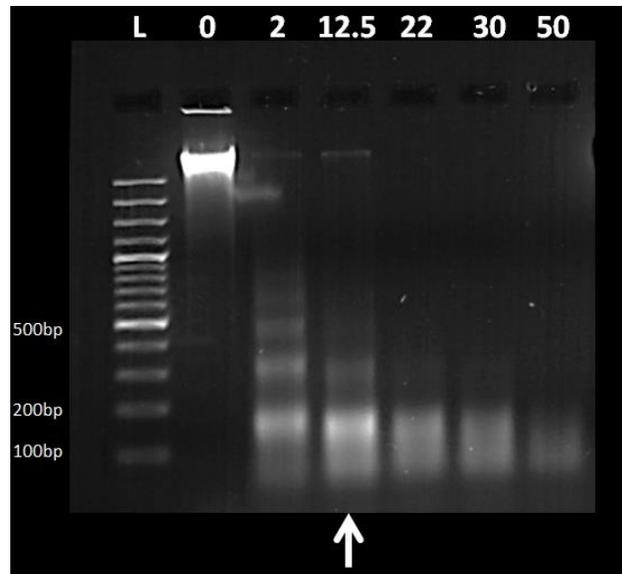


Figure 2. Micrococcal nuclease titrated digested DNA on a 2% agarose gel . The numbers above the lane correspond to enzyme units used for the digestion and L is the ladder. The numbers on the left correspond to the fragment size of the ladder and the arrow is an example of a sample with ~80 % mononucleosomes.

To assess the degree of micrococcal nuclease digestion on naked DNA (DNA without associated proteins), cells from a single mid exponential culture grown in 50 mL 10% NaCl SC in 250 mL flasks were collected. DNA was isolated using Gentra Puregene Yeast kit (Qiagen) according to the manufacturer’s instructions. The naked genomic DNA (gDNA) was digested with Proteinase K to remove any residual proteins. The DNA was aliquoted into five 1.5 mL microcentrifuge tubes and titrated with 0U, 0.01U, 0.03U, 0.06U and 0.1U of micrococcal nuclease. Digestion and DNA precipitation was done as previously described and visualized on a 2% agarose gel. The sample corresponding to a similar band as described for nucleosomal DNA was selected for library preparation.

2.4. Next generation sequencing library preparation of mononucleosomal DNA fragments

Nucleosome libraries were prepared using NEBNext® Ultra™ DNA Library Prep Kit (New England Biolabs) directly from the DNA isolated from MNase digestion according to manufacturer’s instructions. The size selection for mononucleosomes was done via AMPure beads

selecting for 150 base pair (bp) insert size (270 bp insert and adaptors) as described in the NEBNext® Ultra™ DNA Library Prep manual. Next, 65 µL of AMPure beads was used for the first size selection step and 25 µL for the second resulting in 270 bp fragments. The library size was confirmed using a 2100 Agilent bioanalyzer (Agilent Genomics) which checks DNA fragment size distribution within a DNA sample. The resulting average fragments had an average peak of 260 bp, corresponding to the desired size. The samples were then sequenced using the Illumina HiSeq (Illumina) with paired end reads of 101 bp read length.

2.5. RNA-Isolation and sequencing

Using the same salinity conditions (0% NaCl, 10% NaCl, and 20% NaCl) 50 mL cultures of *H. werneckii* were grown to mid-exponential phase in 250 mL flasks. Cells were isolated by filtration using a 0.2 µm pore membrane and immediately flash frozen along with the membrane using liquid nitrogen. The resulting frozen membrane and cells were ground using a ceramic RNase-free mortar and pestle kept at -80°C to minimize degradation and stored at -80°C until RNA isolation. Approximately 80 mg of the frozen grindate was transferred to a 1.5 mL microcentrifuge tube for each condition. Next, 1 ml of TRIzol reagent (Thermo Fisher Scientific) was added to each sample and total RNA was isolated according to manufacturer's instructions [104]. Degradation of RNA was checked on a 2% agarose gel. Samples with no degradation were selected for library preparation. The selected samples were prepared for sequencing using Illumina's 101 TruSeq RNA v2 library preparation kit according to manufacturer's instructions. The resulting libraries were sequenced on a HiSeq 2500 with paired end reads of 101 bp read length averaging 118 million reads. The 0% NaCl condition displayed substantial degradation, and three more samples along with a single 10% NaCl were re-isolated and libraries prepared as previously described. These samples were sequenced on the Illumina MiSeq and only had an average of 3 million reads per sample.

2.6. Hybrid Genome Assembly

A first trial *de novo* genome was assembled using HGAP version 2.3 pipeline in the SMRT Analysis suite provided by Pacific Biosciences. The filtered subreads (quality score ≥ 0.75) outputted from the SMRT portal were used for subsequent *de novo* assembly. These subreads were then used as input for the assembly algorithm Canu and all subsequent long-read based algorithms [73, 105]. Short reads from previous experiments and MNase-seq were corrected using BLESS[106] and assembled into multiple genome assemblies using SPAdes[107] and Platanus[108]. Corrected long reads from Canu output were used for DBG2OLC to merged assemblies[73]. The assemblies were corrected using Sparc and Pilon[109, 110]. Reads were mapped back to the genome and sorted using BWA and samtools respectively [111-113]. BUSCO was then used to assess genome completeness, only genome assemblies with higher completeness than the HwHGAP3 assembly were selected for merging[112]. Assemblies were merged with stringent parameters (overlap cutoff used in selection of anchor contigs (-hco) ≥ 7.0 and overlap cutoff for contigs used for extension of the anchor contig (-c) ≥ 4.0) via quickmerge[114]. The assemblies were then scaffolded using SSPACE-LongRead[115], polished using Pilon three times with corrected short reads and assessed for completeness using BUSCO and QAST[116]. The pipeline is summarized in figure 3 and the corresponding code in appendix 1 (Figure 3, Appendix A).

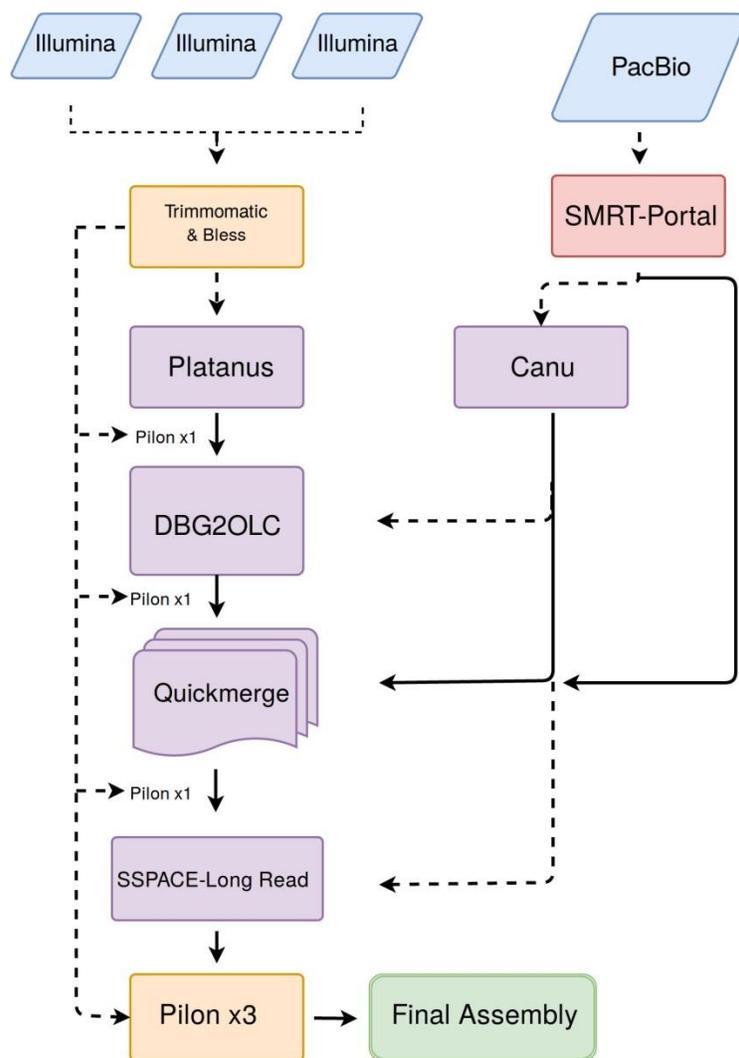


Figure 3. Summary of the bioinformatic assembly pipeline for *H. werneckii*'s genome. Blue shapes represent sequencing reads and where they originated from. Orange shapes are short read and assembly preparation algorithms. Purple shapes are assembly algorithms. Red are long-read preparation algorithms. Dotted lines represent reads being passed through whereas solid lines represent assemblies being passed through the pipeline.

2.7. Transcriptome analysis

A single run of 10 % NaCl RNA-seq reads were assembled using Trinity and annotated using the Maker2 pipeline by an external collaborator, Dr. Jason Stajich [117, 118]. RNA-seq reads were pseudoaligned to the transcriptome via Kallisto with 100 bootstraps and analyzed using Wald's test function in Sleuth for differential expression[119, 120]. Corresponding code for differential

analysis is in Appendix B. Gene ontology terms were assigned to each transcript using InterProScan and analyzed for enrichment using BINGO and GO Slim[121-123].

2.8. MNase-seq analysis

MNase-seq reads were trimmed using Trimmomatic and any reads smaller than 80 bp were discarded[124]. The reads were aligned to the *de novo* reference genome using BWA. Reads were filtered using samtools and bedtools and analyzed for positioning and occupancy using DANPOS2 [125, 126]. Each condition was fold normalized to allow proper comparison between the experiments and all scores were normalized to 10,000,000 reads. Differences between the nucleosome profiles for each salinity were calculated by DANPOS2. The pipeline is available in Appendix C. Heatmaps and genome wide phasograms were generated using DANPOS2 aligned to the transcription start site (TSS) and coding start site (CSS). The nucleosome occupancy profiles were visualized and clustered using NucTools[127]. Gene ontology enrichment was conducted using BiNGO and GO slim[122].

3. Using third and next generation sequencing to assemble of *H. werneckii*'s genome

3.1. Characterizing a halotolerant fungus

To address the underlying molecular mechanisms of salt stress tolerance, *H. werneckii*'s morphology and phenotype must first be assessed. *H. werneckii* was grown in synthetic complete media supplemented with NaCl at concentrations of 0%, 10% and 20% (w/v). Growth was measured on a spectrophotometer, reading every 15 minutes until saturation for all cultures (140 hrs) (Figure 4A). The 20% NaCl condition, had the longest generation time at 15.7 hrs, 0% NaCl at 9.8 hours and 10% NaCl at 9.0 hours although it had a longer lag phase (Figure 4B). In the salt conditions, the pH was lower with 20% NaCl at pH 4.05, 10% NaCl at pH 4.85 and 0% NaCl at pH 5.18 at end of saturation. When *H. werneckii* cells are actively growing, they establish a proton gradient at the plasma membrane to drive secondary active transporters [4]. This observed difference and change of pH between conditions could explain the lag phase observed, because the cell would first have to divert more energy from growth processes to establish the proton gradient, akin to a pH shock. Cultures grown in 20% NaCl were darker at mid exponential phase due to increase disposition of melanin, a known trait of *H. werneckii* in high salinities (data not shown) [51]. The 10% NaCl culture was slightly less pigmented and 0% was the least pigmented at mid exponential phase. Melanin, as previously mentioned has been implicated in halotolerance for *H. werneckii*[51]. It should also be noted that *H. werneckii* is extremely polymorphic and displays differences in colony morphology even on the same condition of 10 % NaCl (Figure 5A), however microscopically the cells look similar (Figure 5B). This polymorphic nature is common in black extremophilic yeasts [128]. *H. werneckii* is far diverged from *S. cerevisiae* as seen in figure 6, and the Capnodiales are quite a diverged group (Figure 6).

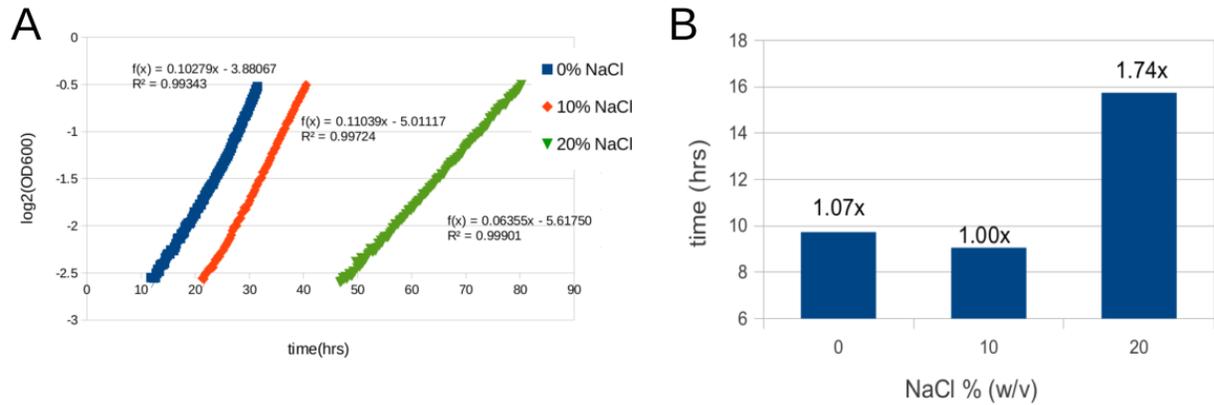


Figure 4 Growth characterization of *H. werneckii*. A) Growth curves of *H. werneckii* grown in SC supplemented with NaCl at 0%, 10%, and 20% (w/v) on a 48 well plate. OD readings were taken every 15 minutes, for 600 readings total. B) Doubling times defined as growth rates in time (hrs), for different 0%, 10% and 20% NaCl concentrations relative to reference (10% NaCl).

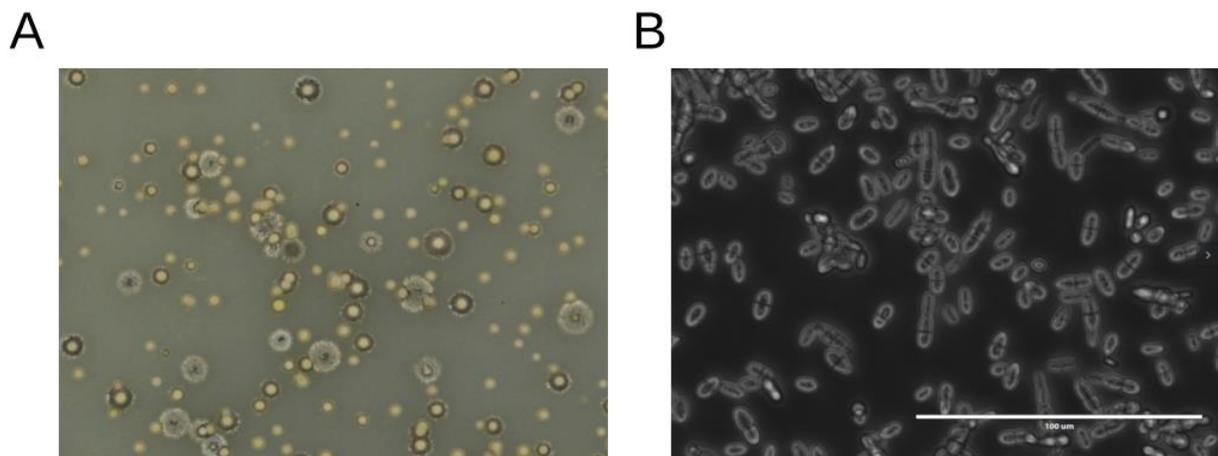


Figure 5. Phenotype of *H. werneckii*. A) Photograph *H. werneckii* colonies grown on YNB agar plate supplemented with 10% NaCl showing the observable marco differences in colony morphology. B) Micrograph of *H. werneckii* cells grown in 10% NaCl YNB broth under 400x magnification showing the yeast stage of the fungus.

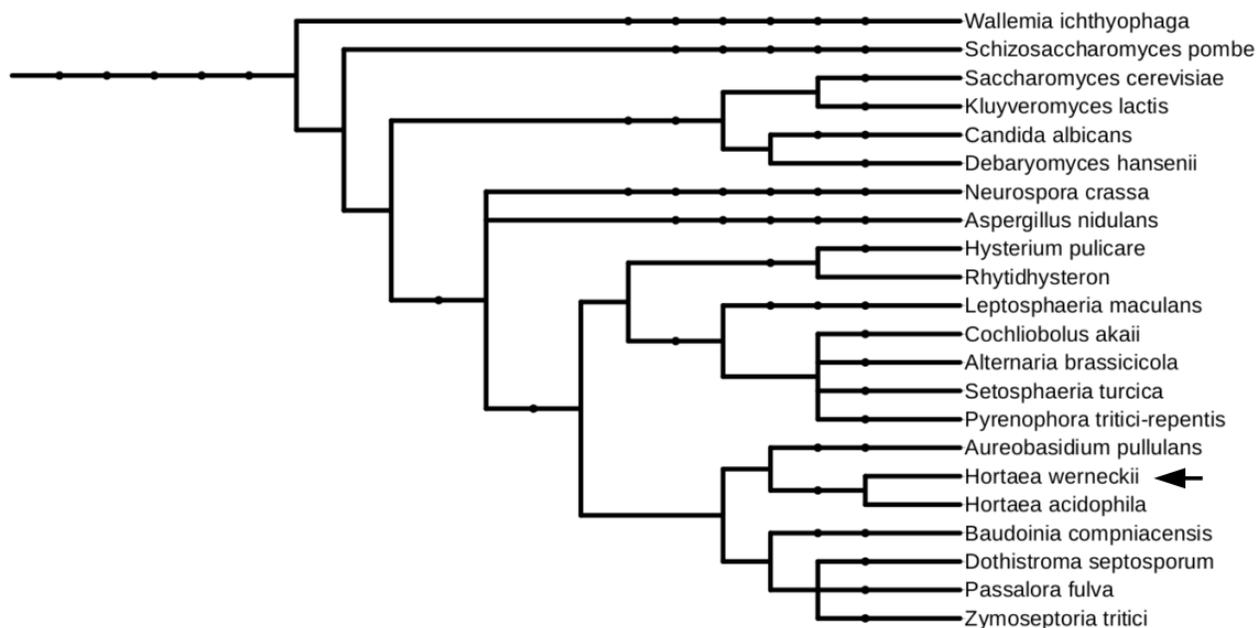


Figure 6. Phylogenetic tree of *H. werneckii* and fungi referenced throughout the thesis. *Wallemia ichthyophaga* was used as outgroup, using phylot from <http://phylot.biobyte.de/> and respective NCBI taxonomy relationships. Dots represent levels of clades and lengths of branches do not represent evolutionary time.

3.2. Combining next generation sequencing and third generation sequencing to assemble a contiguous genome

This first assembly project to sequence *H. werneckii* used a single library with 110x coverage using 75 bp paired end reads and assembled using AbySS, a de Bruijn graph assembler. The result was ~12,000 contigs [35]. This was adequate to demonstrate a whole genome duplication as almost all the predicted proteins within this assembly had a duplicate. Of course, this raised the question of whether this fungus is a haploid or a diploid. The researchers searched the genome for genes associated with the idiomorphs MAT1-1 and MAT1-2 from *Zymoseptoria tritici* but only had two hits, both for MAT1-1. This suggests that this fungus is heterothallic and requires the opposite mating type to initiate sexual reproduction [35]. No sexual state has been described for *H. werneckii*, or an opposite mating type with a *MAT1-2-1* gene. The *H. werneckii* genome is relatively large compared to other Capnodiales at 51 Mb and considering that the proportion of repetitive sequences is very low at 1.02%, this suggests that repetitive regions would

not be a huge contributing factor to the fragmented assembly as prior Illumina sequencing projects with other Capnodiales like *Cladosporium fulvum* resulted in 5899 contigs. Likely the fragmented genome is a result of the WGD. In theory, if there has not been dramatic sequence divergence between the duplicated portions of the genome an assembler cannot confidently predict the links between paths of k-mers in the DBG method resulting in a fragmented genome. Nonetheless, it appears there is enough sequence divergence for the contigs to be assembled into unique individual contigs. The fragmented genome also presents a challenge for other molecular experiments that rely on reference genomes such as MNase-seq and RNA-seq. Therefore, closing and improving this genome is of considerable interest. To address this assembly problem, *H. werneckii* would need more sequencing and in particular long reads using PacBio.

Genomic DNA was isolated from *H. werneckii* grown in 10% NaCl and sequenced on a PacBio sequencer RS II. The long reads were used to assemble the genome at 24x coverage. The first attempt of genome assembly using these reads was done using the Celera assembler which uses MinHash Alignment Process (MHAP and HGAP) constructed for PacBio long reads [129]. This resulted in 651 contigs, 49.8 Mb of sequence and a N50 of 153,735bp. N50 is an assembly statistic in which represents the smallest contig length at which 50% of the genome would be contained. That is, 50% of the entire genome of *H. werneckii* is contained within contigs larger than 153,735bp in this assembly versus the first assembly N50 of 8,354bp. This is relatively high. However, four contigs had very high coverage, with one exceeding the average throughout the genome by ~50 fold. Using BLAST with the NCBI database, the highest coverage contig was designated as a contaminant plasmid, showing that even after assembly, manual curation is required[130]. Another contig also had ~12.5-fold increase in coverage relative the global average. In this case, it mapped to other Capnodiales mitochondrial genomes. This explains the high coverage of this contig as there are more mitochondrial genomes versus nuclear genomes. The other two contigs corresponded to ribosomal regions which are rich in repeats and likely an artifact of sequencing and mapping. The mitochondrial contig was taken out of the assembly for protein prediction as mitochondrial genomes have a different genetic code. The nuclear genome was subsequently annotated using data from RNA-seq for gene prediction, discussed later.

Despite the increase in contiguity, the contig regions were not split at repetitive regions, rather many contigs had genic regions that were fragmented when the transcriptome was mapped to the genome assembly. For example, in the first assembly, four homologs coding for a plasma membrane ATPase (PMA1 and PMA2) similar to *S. cerevisiae* were found. This enzyme has been previously shown to have an important role in halotolerance of *H. werneckii* by actively creating a proton gradient which in turn powers other transporters at the plasma membrane [39]. The plasma membrane ATPase, Pma1 is crucial for establishing the proton gradient at the plasma membrane which in turn provides the necessary proton motive force for secondary ion transporters. *H. werneckii* has four homologs of *S. cerevisiae* Pma1 and Pma2, and their expression is salt dependent [35]. In the new assembly, there are six reported PMA1 genes which could mean that PMA1 actually has six copies or the assembly is flawed. Three of these PMA1 genes were present at the ends of two contigs and all of them had no annotated 5'UTR, meaning that this single gene annotation was split into three. The fragmentation of the genome, even though greatly improved, affected the transcriptome annotation and would affect gene expression analysis.

Complete contiguous fungal genomes are becoming more important as recent studies have demonstrated that some genomic features, particularly non-coding regions and repetitive regions play important roles in the adaptability and life-style of different fungi [131]. *H. werneckii* belongs to the order Capnodiales which contains many plant pathogens that are extremotolerant and can produce effector molecules such as in the barley pathogen *Ramularia collo-cygni* and the wheat pathogen *Mycosphaerella graminicola* (*Zymoseptoria tritici*) [132, 133]. Effector molecule genes in pathogenic and saprophytic fungi usually lie within the subtelomeric regions which have higher mutation rates, allowing rapid adaptation. *H. werneckii* is the causative agent of Tinea Nigra, a superficial skin infection, and effector molecules may play a role in the superficial infection [134]. Additionally, a contiguous genome can also give insight into biosynthetic gene clusters, which code for successive steps of a biosynthetic pathway [135, 136].

Hybrid assembly algorithms offer an alternative approach to short and long read assembly. Over the last three years these methods have been refined and used to complete other fungal genomes[137-139]. Here, using multiple data-sets from different experiments and sequencing technologies combined with the use of different assemblers *H. werneckii*'s genome was reduced to 30 super contigs, with a N50 of 2,868,097.

First, six DBG assemblies were generated from short reads originating from two MNase-seq, three long term evolution samples using the software Platanus and one long term evolution sample using SPAdes[107, 108]. The long-term evolution data-sets were generated from *H. werneckii* cultures that have undergone ~1000 generations in a high salinity condition. SPAdes is a standard microbial short read assembler and was used to compare with Platanus. Platanus was designed for highly heterozygous diploid genomes. As mentioned, *H. werneckii* is not a diploid but the WGD represents a similar problem in assembly for phased genomes (i.e. heterozygous assembly with parental allele information)[108]. The contigs produced from Platanus are high confidence contigs with low error rates. The trade-off in this case is many more contigs versus other de Bruijn assemblers, averaging 33,300 contigs per assembly and low N50 whereas SPAdes produced only 9140 contigs. Although the Platanus assemblies had substantially more contigs than the previous de Bruijn graph assembly of 2013 or SPAdes, they are of high quality which is essential for the hybrid assembler DBG2OLC[73]

DBG2OLC hybrid assembler uses high confidence contigs from a de Bruijn graph assembler and maps long reads to the contigs bridging them and resulting in a more contiguous assembly. This step does not perform error correction because per base error correction and alignment of long reads is computationally difficult and the authors claim that step is not necessary at the initial assembly stage [73]. However, for *H. werneckii* corrected long reads were generated prior to this step using the Canu assembler[105]. Misassemblies are more likely with *H. werneckii* due to the WGD which can confound the mapping of reads with high error rates. DBG2OLC uses the de Bruijn graph NGS assembly as an aid for aligning of the corrected long read sequences while taking advantage of the novel assembly contigs produced with a NGS assembly method. After the

assembly is constructed into a backbone or rough draft, the long reads are then mapped back to the assembly for consensus error correction. This method resulted in more contiguous assemblies than long read only assemblies for *H. werneckii* (Table 1).

Table 1. Comparison of hybrid assemblies versus long read only assemblies statistics and origin of data.

Assembly Name	Hybrid	Short Read assembler	Long/Hybrid Assembler	Total length (bp)	N50 (bp)	Contig Number	Short read origin
Hw1	Yes	Platanus	DBG2OLC	50.5 Mb	458539	205	LTE
SPAdes	Yes	SPAdes	DBG2OLC	49.8 Mb	153736	1313	LTE
Nuc10	Yes	Platanus	DBG2OLC	50.5 Mb	748033	132	MNase-seq
Hw7	Yes	Platanus	DBG2OLC	50.2 Mb	662438	125	LTE
Hw5-7	Yes	Platanus	DBG2OLC	50.0 Mb	784257	101	MNase-seq/LTE
Hw5	Yes	Platanus	DBG2OLC	51.6 Mb	527090	166	MNase-seq
Hw3	Yes	Platanus	DBG2OLC	50.5 Mb	516708	166	LTE
Hw2HGAP3	No	NA	Celera	49.9 Mb	153736	651	NA
Canu	No	NA	Canu (OLC)	49.8 Mb	411618	202	NA

It is important to note that Canu has replaced the Celera assembler in the Pacific Biosciences assembly pipeline, and thus was also used to construct a long read only assembly for comparison. Canu produced superior results when compared to the Hw2HGAP3 assembly, with higher N50 and lower contig number (Table 1). Moreover, Canu and the NGS assemblers were run on a desktop computer whereas HGAP needed a server cluster demonstrating that *de novo* assembly does not need expensive servers for assembling microbial genomes. Although DBG2OLC produced the most contiguous assemblies, Canu provided the corrected long reads which were used within DBG2OLC. The assemblies were then evaluated using a BUSCO which measures genome completeness via presence of near-universal single-copy fungal orthologs[111]

The BUSCO analysis was set to look for shared fungal orthologs within the genome giving a quantitative assessment of genome assembly completeness based on evolutionarily expectations of gene content rather than only the technical measure of N50 and contig number (Figure 7) [111]. Without the use of corrected reads, the missing BUSCOs within the DBG2OLC assemblies were higher than the Canu and Hw2HGAP3 assemblies (data not shown). The most striking result is the amount of duplicated orthologs in all assemblies. This strongly supports the hypothesis of a recent WGD in *H.werneckii*. The Hw5-7 assembly was determined to be the best assembly with 101 contigs, an N50 of 784,287, 50.0 Mb assembly size and 97.9% complete (Figure 7). It did however have lower duplicated orthologs than the other assemblies. Therefore, Hw7 which has more duplicated orthologs and represents a more conservative assembly (50.2Mb) was selected as the base assembly for Quickmerge, a metaassembler that combines different assemblies [114].

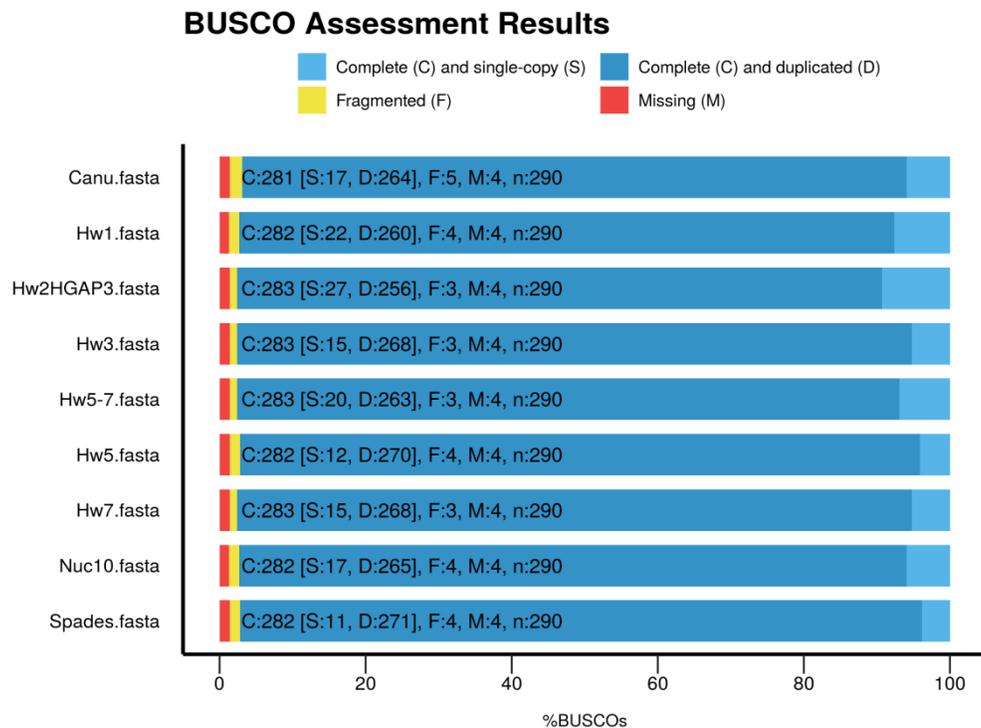


Figure 7. BUSCO assessment of completeness of the hybrid assemblies compared with long read (PacBio) only assemblies. Analysis of completeness of hybrid genome assemblies constructed using DBG2OLC via presence of universal single copy orthologs defined by evolutionary expectations. Red represent missing common orthologs in the BUSCO data base, light blue are complete and single copy orthologs found, dark blue represent duplicated complete orthologs and yellow are fragmented orthologs. The amount of duplicated genes is similar throughout the assemblies designated as dark blue.

Three assemblies, Hw3, Hw5 and Hw2HGAP3, were merged with the base assembly Hw7 using highly stringent parameters to ensure no misassemblies (see methods). The selection was based on the BUSCO completeness, size of the assembly, contig number, percent of short reads mapping back to the assembly and N50. This resulted in a 64 contig assembly with N50 1,047,462. Pilon, a consensus polishing algorithm used to improve microbial genome assemblies was then applied to the new assembly to correct any base pair errors and indels[110]. Lastly, the assembly was scaffolded twice using SSPACE-longread with stringent mapping parameters which were required a 5000 bp overlap, 99% identity and 0.05 ratio between potential links[115] (Appendix A). Pilon was used again, with three iterations for polishing of the assembly. The end result was an assembly with 30 scaffolds, N50 of 2,868,097, and size of 50.0 Mb (Table2). BUSCO and QAST were both used to evaluate the original 2013 Illumina Assembly (AIJO01.1), the Celera PacBio Assembly (Hw2HGAP3) and the new hybrid assembly Merged3.BC3[116]. QAST is another quality assessment tool for genome assemblies and invokes *in silico* gene prediction allowing proper comparison between the older assemblies and the new assembly. Also, instead of the fungal ortholog sets provided, orthologs from pezizomycontina_odb9 were used from the ortholog database (www.orthodb.org)[140]. This allows a more accurate evaluation of genome completeness of *H. werneckii*. The result is that the hybrid assembly is a more complete and contiguous genome of *H. werneckii* (Table 2, Figure 8). Evidence for the WGD is again exemplified by the BUSCO results because most orthologs exist in duplicate (Figure 8). This raises the question of how are these duplicated regions within the genome arranged; are there two similar contigs that exist, or have there been chromosomal rearrangements. Synteny within the genome was evaluated by mapping the genome to itself using Symap[141, 142].

Table 2. Final assembly statistics of the hybrid assembly compared with the Celera and Abyss assemblies via the program QUILT. Genes were predicted using the glimmer option with eukaryotic specification. OLC: Overlap consensus

	AIJO01.1 (Illumina)	Hw2HGAP3	Merged3.BC3
Method	de Bruijin	OLC	Hybrid
Year assembled	2013	2014	2017
Total length	51.6 Mb	49.9 Mb	50.0 Mb
# contigs	10,261	650	30
Largest contig	71,563	787,827	3,7161,197
GC (%)	53.56	53.48	53.52
N50	8,354	153,735	2,868,097
Predicted genes (>= 0 bp)	41,149	36,605	35,408
Predicted genes (>= 300 bp)	31,077	28,228	27,435
Predicted genes (>= 1500 bp)	6,181	6,881	7,222
Predicted genes (>= 3000 bp)	1,102	1,452	1,613
BUSCO (%) Completeness	77.6	96.5	96.9

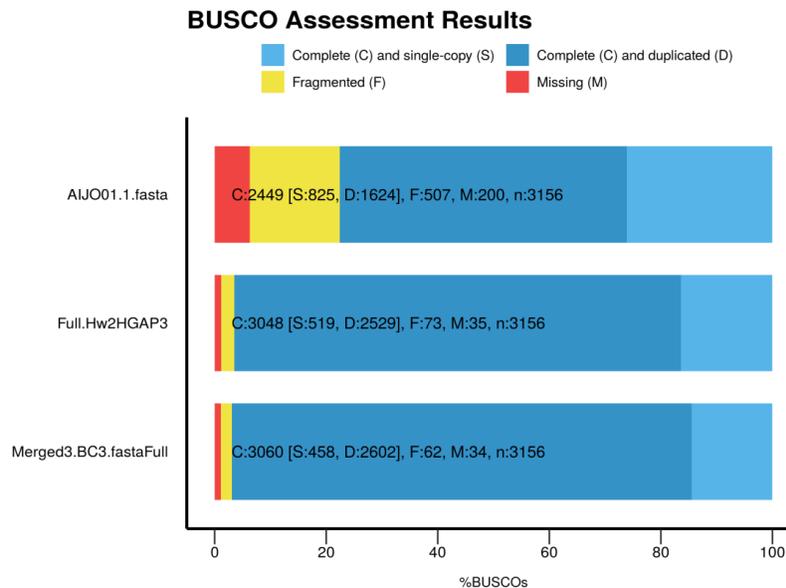


Figure 8. Hybrid BUSCO assembly statistics of *H. werneckii*, compared with previously assembled genomes. Analysis of completeness of current *H. werneckii* genome assemblies via presence of universal single copy orthologs as defined by evolutionary expectations. AIJO01.1.fasta represents the Illumina only assembly, Full.Hw2HGAP3 represents the PacBio only assembly and Merged3.BC3.fasta.Full represents the hybrid assembly. Red represent missing common orthologs in the BUSCO data base, light blue are complete and single copy orthologs found, dark blue represent duplicated complete orthologs and yellow are fragmented orthologs. The amount of duplicated genes is similar throughout the assemblies designated as dark blue.

3.3. The Hybrid Assembly

Each contig has large regions that are highly similar to other portions of the genome, providing support of a recent WGD and possible chromosomal rearrangements (Figure 9). Moreover, only four PMA1 gene locations were found within the genome instead of six in the Hw2HGAP3 meaning that this assembly should be used for the transcriptome annotation as well, because contiguous regions will result in more accurate gene mapping.

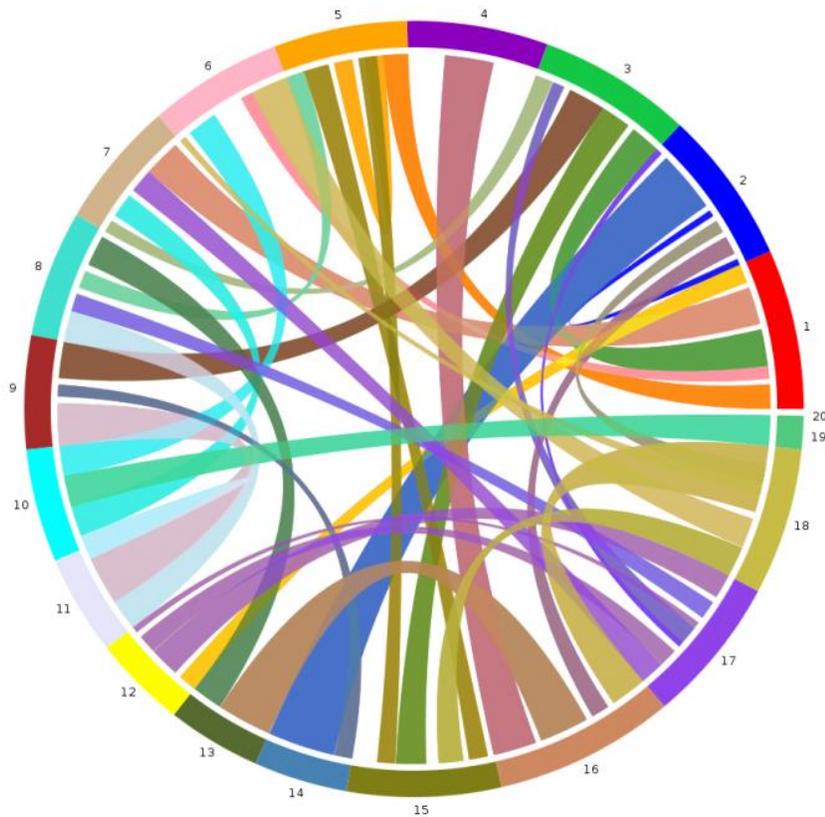


Figure 9. Whole genome duplication represented as a syntenic map of *H. werneckii*'s genome. The genome was aligned to itself and each colored block with a number represents a specific contig within the assembly. The large lines connection contigs represent syntenic matches found within the genome and are at least 50,000 bp in sequence length, which display conservation of segments with similar order and orientation of syntenic blocks.

The GC content of *H. werneckii* was unchanged (~53.52 %) across assemblies and consistent with the Capnopiales order[133]. In the Merged3.BC3 assembly the repeat content is 2.21% and the genome is still relatively large when compared to other related low repeat content

Capnodiales, like *Baudoinia compniacensis* (21.88 Mb, 0.8% repeat content) [133]. Also, Capnodiales with larger genomes such as *Cladosporium fulvum*, *Mycosphaerella graminicola*, and *Mycosphaerella populicola* (61.11, 39.69, 33.19 Mb respectively) all exhibit high repetitive content (44.44%, 12.26%, and 23.37% respectively) this makes *H. werneckii* an exception within the order [35]. However, these fungi are not within the same genus, and more similar genetic compositions to *H. werneckii* may exist within closely related fungi like *Hortaea acidophila* (figure 4). Formally a haploid, *H. werneckii* has the genetic redundancy roughly equivalent to a diploid organism and this has been proposed to contribute to its halotolerance [35]. It has been shown that WGD in plants (autopolyploidy) is associated with increased abiotic stress tolerance and polyploids are more commonly found in extreme environments like the autotetraploid citrus, *Brassica rapa* L. and black locust, *Tobinia pseudoacacia* [143-145]. Indeed, in stressful conditions *S. cerevisiae* can undergo chromosomal duplications resulting in aneuploidy as a “quick” means to alter transcript expression [146]. Nevertheless, chromosomal duplications (aneuploidy) result in genome destabilizing events which negatively impact the fitness of the organism overtime [146-148]. Yona *et al.* demonstrated that aneuploidy in long term evolution experiments was transient and eventually the cells revert to a haploid [146]. The recent WGD of *H. werneckii* provides perspective into gene retention. After a WGD, both gene copies are functionally redundant and if devoid of selective pressures one of the genes copies may decay in a stochastic process. Environments with extreme abiotic stresses may select for duplicate retention as seen in polyploid plants [67, 149] The recent WGD within *H. werneckii* and the new reference genome could be potentially used to study the effect of “Gene Balance Hypothesis” assessing what genes would be retained over evolutionary time (long term evolution experiments) [150, 151].

With an assembled genome, other molecular experiments can be conducted, such as MNase-seq and RNA-seq, however the Hw2HGAP3 version will be used despite the caveats aforementioned as the transcriptome annotation was constructed using this genome.

4. The transcriptome of *H. werneckii*

In response to abiotic and biotic stresses, cells modulate gene expression to withstand the influence of the environment and for cell specification. This is termed transcriptomics and the major aims include the cataloguing the species of transcripts including small RNAs, non-coding RNAs, and mRNAs and revealing the molecular mechanisms behind cell change[152]. Using next-generation sequencing (NGS), researchers have been able to quantify mRNAs coming from a cell population while also discovering novel transcripts which may have important roles in determining the fate of a cell. This is accomplished using a method called RNA-seq. The first step in RNA-seq is the isolation of RNA from cells and then the enrichment for mRNA via 3' polyadenylated enrichment. mRNA only accounts for 1-2% of the cell's total RNA, with the bulk being rRNA. Isolated mRNAs are fragmented and reverse transcribed into cDNA. The cDNA then has adapters ligated onto its ends and amplified using PCR [152, 153]. The corresponding collection of cDNA with adapters is termed a library.

To date, no total RNA sequencing has been done for this organism, and the gene prediction previously described was done *in silico* without transcriptional evidence. Using RNA-seq enriched for messenger RNA, a *de novo* assembly had been constructed and from this, differential expression analysis was performed between different conditions to ascertain the transcriptional response to salinity. A *de novo* assembly should be done for each condition, as novel transcripts upregulated within a treatment may not be present within a reference transcriptome or annotation. By way of example, Ramos et al (2014,2016) looked at the developmental transcriptional differences between wild and domesticated grapevine flowers[154, 155]. In their first study they used a general annotation for mapping their RNA-seq reads and noticed a large portion of reads mapping to intergenic regions not previously annotated[154]. Subsequently in 2016, using the same data set they constructed a *de novo* transcriptome for each condition (flower type). They identified novel *loci* not within the previous annotation which are thought to confer to the differences in flower development[155]. This highlights the importance of constructing a *de novo* assembly for each condition in an RNA-seq experiment. Although assembling a transcriptome is possible to do on desktop computer the annotation of these transcripts requires extensive

computational power as each transcript is six frame translated and individually pair-wised searched within vast data-bases for homology. However in the last two years, new algorithms using pseudoalignment, have allowed transcripts to be rapidly quantified [119, 156].

4.1. Insights from a transcriptome assembly

RNA-seq was used to examine gene expression changes associated with chronic exposure to salt stress in *H. werneckii*. Three conditions were selected, 0%, 10% and 20%. This range of salt encompasses the broad optimum defined for *H. werneckii*. One sample was used for construction of the transcriptome, as previously mentioned in the future *de novo* assembly should be done for all conditions to ensure that all transcripts are accounted for. The RNA-seq assembly was performed using Trinity[117] and assembled transcripts annotated using MAKER and the Hw2HGAP3 genome.

The annotation showed that 66.89 % of the genome is covered by genic regions, 10 % higher than the average of the Capnodiales genomes evaluated by Ohm[133] (Table 3). The average GC content of the coding regions was similar to other Capnodiales, with *Baudoinia compniacensis* the most similar at 55.8 %. *H. werneckii* also had more exons per gene than any of the other Capnodiales at 2.5, and even with the mRNA-seq based MAKER annotation the gene count of 16,142 was higher in *H. werneckii* [133]. The number of genes with a functional KOG annotation (eukaryotic clusters of orthologous groups) was relatively low compared with other Capnodiales which had an average of 60% genes annotated [157]. This suggests that some of the genes 1) may have novel functions, 2) are too diverged to be annotated or 3) are a misassembly artefact. However, using BUSCO analysis, the transcriptome was estimated to be 96% complete, and 33% duplicated. This result is contrary to the highly-duplicated number of genes predicted within the genome using BUSCO, however this transcriptome is only from one condition and likely a subset of the true gene number. Moreover, 61.65 % of the genes had a Gene Ontology (GO) annotation from the GO universe. In terms of gene loci, 1,806 (903 pairs) of the predicted genes overlapped each other by 50% with the direction of transcription being the opposite direction.

This could be a transcriptome assembly error because within 82.6 % of the pairs, one gene had no annotated 5'UTR or 3'UTR, while 154 pairs both had a missing 5'UTR and 3'UTR annotation. A total of 8510 genes had annotated 5'UTRs, 8,206 had annotated 3'UTRs and 7,226 had both (Table 3). These regions are important for elucidating the position of transcription start sites (TSS) and transcription termination sites (TTS) of transcripts[158]. Regulatory regions fall upstream of the TSS, therefore knowing their location can aid in promoter identification and subsequent motif enrichment analysis. The annotation also revealed novel insights into the life strategy of *H. werneckii*, as some gene annotations have multiple associated transcripts (more than the expected two from a genome duplication) such as *DAK1*, *STL1* and *ZRT1* (discussed later).

Table 3. Transcriptome assembly statistics of from the Trinity assembler of *H. werneckii*. Compared with *in silico* gene predictions of Capnodiales from genome assemblies.

Transcriptome Metrics	<i>Hortaea werneckii</i>	Capnodiales Average
Predicted gene count	16,142	11,605
mRNA count	15,987	-
Exon count	40,528	-
Exons per mRNA (mean)	2.5	2
rRNAs	2	-
tRNAs	1	-
tRNAs	153	-
3' UTRs	8,206	-
5' UTRs	8,510	-
Intron count	24,541	-
Median intron size (bp)	119.1	-
Gene sum length (bp)	34,922,966	-
Total genome covered (%)	66.89	-
GC content of coding DNA (%)	55.9	54.7
Genes with KOG annotation	6,333 (39.23 %)	6,885 (59.97%)
Genes with GO annotation	9,952 (61.65 %)	-

4.2. Analyzing abundance of gene expression

Expression analysis was done using Kallisto, a program which quantifies abundance of transcripts from RNA-seq and based on the pseudoalignment algorithm. This algorithm skips read mapping, greatly reducing run times, while also being extremely accurate when compared to read

mapping based RNA-seq abundance[156] [119]. The read counts are then read into Sleuth, a program that performs differential expression analysis on Kallisto's output. Sleuth and Kallisto use the normalized metric, Transcripts Per Kilobase Million (TPM), to estimate the abundance of a transcript within a sample. These two programs are relatively new, but pseudoalignment of RNA-seq reads is used, due to the ease of use, efficiency and the accuracy of the quantification[156]. Sleuth uses Walds testing to estimate abundance change, resulting in a beta value that is analogous to log2FC but as a bias estimator taking into account covariates applied [120] [159].

The 300 highest and lowest (TPM > 1.0)expressed genes in the reference condition of 10% NaCl (3 biological replicates) were analyzed for any biological process enrichment with BiNGO[122] using classical hypergeometric testing[123]. The most highly expressed genes in *H. werneckii* were associated with GO terms 1) ribosomal genes, 2) the ribonucleoprotein complex, 3) gene expression and 4) cellular biosynthetic process. Low abundance genes were enriched for transporter activity. Interestingly, 14 % of all genes had the GO term "transport" which is an increase from the 9.6 % in the previous GO annotation of *H. werneckii* [35].

To analyze differential expression between the salinities, the Wald test from the Sleuth package was used for all pair-wise combinations. Although 10% NaCl is the reference condition, a comparison between 0% NaCl and 20%NaCl was also done to see if there are any changes between 0% and a high salinity condition. It is important to note, that samples ran on the MiSeq had substantially lower coverage for transcripts then other conditions (all of the 0% NaCl samples and one 10% NaCl sample) resulting in the absence of counts for approximately 3000 transcripts. Adequate coverage is a crucial aspect for RNA-seq to see subtle changes between treatment and control, while also being important for normalization measures. This is because low coverage influences the relative abundance measurement of transcripts within the sample. As this influences RNA-seq analysis, another covariate, the batch of the sample, was included together with condition for the generalized linear model in Sleuth. Despite the lower coverage, there is significant differential expression between the conditions and the 10% NaCl compared with the 20% NaCl had the most statistical power due to the sheer amount of reads from each sample.

Differentially expressed transcripts were filtered for significance ($q < 1e-3$, $b > 0.86$) and a these were used further analysis (Table 4).

Table 4. Summary numbers of up and down regulated genes in each condition relative to arbitrarily set reference condition.

Reference Condition	Treatment	Up Regulated	Down Regulated
10% NaCl	0% NaCl	274	281
10% NaCl	20% NaCl	513	574
20% NaCl	0% NaCl	250	308

4.3. Differential expression of *H. werneckii* grown in 0% NaCl versus 10% NaCl media

These conditions (0% NaCl and 10% NaCl) both represent optimal conditions in terms of growth rate (Figure 4) when compared to the 20% NaCl condition. Additionally, the 10% had a longer lag phase but when the cells adjusted to the new medium their growth rate was highest. This is due to the changes in pH of the medium, when the cells are growing, the media become more acidic with 20% having the largest overall decrease in pH when measured. The cells are brought to mid exponential then back diluted in fresh media and although the salinity remains the same, the pH is back to neutral (7.0). Cells were harvested at mid exponential phase are considered adapted cells and not under hyperosmotic shock but under a continuous stress.

The differential expression analysis between conditions revealed many up and down regulated genes (Figure 10). In the 0% NaCl condition many of the top upregulated genes had the annotation “Protein of unknown function” and fell within the 3’ end of other genes within the genome. Of the top 6 unknown proteins, three of them did not fall within another gene and were analyzed for any biological function using blastx[160]. Two of the most upregulated genes are similar to gel1, a 1,3-beta-glucanosyltransferase, a protein involved in fungal cell wall maintenance(Figure 11)[161]. Beta-1,3-glucan is the main component of the fungal cell wall and gel1 is involved in the polymer elongation.[162]. The up regulation of this transcript likely represents changes in cell wall composition. Moreover, gel1 in 20% NaCl is also upregulated

points $q < (0.001)$ and $\text{abs}(b) > 1$ Labels were applied only to values with $q < 1e-3$ and $\text{abs}(b) > 2$. Significant points without labels did not have a known protein annotation (Protein Unknown).

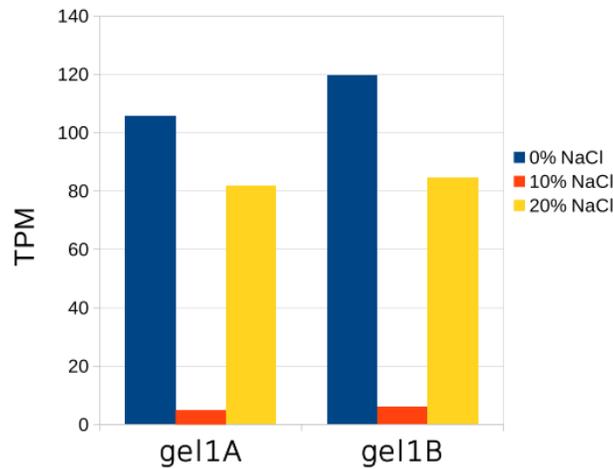


Figure 11 Normalized transcripts per kilobase million (TPM) of 1,3-beta-glucanosyltransferase (Gel1) like transcripts in *H. werneckii* across a range of salinities. Blue is TPM in 0% NaCl, red in 10% NaCl and yellow in 20% NaCl.

For the 10% NaCl condition the top GO biological process terms were respiration and oxidation reduction. The 10% NaCl also had increase of genes associated with oxidative stress and peroxidase activity. When exposed to salt stress *H. werneckii* has been shown to increase respiration providing the necessary potential energy for secondary active transport pumps. Moreover, genes with the GO term peroxidase activity were also enriched in the 10% NaCl subset. These pathways are known to interact in ROS detoxification generated from respiration, and therefore they may contribute to the upper salinity range of *H. werneckii* as proposed by Petrovic[45] in which *H. werneckii* was exposed to high salinity (25%) and H_2O_2 . The oxidative stress response system is known to be induced under conditions of salt stress in fungi and plants [20, 163, 164]. In the 10% NaCl condition genes associated with these processes were upregulated suggesting a role in salt tolerance as this concentration is lethal to most microbes but within *H. werneckii*'s natural broad range optimum[165]. Two of the most upregulated transcripts in the 10% NaCl condition relative to 0% NaCl, were apoptosis inducing factor 1 (*aif1* from *S. pombe*) ($b > 1.20$, $q < 1e-3$) (Figure 10). The same genes were also significantly upregulated relative to 20% NaCl ($b > 1.42$, $q < 1e-66$). To check if this annotation was accurate, the gene was translated and searched within the NCBI non-redundant protein database[160]. The best matches were similar to

rubredoxin-nad (+) reductase, an iron-sulfur protein ($e < 1e-10$), which is involved in alkane degradation in bacteria. This suggests a difference in energy-producing systems going on at 10% NaCl as Petrovic suggested. The observation was based on the low CO₂ production from *H. wenreckii*'s at its salt optimum (17%) relative to other salinities[41]. The difference between 20% NaCl and 10% NaCl was more substantial, consistent with the observed changes of specific cellular adaptations around 17% NaCl where CO₂ production was the lowest and increased at 25%[165].

4.4. The gene expression profile of high salt relative to optimum growth conditions

Although the upper salinity range of *H. wenreckii* was not included in this study, 20% NaCl represents a valuable test resource for halotolerant genes as this salinity is toxic for most eukaryotes[165]. These two salinities had high read counts per sample and thus very strong differential expression statistical power resulting in higher q values (Figure 12A). In 20% the top three most upregulated fell within the 3' end of another gene and had no annotated TSS or TTS. This suggests that 1) there is a high prevalence of anti-sense transcripts, 2) degradation or 3) that these are real gene transcripts encoding proteins similar to their annotation. Two of these genes had very high beta values in the high salt condition ($b > 7.9$, $q < 1e-135$). These genes were annotated as *erg26*, sterol-4-alpha-carboxylate 3-dehydrogenase and *pvdA* L-orthithine 5-monooxygenase. Both were searched within the NCBI database to verify annotations (and both had accurate annotations). Moreover, *ERG26* is an essential gene required for ergosterol synthesis in fungi. Repressible mutants in yeast have abnormal mitochondrial morphology [166].

driving the expression of zinc transporters or it may be involved in coping with oxidative stress from respiration. Zinc deficiency is known to be associated with oxidative stress in yeast and humans and may have a role in conferring salt tolerance by affecting these pathways[48, 167]. The ZRT1 annotated genes had the highest expression levels than any of the other zinc transporters and these transporters have the highest affinity for zinc. The increase of vacuole like zinc transporter (Zrt3) expression relative to the other conditions suggests that the cell mobilizes zinc from its storage organelles. The expression of these transporters is controlled by transcription factor Zap1p which induces its own expression under zinc deficiency via autoregulation. *H. werneckii* has two ZAP1 like transcripts that are upregulated and induced in high salinity (Figure 12). The high expression and prevalence of these transporters and their transcription factor suggest a predominant role for zinc in the biological response to salt stress[168]. Also, zinc is an essential nutrient and required as a structural or catalytic cofactor by many proteins within eukaryotes[48].

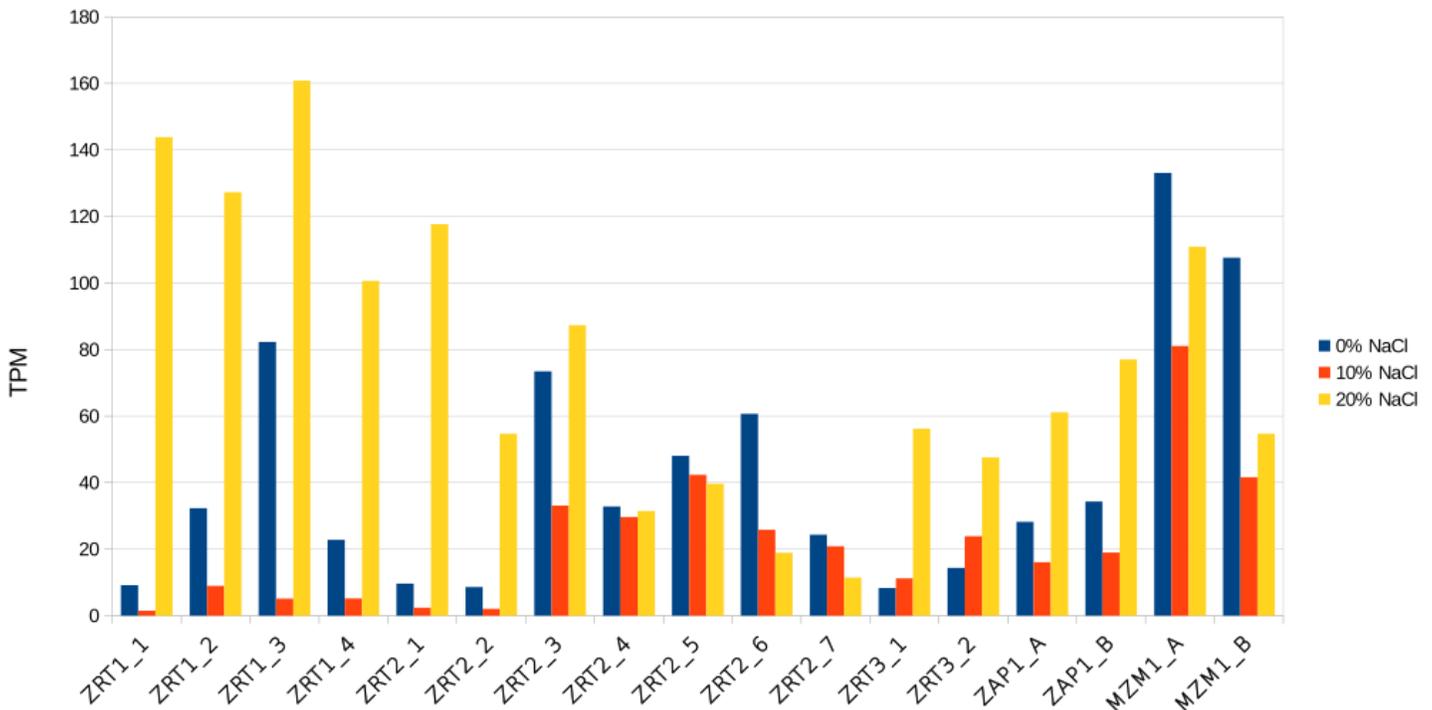


Figure 13. Normalized TPM of zinc transporters (Zrt1, Zrt2, Zrt3), zinc deficiency transcription factor (Zap1) and mitochondrial zinc maintenance (Mzm1) like transcripts in the salinity range of 0%NaCl to 20% NaCl.

A number of mechanisms have been proposed for zinc deficiency and the association of oxidative stress. Two ways it can influence oxidative stress is via the antioxidant Cu/Zn superoxide dismutase (SodC) which requires Zn^{+2} to catalyse the conversion of superoxide to H_2O_2 [169]. Zn^{+2} cofactor proteins are prevalent in mitochondria such as alcohol dehydrogenase and major components of the electron transport chain[48]. In yeast, zinc deficiency also induces a principal cytosolic peroxiredoxin [170], *TSA1* via Zap1p. Tsa1 is responsible for the degradation of H_2O_2 to H_2O and acts synergistically with cytosolic SodC Cu/Zn ($O_2^- \rightarrow H_2O_2$). This particular SOD had a single annotation within the transcriptome. However in the genome it had two hits with one less statistically significant. Nonetheless, the transcripts for Tsa1 (*tpx1*), were not induced in 20% NaCl and had very low expression overall, suggesting the increase of zinc transporters are not involved with Tsa1 (Tpx). Since there are other antioxidant proteins involved in the degradation of ROS, such as glutathione peroxidases and thioredoxin-dependent peroxiredoxins, it may be that a different catalase is responsible for cytosolic detoxification of H_2O_2 . Moreover, MZM1 a mitochondrial zinc maintenance protein responsible for establishing the zinc pool in the mitochondria was also upregulated in 20% relative to 10% ($b > 0.27$, $q < 1e-3$) (Figure 13) [171].

The increase of Zinc transporters coincides with increase of mitochondrial associated respiratory genes in *H. werneckii* and this is the predominant form for energy production in high salt concentrations. Supporting this is the upregulation of two transcripts similar to mitochondrial superoxide dismutase (SodA [Mn^{2+}]) ($b > 3.8$, $q < 1e-16$) (Figure 12B, Figure 14A). These transcripts were only highly transcribed in 20% NaCl and use Mn^{2+} as a cofactor. Interestingly, addition of $MnSO_4$ to sodium chloride salt stressed rice seedlings alleviated their salt tolerance [20]. This fits a model where salt induces a major increase of aerobic respiration and the detoxification of resulting ROS is a key determinant of salinity stress. Moreover, transcripts similar to PRX1, a mitochondrial peroxiredoxin, had higher expression in 20% NaCl whereas peroxisomal catalases like POX9 were down regulated (Figure 14A). Taken together, this suggests a shift in energy metabolism to the mitochondria in high salt.

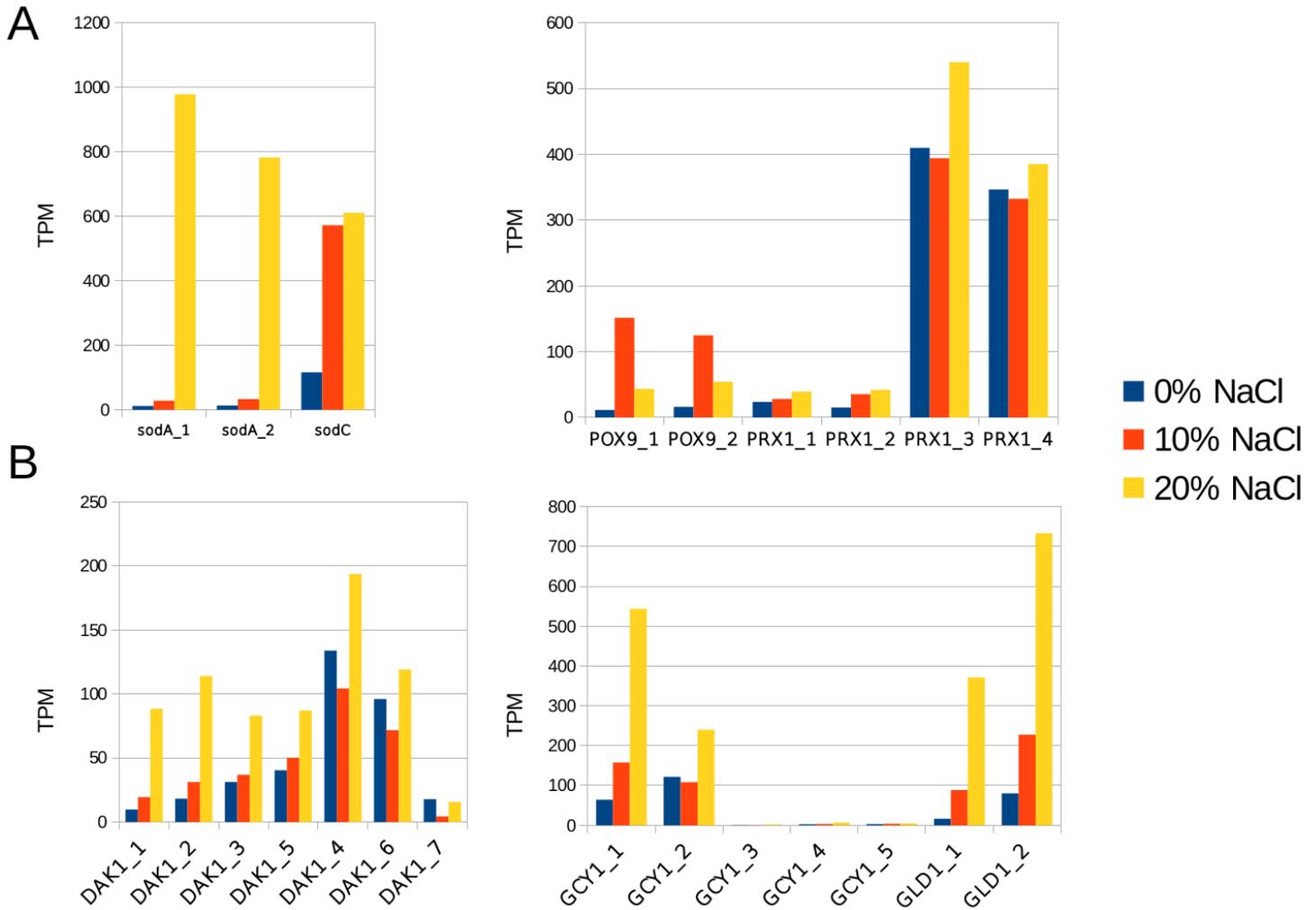


Figure 14. Normalized transcriptional abundance (TPM) of gene families involved in carbon metabolism and oxidative stress. A) Superoxide dismutase [Mn] like transcripts and peroxisomal like transcriptional abundance in different salinities. B) Dihydroxyacetone kinase and glycerol dehydrogenases transcriptional abundance in 0%NaCl through 20% NaCl.

Seven dihydroxyacetone kinase like transcripts were annotated in the transcriptome. All of the transcripts were upregulated in 20 % NaCl with three highly upregulated ($b > 1.2$, $q < 1e-8$) and the other four upregulated though less pronounced ($b > 0.5$, $q < 1e-4$) (Figure 14B). *Dak1* and *Dak2* are involved in the metabolism of glycerol, the main compatible solute produce in *H. werneckii*. They catalyze the phosphorylation of both dihydroxyacetone (DHA) and glyceraldehyde and in glycerol assimilation they are the last step in the two part conversion of glycerol to dihydroxyacetone phosphate [172]. In *S. cerevisiae*, glycerol is first oxidized to DHA via *Gyc1* (glycerol 2-dehydrogenase) but in *S. pombe* this role is performed by the enzyme *Gld1* [173]. The glycerol dehydrogenase of *S. pombe* has a zinc catalytic subunit and this is not seen

in *S. cerevisiae* or *H. werneckii*. Interestingly two transcripts similar to GLD1 and five transcripts like GYC1 were upregulated in high salinity (Figure 14B). However, when these transcripts were searched in the BLAST database the zinc catalytic unit could not be accounted for but still had significant homology to other glycerol hydrogenases. Likely, *S. pombe* Gld1 has undergone extensive evolutionary divergence and has alternate proteins for glycerol metabolism[173]. The increase of glycerol assimilation pathway protein transcripts is contrary to what would be expected in high salt concentrations. However, in *S. cerevisiae* Dak1 has been shown to increase 4-fold during salt stress growth(1.4M) while also having an increase of transcripts encoding Gcy1. The authors suggested this process is the result of metabolic overflow path, fine tuning the amount of intracellular glycerol[174, 175]. It is interesting to observe this level of glycerol metabolism in *H. werneckii* in response to high salt (Figure 14B). This could be due to increased energy demands of growing in high salt and glycerol is acting as an additional energy and carbon source or because glycerol is being shunted into other pathways necessary for growth at this salinity. Regardless of glycerol's fate within the cell, genes associated with its synthesis and assimilation, are upregulated in salt stress and therefore glycerol has a crucial role within the cell.

Stl1 is a high affinity glycerol transporter and involved in glycerol reuptake from the environment[176, 177]. Shockingly, 49 transcripts were annotated as Stl1 transporters. However, in the 20% NaCl only two STL1 transporters had high expression relative to the other conditions despite Stl1 known to have a role in salt stress[178]. The STL1 transcripts were aligned using ClustalW[179, 180]. Three of the pairs had substantial increase of expression depending on salinity. This suggests that a set of transporters are upregulated in response to salt stress and that they each may have specialized structure and function.

4.5. Comparison of high salt (20% NaCl) and no salt (0% NaCl) gene expression in *H. werneckii*

Gene expression in 20% NaCl was compared to 0%. The 10% NaCl dataset was included to construct the generalized linear model for Sleuth because the 20% contained no MiSeq data.

Highly differentially expressed genes should still have low q-values and this data-set can still be used to understand mechanisms of salt tolerance (Figure 15A). This comparison is important to scrutinize shared up/down regulated genes of 10% NaCl and 20% NaCl which aid in understanding mechanisms of salt tolerance. The top upregulated pathways of 20% NaCl relative to 0% NaCl were sulfur metabolic processes and superoxide metabolic processes. This is in line with the high expression of SodA observed in 20% NaCl relative to other salinities. Transcripts similar to tryihydroxynapthalene reductase, TRH1, were also upregulated in response to high salinity and moderate salinity. This is a crucial gene in DHN-melanin synthesis and corresponds to the dark pigmentation observed in the higher salinities (Figure 15B)[52]. It is also worth mentioning that in 20% NaCl there was also up regulation of Pmp20, a peroxiredoxin pump involved in detoxification of ROS [181]

number of introns and exons found within the genome[183, 184]. In *S. cerevisiae*, very few genes have introns (~4.7%) whereas in *H. werneckii* almost all of the genes have introns[185]. Additionally, RNA-seq can be combined with MNase-seq to explore changes of both chromatin and gene expression in response to environmental stress [84, 98]. Over the last decade many advances have been made in regards to understanding this biological process and some challenge the current chromatin paradigms. One of the major molecular questions surrounding transcriptional dynamics is how and to what extent, does chromatin influence transcription. The major findings of recent studies are that this regulation is more complex and no single model fits all the data[77, 84].

5. Evaluating the chromatin landscape of *H. werneckii*

5.1. The influence of chromatin and nucleosomes

Nucleosome positioning maps can be generated using the enzyme micrococcal nuclease which preferentially cleaves linker DNA in between nucleosomes resulting in DNA bound to nucleosomes[186]. The nucleosome-bound DNA is isolated and then sequenced using NGS to infer nucleosome locations within the genome within a population. Large datasets of nucleosome maps in model organisms can address questions surrounding transcriptional regulation on the level of chromatin[187]. Nucleosome maps for non-model organisms are less common, but as the cost of sequencing goes down they are becoming more available. These maps facilitate understanding of *cis* regulatory elements of DNA sequence that disfavor nucleosome formation and are important because *H. werneckii* has relatively higher GC content versus the commonly studied baker and fission yeast (36% and 38% respectfully)[85, 186, 188]. Therefore, examining the nucleosome configuration in *H. werneckii* can address many questions such as what is the typical nucleosome architecture surrounding TSS and the TTS and does it differ in response to environmental salt differences on global and local scale? Is nucleosome occupancy correlated with gene expression? What type of genes display stereotypical NDRs? Are there shared *cis* determinants between yeast species? This portion of the study aims to address these questions in *H. werneckii*.

5.2. Inferring nucleosome positions with MNase-seq

MNase-seq was performed for three salt conditions with three biological replicates. It is important to note that MNase-seq data sets are inherently noisy due to experimental variation of cross-linking, digestion, isolation and the *in vivo* dynamics of nucleosomes [90, 103, 189]. To reduce the potential bias reported with gel isolation, nucleosome fragments with ~80% mononucleosomes were directly used for library preparation without gel isolation [103, 189, 190]. Libraries were then sequenced on the Illumina HiSeq and the reads aligned to the Hw2HGAP3 genome using BWA[112]. This assembly was used because it has the TSS annotation in which RNA-seq was used despite the caveats associated with it aforementioned. DANPOS2 was then used to calculate nucleosome positions and occupancy for 8394 genes with a defined TSS (some genes were split by ends of contigs and not used)[126].

5.3. *H. werneckii* genome *cis* determinants of nucleosome positioning

To uncover the *cis* genetic determinants of nucleosome positioning in *H. werneckii*, the 10% NaCl data set (three biological replicates) was used to find well positioned nucleosomes. Relative nucleotide frequency plots were then constructed to examine the dinucleotide and polynucleotide frequencies with respect to well positioned nucleosome dyads.

DANPOS2 predicted 273,256 nucleosomes covering 80.33 % of *H. werneckii*'s genome with an average summit to summit distance of 165 bp (n= 263,228) and a linker region of 18 bp similar to other yeast species [191] (Figure 16A). DANPOS2 also assigns nucleosomes with a "fuzziness" metric which measures the extent of nucleosome positioning, with well positioned nucleosomes having low fuzziness scores [126]. The positions are based on nucleosome occupancy summit scores and broadly correspond to the location of the nucleosome dyad. DNA sequences within 500 bp of the summit (at the centre) were analyzed for nucleotide frequencies. Linker regions were enriched for poly(dA:dT) (as seen in other fungi), however they were also enriched for poly(dG:dC) tracts which has not been described in other fungi (Figure 16A). Poly(dG:dC) tracts are rare in fission and baker's yeasts and no depletion of nucleosomes has been observed at this level [188]. However, *in vitro* studies have shown that poly(dG:dC) tracts are inhibitory to nucleosome formation, which could explain enrichment within these regions, considering the high GC content of *H. werneckii*'s genome (54%) [192]. These regions may reflect evolutionary divergence in terms of sequence between *S. pombe*, *S. cerevisiae* and the Capnodiales.

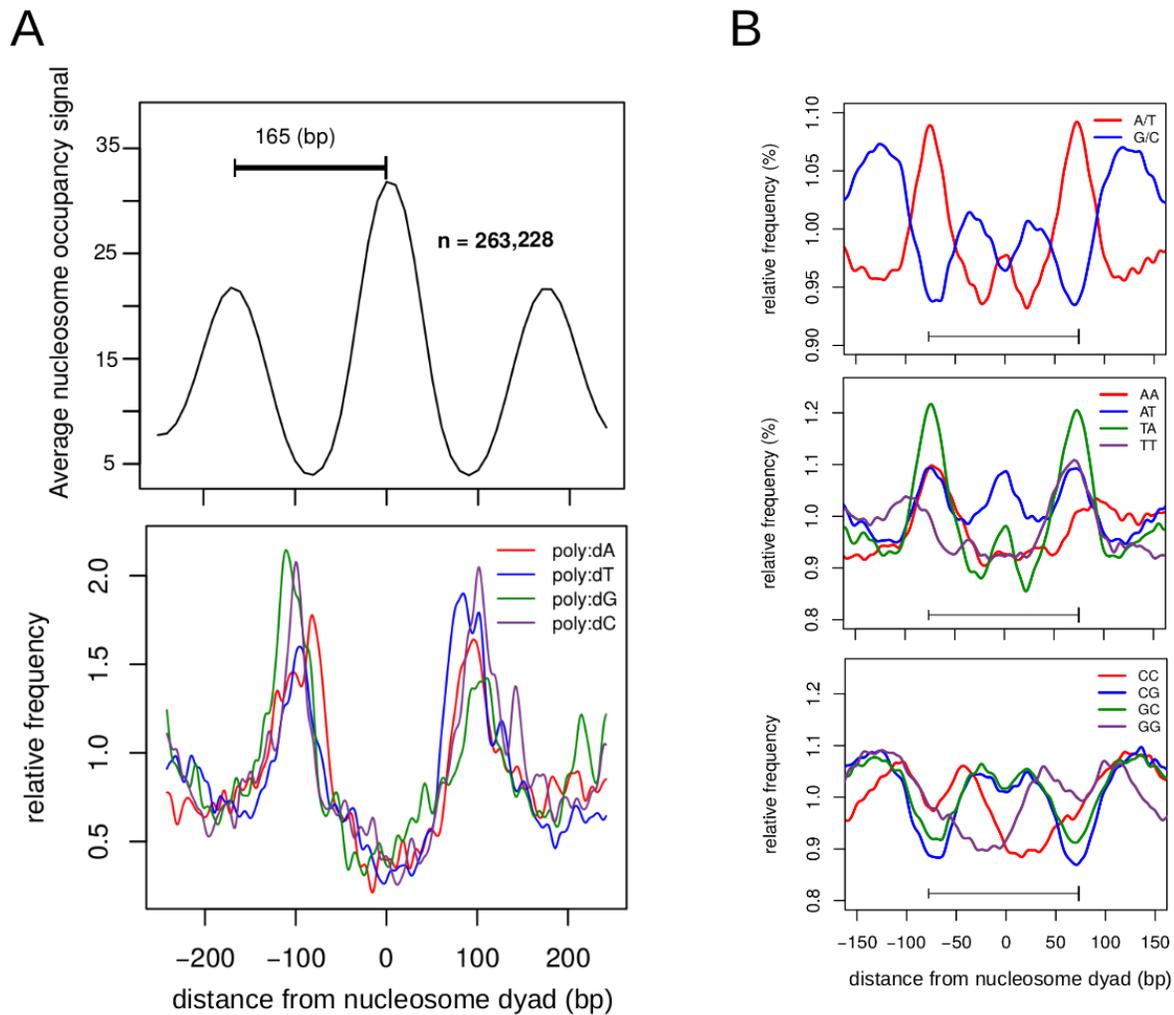


Figure 16. Sequence (*cis*) composition of well positioned nucleosomes in *H. werneckii*. A) Aligned nucleosomes and their occupancy profile with respect to their summit (dyad). The average summit to summit length is 165 bp. Poly nucleotide enrichment of the same nucleosomes within the linker regions. B) Dinucleotide enrichment of well positioned nucleosomes.

Dinucleotide frequencies were calculated from nucleosomes above fuzziness score of 50, resulting in 43,672 nucleosomes (fuzzy nucleosomes have ambiguous summit calls). MNase-seq experiments do not have the resolution to observe the 10 bp dinucleotide oscillation except if combined with extensive computation or chemical cleavage maps [188, 193]. Despite this, nucleosomes displayed a significant pattern of dinucleotide frequency (Figure 16B). Nucleosome dinucleotide patterns are known to be species specific and of interest in synthetic biology [87, 97, 194]. The linker regions of well positioned nucleosomes have increased A/T dinucleotide but likely this is a consequence of the poly(dA:dT) tracts, as GC dinucleotides are also enriched towards the

linker regions(Figure 16). On the individual dinucleotide level, the dinucleotides adopt a different pattern. The dyad is enriched for AT, then CG/GC towards the centre of the dyad followed by GG/CC. Enrichment of A/T residues at the dyad is seen within *S. pombe* and *S. cerevisiae* [193]. These patterns reflect that the maps generated are representative of nucleosome positions and the summits reflect dyads while also providing the basis for sequence affinity manipulation in *H. werneckii*. These results are important in determining the “nucleosome signature” of sequence contributing to well positioned nucleosomes in *H. werneckii* although the extent of DNA sequence impact on nucleosome positioning is highly controversial [77, 91, 188, 195-197]. Nevertheless, González et al. demonstrated that manipulation of these base pair sequences can alter the nucleosome landscape of genes and thus alter its expression within the same species [87]. As nucleosome occupancy influences transcription, nucleosomes proximal to the TSS were profiled in all salinities.

5.4. Evaluating nucleosome occupancy and positioning at the transcription start site

The genome wide occupancy profile illustrates a typical eukaryotic nucleosome phasing with a strongly positioned +1 nucleosome immediately downstream of the transcription start site (TSS), a nucleosome depleted region (NDR) upstream of the TSS and a -1 positioned nucleosome upstream of the NDR (Figure 17). Nucleosome phasing is consistent throughout the genic region but significant fuzziness is observed toward the 3' end of genes. The transcription termination site (TTS) also displays significant nucleosome depletion (as in other eukaryotes). Additionally, digested naked gDNA shows no phasing throughout the genic region but does exhibit a small increase of occupancy upstream of the TSS consistent with other studies [198]. No major nucleosome occupancy changes are seen between the salinity conditions although the 10% NaCl does display slightly higher -1 and +1 nucleosome occupancy. In other eukaryotes these regions (NDRs) are indicative of regulatory promoter regions. Taken together, this data defines the regulatory regions upstream of the TSS in *H. werneckii* which now can be explored in terms of motifs.

The nucleosome depleted regions of *H. werneckii* are enriched for dinucleotides G/C but there was no enrichment for dinucleotides A/T (data not shown). The nucleosome depleted region

upstream of these genes did not have any enrichment for poly (dA:dT) tracts, but did have a enrichment for poly(dG:dC). When all NDRs in the genome were analyzed, independent of TSS location, both poly(dA:dT) and poly (dG:dC) tracts were enriched, suggesting that other NDRs within the genome do not always share the same sequence specificity which could be related to their biological function such as enhancers or origins of replication as both of these regulatory elements are known to be depleted of nucleosomes [100]. The GC enrichment within nucleosome depleted regions is not observed within *S. cerevisiae*, but does occur in nucleosome depleted promoters of *Arabidopsis thaliana* and mammalian cells [85, 199, 200].

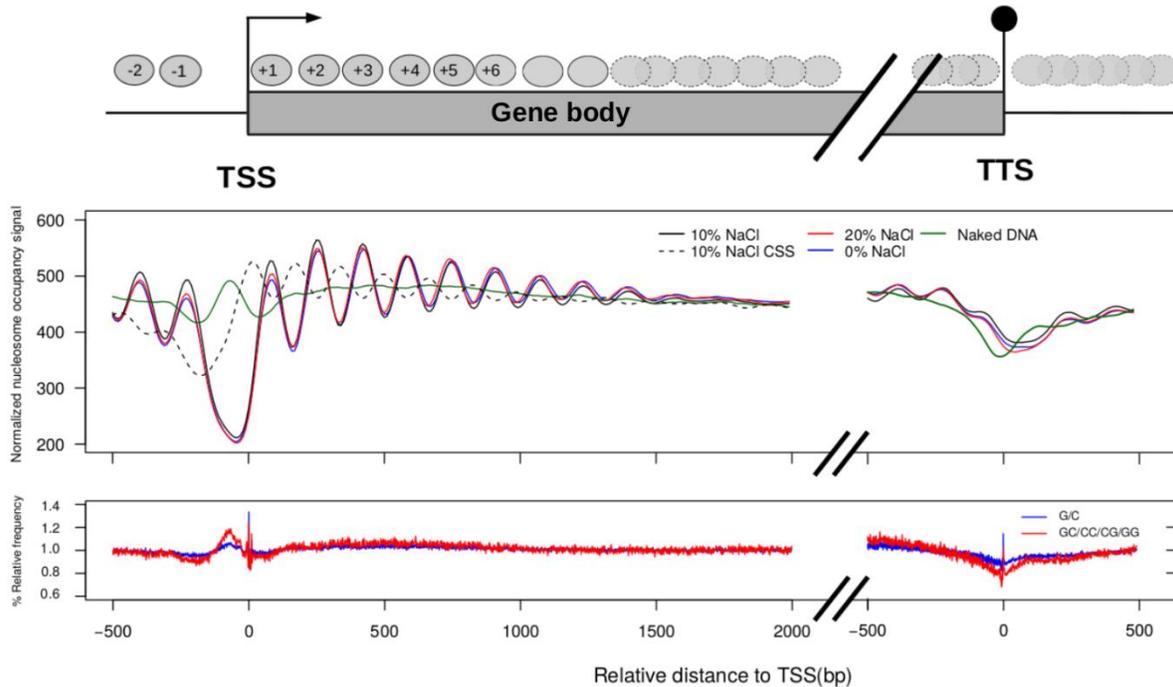


Figure 17. Nucleosome configuration surrounding the transcription start site (TSS) of annotated genes in *H. werneckii*. The grey bar represents region of the gene body. The phasogram shows oscillating nucleosomal peaks along the sequence region with a well-defined nucleosome depleted region (NDR) upstream of the transcriptional start site. Nucleosome occupancy was profiled for all conditions, and dinucleotide content was also analyzed for enrichment along the DNA region.

The prevalence of high GC content within the NDR suggests GC rich promoter motifs. To quantify this, a k-means clustering algorithm from the NucTools suite was applied to the 300 bp region upstream of the plus one nucleosome instead of the TSS, resulting in a four distinct clusters (Figure 17A)[127]. This was to ensure that the GC content within the NDR was not being

influenced by other nucleosomes because the position of the TSS varies with respect to the +1 nucleosome. The clusters show well phased nucleosomes downstream corresponding to the genic region and a predominant NDR upstream for clusters 1-3. Cluster 4 had very few NDRs upstream and nucleosomes present where an NDR would normally be and had lower nucleosome occupancy scores. The +1 nucleosome in all of the clusters had the highest occupancy which was not evident in the TSS aligned data set. The occupancy of the plus one nucleosome is a common trait within eukaryotes[77, 79, 82, 201, 202] . The individual clusters were then analyzed for dinucleotide enrichment throughout the NDR. Clusters 1-3 showed substantial increase of GC dinucleotide content (Figure 17B) which was largely driven by the dinucleotides CG and GC (data not shown). To date, this has not been described in any fungi, and likely is a result of the Capnodiales higher GC content. However, because these genes are only a subset, the predominant NDR could be majorly driven by *trans*-factors binding to specific motifs and inducing an open chromatin conformation. The high GC content within the promoter NDR is similar to higher order eukaryotes thereby making *H. werneckii* an attractive model to examine the effect of *cis* determinants of nucleosome positioning in eukaryotes[187, 195, 199, 203, 204].

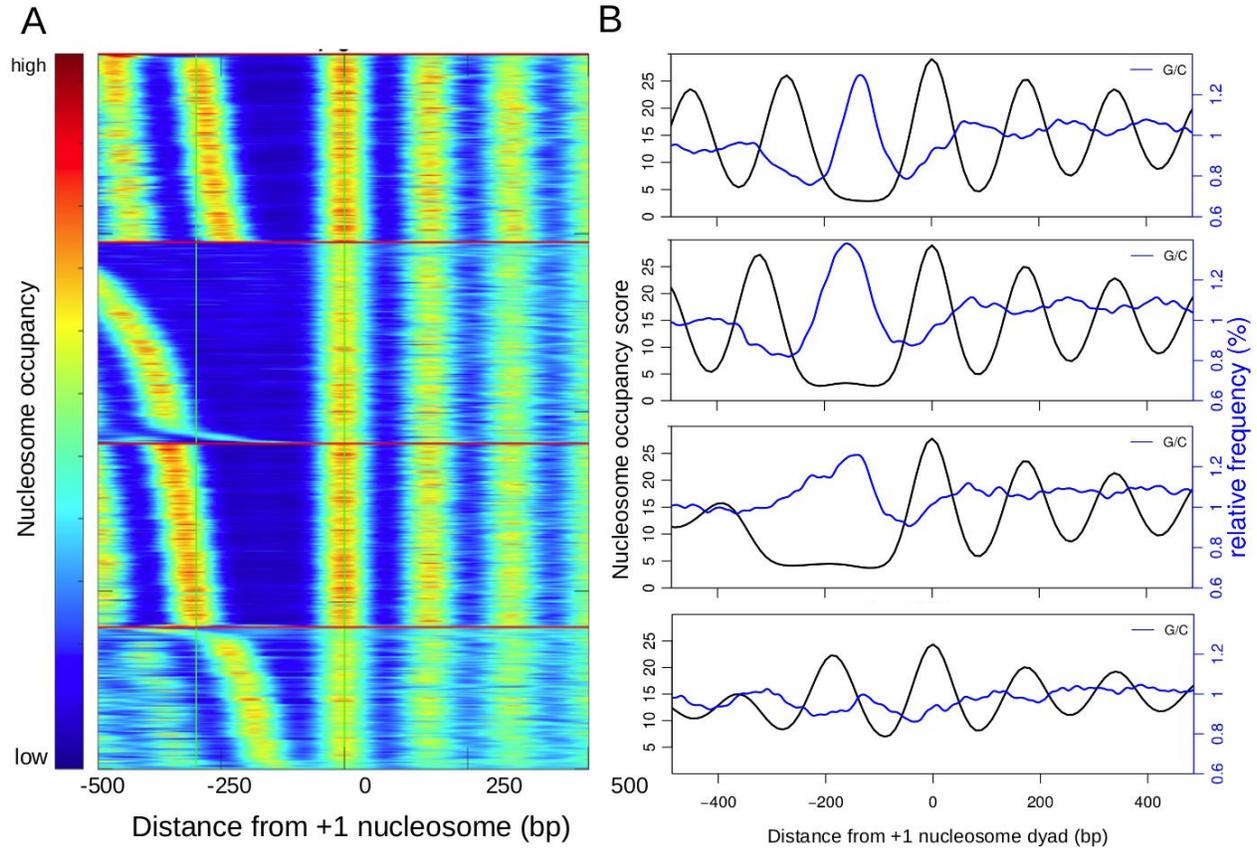


Figure 18. Nucleosome cluster profiles along the TSS. A) Nucleosome occupancy data was clustered after being normalized to a global value of 1. Blue represents low nucleosome occupancy and red represents high nucleosome occupancy. The profiles were clustered into 4 distinct clusters using a k-means clustering algorithm after constructing silhouette plots. B) Phasograms of the nucleosome clusters and the corresponding GC content enrichment (blue).

To evaluate if there are any common known motifs within the clusters, the 200 bp region upstream of the plus one nucleosome was the used for motif enrichment using the program Homer and the top three enriched motifs with a known functional yeast motif are shown as positional weight matrices [205] (Figure 19). GO enrichment of the cluster genes using BiNGO [122] was performed to query any biological relationship between promoter and predicted function. It is important to note that yeast motif matches are only best matches to known motifs, and these *de novo* motifs may represent binding sites for transcription factors (Figure 19). Many of the motifs were shared among the NDR clusters and this would be expected considering degree of GC content, and there were some *de novo* motifs with poor match pair to known yeast motifs but nonetheless exist. These included motifs for transcription factors Mig2, Swi6, and Stp3. The Mig2 motif is likely a false positive as it binds a poly(dG:dC) motif which are common sequence

throughout *H. werneckii*'s genome. Swi6 is associated with growth phase transition and associated with NDRs of *S. pombe* of growth genes[206, 207].

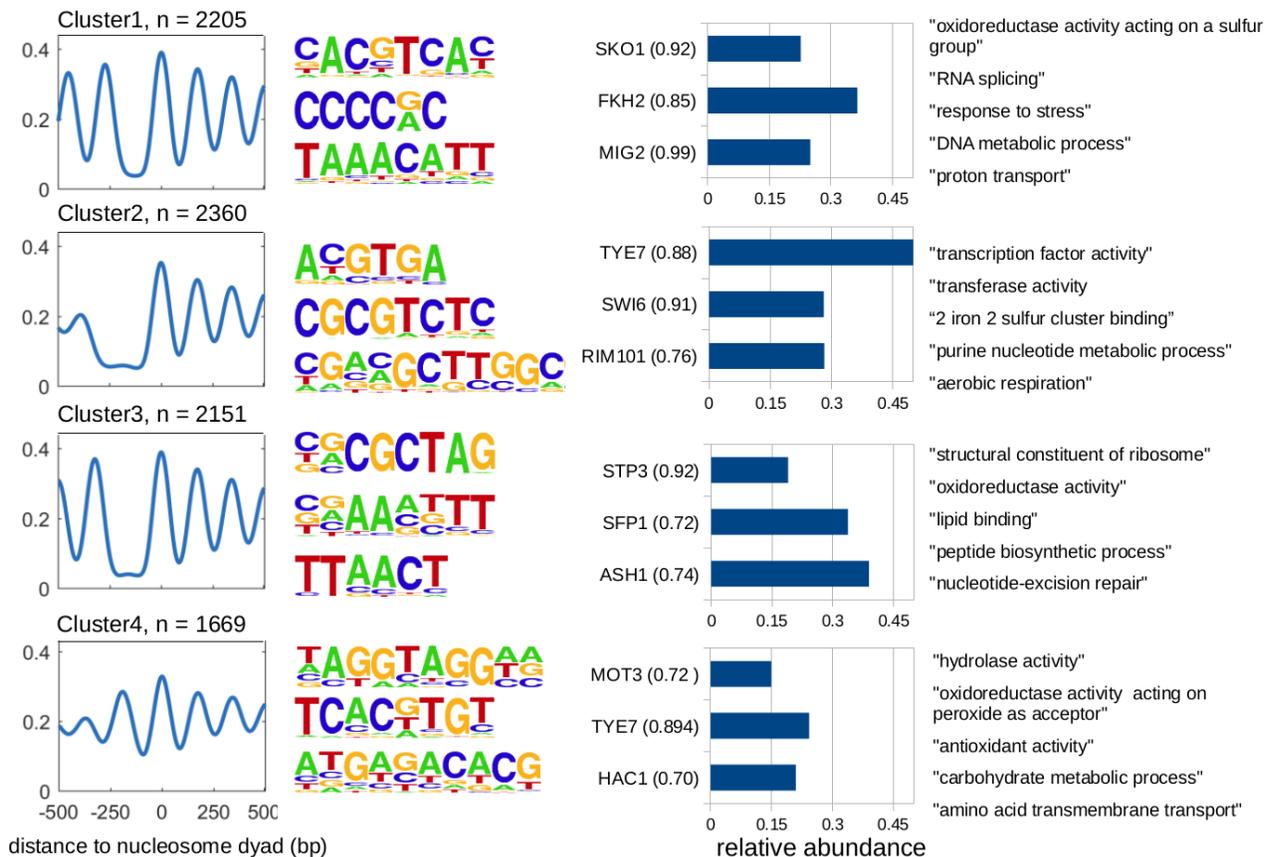


Figure 19. Cluster *de novo* motifs upstream of the TSS and gene ontology (GO) enrichment. The phasograms represent the nucleosome profile of the cluster. The positional weighted matrices were constructed from *de novo* motif enrichment. The gene names correspond with the closest yeast transcription factor with the same binding site and represented as relative abundance within its respective cluster. The last column are the predominant GO enrichments for each cluster.

The clusters in combination with the motif analysis were analyzed for GO term enrichment. Cluster 1 genes were associated with the GO term "response to stress", which is surprising given that "stress" genes in yeast usually have a closed promoter conformation [191](Figure 19). This could mean that an *H. werneckii* may have evolved open promoters at certain genes allowing rapid adaptation to fluctuating environments common in its natural niche[128]. However, such a conclusion would need to be addressed on a single gene basis level between related fungi with different life strategies. Cluster 2 also exhibited GO enrichment for "DNA-repair" which coincides with motif enrichment similar to Fkh2, a motif unique to this cluster. Fkh2 is involved in variety of

cellular functions including DNA-repair and stress resistance, both of which are GO terms enriched in Cluster 2 [208]. Therefore, some of these motifs reflect the associated enriched biological processes.

Differences of nucleosome depletion have been coordinated to lifestyle strategies within two lineages of hemiascomycota yeast. Yeasts closely related to *S cerevisiae* commonly undergo the Crabtree effect and genes associated with mitochondrial respiration have NDRs occluded by nucleosomes[191, 209]. Whereas yeasts that have diverged before the ancient WGD of *S. cerevisiae* preferably undergo aerobic respiration in the presence of high sugar and genes associated with respiration are therefore depleted of nucleosomes[191]. Therefore, the pervasive and wide NDR observed in Cluster 2 may reflect the *H. werneckii*'s extensive aerobic requirements as this cluster was the only one associated with GO term "aerobic respiration". Cluster 4 had very few *de novo* motifs with greater than 15% representation and had very low p-values relative to the other clusters with and very few motifs (Figure 19). The closed conformation of the cluster may be reminiscent of "stress"-like genes as the GO term "antioxidant" was enriched. Tye7 is a motif associated with the adenylic stress response in *S. cerevisiae*, however this motif was also found in Cluster 2 [210]. After careful examination of the Tye7 motif, it was apparent that Cluster 2 was a false positive as it had one less nucleotide match than in Cluster 4. This illustrates the importance of examining motif results more closely (as recommend by Homer's authors) because even though some of these motifs may associate with similar biological process reported in yeast, their regulatory function in *H. werneckii* remains to be determined [136].

Taken together, the data-set produced using MNase-seq and RNA-seq reveal novel insights into functional elements of *H. werneckii*'s genome and could potentially be used for future molecular manipulations. It should be noted that motif function derived from association with known motifs, should always be taken with a grain of salt because motifs differ from species to species. This is especially true for *H. werneckii* where the genomic DNA sequence and associated composition of regulatory regions are extremely different from model yeasts. However, motif enrichment analysis can provide useful information in inferring common motifs within promoter elements of non-model organisms because the enrichment is *de novo*. This deep characterization

of the regulatory regions leads to the question “does salinity affect chromatin structure in regulatory regions of halotolerant fungi?”

6. The relationship of gene expression and chromatin states across salt conditions

6.1. Steady transcription and chromatin states, growing in high salt.

The halotolerance of *H. werneckii* presents an interesting model for extreme environmental adaptation at both the chromatin and transcriptional level. Halotolerance confers salt tolerance, but also allows organisms to thrive in conditions where obligate halophiles have reduced fitness such as hypo-osmotic environments. This trait allows a broader comparison between eukaryotes, because this condition is detrimental (hypo-osmotic stress) to halophiles while close to optimum in *H. werneckii*. Stress shocks in *S. cerevisiae* have been molecularly characterized and the influence of chromatin is beginning to be uncovered [101, 211]. Hyperosmotic shock in yeast induces three major phases, shock, induction and recovery [82, 211, 212]. Chromatin remodelers are more involved in gene induction and repression during the shock phase rather than in steady state cells and this chromatin remodeling is associated with transcriptional changes [213-215]. The exact mechanisms driving this are not yet clear [216]. Generally, nucleosomes present within the 5' NDR are repressive to gene expression but not all chromatin remodeling is associated with expression changes [214]. A more complex paradigm is emerging because nucleosomes within TSS sites have also been shown to induce transcription [217]. Here, acclimated cells, grown across a range of extreme salt conditions are examined for changes on the chromatin and transcriptional level. The shock associated transient transcriptional changes at gene loci are not examined, rather this study examines chromatin and transcription in the steady state, which is essential to better understand *H. werneckii's* halotolerance.

6.2. Gene expression and nucleosome dynamics

Nucleosome maps were generated for all three salinity conditions from mid exponential cells totalling three biological replicates. It is known that large nucleosome depleted regions are associated with constitutively expressed genes. In the 10% reference condition, there was modest correlation between the size of the NDR and expression abundance (Spearman's $\rho = 0.356$, $p < 2.2e-16$). This suggests that an open configuration is associated with higher levels of

transcription. Moreover, in the previous clusters based on the NDR, the average TPM for Cluster 4 (with the closed conformation) was much lower than any of the other clusters (Table 5). Thus, 5' NDR of *H. werneckii*'s genes represent the typical nucleosome "on" conformation. With this relationship established, now changes of transcription between the salinities can be compared with changes in nucleosome structure.

Table 5. Gene expression abundance (TPM) of the genes within nucleosome profile clusters.

	TPM (mean)	TPM (median)
Cluster1	69.04	21.23
Cluster2	105.25	25.94
Cluster3	102.38	29.98
Cluster4	34.66	6.617

6.3. Changes of nucleosome positioning and occupancy in response to salt

To first evaluate any chromatin changes with respect to different salinities, the summits of nucleosomes from each condition were compared using DANPOS2. This resulted in a log₂FC and q-value for each nucleosome summit. Only single pair wise comparisons can be done in DANPOS2 which compares a treatment and control. The 10% NaCl condition was the reference when compared with 0% and 20% and 20% NaCl the reference when compared with 0%. The results illustrate that there are individual nucleosome changes throughout the genome. However, significant ($q < 0.012$, $\log_2FC > \text{abs}(0.56)$) changes only occur in a fraction of the nucleosome population. The 0% versus 10% NaCl comparison had the lowest percentage of change (3.4%), 10% versus 20% NaCl the second highest (3.7%) and 0% versus 20% NaCl had the greatest change of nucleosomes (4.0%). The low fraction of nucleosome changes demonstrates that chromatin changes very little within acclimatized cells.

The 3.7% nucleosomes that showed significant changes were analyzed for proximity to a TSS. Only 634 genes with changing nucleosomes were within 300 bp of their TSS. Changes in nucleosome occupancy had a small negative correlation with expression (spearman's $\rho = -0.29$, $p < 2.8e-14$). However, it has been shown that dynamic changes in occupancy are not very predictive of expression [215]. Additionally, many genes have large differences in expression but

nonetheless are highly expressed in both conditions. Nucleosome changes have been reported to show a greater degree of change with respect to genes that have an on or off expression profile [83, 211]. Therefore, genes with high fold change and large absolute differences were selected for examination of nucleosome occupancy at their plus one nucleosome. No major chromatin changes were observed for these selected genes, however there is a trend of nucleosome depletion at the plus one nucleosomes of upregulated genes in 20% NaCl condition (Figure 20A). Upregulated genes in 20% salinity have decreased nucleosome occupancy at plus-one nucleosomes and down regulated genes relative to 10% show a slight increase of occupancy within the NDR albeit subtle(Figure 20B). Taken together, this data represents a great resource to probe for genes that have different chromatin configurations and how they may relate to transcription.

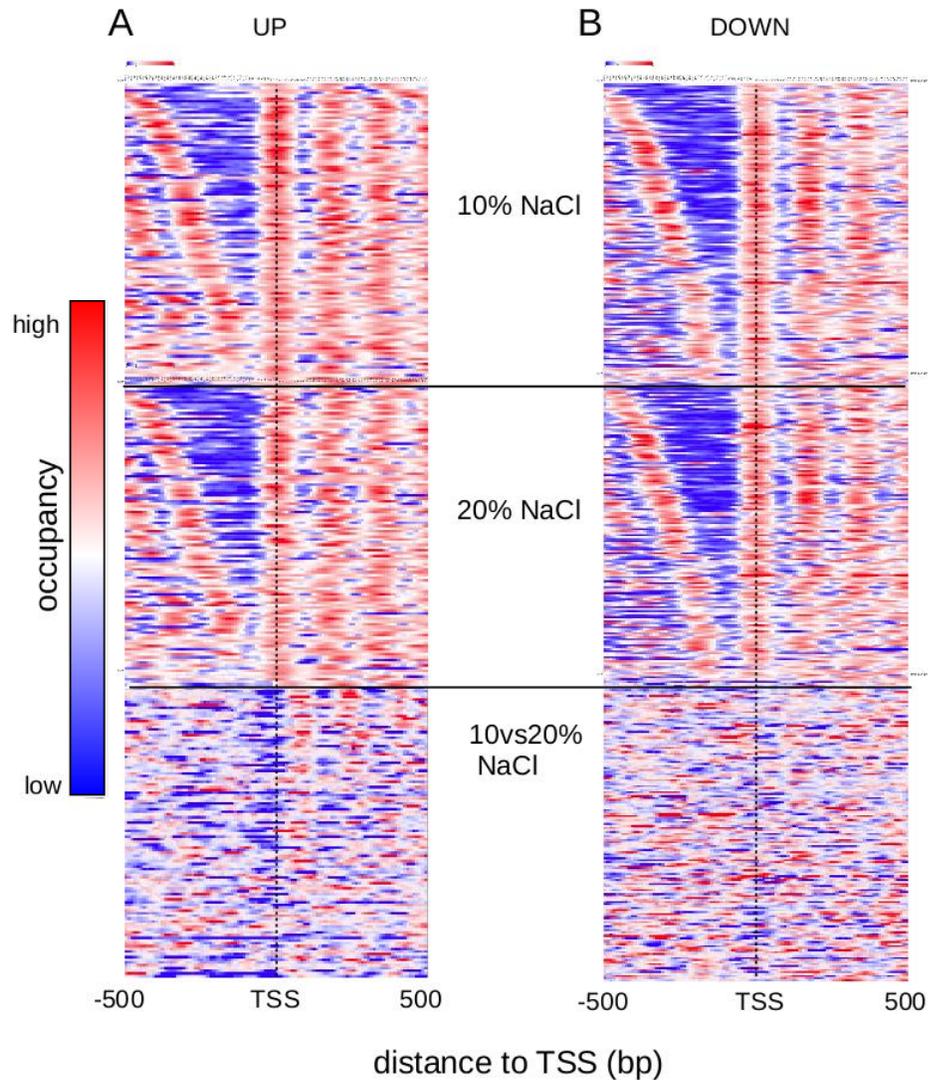


Figure 20. Heatmap nucleosome profiles of both 10% NaCl and 20% NaCl of the up and down regulated genes in 20% NaCl relative to 10% NaCl. The genes included are only those that are up (A) or down (B) regulated in 20% NaCl relative to 10% NaCl. Blue represents nucleosome depleted regions while red shows areas of high occupancy. The blue in the versus profile shows loss of nucleosomes relative to 20% whereas red shows gain of occupancy. The nucleosomes are aligned to the summit of their genes respective plus one nucleosome.

In *H. werneckii* there were some genes with obvious changes in nucleosome occupancy upstream of the TSS. One such example was the up regulation of two transcripts similar to *Int1* of *A. thaliana* in high salt (20% NaCl) (Figure 21B). They both showed similar expression profiles with 20% NaCl having the highest and the other two salinities had similar abundance. The loci of the

genes have different nucleosome architectures at the TSS which reflect the abundance of the transcripts (Figure 21A). The 0% NaCl condition shows a highly occupied TSS with two nucleosomes occluding the supposed regulatory region. The 10% NaCl condition has lower occupancy for both transcripts and differences in positioning of the summit. The 20% NaCl shows the largest perturbation of nucleosome structure with obvious depletion at the TSS and a nucleosome shift in the 3' direction. The changes reflect the differences in expression abundance. This raises the question of how important are these transcripts for high salinity tolerance. When the gene was searched in a protein data base, sugar transporters were the top hits with hexose transporter being the most specific. This likely reflects the import of sugars for the increased energy expenditure necessary in high salt. Although this gene reflects the typical repression by occlusion of a nucleosome, other differentially expressed genes do not [99, 218].

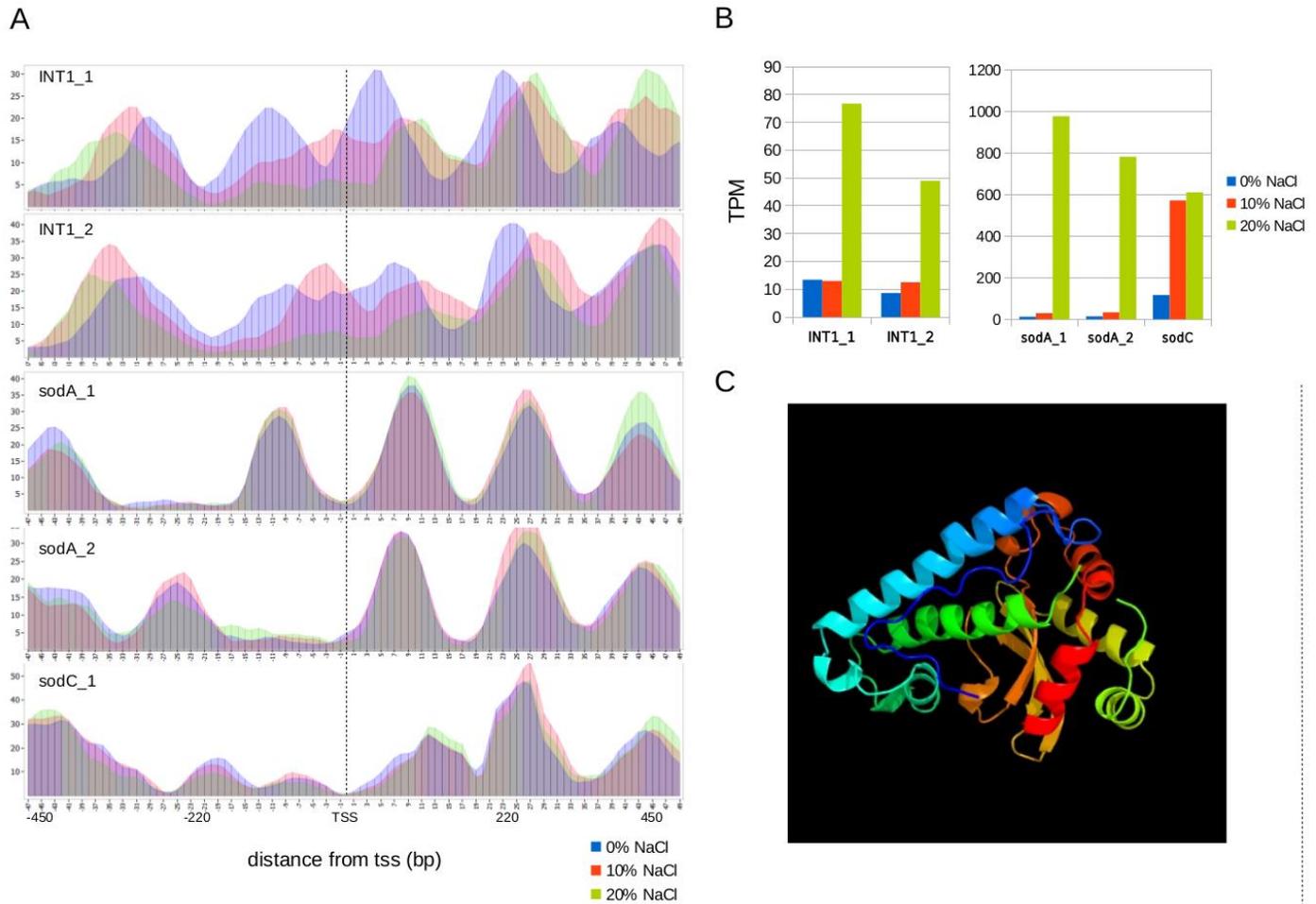


Figure 21. Comparison of the nucleosome landscape and expression profiles of Int1, SodA and SodC.

A) Phasograms of nucleosome occupancy at gene loci encoding for transcripts similar to INT1 Inositol transporter 1, superoxide dismutase A and C. Blue represents 0% NaCl nucleosome occupancy, red represents 10% NaCl nucleosome occupancy and green represents 20% NaCl nucleosome occupancy. B) Gene expression profile of the transcripts measured in TPM. C) Predictive protein modeling of the SodA-like protein showing that its 3D structure can be modeled and is similar to structure of SodA in related species.

The SodA-like transcripts were among the most differential expressed genes with the highest expression in 20% NaCl (Figure 21B). Due to their high differential expression, they provide excellent candidates to study the difference in chromatin configuration. Both genes have a 5' NDR with the distances varying with respect to the TSS. Contrary to expectation, there were no significant changes on the chromatin level in response to high salt for this gene (Figure 21A). However, the SodA_2 had decreased nucleosome occupancy at the -1 nucleosome in the high salinity condition. When environmental salinity rapidly increases, *H. werneckii* undergoes

excessive respiration to provide the required energy for adaptation. The open promoter conformation may allow rapid transcriptional response to the increased ROS generated within the mitochondria without the need to evict histone octamers. *H. werneckii*'s natural environment is prone to fluctuating abiotic stress such as salinity, heat and UV radiation. To confirm that this transcript codes for a SodA-like molecule, the coding sequencing (CDS) of SodA_1 was translated and structurally modeled using Phyre2, a protein homology recognition engine (Figure 21C) [182]. The major hits were all SodA-like proteins with 206 residues (80%) of the sequence being modeled with 100% confidence. Oxidative stress after exposure to abiotic stress is common in eukaryotes, and the open chromatin configuration for the SodA-like gene may aid in its remarkable ability to adapt to diverse environments. Moreover, the increased demand for mitochondrial respiration may induce the cell to use other carbon substrates in the environment for energy.

Transcripts encoding proteins needed for glycerol assimilation were upregulated in high salt conditions, especially transcripts encoding proteins similar to dihydroxyacetone kinase (dak1) (Figure 22C). The increase of these enzymes in high salt may reflect the cell employing alternate mechanisms to increase substrate entering the Krebs cycle in mitochondrial respiration. Glycerol can be used as an alternative carbon source through the glycerol catabolic pathway and the increase of dihydroxyacetone kinase-like transcripts and other enzymes involved in the pathway make it of interesting note [173, 175, 219]. Dak1 like genes were profiled along their respective TSS regions. Three of the dak1 proteins (Dak1_3,5,7) have a pronounced nucleosome depleted region (Figure 22A). The expression level of Dak1_7 was much lower than the other dak1 like transcripts, however Dak1_3 and Dak_5 showed high expression levels with nucleosome depletion for the higher salinities (Figure 22C), evident of an open promoter. Dak1_1, Dak_2, Dak1_3, Dak1_4 all had closed configurations but at 0% NaCl they had higher nucleosome occupancy at the TSS, while 20% NaCl had a 3' position shift, suggestive of a change in transcription. Dak1_1 and Dak1_2 have similar homology to yeast Dak1 which have similar closed conformation in wild type cells (Figure 20B & D). The difference in chromatin structure, expression, and copy numbers of the Dak1-like genes may reflect the differences in the environments between fungi. Some yeasts can thrive using glycerol as a carbon source, however *S. cerevisiae* growth is noticeably reduced (~0.11 in *S. cerevisiae* vs 0.42 h⁻¹ in *P. tannophilus*) [219]. Extracellular glycerol in hypersaline environments

can reach very high levels due to accumulation from multiple halophiles employing the same compatible solute strategy [220]. The enrichment of genes involved in glycerol catabolism may allow *H. werneckii* to utilize it as an additional carbon source in hypersaline environments. The extracellular glycerol content of *H. werneckii* has previously been shown to increase considerably in high salinity (> 17%) whereas intracellular levels remain constant throughout a broad range (7%-25% NaCl)[41]. This is consistent with the number of Stl1 like transporters in the genome and is suggestive of an efficient glycerol uptake system. This type of gene enrichment is also seen in glycerol efficient yeast and is the rate limiting step for *S. cerevisiae* in glycerol catabolism [178, 219].

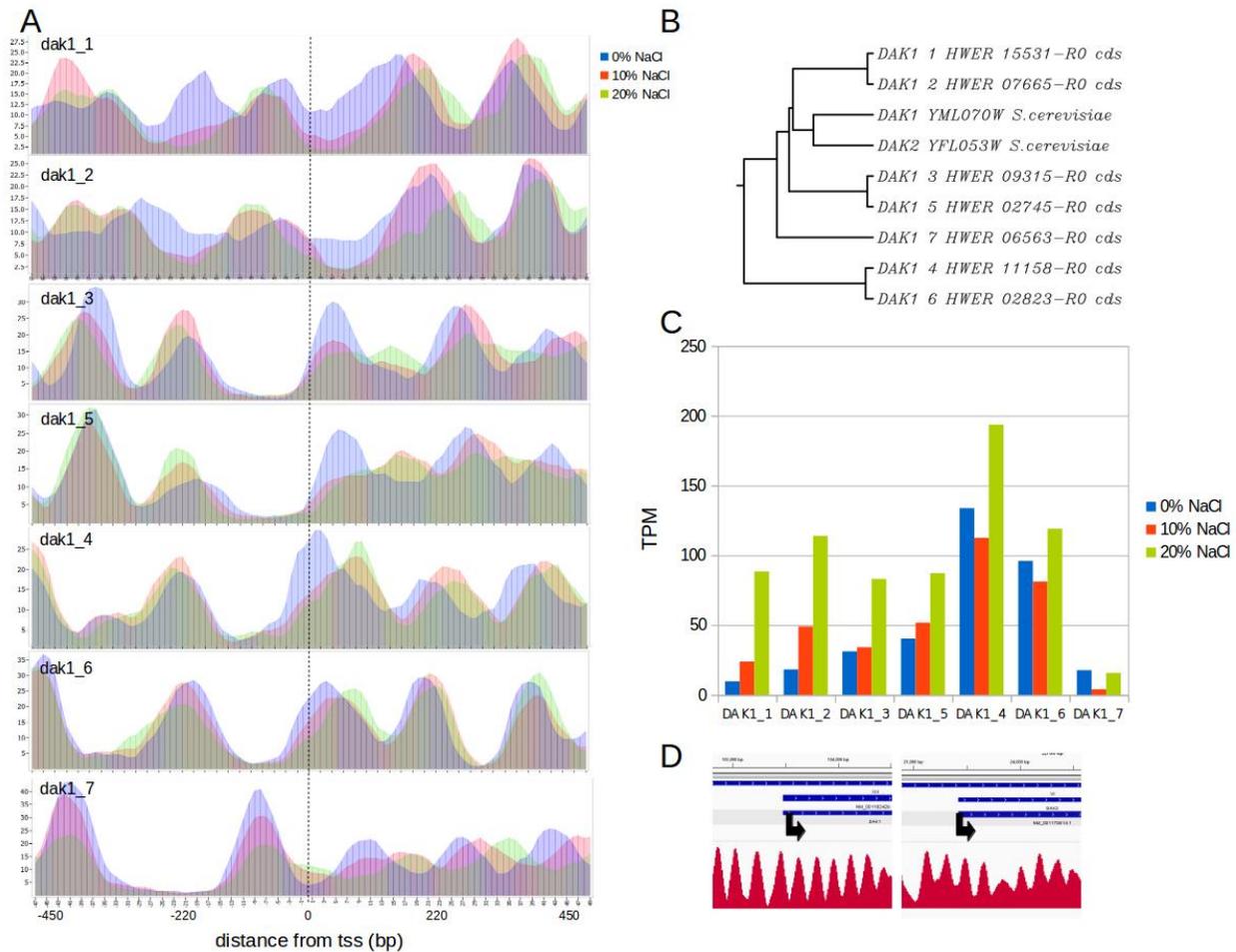


Figure 22. Comparison of nucleosome landscape and expression of Dihydroxyacetone kinase-like transcripts in *H. werneckii* across a range of salinities. A) Dihydroxyacetone kinase (dak)-like transcripts profiled with respect to their TSS, with 0% NaCl as blue, 10% as red and 20% as green. B) Comparison of homology based on amino-acid sequence similarity with *S. cerevisiae* dihydroxyacetone kinases and *H. werneckii*. C) Expression profile of the dak1 like transcripts in different salt concentrations. D) The nucleosome landscape of *S. cerevisiae* dihydroxyacetone genes (*DAK1*, *DAK2*).

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is an essential enzyme in glycolysis and gluconeogenesis. It has been also shown to be upregulated on the protein and transcript level in response to abiotic stress in plants and fungi[221, 222] . It has been shown that transgenic plants and fungi expressing heterologous GAPDH and plants overexpressing their endogenous GAPDH have improved salt tolerance[25, 223-225]. This suggests an important role of this enzyme in salt stress though the mechanisms remain unclear[224]. *H. werneckii* has two transcripts encoding glyceraldehyde-3-phosphate (THR) and they are upregulated in the salt conditions (Figure 23A.) The nucleosome configuration surrounding the TSS is very different (Figure 23B). The

5' NDR of *HwTDH1a* has an occupied promoter versus its ohnolog (*HwTDH1b*) where there is prevalent NDR despite similar gene expression levels (Figure 23 A). This may reflect gene divergence on the chromatin level however more experiments would be needed. This gene could be a potential transgene to confer salt tolerance in other organisms as it is also upregulated in response to salt.

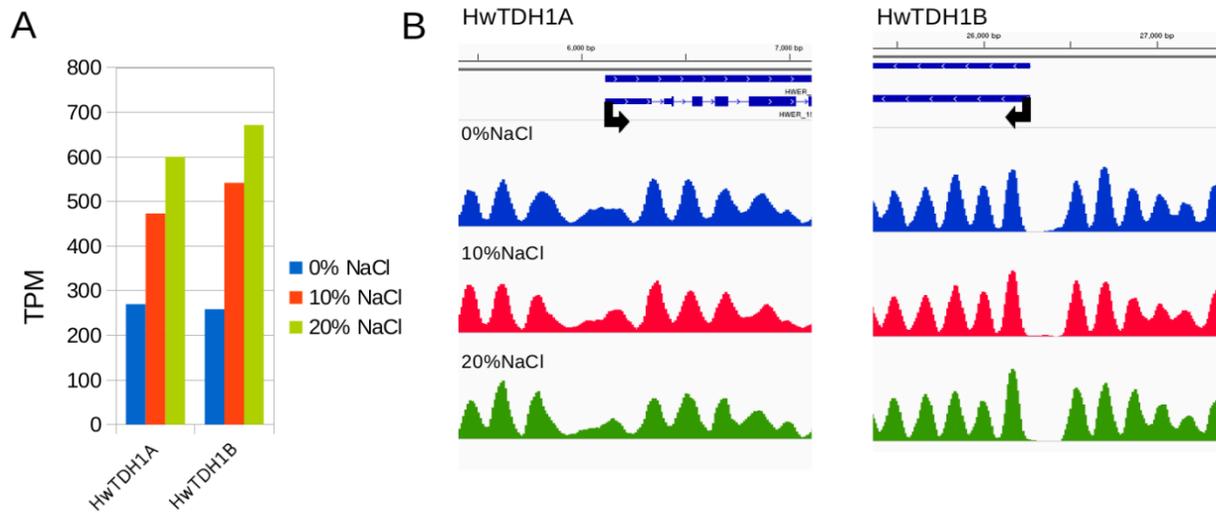


Figure 23. Expression and nucleosome landscape of glycerol dehydrogenase like transcripts in *H. werneckii*. A) TPM gene expression of transcripts of glycerol dehydrogenase like transcripts (*HwTDH1*). B) Nucleosome profile surrounding the TSS for *HwTDH1* genes.

These data sets, when combined, offer a high-resolution picture into transcriptional regulation of the steady state adaptation to hypersalinity. Assessing the stable parameters of chromatin across salinities has revealed that glycerol catabolism and mitochondrial associated processes are key elements in response to salt. This information may be applied to other organisms to modulate increases of osmotic, ionic and oxidative stress tolerance. Additionally, this data can also be used to address evolutionary questions of gene divergence and phylogeny when coupled with molecular characterization of related fungi.

7. The genome of *Hortaea acidophila*

There are only two other described members of the genus *Hortaea*, *Hortaea thailandica* and *Hortaea acidophila* although taxonomic classification of these two members is not definitive [226]. *H. acidophila* has recently had its genome sequenced and the raw short reads deposited in the NCBI. A genome assembly, like that for *H. werneckii*, could provide insights into the acidophilic lifestyle. *H. acidophila*, is a melanised extremophile which can grow at a pH as low as 0.6 though very few studies have examined this remarkable adaptation [70]. The only molecular characterization has been focused on laccases, enzymes involved in the last step of DHN-melanin synthesis [71, 227]. In *H. acidophila*, these laccases can function at extremely low pH [71]. Moreover, *H. aciophila* is not halotolerant and only able to grow at concentrations of NaCl less than 2%.

7.1. *H. acidophila* genome assembly and comparison

As no genome assembly existed, here, a new assembly was constructed using Platanus, a short read assembler [108]. Remarkably, the genome was assembled into 171 contigs with a N50 of 247, 367 and a total length of 20 Mb (Table 6). This relatively small genome strongly suggests that the WGD of *H. werneckii* occurred after the lineages diverged and is recent. The genome is 95.7 % complete (using the full pezizomycotina BUSCO set) with only 5 duplicated BUSCOs versus *H. werneckii* which has 2,602 (Figure 24A) [111]. The relatively small number of contigs from a short read assembler, likely, reflects a combination of the read length (151 bp) and the deep sequencing coverage. It had had a L50 of 25, which means that 50% of the totally size of the genome exists on 25 contigs, which is extremely high for a short read assembly with no subsequent scaffolding (Figure 22A). *H. acidophila* also had higher GC (56.56) content versus its relatives *Zymoseptoria tritici* (52.14) and *H. werneckii* (53.52). Like in the assembly of *H. werneckii* there was a single contig around 25 kb with very high coverage ($\sim 1e4$). When this contig was searched in the nucleotide database, the best hits were other full mitochondria. Both *H. werneckii* and *H. acidophila* mitochondrial genomes were annotated using MFannot and visualized using OGdraw [228, 229] (Appendix 4). To address synteny, comparisons between the three species were

evaluated using Symap[142]. Surprisingly, *H. werneckii* had more syntenic blocks and higher genome coverage when compared with *Z. tritici* (62 blocks, 24% coverage) than *H. acidophila* (17 blocks, 12% coverage) (Figure 24B). This suggests that *H. acidophila* is in the wrong genus, as ribosomal DNA sequence comparison also alludes to this (own data). The genome however, can still be used to characterize extremophile adaptation.

Table 6. Assembly statistics of *Hortaea acidophila* compared with phylogenetically related fungi.

	<i>Hortaea acidophila</i>	<i>Hortaea werneckii</i>	<i>Zymoseptoria tritici</i>
Contigs	187	30	21
# contigs > 50000	98	21	20
Total length	20.5 Mb	50.0 Mb	39.7 Mb
Largest Contig	689 329	3 716 197	6 088 797
GC (%)	56.56	53.52	52.14
N50	247 367	2 868 097	2 674 951
# predicted genes	597	744	1613
# Predicted genes	17 868	35 165	30 513

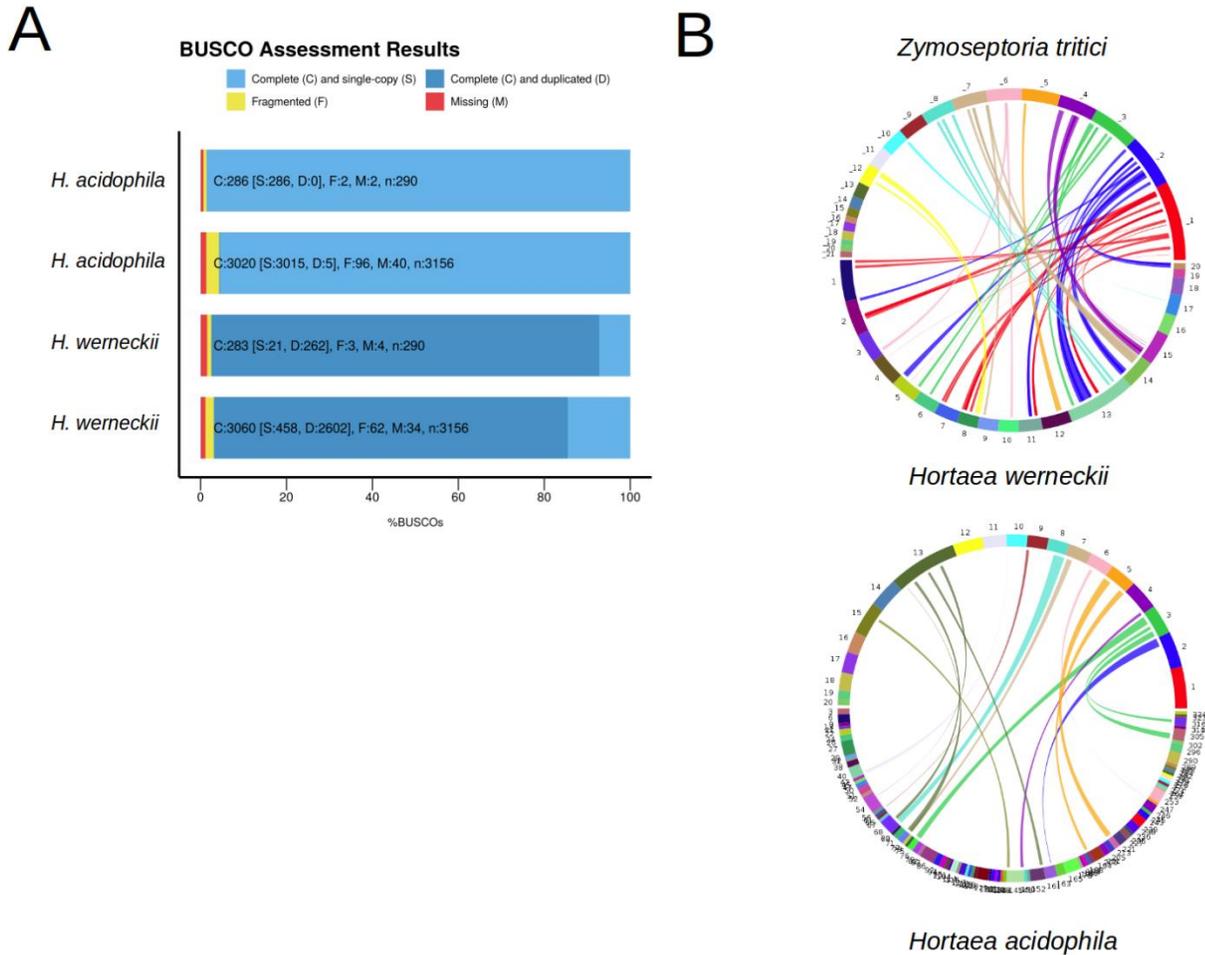


Figure 24. Completeness, comparison and synteny of *H. acidophila* versus *H. werneckii* and *Z. tritici*. A) BUSCO assessment of completeness of *H. acidophila* compared with *H. werneckii* using both fungal and pezizomycotina BUSCO databases. B) Synteny circle plots of *H. werneckii*'s genome and its relationship to *Z. tritici* (top) and *H. acidophila* genomes (bottom). Colored blocks represent contigs and lines across represent syntentic regions.

The previously identified laccases were searched in the genome to confirm their identity and to query if there is duplication of this family of enzymes. There were multiple hits, but only one had reasonable query cover (which reflects the portion of protein sequence (the query) with a match in the genome). Interestingly when *H. werneckii*'s genome was searched for the same protein, there were multiple hits with more query cover (433-674 score) versus the second hit in *H. acidophila* (394 score). One of the hits within the transcriptome assembly was upregulated in 20% and 10% NaCl relative to 0% and annotated as "Similar to lcc2 Laccase-2 (*Botryotinia fuckeliana*)". The genome comparisons allow investigation into "dosage balance hypothesis"

which states that genes that are involved with macromolecular complexes are less likely to diverge due to their stoichiometric restraint such as ribosomal units, transcription factors and core histones[151]. Therefore, inquiries into gene retention after whole genome duplication can be assessed by comparing yeasts [230, 231]. For example, the lineage of yeasts (*S. cerevisiae*) that diverged after an ancient whole genome duplication have two copies each of the histone proteins but only one histone variant H2A.Z[232].

H. werneckii has two copies of the core histone proteins and two copies of the H2A.Z (*htz1*) variant, whereas within the genomes of *Z. tritici* and *H. acidophila* have only a single hit. The absence of a duplicated copy of the H2A.Z variant in *S. cerevisiae* and related yeasts suggests that contrary to core histones, this histone underwent subsequent gene loss. Evaluating this histone variant is therefore interesting in the context of sequence divergence and can be further analyzed at the level of chromatin, expression and amino acid similarity. The histone variants had different expression levels in the reference condition with *htz1_B* having a 1.75 fold increase relative to its ohnolog (Figure 25A). This result is surprising because this protein functions as a subunit within the histone octamer; however, it may be that this variant is explained by the fact that this variant is only enriched at the plus one nucleosome. This could mean that there is less dosage imbalance pressure compared with the core histones which are present in nearly all nucleosomes [233, 234]. The ohnologs were then analyzed at the chromatin level. When aligned to the plus one nucleosome there is a 5' NDR upstream, a higher occupied +2 nucleosome in *htz1_B* and a higher occupied nucleosome at the +3 position in *htz1_A* (Figure 25A). Taken together, this is a modest difference on the individual nucleosome level. When the entire gene was evaluated, there was a NDR within the genic region of *htz1_B* but not *htz1_A*. Also, there was an NDR further upstream of the TSS for *htz1_A* (Figure 25B). Whether these differences affect gene expression remains to be determined. The translated amino acid composition of the ohnologs was 96.43% similar; however, the N termini show low conservation, which could reflect the differences in start sites within the gene (Figure 25C). In terms of salinity, H2A.z is involved in gene regulation in response to environmental perturbation making it an interesting protein for future studies in *H. werneckii*[234].

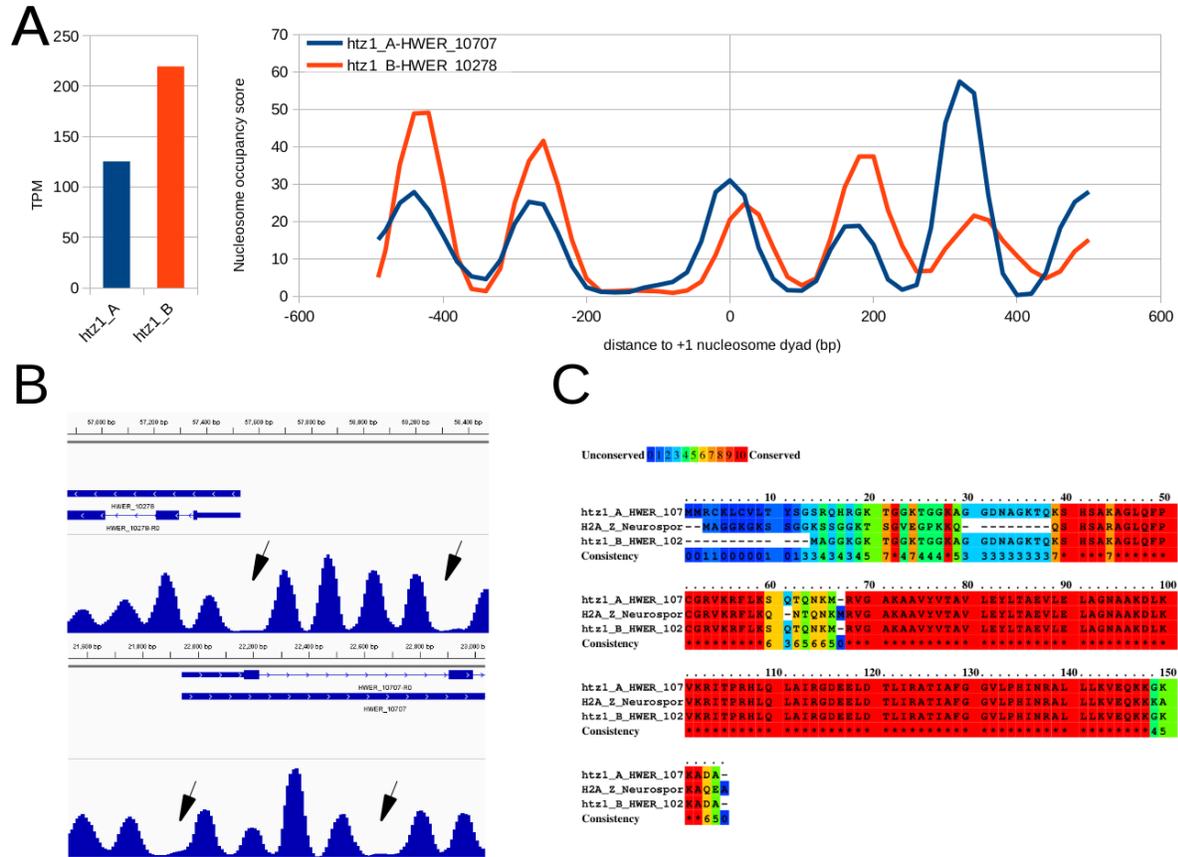


Figure 25. Analysis of the histone variant gene loci, Htz1 (H2A.z), in *H. werneckii*. A) Expression profiles of the ohnolog at 10% NaCl and the corresponding phasogram. B) Genome browser snap shots of the regions surrounding the TSS of Htz1 like transcripts. C) Amino acid sequence comparison between *H. werneckii* histone variant Htz1 (H2A.Z) and the histone variant (H2A.Z) of *N. crassa*

Together, these observations show the usefulness of mining bioinformatic data to ask questions of evolution and stress adaptation. The increase of available sequencing data can actually outpace the speed at which analyses are performed. Extremophile data sets in particular should be given more attention because synthetic biology and bioengineering could greatly benefit from such knowledge [40].

8. Conclusion

High salinity exposure invokes substantial ROS production in eukaryotes through mechanisms related to increased energy expenditure [48, 51]. This type of response is seen in yeast, where the ROS are attributed to a surge of aerobic respiration [164, 235]. *H. werneckii*, when exposed to hyperosmotic shock, upregulates many mitochondrial metabolism and biogenesis proteins, indicative of a mitochondrial role in salt tolerance [235]. Moreover, correlated changes of mitochondrial morphology have been reported [235]. This intensification of respiration and *H. werneckii*'s ability to degrade exogenous H₂O₂ led Petrovic to postulate that antioxidant pathways are involved in halotolerance and may determine its upper salinity range [45]. The presented data here supports this hypothesis while also adding alternative explanations for the role for carbon metabolism in *H. werneckii*'s adaptation to salt stress.

H. werneckii's genome was first assembled into a more contiguous assembly which allowed analysis of both expression and chromatin architecture in response to salinity stress. The combined data-sets provide a rich resource for data mining and will help in the quest for stress tolerant genes. Two such genes that could be of considerable interest encode for a superoxide dismutase with Mn²⁺ as its cofactor. The large difference in abundance between salinities suggests a strong role in halotolerance. In fact, transgenic organisms expressing heterologous SodA genes had a substantial increased tolerance to salt [20, 46, 236]. Therefore, this gene could be a useful transgene for developing salt-tolerant organisms.

Transcripts encoding for plasma membrane ATPase (Pma1,2) were highly expressed in all conditions, which is not surprising because this is the most abundant plasma membrane protein required for establishing the proton gradient [39]. In high salinities, this is a crucial factor that requires more energy. Additionally, the rapid synthesis of glycerol is also associated with osmotic shock via the protein glycerol-3-phosphate dehydrogenase. Transcripts for this gene were upregulated in high salt, which is in line with previous literature [237]. However, contrary to expectation, genes associated with glycerol catabolism such as dihydroxyacetone kinase, were upregulated substantially in 20% NaCl. An explanation for this phenomenon could be that when cells are growing at a steady state at a concentration of 20% NaCl, they are adapted. They do not

need to produce as much glycerol because it is available in the environment and these cells are osmotically balanced. However, the cell still needs to balance the ion concentration gradient at the plasma membrane which requires energy[24]. This, in combination with the required energy for growth, drives the reuptake of glycerol as an additional carbon source from the external environment. This would explain the enrichment of Stt1 transporters and the upregulation of certain family members in response to salt. Glycerol would then be converted to pyruvate via the aerobic glycerol pathway. Many transcripts encoding proteins in glycerol catabolism pathway were upregulated in *H. werneckii* in response to salt. These include transcripts similar to DAK1, GYC1, TPI1, GLD1 and glycerol kinase (Gk). Although, GPD1 catalyzes glycerol production, it also catalyzes the reverse reaction of glycerol-3-phosphate to dihydroxacetone phosphate. Therefore, up regulation of these genes could reflect the cell's energy requirements and the increase of respiration at the mitochondria. The use of glycerol as a carbon source may be a consequence of its availability in saline conditions such as in niche of salterns where resident halophiles produce glycerol[219, 220]. Lastly, the upregulation of mitochondria metabolic and biogenesis genes, in combination with previous literature, would explain the increased zinc transporters and supposed mobilization of the ion. Zinc is a major cofactor in mitochondrial pathways and is essential for biogenesis, proper function and redox reactions [48, 167, 171].

The halotolerance of *H. werneckii* has led to many studies regarding its response to hyperosmotic shock yet few address the adapted steady state [24, 35, 42]. That said, hyperosmotic shock still allows a glimpse into transcriptional stress response bursts and chromatin upheaval as the cell acclimates to its new environment [101, 212]. In these experiments, the cells are examined after they have adapted to hyper salinity and are in steady state of growth. Changes in chromatin likely reflect static compositions and are not as transient as seen in hyperosmotic shock [101, 211, 212]. These experiments are crucial to understand extremophiles and how they remain adapted to such environments on the molecular level. Extremophiles thrive in such environments and very few can survive the broad salinity range of *H. werneckii*. Insights into these relatively static molecular differences can reveal insights for genetic engineering of osmotolerant yeast strains without the introduction of foreign DNA[194]. For example, nucleosome depleted regions upstream of homologous genes that respond to hyperosmotic shock could be compared to

the nucleosome conformation exhibited in *S. cerevisiae*. Base pair alterations could be made at these candidate loci, to induce or repress nucleosome formation in which they mimic *H. werneckii*'s conformation leading to a more osmotolerant yeast strain. To validate these genes, future experiments should include knock-out mutants of salt relevant genes, to confirm if these genes do in fact confer salt tolerance. Taken together this data provides insights into possible mechanisms for increasing salt tolerance in other eukaryotes such as increasing the antioxidant defense system for increases in respiration and provides an extensive data-base for future work.

References

1. Rhoads, A. and K.F. Au, *PacBio Sequencing and Its Applications*. Genomics Proteomics Bioinformatics, 2015. **13**(5): p. 278-89.
2. Koboldt, D.C., et al., *The next-generation sequencing revolution and its impact on genomics*. Cell, 2013. **155**(1): p. 27-38.
3. Choi, S.C., *On the study of microbial transcriptomes using second- and third-generation sequencing technologies*. J Microbiol, 2016. **54**(8): p. 527-36.
4. Gostincar, C., et al., *Fungal adaptation to extremely high salt concentrations*. Adv Appl Microbiol, 2011. **77**: p. 71-96.
5. Walsh, J., *Encyclopedia of Environment and Society*. 2007, SAGE Publications, Inc.: Thousand Oaks.
6. Mateo-Sagasta, J. and J. Burke, *Report 8 - Agriculture and water quality interactions: a global overview*. FAO, 2012. Retrieved April 12, 2017; from: http://www.fao.org/fileadmin/templates/solaw/files/thematic_reports/TR_08.pdf.
7. Negrão, S., S.M. Schmöckel, and M. Tester, *Evaluating physiological responses of plants to salinity stress*. Annals of Botany, 2017. **119**(1): p. 1-11.
8. Petronia Carillo, M.G.A., Giovanni Pontecorvo, Amodio Fuggi, Pasqualina Woodrow, *Salinity Stress and Salt Tolerance*, in *Abiotic Stress in Plants - Mechanisms and Adaptations*, A. Shanker, Editor. 2011.
9. Shabala, S., et al., *On a quest for stress tolerance genes: membrane transporters in sensing and adapting to hostile soils*. J Exp Bot, 2015.
10. Acosta-Motos, J.R., et al., *Plant Responses to Salt Stress: Adaptive Mechanisms*. Agronomy, 2017. **7**(1): p. 18.
11. Zhu, J.-K., *Plant Salt Stress*, in *eLS*. 2007, John Wiley & Sons, Ltd.
12. Munns, R. and M. Tester, *Mechanisms of salinity tolerance*. Annu Rev Plant Biol, 2008. **59**: p. 651-81.
13. Roy, S.J., S. Negrão, and M. Tester, *Salt resistant crop plants*. Curr Opin Biotechnol, 2014. **26**: p. 115-24.
14. Deinlein, U., et al., *Plant salt-tolerance mechanisms*. Trends in plant science, 2014. **19**(6): p. 371-379.
15. Mühlhling, K.H. and E. Läuchli, *Physiological traits of sodium toxicity and salt tolerance*, in *Plant Nutrition: Food security and sustainability of agro-ecosystems through basic and applied research*, W.J. Horst, et al., Editors. 2001, Springer Netherlands: Dordrecht. p. 378-379.
16. Bose, J., A. Rodrigo-Moreno, and S. Shabala, *ROS homeostasis in halophytes in the context of salinity stress tolerance*. J Exp Bot, 2014. **65**(5): p. 1241-57.
17. Cabot, C., et al., *Lessons from crop plants struggling with salinity*. Plant Sci, 2014. **226**: p. 2-13.
18. Plett, D., et al., *Improved Salinity Tolerance of Rice Through Cell Type-Specific Expression of AtHKT1;1*. PLOS ONE, 2010. **5**(9): p. e12571.
19. Sun, Y., et al., *Potassium Retention under Salt Stress Is Associated with Natural Variation in Salinity Tolerance among Arabidopsis Accessions*. PLOS ONE, 2015. **10**(5): p. e0124032.
20. Rahman, A., et al., *Manganese-induced salt stress tolerance in rice seedlings: regulation of ion homeostasis, antioxidant defense and glyoxalase systems*. Physiol Mol Biol Plants, 2016. **22**(3): p. 291-306.
21. Wang, M., et al., *The critical role of potassium in plant stress response*. Int J Mol Sci, 2013. **14**(4): p. 7370-90.

22. Hanin, M., et al., *New Insights on Plant Salt Tolerance Mechanisms and Their Potential Use for Breeding*. *Frontiers in Plant Science*, 2016. **7**(1787).
23. Ashraf, M. and N.A. Akram, *Improving salinity tolerance of plants through conventional breeding and genetic engineering: An analytical comparison*. *Biotechnology for the Sustainability of Human Society*, 2009: p. Volume 27, Issue 6, Pages 744-752.
24. Plemenitas, A., et al., *Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective*. *Front Microbiol*, 2014. **5**: p. 199.
25. Cho, J.I., et al., *Over-expression of PsGPD, a mushroom glyceraldehyde-3-phosphate dehydrogenase gene, enhances salt tolerance in rice plants*. *Biotechnol Lett*, 2014. **36**(8): p. 1641-8.
26. Liang, X., et al., *A ribosomal protein AgRPS3aE from halophilic Aspergillus glaucus confers salt tolerance in heterologous organisms*. *Int J Mol Sci*, 2015. **16**(2): p. 3058-70.
27. Dodd, I.C. and F. Pérez-Alfocea, *Microbial amelioration of crop salinity stress*. *Journal of Experimental Botany*, 2012. **63**(9 3415-3428): p. doi: 10.1093/jxb/ers033.
28. Gostincar, C., et al., *Fungal Adaptation to Extremely High Salt Concentrations*. 2011. **77**(71-96).
29. Steensels, J., et al., *Improving industrial yeast strains: exploiting natural and artificial diversity*. *FEMS Microbiol Rev*, 2014. **38**(5): p. 947-95.
30. Burg, D., et al., *Proteomics of extremophiles*. *Environmental Microbiology*, 2011: p. 13 (8), 1934-1955.
31. Lenassi, M., et al., *Whole Genome Duplication and Enrichment of Metal Cation Transporters Revealed by De Novo Genome Sequencing of Extremely Halotolerant Black Yeast *Hortaea werneckii**. *Plos ONE*, 2013: p. (8), 1-18.
32. Gunde-Cimerman, N. and A. Plemenitaš, *Ecology and molecular adaptations of the halophilic black yeast *Hortaea werneckii**. *Life in Extreme Environments*, 2006: p. pp 177-185.
33. Plemenitaš, A., et al., *Adaptation to high salt concentrations in halotolerant/halophilic fungi: a molecular perspective*. *Frontiers in Microbiology*, 2014: p. Volume 5, Article 199.
34. Gunde-Cimerman, N., et al., *Hypersaline waters in salterns - natural ecological niches for halophilic black yeasts*. 2000. **32**(235-240).
35. Lenassi, M., et al., *Whole genome duplication and enrichment of metal cation transporters revealed by de novo genome sequencing of extremely halotolerant black yeast *Hortaea werneckii**. *PLoS One*, 2013. **8**(8): p. e71328.
36. Volkov, V., *Salinity tolerance in plants. Quantitative approach to ion transport starting from halophytes and stepping to genetic and protein engineering for manipulating ion fluxes*. *Front Plant Sci*, 2015. **6**: p. 873.
37. Yenush, L., *Potassium and Sodium Transport in Yeast*. *Adv Exp Med Biol*, 2016. **892**: p. 187-228.
38. Arino, J., J. Ramos, and H. Sychrova, *Alkali metal cation transport and homeostasis in yeasts*. *Microbiol Mol Biol Rev*, 2010. **74**(1): p. 95-120.
39. Plemenitas, A., et al., *Transport Systems in Halophilic Fungi*. *Adv Exp Med Biol*, 2016. **892**: p. 307-25.
40. Kogej, T., et al., *The halophilic fungus *Hortaea werneckii* and the halotolerant fungus *Aureobasidium pullulans* maintain low intracellular cation concentrations in hypersaline environments*. *Appl Environ Microbiol*, 2005. **71**(11): p. 6600-5.
41. Petrovic, U., N. Gunde-Cimerman, and A. Plemenitas, *Cellular responses to environmental salinity in the halophilic black yeast *Hortaea werneckii**. *Mol Microbiol*, 2002. **45**(3): p. 665-72.
42. Vaupotic, T. and A. Plemenitas, *Differential gene expression and Hog1 interaction with osmoresponsive genes in the extremely halotolerant black yeast *Hortaea werneckii**. *BMC Genomics*, 2007. **8**: p. 280.

43. Murakami, K. and M. Yoshino, *Effect of fructose 1,6-bisphosphate on the iron redox state relating to the generation of reactive oxygen species*. *Biometals*, 2015. **28**(4): p. 687-91.
44. Breitenbach, M., et al., *Oxidative stress in fungi: its function in signal transduction, interaction with plant hosts, and lignocellulose degradation*. *Biomolecules*, 2015. **5**(2): p. 318-42.
45. Petrovic, U., *Role of oxidative stress in the extremely salt-tolerant yeast *Hortaea werneckii**. *FEMS Yeast Res*, 2006. **6**(5): p. 816-22.
46. Wang, Y.C., et al., *Enhanced salt tolerance of transgenic poplar plants expressing a manganese superoxide dismutase from *Tamarix androssowii**. *Mol Biol Rep*, 2010. **37**(2): p. 1119-24.
47. Shafi, A., et al., *Expression of SOD and APX genes positively regulates secondary cell wall biosynthesis and promotes plant growth and yield in *Arabidopsis* under salt stress*. *Plant Mol Biol*, 2015. **87**(6): p. 615-31.
48. Eide, D.J., *The oxidative stress of zinc deficiency*. *Metallomics*, 2011. **3**(11): p. 1124-9.
49. Wan, C., et al., *The impact of zinc sulfate addition on the dynamic metabolic profiling of *Saccharomyces cerevisiae* subjected to long term acetic acid stress treatment and identification of key metabolites involved in the antioxidant effect of zinc*. *Metallomics*, 2015. **7**(2): p. 322-32.
50. Sun, J., et al., *Melanization of a meristematic mutant of *Fonsecaea monophora* increases tolerance to stress factors while no effects on antifungal susceptibility*. *Mycopathologia*, 2011. **172**(5): p. 373-80.
51. Kejzar, A., et al., *Melanin is crucial for growth of the black yeast *Hortaea werneckii* in its natural hypersaline environment*. *Fungal Biology*, 2013. **117**(5): p. 368-379.
52. Eisenman, H.C. and A. Casadevall, *Synthesis and assembly of fungal melanin*. *Appl Microbiol Biotechnol*, 2012. **93**(3): p. 931-40.
53. Kogej, T., et al., *Evidence for 1,8-dihydroxynaphthalene melanin in three halophilic black yeasts grown under saline and non-saline conditions*. *FEMS Microbiol Lett*, 2004. **232**(2): p. 203-9.
54. Ngamskulrungraj, P., et al., *Cryptococcus gattii virulence composite: candidate genes revealed by microarray analysis of high and less virulent Vancouver island outbreak strains*. *PLoS One*, 2011. **6**(1): p. e16076.
55. Taylor, J.S. and J. Raes, *Duplication and divergence: the evolution of new genes and old ideas*. *Annu Rev Genet*, 2004. **38**: p. 615-43.
56. Ohno, S., *Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999*. *Semin Cell Dev Biol*, 1999. **10**(5): p. 517-22.
57. Wolfe, K.H., *Origin of the Yeast Whole-Genome Duplication*. *PLoS Biology*, 2015. **13**(8): p. e1002221.
58. Marcet-Houben, M. and T. Gabaldon, *Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage*. *PLoS Biol*, 2015. **13**(8): p. e1002220.
59. Ceccarelli, M., et al., *Chromosome endoreduplication as a factor of salt adaptation in *Sorghum bicolor**. *Protoplasma*, 2006. **227**(2-4): p. 113-8.
60. De Storme, N. and A. Mason, *Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance*. *Current Plant Biology*, 2014. **1**: p. 10-33.
61. Albertin, W. and P. Marullo, *Polyploidy in fungi: evolution after whole-genome duplication*. *Proceedings of the Royal Society B: Biological Sciences*, 2012. **279**(1738): p. 2497-2509.
62. Zielke, N., B.A. Edgar, and M.L. DePamphilis, *Endoreplication*. *Cold Spring Harb Perspect Biol*, 2013. **5**(1): p. a012948.
63. Fox, D.T. and R.J. Duronio, *Endoreplication and polyploidy: insights into development and disease*. *Development*, 2013. **140**(1): p. 3-12.

64. De Veylder, L., J.C. Larkin, and A. Schnittger, *Molecular control and function of endoreplication in development and physiology*. Trends Plant Sci, 2011. **16**(11): p. 624-34.
65. Van de Peer, Y., E. Mizrachi, and K. Marchal, *The evolutionary significance of polyploidy*. Nat Rev Genet, 2017. **advance online publication**.
66. John, P.C. and R. Qi, *Cell division and endoreduplication: doubtful engines of vegetative growth*. Trends Plant Sci, 2008. **13**(3): p. 121-7.
67. Gerstein, A.C. and S.P. Otto, *Ploidy and the causes of genomic evolution*. J Hered, 2009. **100**(5): p. 571-81.
68. Larkins, B.A., et al., *Investigating the hows and whys of DNA endoreduplication*. J Exp Bot, 2001. **52**(355): p. 183-92.
69. Edger, P.P. and J.C. Pires, *Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes*. Chromosome Res, 2009. **17**(5): p. 699-717.
70. Holker, U., et al., *Hortaea acidophila, a new acid-tolerant black yeast from lignite*. Antonie Van Leeuwenhoek, 2004. **86**(4): p. 287-94.
71. Tetsch, L., J. Bend, and U. Holker, *Molecular and enzymatic characterisation of extra- and intracellular laccases from the acidophilic ascomycete Hortaea acidophila*. Antonie Van Leeuwenhoek, 2006. **90**(2): p. 183-94.
72. Kuraku, S. and A. Meyer, *Detection and phylogenetic assessment of conserved synteny derived from whole genome duplications*. Methods Mol Biol, 2012. **855**: p. 385-95.
73. Ye, C., et al., *DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies*. Sci Rep, 2016. **6**: p. 31900.
74. Sohn, J.I. and J.W. Nam, *The present and future of de novo whole-genome assembly*. Brief Bioinform, 2016.
75. Li, Z., et al., *Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph*. Brief Funct Genomics, 2012. **11**(1): p. 25-37.
76. Simpson, J.T. and M. Pop, *The Theory and Practice of Genome Sequence Assembly*. Annu Rev Genomics Hum Genet, 2015. **16**: p. 153-72.
77. Struhl, K. and E. Segal, *Determinants of nucleosome positioning*. Nature structural & molecular biology, 2013. **20**(3): p. 267-273.
78. Ho, J.W., et al., *Comparative analysis of metazoan chromatin organization*. Nature, 2014. **512**(7515): p. 449-52.
79. Teves, S.S., C.M. Weber, and S. Henikoff, *Transcribing through the nucleosome*. Trends Biochem Sci, 2014. **39**(12): p. 577-86.
80. Bai, L. and A.V. Morozov, *Gene Regulation by nucleosome positioning*. Trends in Genetics, 2010. **26**(11): p. Volume 26, Issue 11, 476-483.
81. Iyer, V.R., *Nucleosome positioning: bringing order to the eukaryotic genome*. 2012. **22**(5).
82. Chereji, R.V. and A.V. Morozov, *Functional roles of nucleosome stability and dynamics*. Brief Funct Genomics, 2015. **14**(1): p. 50-60.
83. van Bakel, H., et al., *A Compendium of Nucleosome and Transcript Profiles Reveals Determinants of Chromatin Architecture and Transcription*. PLoS Genetics, 2013. **9**(5): p. e1003479.
84. Lieleg, C., et al., *Nucleosome positioning in yeasts: methods, maps, and mechanisms*. Chromosoma, 2015. **124**(2): p. 131-51.
85. Liu, M.J., et al., *Determinants of nucleosome positioning and their influence on plant gene expression*. Genome Res, 2015. **25**(8): p. 1182-95.
86. Liu, G., et al., *A deformation energy-based model for predicting nucleosome dyads and occupancy*. Sci Rep, 2016. **6**: p. 24133.
87. González, S., et al., *Nucleosomal signatures impose nucleosome positioning in coding and noncoding sequences in the genome*. Genome Research, 2016. **26**(11): p. 1532-1543.

88. Turner, B., *The Adjustable nucleosome: an epigenetic signaling module*. Trends in Genetics, 2012. **28**(9): p. 436-444.
89. Yen, K., et al., *Genome-wide nucleosome specificity and directionality of chromatin remodelers*. Cell, 2012. **149**(7): p. 1461-1473.
90. Flores, O., et al., *Fuzziness and noise in nucleosomal architecture*. Nucleic Acids Res, 2014. **42**(8): p. 4934-46.
91. Jiang, C. and B.F. Pugh, *Nucleosome positioning and gene regulation: advances through genomics*. Nat Rev Genet, 2009. **10**(3): p. 161-172.
92. Zhang, T., W. Zhang, and J. Jiang, *Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants*. Plant Physiol, 2015. **168**(4): p. 1406-16.
93. Mavrich, T.N., et al., *Nucleosome organization in the Drosophila genome*. Nature, 2008. **453**(7193): p. 358-362.
94. Valouev, A., et al., *Determinants of nucleosome organization in primary human cells*. Nature, 2011. **474**(7352): p. 516-520.
95. Wu, Y., W. Zhang, and J. Jiang, *Genome-Wide Nucleosome Positioning Is Orchestrated by Genomic Regions Associated with DNase I Hypersensitivity in Rice*. PLOS Genetics, 2014. **10**(5): p. e1004378.
96. Keung, A.J., et al., *Chromatin regulation at the frontier of synthetic biology*. Nature reviews. Genetics, 2015. **16**(3): p. 159-171.
97. Portela, R.M. and T. Vogl, *Synthetic Core Promoters as Universal Parts for Fine-Tuning Expression in Different Yeast Species*. 2017. **6**(3): p. 471-484.
98. Zhang, P., et al., *Genome-wide mapping of nucleosome positions in Saccharomyces cerevisiae in response to different nitrogen conditions*. Sci Rep, 2016. **6**: p. 33970.
99. Wang, X., et al., *Nucleosomes and the accessibility problem*. Trends Genet, 2011. **27**(12): p. 487-92.
100. Deniz, O., et al., *Nucleosome architecture throughout the cell cycle*. Sci Rep, 2016. **6**: p. 19729.
101. Weiner, A., et al., *High-resolution chromatin dynamics during a yeast stress response*. Mol Cell, 2015. **58**(2): p. 371-86.
102. Tsui, K., et al., *Genomic approaches for determining nucleosome occupancy in yeast*. Methods Mol Biol, 2012. **833**: p. 389-411.
103. Henikoff, J.G., et al., *Epigenome characterization at single base-pair resolution*. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(45): p. 18318-18323.
104. Chomczynski, P., *A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples*. Biotechniques, 1993. **15**(3): p. 532-4, 536-7.
105. Koren, S., et al., *Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation*. bioRxiv, 2016.
106. Heo, Y., et al., *BLESS: bloom filter-based error correction solution for high-throughput sequencing reads*. Bioinformatics, 2014. **30**(10): p. 1354-62.
107. Bankevich, A., et al., *SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing*. Journal of Computational Biology, 2012. **19**(5): p. 455-477.
108. Kajitani, R., et al., *Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads*. Genome Res, 2014. **24**(8): p. 1384-95.
109. Ye, C. and Z. Ma, *Sparc: a sparsity-based consensus algorithm for long erroneous sequencing reads*. PeerJ, 2016. **4**: p. e2016.
110. Walker, B.J., et al., *Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement*. PLoS One, 2014. **9**(11): p. e112963.

111. Simão, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs*. *Bioinformatics*, 2015.
112. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. *Bioinformatics*, 2009. **25**(14): p. 1754-60.
113. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. *Bioinformatics*, 2009. **25**(16): p. 2078-9.
114. Chakraborty, M., et al., *Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage*. *Nucleic Acids Research*, 2016. **44**(19): p. e147-e147.
115. Boetzer, M. and W. Pirovano, *SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information*. *BMC Bioinformatics*, 2014. **15**: p. 211.
116. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies*. *Bioinformatics*, 2013. **29**(8): p. 1072-1075.
117. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity*. *Nature protocols*, 2013. **8**(8): p. 10.1038/nprot.2013.084.
118. Holt, C. and M. Yandell, *MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects*. *BMC Bioinformatics*, 2011. **12**: p. 491.
119. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification*. *Nat Biotech*, 2016. **34**(5): p. 525-527.
120. Pimentel, H.J., et al., *Differential analysis of RNA-Seq incorporating quantification uncertainty*. *bioRxiv*, 2016.
121. Jones, P., et al., *InterProScan 5: genome-scale protein function classification*. *Bioinformatics*, 2014. **30**(9): p. 1236-40.
122. Maere, S., K. Heymans, and M. Kuiper, *BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks*. *Bioinformatics*, 2005. **21**(16): p. 3448-9.
123. Falcon, S. and R. Gentleman, *Using GOstats to test gene lists for GO term association*. *Bioinformatics*, 2007. **23**(2): p. 257-8.
124. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data*. *Bioinformatics*, 2014. **30**(15): p. 2114-20.
125. Quinlan, A.R., *BEDTools: The Swiss-Army Tool for Genome Feature Analysis*. *Curr Protoc Bioinformatics*, 2014. **47**: p. 11.12.1-34.
126. Chen, K., et al., *DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing*. *Genome Res*, 2013. **23**(2): p. 341-51.
127. Vainshtein, Y., K. Rippe, and V.B. Teif, *NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data*. *BMC Genomics*, 2017. **18**(1): p. 158.
128. Gunde-Cimermana, N., et al., *Hypersaline waters in salterns - natural ecological niches for halophilic black yeasts*. *FEMS Microbiol Ecol*, 2000. **32**(3): p. 235-240.
129. Berlin, K., S. Koren, and C.S. Chin, *Assembling large genomes with single-molecule sequencing and locality-sensitive hashing*. 2015. **33**(6): p. 623-30.
130. Camacho, C., et al., *BLAST+: architecture and applications*. *BMC Bioinformatics*, 2009. **10**: p. 421-421.
131. Thomma, B.P., et al., *Mind the gap; seven reasons to close fragmented genome assemblies*. *Fungal Genet Biol*, 2016. **90**: p. 24-30.
132. Havis, N.D., et al., *Ramularia collo-cygni--An Emerging Pathogen of Barley Crops*. *Phytopathology*, 2015. **105**(7): p. 895-904.
133. Ohm, R.A., et al., *Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi*. *PLoS Pathog*, 2012. **8**(12): p. e1003037.
134. Schwartz, R.A., *Superficial fungal infections*. *Lancet*, 2004. **364**(9440): p. 1173-82.

135. van der Lee, T.A. and M.H. Medema, *Computational strategies for genome-based natural product discovery and engineering in fungi*. Fungal Genet Biol, 2016. **89**: p. 29-36.
136. Martin, J.F. and P. Liras, *Evolutionary formation of gene clusters by reorganization: the meleagrins/roquefortine paradigm in different fungi*. Appl Microbiol Biotechnol, 2016. **100**(4): p. 1579-87.
137. Linde, J., et al., *De Novo Whole-Genome Sequence and Genome Annotation of Lichtheimia ramosa*. Genome Announc, 2014. **2**(5).
138. Kumar, A., et al., *Draft genome sequence of Karnal bunt pathogen (Tilletia indica) of wheat provides insights into the pathogenic mechanisms of quarantined fungus*. PLoS ONE, 2017. **12**(2): p. e0171323.
139. Seidl, M.F., et al., *The Genome of the Saprophytic Fungus Verticillium tricorpus Reveals a Complex Effector Repertoire Resembling That of Its Pathogenic Relatives*. Mol Plant Microbe Interact, 2015. **28**(3): p. 362-73.
140. Zdobnov, E.M., et al., *OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs*. 2017. **45**(D1): p. D744-d749.
141. Soderlund, C., et al., *SyMAP: A system for discovering and viewing syntenic regions of FPC maps*. Genome Res, 2006. **16**(9): p. 1159-68.
142. Soderlund, C., M. Bomhoff, and W.M. Nelson, *SyMAP v3.4: a turnkey synteny system with application to plant genomes*. Nucleic Acids Res, 2011. **39**(10): p. e68.
143. del Pozo, J.C. and E. Ramirez-Parra, *Whole genome duplications in plants: an overview from Arabidopsis*. J Exp Bot, 2015. **66**(22): p. 6991-7003.
144. Madlung, A., *Polyploidy and its effect on evolutionary success: old questions revisited with new tools*. Heredity (Edinb), 2013. **110**(2): p. 99-104.
145. Panchy, N., M. Lehti-Shiu, and S.-H. Shiu, *Evolution of Gene Duplication in Plants*. Plant Physiology, 2016. **171**(4): p. 2294-2316.
146. Yona, A.H., et al., *Chromosomal duplication is a transient evolutionary solution to stress*. Proc Natl Acad Sci U S A, 2012. **109**(51): p. 21010-5.
147. Torres, E.M., B.R. Williams, and A. Amon, *Aneuploidy: cells losing their balance*. Genetics, 2008. **179**(2): p. 737-46.
148. Sheltzer, J.M., et al., *Aneuploidy drives genomic instability in yeast*. Science, 2011. **333**(6045): p. 1026-30.
149. Tu, Y., et al., *Genome duplication improves rice root resistance to salt stress*. Rice, 2014. **7**(1): p. 15-15.
150. Birchler, J.A. and R.A. Veitia, *The gene balance hypothesis: from classical genetics to modern genomics*. Plant Cell, 2007. **19**(2): p. 395-402.
151. Birchler, J.A. and R.A. Veitia, *Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines*. Proc Natl Acad Sci U S A, 2012. **109**(37): p. 14746-53.
152. Hrdlickova, R., M. Toloue, and B. Tian, *RNA-Seq methods for transcriptome analysis*. Wiley Interdiscip Rev RNA, 2016.
153. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nat Rev Genet, 2009. **10**(1): p. 57-63.
154. Ramos, M.J.N., et al., *Flower development and sex specification in wild grapevine*. BMC Genomics, 2014. **15**(1): p. 1095.
155. Ramos, M.J.N., et al., *Deep analysis of wild Vitis flower transcriptome reveals unexplored genome regions associated with sex specification*. Plant Molecular Biology, 2016: p. 1-20.
156. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression*. Nat Methods, 2017.

157. Tatusov, R.L., et al., *The COG database: an updated version includes eukaryotes*. BMC Bioinformatics, 2003. **4**: p. 41.
158. Jin, H., H.T. Rube, and J.S. Song, *Categorical spectral analysis of periodicity in nucleosomal DNA*. Nucleic Acids Research, 2016. **44**(5): p. 2047-2057.
159. Gordon, P. *Using Kallisto & Sleuth for RNASeq analysis*. 2016; Retrieved February 24, 2017; from: <http://achri.blogspot.ca/2016/10/using-kallisto-sleuth.html>.
160. Coordinators, N.R., *Database resources of the National Center for Biotechnology Information*. Nucleic Acids Research, 2016. **44**(Database issue): p. D7-D19.
161. Zhao, W., et al., *The Aspergillus fumigatus beta-1,3-glucanosyltransferase Gel7 plays a compensatory role in maintaining cell wall integrity under stress conditions*. Glycobiology, 2014. **24**(5): p. 418-27.
162. Zhao, W., et al., *N-Glycosylation of Gel1 or Gel2 is vital for cell wall beta-glucan synthesis in Aspergillus fumigatus*. Glycobiology, 2013. **23**(8): p. 955-68.
163. Melamed, D., L. Pnueli, and Y. Arava, *Yeast translational response to high salinity: Global analysis reveals regulation at multiple levels*. RNA, 2008. **14**(7): p. 1337-1351.
164. Pastor, M.M., M. Proft, and A. Pascual-Ahuir, *Mitochondrial Function Is an Inducible Determinant of Osmotic Stress Adaptation in Yeast*. The Journal of Biological Chemistry, 2009. **284**(44): p. 30307-30317.
165. Petrović, U., N. Gunde-Cimerman, and A. Plemenitas, *Cellular responses to environmental salinity in the halophilic black yeast Hortaea werneckii*. Molecular Microbiology, 2002. **45**(3): p. 665-672.
166. McCulley, A., et al., *Chemical suppression of defects in mitotic spindle assembly, redox control, and sterol biosynthesis by hydroxyurea*. G3 (Bethesda), 2014. **4**(1): p. 39-48.
167. Wu, C.Y., et al., *Regulation of the yeast TSA1 peroxiredoxin by ZAP1 is an adaptive response to the oxidative stress of zinc deficiency*. J Biol Chem, 2007. **282**(4): p. 2184-95.
168. Wilson, S. and A.J. Bird, *Zinc sensing and regulation in yeast model systems*. Arch Biochem Biophys, 2016. **611**: p. 30-36.
169. Hernandez-Saavedra, N.Y. and R. Romero-Geraldo, *Cloning and sequencing the genomic encoding region of copper-zinc superoxide dismutase enzyme from several marine strains of the genus Debaryomyces (Lodder & Kreger-van Rij)*. Yeast, 2001. **18**(13): p. 1227-38.
170. MacDiarmid, C.W., et al., *Peroxiredoxin chaperone activity is critical for protein homeostasis in zinc-deficient yeast*. J Biol Chem, 2013. **288**(43): p. 31313-27.
171. Atkinson, A., et al., *Mzm1 Influences a Labile Pool of Mitochondrial Zinc Important for Respiratory Function*. The Journal of Biological Chemistry, 2010. **285**(25): p. 19450-19459.
172. Zhang, L., et al., *Engineering of the glycerol decomposition pathway and cofactor regulation in an industrial yeast improves ethanol production*. J Ind Microbiol Biotechnol, 2013. **40**(10): p. 1153-60.
173. Matsuzawa, T., et al., *The gld1+ gene encoding glycerol dehydrogenase is required for glycerol metabolism in Schizosaccharomyces pombe*. Appl Microbiol Biotechnol, 2010. **87**(2): p. 715-27.
174. Norbeck, J. and A. Blomberg, *Metabolic and regulatory changes associated with growth of Saccharomyces cerevisiae in 1.4 M NaCl. Evidence for osmotic induction of glycerol dissimilation via the dihydroxyacetone pathway*. J Biol Chem, 1997. **272**(9): p. 5544-54.
175. Molin, M., J. Norbeck, and A. Blomberg, *Dihydroxyacetone kinases in Saccharomyces cerevisiae are involved in detoxification of dihydroxyacetone*. J Biol Chem, 2003. **278**(3): p. 1415-23.
176. Posas, F., et al., *The transcriptional response of yeast to saline stress*. J Biol Chem, 2000. **275**(23): p. 17249-55.
177. Kayingo, G., et al., *A permease encoded by STL1 is required for active glycerol uptake by Candida albicans*. Microbiology, 2009. **155**(Pt 5): p. 1547-1557.

178. Duskova, M., et al., *Two glycerol uptake systems contribute to the high osmotolerance of Zygosaccharomyces rouxii*. Mol Microbiol, 2015. **97**(3): p. 541-59.
179. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0*. Bioinformatics, 2007. **23**(21): p. 2947-2948.
180. Goujon, M., et al., *A new bioinformatics analysis tools framework at EMBL–EBI*. Nucleic Acids Research, 2010. **38**(suppl_2): p. W695-W699.
181. Bener Aksam, E., et al., *Absence of the peroxiredoxin Pmp20 causes peroxisomal protein leakage and necrotic cell death*. Free Radic Biol Med, 2008. **45**(8): p. 1115-24.
182. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis*. 2015. **10**(6): p. 845-58.
183. Marshall, A.N., et al., *Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes*. PLoS Genet, 2013. **9**(3): p. e1003376.
184. Grutzmann, K., et al., *Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study*. DNA Res, 2014. **21**(1): p. 27-39.
185. Parenteau, J., et al., *Deletion of many yeast introns reveals a minority of genes that require splicing for function*. Mol Biol Cell, 2008. **19**(5): p. 1932-41.
186. Hughes, A.L. and O.J. Rando, *Mechanisms underlying nucleosome positioning in vivo*. Annu Rev Biophys, 2014. **43**: p. 41-63.
187. Pugh, B.F., *A preoccupied position on nucleosomes*. Nat Struct Mol Biol, 2010. **17**(8): p. 923-923.
188. Moyle-Heyrman, G., et al., *Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning*. Proc Natl Acad Sci U S A, 2013. **110**(50): p. 20158-63.
189. Mieczkowski, J., et al., *MNase titration reveals differences between nucleosome occupancy and chromatin accessibility*. 2016. **7**: p. 11485.
190. Meyer, C.A. and X.S. Liu, *Identifying and mitigating bias in next-generation sequencing methods for chromatin biology*. Nat Rev Genet, 2014. **15**(11): p. 709-21.
191. Tsankov, A.M., et al., *The role of nucleosome positioning in the evolution of gene regulation*. PLoS Biol, 2010. **8**(7): p. e1000414.
192. Soumpasis, D.M., *Effects of DNA sequence and conformation on nucleosome formation*. J Biomol Struct Dyn, 1985. **3**(1): p. 1-10.
193. Zhou, X., et al., *A computational approach to map nucleosome positions and alternative chromatin states with base pair resolution*. eLife, 2016. **5**: p. e16970.
194. Curran, K.A., et al., *Design of synthetic yeast promoters via tuning of nucleosome architecture*. Nat Commun, 2014. **5**: p. 4002.
195. Tillo, D. and T.R. Hughes, *G+C content dominates intrinsic nucleosome occupancy*. BMC Bioinformatics, 2009. **10**: p. 442.
196. Quintales, L., et al., *A species-specific nucleosomal signature defines a periodic distribution of amino acids in proteins*. Open Biol, 2015. **5**(4): p. 140218.
197. Hughes, A.L., et al., *A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern*. Mol Cell, 2012. **48**(1): p. 5-15.
198. Deniz, O., et al., *Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast*. BMC Genomics, 2011. **12**: p. 489.
199. Fenouil, R., et al., *CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters*. Genome Research, 2012. **22**(12): p. 2399-2408.
200. Lubliner, S., L. Keren, and E. Segal, *Sequence features of yeast and human core promoters that are predictive of maximal promoter activity*. Nucleic Acids Research, 2013. **41**(11): p. 5569-5581.

201. Kaplan, N., et al., *Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology*. *Genome Biol*, 2010. **11**(11): p. 140.
202. Ballare, C., et al., *More help than hindrance: nucleosomes aid transcriptional regulation*. *Nucleus*, 2013. **4**(3): p. 189-94.
203. Peckham, H.E., et al., *Nucleosome positioning signals in genomic DNA*. *Genome Research*, 2007. **17**(8): p. 1170-1177.
204. Segal, E., et al., *A genomic code for nucleosome positioning*. *Nature*, 2006. **442**.
205. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities*. *Mol Cell*, 2010. **38**(4): p. 576-89.
206. Bai, L., et al., *Nucleosome-Depleted Regions in Cell-Cycle-Regulated Promoters Ensure Reliable Gene Expression in Every Cell Cycle*. *Developmental Cell*. **18**(4): p. 544-555.
207. Zentner, G.E. and S. Henikoff, *Regulation of nucleosome dynamics by histone modifications*. *Nat Struct Mol Biol*, 2013. **20**(3): p. 259-266.
208. Linke, C., et al., *Fkh1 and Fkh2 associate with Sir2 to control CLB2 transcription under normal and oxidative stress conditions*. *Front Physiol*, 2013. **4**: p. 173.
209. Hagman, A., et al., *Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication*. *PLoS One*, 2013. **8**(7): p. e68734.
210. Servant, G., et al., *Tye7 regulates yeast Ty1 retrotransposon sense and antisense transcription in response to adenylc nucleotides stress*. *Nucleic Acids Res*, 2012. **40**(12): p. 5271-82.
211. Klopf, E., et al., *INO80 represses osmostress induced gene expression by resetting promoter proximal nucleosomes*. 2016.
212. Miller, C., et al., *Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast*. *Molecular Systems Biology*, 2011. **7**(1).
213. Weiner, A., et al., *Systematic dissection of roles for chromatin regulators in a yeast stress response*. *PLoS Biol*, 2012. **10**(7): p. e1001369.
214. Shivaswamy, S., et al., *Dynamic Remodeling of Individual Nucleosomes Across a Eukaryotic Genome in Response to Transcriptional Perturbation*. *PLOS Biology*, 2008. **6**(3): p. e65.
215. Huebert, D.J., et al., *Dynamic Changes in Nucleosome Occupancy Are Not Predictive of Gene Expression Dynamics but Are Linked to Transcription and Chromatin Regulators*. *Molecular and Cellular Biology*, 2012. **32**(9): p. 1645-1653.
216. Nocetti, N. and I. Whitehouse, *Nucleosome repositioning underlies dynamic gene expression*. *Genes Dev*, 2016. **30**(6): p. 660-72.
217. Nagai, S., et al., *Chromatin potentiates transcription*. *Proc Natl Acad Sci U S A*, 2017. **114**(7): p. 1536-1541.
218. Choi, J.K. and Y.-J. Kim, *Intrinsic variability of gene expression encoded in nucleosome positioning sequences*. *Nature genetics*, 2009. **41**.
219. Klein, M., et al., *Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities*. *Environ Microbiol*, 2017. **19**(3): p. 878-893.
220. Oren, A., *Glycerol metabolism in hypersaline environments*. *Environ Microbiol*, 2017. **19**(3): p. 851-863.
221. Zeng, L., et al., *Genome-wide identification and characterization of Glyceraldehyde-3-phosphate dehydrogenase genes family in wheat (*Triticum aestivum*)*. *BMC Genomics*, 2016. **17**(1): p. 240.
222. Guo, L., et al., *Cytosolic Glyceraldehyde-3-Phosphate Dehydrogenases Interact with Phospholipase D δ to Transduce Hydrogen Peroxide Signals in the Arabidopsis Response to Stress*. *The Plant Cell*, 2012. **24**(5): p. 2200-2212.
223. Jeong, M.J., S.C. Park, and M.O. Byun, *Improvement of salt tolerance in transgenic potato plants by glyceraldehyde-3 phosphate dehydrogenase gene transfer*. *Mol Cells*, 2001. **12**(2): p. 185-9.

224. Zeng, L., et al., *Genome-wide identification and characterization of Glyceraldehyde-3-phosphate dehydrogenase genes family in wheat (Triticum aestivum)*. BMC Genomics, 2016. **17**: p. 240.
225. Chang, L., et al., *The beta subunit of glyceraldehyde 3-phosphate dehydrogenase is an important factor for maintaining photosynthesis and plant development under salt stress-Based on an integrative analysis of the structural, physiological and proteomic changes in chloroplasts in Thellungiella halophila*. Plant Sci, 2015. **236**: p. 223-38.
226. Crous, P.W., et al., *Phylogenetic lineages in the Capnodiales*. Studies in Mycology, 2009. **64**: p. 17-47-S7.
227. Tetsch, L., et al., *Evidence for functional laccases in the acidophilic ascomycete Hortaea acidophila and isolation of laccase-specific gene fragments*. FEMS Microbiol Lett, 2005. **245**(1): p. 161-8.
228. Beck N, L.B. MFannot, *organelle genome annotation webserver*. 2010. Retrieved March 27, 2017; from: <http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>.
229. Lohse, M., et al., *OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets*. Nucleic Acids Research, 2013. **41**(W1): p. W575-W581.
230. Veitia, R.A. and M.C. Potier, *Gene dosage imbalances: action, reaction, and models*. Trends Biochem Sci, 2015. **40**(6): p. 309-17.
231. Veitia, R.A., S. Bottani, and J.A. Birchler, *Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects*. Trends Genet, 2008. **24**(8): p. 390-7.
232. Scienski, K., J.C. Fay, and G.C. Conant, *Patterns of Gene Conversion in Duplicated Yeast Histones Suggest Strong Selection on a Coadapted Macromolecular Complex*. Genome Biology and Evolution, 2015. **7**(12): p. 3249-3258.
233. Guillemette, B. and L. Gaudreau, *Reuniting the contrasting functions of H2A.Z*. Biochem Cell Biol, 2006. **84**(4): p. 528-35.
234. Talbert, P.B. and S. Henikoff, *Environmental responses mediated by histone variants*. Trends Cell Biol, 2014. **24**(11): p. 642-50.
235. Vaupotic, T., et al., *Mitochondrial mediation of environmental osmolytes discrimination during osmoadaptation in the extremely halotolerant black yeast Hortaea werneckii*. Fungal Genet Biol, 2008. **45**(6): p. 994-1007.
236. Kaouthar, F., et al., *Responses of transgenic Arabidopsis plants and recombinant yeast cells expressing a novel durum wheat manganese superoxide dismutase TdMnSOD to various abiotic stresses*. J Plant Physiol, 2016. **198**: p. 56-68.
237. Kejzar, A., et al., *HwHog1 kinase activity is crucial for survival of Hortaea werneckii in extremely hyperosmolar environments*. Fungal Genet Biol, 2015. **74**: p. 45-58.

Appendices

Appendix A

1) Read preparation using Trimmomatic and Bless

```
# BLESS
```

```
./bless -read1 <forward fastq> -read2 <reverse fastq> -prefix <output prefix>  
-kmerlength 31
```

```
# Trimmomatic
```

```
java -jar trimmomatic-0.32.jar PE -baseout <output prefix> -basein <fastq in  
prefix> ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:7:2:true LEADING:3 TRAILING:3  
SLIDINGWINDOW:4:15 MINLEN:80
```

```
# 2) Canu, generate two sets of corrected long reads reads and two assemblies  
based on read length
```

```
canu \
```

```
-p Hw -d Hw-auto2 \
```

```
genomeSize=50m \
```

```
errorRate=0.035 \
```

```
maxMemory=30 \
```

```
maxThreads=6 \
```

```
-pacbio-raw ../filtered_subreads.fastq
```

```
canu \
```

```
-p Hw -d Hw-auto3 \
```

```
genomeSize=50m \
```

```
corMinCoverage=0 errorRate=0.035 \
```

```
maxMemory=30 \
```

```

maxThreads=6 \

-pacbio-raw ../filtered_subreads.fastq

# 3) Spades Assembly
spades.py -o <output prefix> \
--pe1-1 <fastq in prefix> \
--pe1-2 <fastq in prefix> \
-t 6 \
-m 28 \

# 4) Platanus Assemblies
platanus assemble -o <output prefix> -f <fastq in prefix> -t 4 -m 28 2>
assemble.log

# 5) Hybrid Assemblies using DBG2OLC for each short read assembly
../DBG2OLC k 17 AdaptiveTh 0.004 KmerCovTh 4 MinOverlap 20 PathCovTh 1
RemoveChimera 1 Contigs <input short read assembly contigs> f <corrected long
reads>

# 6) Sparc Consensus
../split_and_run_sparc.sh backbone_raw.fasta DBG2OLC_Consensus_info.txt
<corrected long reads used in DBG2OLC> ./consensus 2 >cns_log.txt

# 7) Analyzing completeness using BUSCO
python3 BUSCO.py -o <output prefix> -i <input assembly> -l fungi_odb9 -m
genome -sp neurospora_crassa -c 7

# 8) Pilon Polishing/Consensus, 1 iteration example:
bwa index -p T6 -a is /media/sean/backup/HwCanu/Hw-auto3/Hw.contigs.fasta &&
\

bwa mem -t 6 -B 6 T6 /media/sean/backup/blesscortectedreads/Hw7_1P.cor.cor.fq
/media/sean/backup/blesscortectedreads/Hw7_2P.cor.cor.fq > Hw7-T6.bwa.sam &&
\

samtools view -Sb Hw7-T6.bwa.sam > Hw7-T6.bwa.bam && samtools sort -m 8G -@ 6
Hw7-T6.bwa.bam Hw7-T6_sorted.bwa && \

samtools index Hw7-T6_sorted.bwa.bam && \

bwa mem -t 6 -B 6 T6 /media/sean/backup/blesscortectedreads/R2-10_1P.cor.fq
/media/sean/backup/blesscortectedreads/R2-10_2P.cor.fq > Nuc10-T6.bwa.sam &&
\

samtools view -Sb Nuc10-T6.bwa.sam > Nuc10-T6.bwa.bam && samtools sort -m 8G
-@ 6 Nuc10-T6.bwa.bam Nuc10-T6_sorted.bwa && \

```

```

samtools index Nucl0-T6_sorted.bwa.bam && \

bwa mem -t 6 -B 6 T6 /media/sean/backup/blessccorrectedreads/Hw5_1P.cor.cor.fq
/media/sean/backup/blessccorrectedreads/Hw5_2P.cor.cor.fq > Lib5-T6.bwa.sam &&
\

samtools view -Sb Lib5-T6.bwa.sam > Lib5-T6.bwa.bam && samtools sort -m 8G -@
6 Lib5-T6.bwa.bam Lib5-T6_sorted.bwa && \

samtools index Lib5-T6_sorted.bwa.bam && \

bwa mem -t 6 T6
/media/sean/backup/Danpos2016/RandomSequenceforconsensus3.fasta >
MappedReads3.sam && \

samtools view -q 30 -Sb MappedReads3.sam > PB-T6.2.bwa.bam && samtools sort -
m 8G -@ 6 PB-T6.2.bwa.bam PB-T6.2_sorted.bwa && \

samtools index PB-T6.2_sorted.bwa.bam && \

java -Xmx26g -Xms16g -jar /media/sean/backup/Pilon/pilon-1.20.jar --genome
/media/sean/backup/HwCanu/Hw-auto3/Hw.contigs.fasta --frags Nucl0-
T6_sorted.bwa.bam --frags Hw7-T6_sorted.bwa.bam --frags Lib5-
T6_sorted.bwa.bam --bam PB-T6.2_sorted.bwa.bam && \

awk '/^>/{print ">unitig1_" ++i; next}{print}' < pilon.fasta >
Hw.canu3.consensus.fasta

# 9) Analyzing completeness using BUSCO

python3 BUSCO.py -o <output prefix> -i <input assembly> -l fungi_odb9 -m
genome -sp neurospora_crassa -c 7

# 10) Assembly merging using quickmerge x3

nucmer -l 100 -prefix T1 <PacBio only assembly>.consensus.fasta 7.feb2.fasta
&& \

delta-filter -i 99 -r -q T1.delta > T1.rq.delta && \

./merger/quickmerge -d T1.rq.delta -q 7.feb2.fasta -r <PacBio only
assembly>.consensus.fasta -hco 8.0 -c 4.0 -l 100000 -ml 10000 && \

AssemblyStatistics contigs merged.fasta && \

awk '/^>/{print ">Merged1_" ++i; next}{print}' < merged.fasta > merged1.fasta

nucmer -l 100 -prefix T1 Hw3.consensus.fasta merged1.fasta && \

delta-filter -i 99 -r -q T1.delta > T1.rq.delta && \

./merger/quickmerge -d T1.rq.delta -q merged1.fasta -r Hw3.consensus.fasta -
hco 8.0 -c 5.0 -l 200000 -ml 20000 && \

AssemblyStatistics contigs merged.fasta && \

awk '/^>/{print ">Merged1_" ++i; next}{print}' < merged.fasta > merged2.fasta

```

```

nucmer -l 100 -prefix T4 Hw5.consensus.fasta merged2.fasta && \

delta-filter -i 99 -r -q T4.delta > T4.rq.delta && \

./merger/quickmerge -d T4.rq.delta -q merged2.fasta -r Hw5.consensus.fasta -
hco 7.0 -c 4.0 -l 200000 -ml 50000 && \

AssemblyStatistics contigs merged.fasta && \

awk '/^>/{print ">Merged3_" ++i; next}{print}' < merged.fasta > merged3.fasta

# 11) Assembly polishing with Pilon

java -Xmx26g -Xms16g -jar /media/sean/backup/Pilon/pilon-1.20.jar --genome
merged3.fasta --frags Nucl0-T1_sorted.bwa.bam --frags Hw7-T1_sorted.bwa.bam -
-frags Lib5-T1_sorted.bwa.bam --bam PB-T1.2_sorted.bwa.bam && \

awk '/^>/{print ">unitig1_" ++i; next}{rint}' < pilon.fasta >
merged3.consensus.fasta

# 12) Two iterations of scaffolding using SSPACE-Longread

perl /media/sean/backup/SSPACE-LongRead_v1-1/SSPACE-LongRead.pl -c
merged3.consensus.fasta -r 0.05 -t 6 -o 5000 -b Merged3.1 -i 99 -p
/media/sean/backup/HwCanu/Canu_Corrected_Reads/Hw.correctedReads.fasta && \

perl /media/sean/backup/SSPACE-LongRead_v1-1/SSPACE-LongRead.pl -c
./Merged3.1/scaffolds.fasta -r 0.10 -t 6 -o 5000 -b Merged3.2 -i 99 -p
/media/sean/backup/HwCanu/Canu_Corrected_Reads/Hw.correctedReads.fasta

# 13) polishing with Pilon x 3 iterations and with short read libraries

java -Xmx26g -Xms16g -jar /media/sean/backup/Pilon/pilon-1.20.jar --genome
./Merged3.2/scaffolds.fasta --frags Hw7-T1_sorted.bwa.bam --frags Hw5-
T1_sorted.bwa.bam --frags Nucl0-T1_sorted.bwa.bam --bam PB-
T1.2_sorted.bwa.bam --fix amb,all && \

# 14) Final assessment of assembly using BUSCO and QUASt

python3 /media/sean/backup/BuscoV2Run/BUSCO.py -o Merged3.BC2.fasta -i
Merged3.BC2.fasta -l pezizomycotina_odb9 -m genome -sp neurospora_crassa -c 7

python quast.py < assemblies to compare > -f --eukaryote -o <output> --
glimmer

```

Appendix B

RNA-seq Pipeline

```
# 1) Abundance estimation using RNA-seq reads pseudoaligned to the transcriptome
assembly using Kallisto
```

```
kallisto quant \  
-i Hw2tran \  
-o <output prefix> \  
--bias \  
--threads 12 \  
-b 100 \  
<reads in>
```

```
# 2) Differential expression analysis of 10% vs 20% NaCl in R
```

```
library("sleuth")  
base_dir <- "/media/sean/backup/Kallisto"  
sample_id <- dir(file.path(base_dir, "Results1B"))  
kal_dirs <- sapply(sample_id, function(id) file.path(base_dir, "Results1", id,  
"kallisto"))  
kal_dirs  
s2c <- read.table(file.path(base_dir, "10v20.sampleinfoB.txt"), header = TRUE,  
stringsAsFactors=FALSE)  
s2c <- dplyr::select(s2c, sample = sampleNum, condition)  
s2c <- dplyr::mutate(s2c, path = kal_dirs)  
ids <- c('target_id', 'ens_gene', 'ext_gene')  
t2g <- read.csv('/media/sean/backup/Kallisto/geneID.tsv', header = FALSE, sep =  
'\t')  
colnames(t2g) <- ids  
s1 <- sleuth_prep(s2c, ~condition, target_mapping = t2g, min_prop = 0.40 )  
s1 <- sleuth_fit(s1)  
s1 <- sleuth_fit(s1, ~1, 'reduced')  
s1 <- sleuth_wt(s1, "conditionC20-NaCl")
```

```

# 3) Differential expression analysis of 0% vs 10% NaCl in R

library("sleuth")

base_dir <- "/media/sean/backup/Kallisto"

sample_id <- dir(file.path(base_dir, "Results"))

kal_dirs <- sapply(sample_id, function(id) file.path(base_dir, "Results", id,
"Kallisto"))

kal_dirs

s2c <- read.table(file.path(base_dir, "sampleinfo.txt"), header = TRUE,
stringsAsFactors=FALSE)

s2c <- dplyr::select(s2c, sample = sampleNum, condition, batch)

s2c <- dplyr::mutate(s2c, path = kal_dirs)

ids <- c('target_id', 'ens_gene', 'ext_gene')

t2g <- read.csv('/media/sean/backup/Kallisto/geneID.tsv', header = FALSE, sep =
'\t')

colnames(t2g) <- ids

s2 <- sleuth_prep(s2c, ~condition + batch, target_mapping = t2g, min_prop =
0.75)

s2 <- sleuth_fit(s2)

s2 <- sleuth_fit(s2, ~batch, 'reduced')

s2 <- sleuth_wt(s2, "conditionB0-NaCl")

# 4) Differential expression analysis of 0% vs 20% NaCl in R

library("sleuth")

base_dir <- "/media/sean/backup/Kallisto"

sample_id <- dir(file.path(base_dir, "ResultsB"))

kal_dirs <- sapply(sample_id, function(id) file.path(base_dir, "ResultsB", id,
"Kallisto"))

kal_dirs

s2c <- read.table(file.path(base_dir, "sampleinfoB.txt"), header = TRUE,
stringsAsFactors = FALSE)

```

```
s2c <- dplyr::select(s2c, sample = sampleNum, condition, batch)
s2c <- dplyr::mutate(s2c, path = kal_dirs)
ids <- c('target_id', 'ens_gene', 'ext_gene')
t2g <- read.csv('/media/sean/backup/Kallisto/geneID.tsv', header = FALSE, sep =
'\t', stringsAsFactors = FALSE)
colnames(t2g) <- ids
so <- sleuth_prep(s2c, ~condition+batch, target_mapping = t2g, min_prop = 0.60)
so <- sleuth_fit(so)
so <- sleuth_fit(so, ~batch , 'reduced')
so2 <- sleuth_wt(so, which_beta = "conditionB0-NaCl", which_model = "full")
```

Appendix C

Nucleosome Pipeline

```
# 1) Read preparation
```

```
# trimming reads
```

```
trimmomatic PE -baseout R1-0.fastq -basein  
/teamshare/Illumina/HiSeqruns/FastqConversions/150414_SN7001370_0100_Bhkng2ad  
xx_Unaligned_SF4/Project_SF4/Sample_SF0/SF0_ATCACG_L001_R1_001.fastq.gz  
ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:7:2:true LEADING:3 TRAILING:3  
SLIDINGWINDOW:4:15 MINLEN:80
```

```
#BWA index
```

```
bwa index -p hwgenome -a is Hw2HGAP3.fasta
```

```
#BWA mapping
```

```
bwa mem -t 4 -B 6 hwgenome R1_1P.fastq R1_2P.fastq > R1.bwa.sam
```

```
#Flag and filter only unique reads-q > 30 means only reports best alignments  
-> Samtools
```

```
samtools view -F 0x04 -q 30 -Sb R1-0.bwa.sam > R1-0.bwa.bam
```

```
#Sort samfiles
```

```
samtools sort -n R1-0.bwa.bam R1-0_bwa_sorted
```

```
# Convert to bed
```

```
bamToBed -i R1-0_bwa_sorted.bam > R1-0_bwa_sorted.bed
```

```
# 2) Analysis of nucleosome positioning and occupancy using Danpos2
```

```
python danpos.py dpos /media/sean/My\ Passport/Wig/0-bed-2016 -o 0-Step10-  
70extend -m 1 -u 1e-15 -p 1e-10 --mifrsz 110 --mafrsz 200 -c 10000000 --  
extend 70 && \
```

```
python danpos.py dpos /media/sean/My\ Passport/Wig/10-bed-2016 -o 10-Step10-  
70extend -m 1 -u 1e-15 -p 1e-10 --mifrsz 110 --mafrsz 200 -c 10000000 --  
extend 70 && \
```

```
python danpos.py dpos /media/sean/My\ Passport/Wig/20-bed-2016 -o 20-Step10-  
70extend2 -m 1 -u 1e-15 -p 1e-10 --mifrsz 110 --mafrsz 200 -c 10000000 --  
extend 70
```

3) Comparison

```
python danpos.py dpos \
```

```
/media/sean/My\ Passport/Wig/0-bed-2016/:\
```

```
/media/sean/My\ Passport/Wig/10-bed-2016/ \
```

```
-o 10v0-pos4c -u 1e-15 -t 1e-5 --mifrsz 120 --mafrsz 200 -c 40000000 --  
extend 70 -m 1 && \
```

```
python danpos.py dpos \
```

```
/media/sean/My\ Passport/Wig/20-bed-2016/:\
```

```
/media/sean/My\ Passport/Wig/10-bed-2016/ \
```

```
-o 10v20-pos4c -u 1e-15 -t 1e-5 --mifrsz 120 --mafrsz 200 -c 40000000 --  
extend 70 -m 1 && \
```

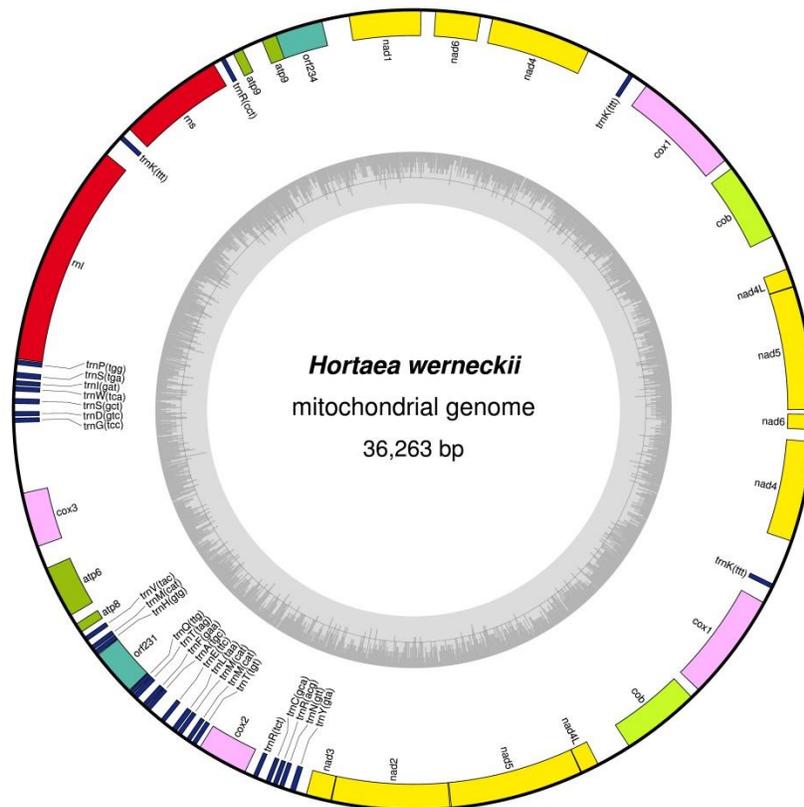
```
python danpos.py dpos \
```

```
/media/sean/My\ Passport/Wig/0-bed-2016/:\
```

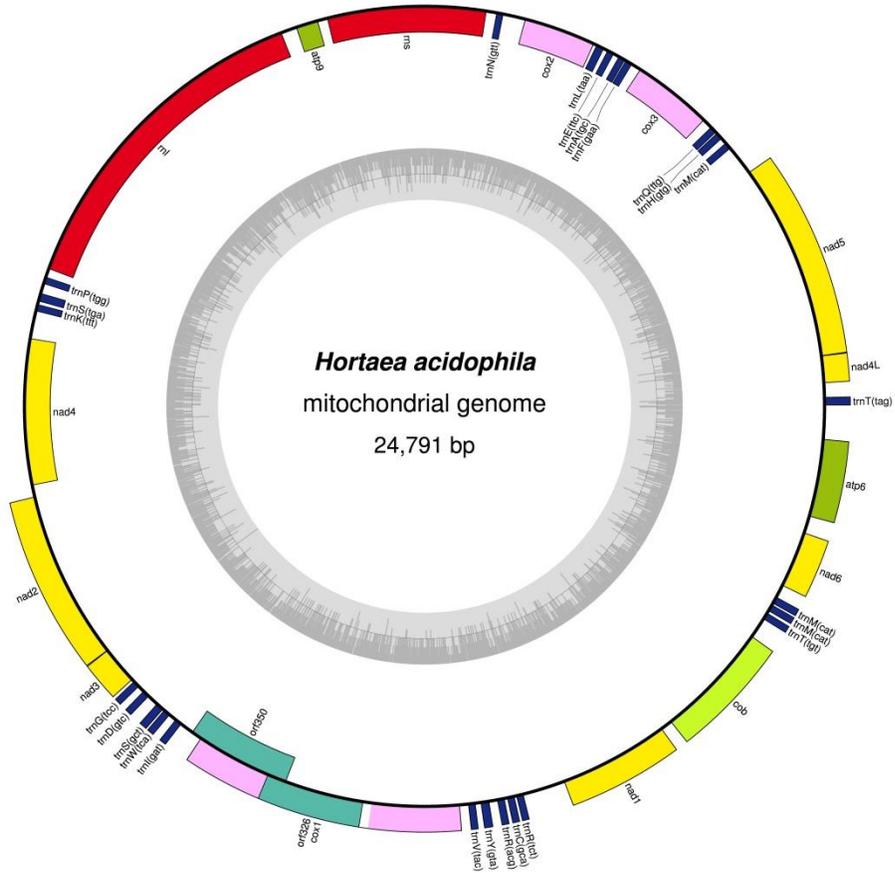
```
/media/sean/My\ Passport/Wig/20-bed-2016/ \
```

```
-o 20v0-pos4c -u 1e-15 -t 1e-5 --mifrsz 120 --mafrsz 200 -c 40000000 --  
extend 70 -m 1
```

Appendix D.



- complex I (NADH dehydrogenase)
- complex III (ubichinol cytochrome c reductase)
- complex IV (cytochrome c oxidase)
- ATP synthase
- ORFs
- transfer RNAs
- ribosomal RNAs



- complex I (NADH dehydrogenase)
- complex III (ubichinol cytochrome c reductase)
- complex IV (cytochrome c oxidase)
- ATP synthase
- ORFs
- transfer RNAs
- ribosomal RNAs
- introns