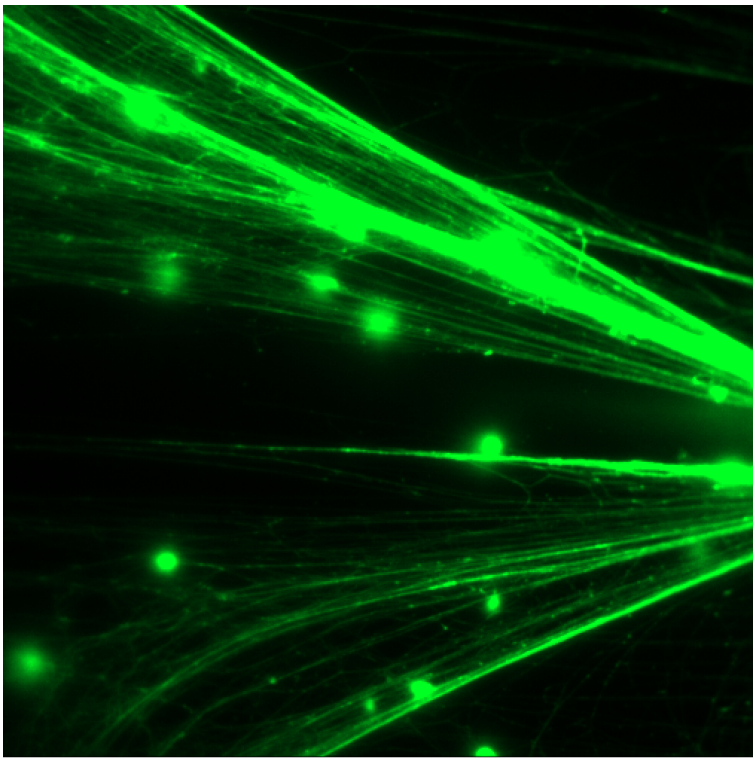

Characterization Of Oxidative Stress-induced Genomic Structural Variations In *Saccharomyces cerevisiae* Through Optical Genome Mapping

Master's thesis in Biotechnology

MOA CARLSSON

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

Characterization Of Oxidative Stress-induced Genomic Structural Variations In *Saccharomyces cerevisiae* Through Optical Genome Mapping

MOA CARLSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Life Sciences
Division of Chemical Biology
Westerlund BioNanoFluidics Research Group
Chalmers University of Technology
Gothenburg, Sweden 2025

Characterization Of Oxidative Stress-induced Genomic Structural Variations In *Saccharomyces cerevisiae* Through Optical Genome Mapping
MOA CARLSSON

Supervisor: Luis Mario Leal Garza, Department of Life Sciences, Chalmers
Examiner: Fredrik Westerlund, Department of Life Sciences, Chalmers

Master's Thesis 2025
Department of Life Sciences
Division of Chemical Biology
Westerlund BioNanoFluidics Research Group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Image showing fluorescently labeled strands of DNA together with yeast cells.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Characterization Of Oxidative Stress-induced Genomic Structural Variations In *Saccharomyces cerevisiae* Through Optical Genome Mapping

MOA CARLSSON
Department of Life Sciences
Chalmers University of Technology

Abstract

The genome of every known organism is composed of deoxyribonucleic acid (DNA). The DNA in organisms like humans, is organized into chromosomes and encodes the genetic instruction necessary for survival. When alterations occur in the genome, they can lead to disease. Structural variations (SVs) are large-scale genomic rearrangements, which can span several megabases, and are difficult to detect using traditional short-read sequencing methods. SVs have been observed in the widely used eukaryotic model organism *Saccharomyces cerevisiae* (*S. cerevisiae*). Particularly in strains with a compromised defense system against endogenous reactive oxygen species (ROS). In this project, optical genome mapping (OGM) was used to study SVs in a ROS-sensitive *S. cerevisiae* strain lacking the TSA1 gene. OGM is a technique that enables sequence-related information to be extracted across long DNA segments. OGM was performed using competitive binding of YOYO-1 and netropsin to fluorescently label DNA, followed by stretching od DNA via a nanofluidic approach. The DNA was then imaged using fluorescence microscopy. A cultivation scheme was developed for both wild-type and ROS-sensitive *S. cerevisiae* strains suitable for the purpose of this study. Additionally, chromosomal DNA extraction methods were evaluated to obtain long DNA fragments suitable for OGM. Fluorescence imaging data was processed to generate single molecule intensity profiles (barcodes). Barcodes where then compared to theoretical intensity profiles of the *S. cerevisiae* reference genome to assess coverage and detect structural changes. Amongst the samples, enough data was obtained to cover the entirety of the *S. cerevisiae* genome approximately 4-11 times. By focusing on chromosome II, comparisons between non-stressed and oxidatively stressed strains revealed intensity profile differences that may indicate SVs. However, further validation, preferably with higher coverage, is necessary to further characterize the genomic changes.

Keywords: DNA, Structural variations, Optical Genome Mapping, *Saccharomyces cerevisiae*, Competitive binding, nanofluidics, reactive oxygen species, TSA1

Acknowledgements

I would like to express my deepest gratitude to my supervisor Luis Mario Leal Garza. Having your support and believing in me during challenging times and good times has made me grow immensely as a scientist. It has been a great experience having your guidance, truly you are an exceptional supervisor.

I would also like to thank Fredrik Westerlund for welcoming me into your research group and giving me the opportunity to carry out my master's thesis there. I am also grateful to the rest of the group for making me feel so welcome and for their help along the way. A special thanks to Radhika Nambannor Kunnath, your encouragement and assistance during key moments truly made a difference.

Additionally, a big thank you to Mikael Molin for giving input and providing the *S. cerevisiae* strains for this project.

Moa Carlsson, Gothenburg, June 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

BME	β-mercaptoethanol
CB	Competitive Binding
DBM	DNA Barcode Matchmaker
HCA	Human Chromosome Alignment
OGM	Optical Genome Mapping
SV	Structural Variation
TBE	Tris-Borate-EDTA
YPD	Yest extract Peptone Dextrose

Contents

List of Figures

List of Tables

1

Introduction

Deoxyribonucleic acid (DNA) carries the genetic information essential for the function, survival, and evolution of all known organisms [1]. The information is packaged in the form of chromosomes, which together constitute the genome. When changes occur to the genome, though mutations or structural variations (SVs), they can lead to disease. Detecting and characterizing these changes is therefore of great importance. Structural variations (SVs) refer to large-scale alterations to the genome, such as deletions, duplications, inversions, or translocations [2]. They have been shown to be associated with numerous types of cancer as well as genetic disorders [3] [4]. Due to their size, which can span up to several megabases, SVs are difficult to detect using traditional short-read sequencing methods. Therefore, techniques that provide long-read sequence information are of interest, such as optical genome mapping (OGM).

OGM is a single-molecule technique that involves fluorescent labeling and stretching of long, intact DNA molecules [5]. The DNA molecules used for OGM are typically over 150 kilobase pairs, which are then imaged using fluorescence microscopy. This enables the extraction of sequence-related information across longer length scales than the currently possible with traditional sequencing.

The widely used eukaryotic model organism *Saccharomyces cerevisiae* (*S. cerevisiae*), has a genome consisting of 16 chromosomes [6]. *S. cerevisiae* has previously shown SVs associates with a compromised oxidative stress defene system [7]. Specifically, a deletion of the TSA1 gene which codes for a peroxiredoxin, impairs the organism's ability to manage endogenous reactive oxygen species (ROS), leading to genomic instability.

However, detailed characterization of these SVs has been limited by the lack of suitable techniques. OGM now provides the potential to detect such variations. Applying OGM to study SVs in *S. cerevisiae* could serve as a foundation for extending this approach to investigate similar mechanisms in human diseases.

1.1 Aim

The aim of this project is to characterize structural variations using competitive binding optical genome mapping in a *Saccharomyces cerevisiae* strain with a compromised defense system against endogenous reactive oxygen species.

To achieve the aim, it can be divided into three parts:

- Develop a *S. cerevisiae* cultivation scheme for studying the development of SVs due to endogenous ROS.
- Evaluate DNA extraction methods to ensure that the resulting DNA is suitable for OGM.
- Perform OGM and identify SVs.

1.2 Limitations

This project will focus on a single strain, wild-type BY4742, and BY4742 with a specific mutation, no other strains will be used throughout the study. Any SVs detected will be attributed to oxidative stress rather than to the strain-specific mutation. While different ROS-sensitive strains with different underlying mutations could result into different types of SVs, it is beyond the scope of this project to investigate. The use of OGM has limitations regarding the size of SVs that can be detected, as it is not possible to obtain base-pair resolution and smaller SVs may therefore go undetected. In addition, the project is limited to a single approach to inducing oxidative stress in cells.

Since the yeast will undergo limited growth in liquid culture (fewer than five generations), it is not entirely possible to exclude genomic changes occurring during this phase. The potential changes to the genomes and the possibility for heterogeneity within a DNA sample represent a limitation in the experimental design.

1.3 Ethical and Societal aspects

At its current stage, this project does not involve any ethical concerns, as no experiments will be conducted on organisms other than *S. cerevisiae*. However, future applications may involve utilizing OGM to detect SVs in the human genome. When the method is ready for human genome analysis, the ethical considerations will need to be reassessed.

2

Theory

The following chapter provides theoretical information on DNA, Optical Genome Mapping (OGM), Structural Variations (SVs), *S. cerevisiae* and Reactive Oxygen Species (ROS), to provide context and understanding of the project.

2.1 DNA

The Central Dogma of molecular biology describes the flow of genetic information within a biological system [1]. It illustrates how genetic instructions are transferred from DNA to RNA, ultimately leading to protein synthesis. At the core of this process is DNA, the molecule that stores all hereditary information. DNA carries genetic information in all known organisms, both eukaryotic and prokaryotic organisms, making it fundamental to life. The genome of an organism is defined as the entirety of DNA present in its cells [8].

2.1.1 DNA Structure

The DNA molecule consists of two antiparallel strands arranged in a double helix structure. Each strand is composed of a linear sequence of nucleotides, which are connected by phosphodiester bonds, forming the sugar-phosphate backbone, as illustrated in Figure 2.1 [9]. Each nucleotide includes a nitrogenous base, either adenine (A), cytosine (C), guanine (G) or thymine (T), which determines the genetic code. These bases pair specifically with complementary bases on the opposite strand through hydrogen bonds. As shown in Figure 2.1, adenine pairs with thymine (A-T) via two hydrogen bonds, while guanine pairs with cytosine (G-C) via three hydrogen bonds. The additional hydrogen bond in G-C pairs contributes to greater stability compared to A-T pairs.

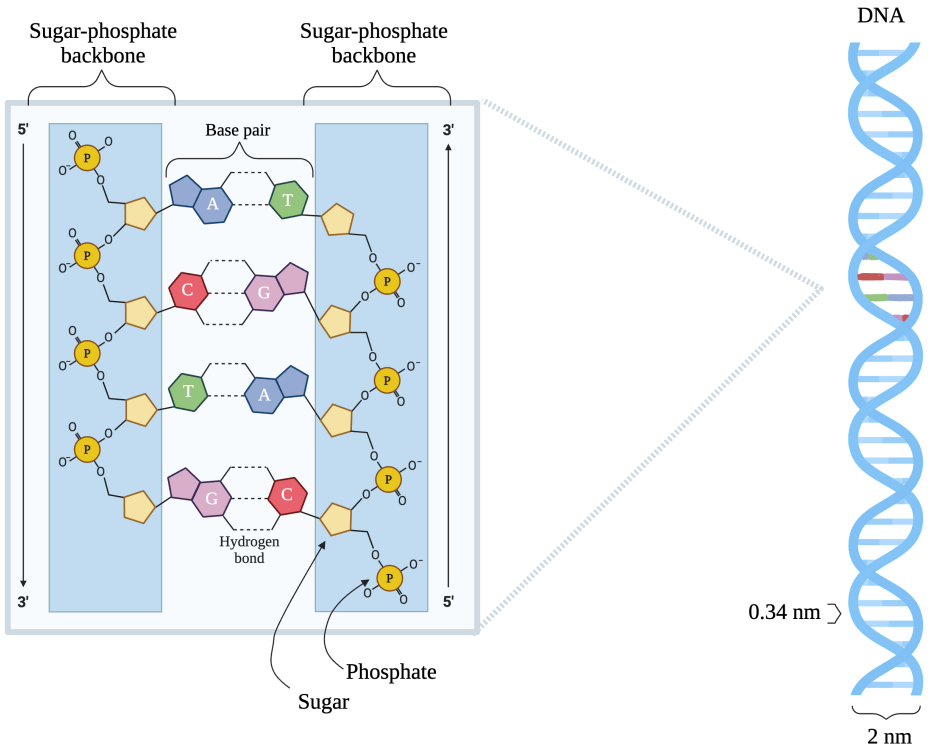

Figure 2.1: Structure of the DNA double helix, highlighting the sugar-phosphate backbone, base pairing, helix width, and the distance between base pairs. Created in https://BioRender.com

Adenine and guanine are classified as purines and consist of two nitrogen-containing rings [9]. Thymine and cytosine are on the other hand classified as pyrimidines and consist of one nitrogen-containing ring. The two strands of DNA are antiparallel and twist around each other to form a double helix. The helical shape of DNA is the most energetically favorable conformation, resulting in the nitrogenous bases being positioned on the inside of the molecule and the sugar-phosphate backbone on the outside.

The phosphate groups in the backbone of the DNA molecule carry negative charges, which gives the entire DNA molecule an overall net negative charge [10]. The diameter of the DNA helix which is determined by the distance between the two backbones across a base pair, is approximately 2 nanometers (nm) [9]. Each base pair is spaced about 0.34 nm apart, and one complete helical turn spans 10 base pairs, which corresponds to a length of approximately 3.4 nm.

2.1.2 Eukaryotic DNA

The organization of DNA differs significantly between eukaryotic and prokaryotic organisms. One difference is where the cells stores the DNA [11]. In eukaryotes, the DNA is enclosed within a membrane-bound nucleus. Prokaryotes, instead of a membrane compartment, have a region where the DNA typically exist, called the

nucleoid. In eukaryotic organisms, the DNA is arranged in chromosomes, where the number of chromosomes depends on the species [12]. The shape of a chromosome is characterized by its arms, with the ends referred to as telomeres and the central region where the arms meet known as the centromere, as illustrated in Figure 2.2.

Due to the limited space within the nucleus, eukaryotic cells must compact their long DNA molecules into a highly condensed and organized structure. The organization of DNA is facilitated by histones, specialized proteins that form histone octamers (complexes composed of eight histone proteins) which package DNA into a more compact structure [12]. Histones are positively charged, which allows them to effectively bind to the negatively charged DNA, facilitating tight packing. DNA wraps around a histone octamer approximately 1.65 times, and thereby forming a structure called nucleosome, as shown in Figure 2.2. A nucleosome represents the fundamental unit of chromatin, which is the complex of DNA and protein that constitutes the chromosomes.

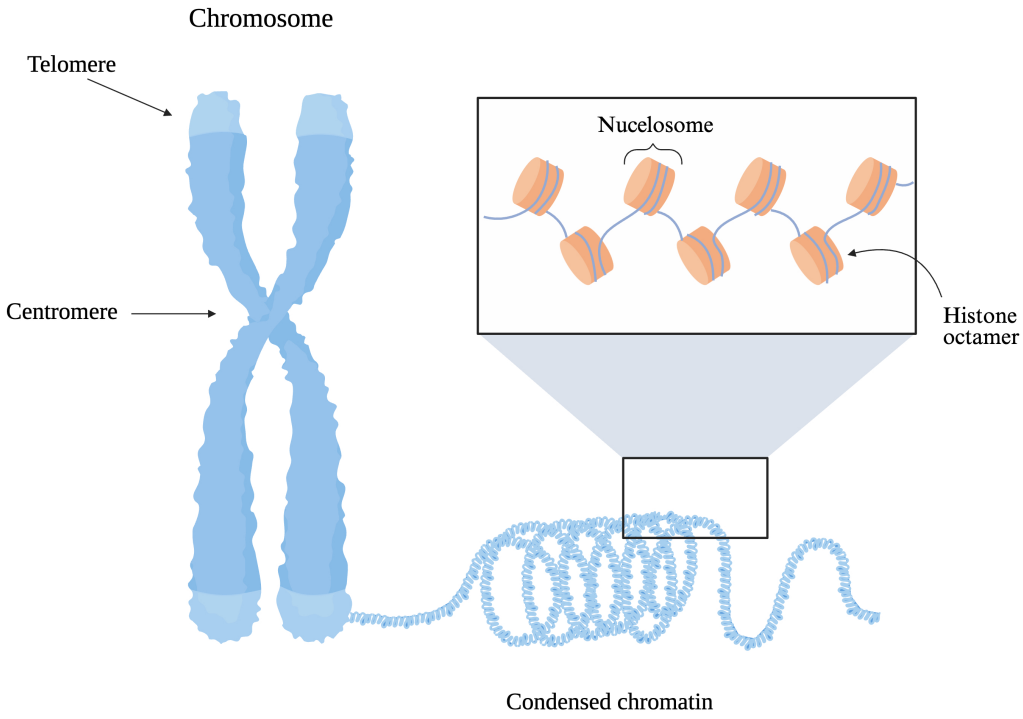

Figure 2.2: Organization of a chromosome, showing condensed chromatin structure and a zoom-in on the nucleosome. The locations of the telomere and centromere are also highlighted on the chromosome. Created in https://BioRender.com

These nucleosomes are connected by linker DNA strands and are further folded into chromatin fibers [12]. Through additional levels of compaction, the chromatin fibers are coiled and condensed into the highly organized structure of a chromosome, enabling efficient storage, regulation, and access to genetic information within the nucleus.

2.2 Advancements in Genomic Mapping Technologies

In recent years, traditional DNA sequencing methods have become increasingly refined, which has been significant for the advancement of genomics as a field [13]. Short-read sequencing refers to technologies that generate DNA fragments of relatively short lengths, typically ranging from 100 to 300 base pairs (bp), which are sequenced individually during a run [14]. Next-generation sequencing (NGS), particularly platforms like Illumina, are a widely used short-read family of methods that has revolutionized genomics by enabling high-throughput, cost-effective sequencing with single-base resolution. NGS is well-established and broadly accessible, and its capacity to generate large amounts of data makes it the preferred method of choice for numerous applications, including detection of single-nucleotide variants (SNV), small insertions and deletions (indels), and copy number variations (CNVs), which require high base-level accuracy [13].

Short-read sequencing has limitations, especially when assembling complex genomic regions or detecting large structural changes that exceed the read length. Fluorescence In Situ Hybridization (FISH) is a method that overcomes some of these limitations which allows for localization of specific DNA sequences directly on chromosomes [15]. However, FISH is most effective for studying genomic changes ranging from several hundred kilobase pairs (kbp) to megabase pairs (Mbp) and has inherent limitations in resolution [16].

To overcome these challenges of short-reads and low resolution, methods that provide information in the kilobase pair (kbp) scale, such as optical genome mapping, may be employed [5].

2.3 Optical Genome mapping

Optical Genome Mapping (OGM) is a technique in which fluorescently labeled individual DNA molecules are analyzed by stretching them out and imaging them using fluorescence microscopy [5]. This method enables the collection of sequence-related information across large genomic regions. OGM provides long-scale optical maps corresponding to a minimum of 150 kbp [17]. This capability allows for the detection of large-scale genomic changes that are often missed with NGS. In addition, OGM can visualize complex regions of the genome, such as repetitive sequences.

Another possibility of OGM is *de novo* genome assembly, which enables the construction of reference maps for entire genomes of organisms [18]. These maps can then be used to detect large-scale genomic variations.

2.3.1 DNA Stretching in OGM

For OGM purposes, DNA can be stretched either by anchoring it to glass or confining it within nanochannels [5]. In the latter approach, an applied force drives the DNA

into nanochannels on a nanofluidic chip, where it becomes elongated. Compared to glass-based stretching methods, nanofluidics offers advantages that include uniform DNA stretching and a higher throughput data collection [19].

Elongation of DNA by confinement occurs due to several factors. DNA molecules carry a negative charge, and when confined within nanochannels that also have negatively charged walls, electrostatic repulsion occurs. The repulsion, combined with the restriction of movement due to the dimensions of the nanochannels, causes the DNA to extend [20].

Some physical aspects of DNA behavior when confined within nanochannels can be described using its persistence length (P) and contour length (L). The persistence length describes the stiffness of DNA, representing its resistance to bending under thermal fluctuations. In contrast, the contour length refers to the total length of the DNA backbone.

The behavior of DNA in nanochannels is also influenced by the dimensions of the channels, specifically the average cross-sectional diameter (D), as this determines the configuratios that the DNA can assume [20]. DNA typically adopts a coil formation when free in solution, as shown in Figure 2.3. However, when confined within a nanochannel, the DNA may take on different regimes depending on the channel dimensions. The de Gennes and the Odijk regime are illustrated in Figure 2.3. The de Gennes regime describes the configuration of DNA as a series of blobs, occurring when the channel size is larger than the DNA's persistence length ($D > P$). In contrast, the Odijk regime occurs when the channel size is smaller than the persistence length ($D < P$). Under these conditions, DNA is less likely to form blob-like structures compared to the de Gennes regime and hence a more extended configuration [20].

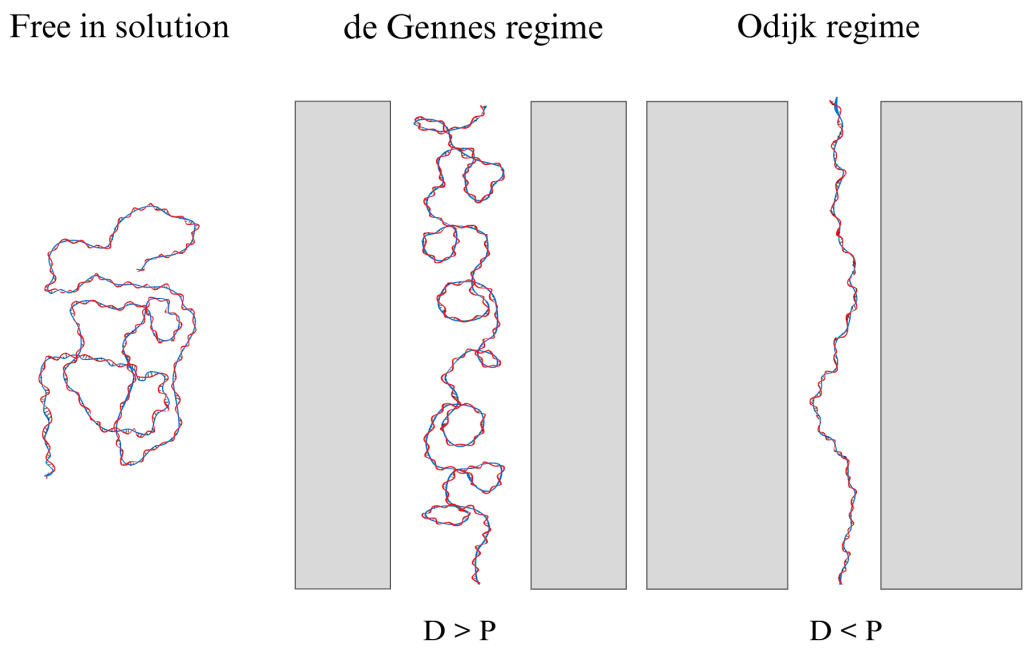

Figure 2.3: Configurations of DNA when free in solution and when confined within nanochannels. In the de Gennes regime (D > P), DNA forms a series of blob-like structures, while in the Odijk regime (D < P) it may adapt a more extended configuration.

The degree of DNA extension can vary depending on solvent conditions, such as salt concentration, which reduces the electrostatic repulsion along the DNA. Therefore, to accurately determine the DNA extension when performing OGM, a size reference is required. λ-DNA is commonly used for this purpose. λ-DNA is linear and double-stranded phage DNA widely used for research purposes and is obtained from an *Escherichia Coli* bacteriophage [21]. When fully extended, λ-DNA has a contour length of around 16.5 μm, consisting of 48.5 kbp [22].

2.3.2 Labeling strategies for OGM

Several DNA labeling strategies have been developed for OGM. The available labeling strategies can be categorized as either sparse labeling or continuous labeling [23]. Sparse labeling involves the introduction of fluorescent tags at specific sequence locations along the DNA molecule. Fluorescent tags can be, for example, incorporated using labeled nucleotides during the repair of controlled single-strand nicks or cuts. These nicks or cuts can be introduced by enzymes such as nicking enzymes or CRISPR-Cas9 systems. The resulting fluorescent pattern creates a barcode-like map where tagged locations correspond to specific DNA sequences. Continuous labeling, on the other hand, results in a barcode represented by a continuous fluorescent intensity profile along the DNA strand. This profile can be achieved by more frequent labeling of the DNA, using denaturation mapping to exploit differences in melting temperatures between sequences or through affinity-based binding with molecules like intercalating dyes [23], [24].

BioNano Genomics has commercialized OGM that uses a sparse enzymatic labeling approach and electrophoretic force to guide DNA into nanochannels. This method enables the collection of sequence-related information across much longer length scales compared to traditional sequencing techniques [25]. Their Direct Labeling and Staining (DLS) approach has allowed for a standardized labeling process and thereby enhances scalability as well as accessibility.

In contrast, Westerlund's BioNanoFluidics research group at Chalmers University of Technology has developed an approach that combines nanofluidics with competitive binding (CB) based labeling [26]. The competitive binding of two ligands results in a sequence-dependent continuous fluorescent signal that can be used for OGM. The two ligands used in Westerlund's CB strategy are YOYO-1 and netropsin.

YOYO-1 is virtually non-fluorescent in solution [27]. However, when bis-intercalated into double-stranded DNA, it exhibits strong fluorescence due to the restriction of its movement. The molecule is a homodimer and belongs to the cyanine dye family and is widely used for probing and visualizing DNA. YOYO-1 exhibits an emission peak at 509 nm when excited at around 470 nm, resulting in green fluorescence. The molecule consists of two oxazole yellow (YO) units, which are aromatic ring structures connected by a methane bridge. When bound to DNA, each YOYO-1 molecule spans approximately 4 base pairs [28]. Studies have shown that YOYO-1 binding causes conformational changes to the DNA, including some extension of the contour length.

Netropsin, on the other hand, is an antibiotic and binds to DNA through hydrogen bonding to adenine (A) and thymine (T) bases [29]. Additionally, netropsin prefers the minor groove of the DNA helix due to it having more possibilities for hydrogen bonding acceptors. For netropsin to bind, it is essential to have at least three consecutive AT base pairs.

By mixing YOYO-1 and netropsin with DNA, netropsin binds to AT-rich regions, blocking these sites and allowing YOYO-1 to primarily bind to GC-rich regions [30]. Since YOYO-1 is fluorescent, this competitive binding produces variations in fluorescence intensity along the DNA molecule, where higher intensity corresponds to GC-rich regions and lower intensity to AT-rich regions, illustrated in Figure 2.4. The fluorescence imaging data is used to generate barcodes for each DNA molecule, corresponding to the fluorescence intensity profile that reflects the GC content at specific positions along the molecule.

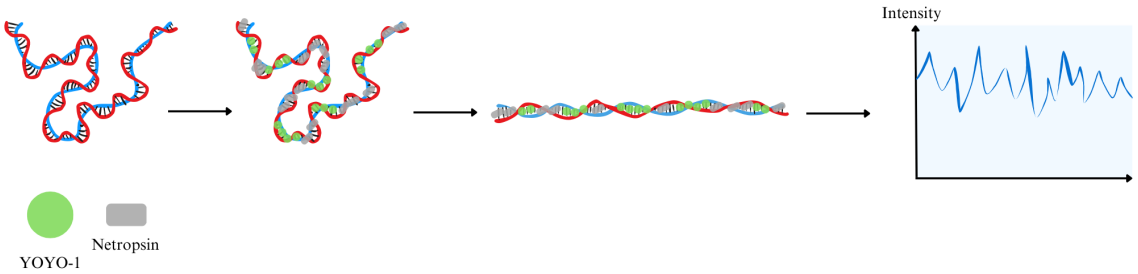

Figure 2.4: Illustration depicting YOYO-1 and netropsin binding to DNA, demonstrating how DNA elongation facilitates the generation of an intensity profile, also referred to as a barcode.

2.4 Structural Variations

In the context of genomics, structural variations (SVs) refer to large genomic rearrangements such as insertions, deletions, translocations and copy number variations [2]. SVs are characterized by a wide range of sizes, typically spanning from 50 base pairs to several megabases. While NGS can detect small SVs, and FISH can resolve larger SVs spanning multiple megabases, there remains a detection gap for SVs in the kilobase to low-megabase range. One such approach is OGM, which enables the detection of SVs in this size range and is commercially available through BioNano Genomics [25][31].

In humans, SVs can arise under diverse conditions and may have significant consequences related to various diseases [32]. For instance, SVs are closely associated with numerous cancers, including breast cancer and hematologic malignancies [3] [4]. Therefore, developing methods to effectively detect SVs and investigating the mechanisms underlying their formation are critical for understanding their role in health and disease.

SVs may result from DNA damage that is not properly repaired by the cell's DNA repair systems. This damage can originate from endogenous factors, meaning that they arise from internal cellular functions [33]. Endogenous damage could for example be improper DNA repair, spontaneous double-stranded brakes or accumulation of reactive oxygen species. Alternatively, exogenous sources of DNA damage could also lead to SVs arising. These include for instance ionizing radiation or chemical exposure. If such damage is not repaired correctly, it may destabilize the genome and potentially leading to SVs.

2.5 *Saccharomyces cerevisiae*

The budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) also known as baker's yeast is a well studied model organism and has been widely used for both research and industrial purposes for many years [6]. The name budding yeast comes from its

particular mechanism of reproduction, where the daughter cell forms as an outgrowth or bud on the parent cell. Once the bud is fully developed, it pinches off and a genetically identical cell has been created.

Its major importance in the world of research stems from it being an eukaryote that shares many fundamental biological processes with higher eukaryotes such as humans [6]. *S. cerevisiae* is also highly susceptible to genetic modifications and therefore excellent for studying gene functions.

2.5.1 The genome of *S. cerevisiae*

Through a collaboration that reached world-wide, the genome of *S. cerevisiae* became the first eukaryotic genome to be completely sequenced [34] [35]. The work was released in 1996 and involved contributions from hundreds of researchers. The achievement was groundbreaking and allowed for valuable insights into the fundamental biology of eukaryotes. The strain S288C was used as a reference strain during its sequencing, which is a widely used strain in laboratory work. The BY4742 strain of *S. cerevisiae* is also commonly used in laboratory settings and is a derivative of S288C, with only minor genetic differences [36].

S. cerevisiae has a genome composed of 16 nuclear chromosomes and a mitochondrial chromosome ranging from 68 to 86 kbp. Together, these comprise approximately 12,200 kbp and encode around 6000 protein coding genes [6]. The 16 chromosomes and their corresponding sizes in kbp are shown in Figure 2.5, ranging from the smallest, chromosome I (∼230 kbp), to the largest, chromosome IV (∼1532 kbp) [37].

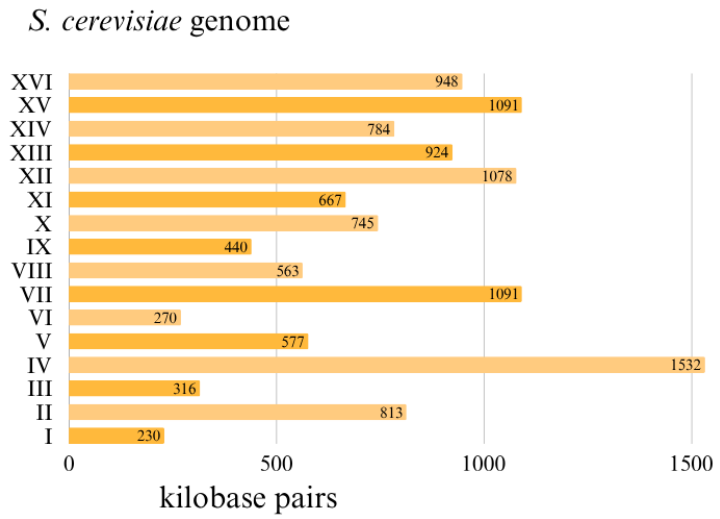

Figure 2.5: All 16 nuclear chromosomes of *S. cerevisiae* shown with their corresponding sizes in kilobase pairs (kbp). Chromosome I is the smallest, at approximately 230 kbp, while chromosome IV is the largest, at around 1532 kbp. Size data obtained from the Saccharomyces Genome Database (SGD)

There are regions of the *S. cerevisiae* genome that are difficult to sequence, leading to limitations in the reference genome [38]. This is particularly true for repetitive regions, such as those found on chromosome XII, where tandem repeats, approximately 150 to 200 in number, make up around 60% of the entire chromosome.

2.5.2 The cell wall of *S. cerevisiae*

A crucial component of the cellular structure of yeast is the cell wall, which for *S. cerevisiae* corresponds to approximately 30% of the dry weight of the cell [39]. The cell wall constitutes four major functions; to maintain the shape of the cell, protection against any physical stress, stabilization of internal osmotic conditions as well as a scaffold for proteins [40]. The cell wall has a width of 110-200 nm where the major components are polysaccharides, mannoproteins and lipids and the main components are illutrated in Figure 2.6 [41]. Out of the polysaccarides, β-glucans compose the major part of the cell wall where β-1,3-glucan forms the primary structural framework of the cell wall and β-1,6-glucan is an essential for maintaining wall integrity and acts as a linker connecting β-1,3-glucan to manoproteins and chitin. Chitin is another polysaccaride of much lower cell wall density, around 1 to 2%, and can be found of the highest concentration in the septum or bud neck, formed as a scar resulted from budding. Mannoproteins, located in the outermost layer of the cell wall, interact with the environment and can exhibit enzymatic activity.

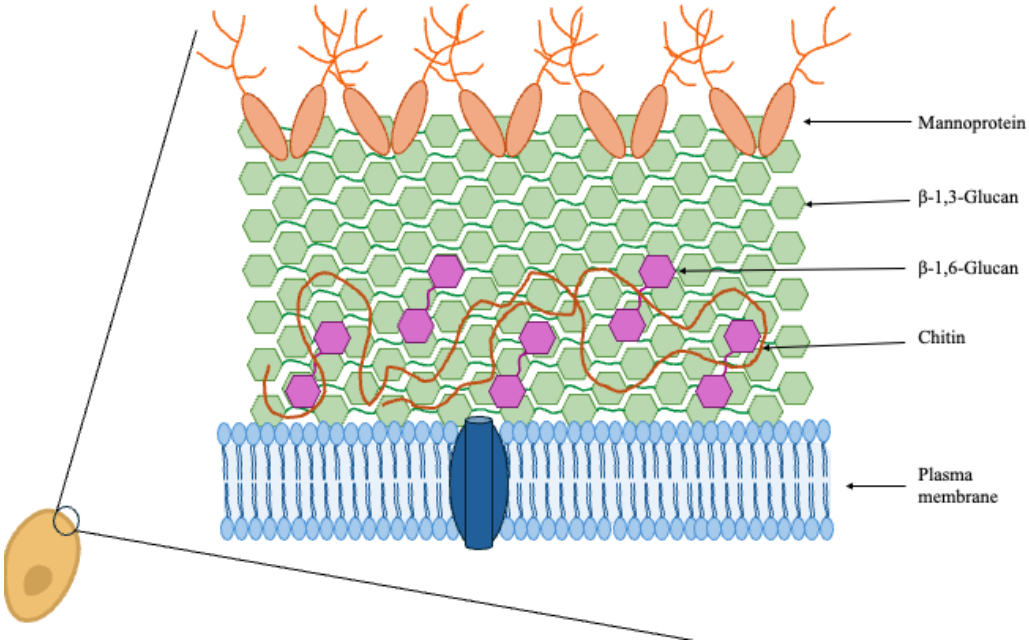

Figure 2.6: llustration of the *S. cerevisiae* cell wall, highlighting its components: mannoproteins, β-1,3-glucans, β-1,6-glucans, chitin and the plasma membrane.

The structural complexity of the yeast cell wall must be considered when planning methods to access intracellular components. Degrading the cell wall results in the creation of spheroplasts, which is what cells are called when the cell wall has

been partially or completely removed [42]. To create spheroplasts, strategies such as mechanical, chemical or enzymatic methods may be employed. For enzymatic degradation, enzyme preparations of lyticase or zymolyase can be used.

Lyticase is a lytic enzyme system purified from *Arthrobacter luteus* and primary consists of β-1,3-glucanase, witch hydrolyses β-1,3-glucan, a major component of the *S. cerevisiae* cell wall [43]. Zymolyase, on the other hand, is a mixture of enzymes in which β-1,3-glucanase is also the dominant source of activity. Like lyticase, zymolyase can be extracted from *Arthrobacter luteus*, but also contains enzymatic activities such as proteases and mannan-degrading enzymes [44]. The more complex composition of zymolyase compared to lyticase, with its ability to target different parts of the cell wall, may lead to a higher success rate in spheroplast generation.

2.5.3 Cultivation of *S. cerevisiae*

Cultivation of *S. cerevisiae* is most commonly performed in liquid medium or on agar plates [45]. As mentioned earlier, the yeast is widely used in laboratory environments and largely because it is easy to cultivate. Yeast extract Peptone Dextrose (YPD) is one of the most commonly used growth media. YPD can be prepared both as liquid broth or solidified with agar to create plates.

In liquid culture, the availability of nutrients is generally more uniform as a result of constant mixing, which reduces the formation of gradients. Cultivation on agar plates on the other hand may create nutrient gradients as cells grow, depleting local nutrients and causing a variation of nutrient availability within a colony [46]. These gradients may subject cells to stress which can result in modifications to the composition of the cell wall. Studies suggest that solid-grown yeast can have a higher content of mannoproteins and glucans compared to liquid-grown yeast, making the cell wall more rigid [47].

The growth of *S. cerevisiae* in a liquid culture can be categorized into three distinct phases: lag, exponential (log) and stationary-phase, all illustrated in Figure 2.7 [48]. The lag phase describes the period before the cells begin exponential growth, during which they adapt to their environment and only minimal cell replication occurs. Under optimal growth conditions, this phase usually lasts between 30 to 90 minutes. once the cells have adapted, they enter the next phase phase, namely the exponential phase, or also often refereed to as the log phase. Here, the the number of cells increase exponentially, and *S. cerevisiae* has a doubling time of approximately 90 minutes. The stationary phase begins when nutrients are depleted and can no longer support exponential growth. To monitor cell growth in liquid culture, a spectrophotometer can be used to measure optical density (OD), which correlates with cell concentration. Continuous OD measurements allow real-time tracking of the growth curve and identification of the current growth phase.

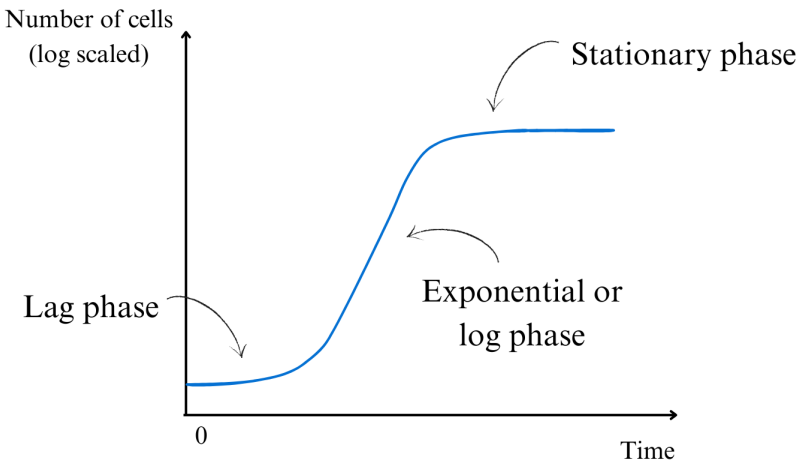

Figure 2.7: Growth curve illustrating the growth of *S. cerevisiae* in liquid culture showing the three main phases: lag, exponential (log) and stationary.

The growth monitoring when cultivating *S. cerevisiae* on solid YPD agar plates differs from how it is done when monitoring a liquid culture. Instead of measuring OD or concentration of cells, growth is observed by the formation and expansion of colonies growing on the plate [49]. Cultivation on solid YPD agar plates takes generally longer time compared to in liquid culture.

2.6 Reactive Oxygen Species (ROS)

Reactive oxygen species (ROS) is a term that is used to describe a group of highly reactive compounds, derived from molecular oxygen, which are formed through reduction-oxidation (redox) reactions [50]. These ROS-generative redox reactions may take place during metabolic processes in the cell (endogenous) or as a result of some sort of exposure from the surrounding environment (exogenous) such as ionizing radiation [51]. A primary source of endogenous ROS production is in the mitochondria as byproducts from the electron transport chain where molecules such as superoxide anion radical (O_2^-) and hydrogen peroxide (H_2O_2) are formed. Within a cell or organism, there exists a phenomenon known as oxidative eustress, where reactive oxygen species (ROS) play essential roles in cellular signaling and homeostasis. However, excessive exposure to ROS can lead to oxidative stress, a condition that causes damage to cellular components and disrupts normal function [50]. ROS have been studied in relation to various diseases, carcinogenesis and in connection to aging.

2.6.1 Endogenous ROS scavenging

The major endogenous ROS hydrogen peroxide, has been shown to have signaling attributes connected to antioxidant defense systems. The cell has developed several

defense mechanisms to excessive amounts of ROS and these systems have been identified in all kingdoms of life. These systems include peroxiredoxins, a family of small antioxidant proteins [52]. Peroxiredoxins scavenge peroxide through reactive cystine residues that can catalyze the reduction of H_2O_2 and using a thioredoxin as electron donor. In *S. cerevisiae*, the TSA1 gene encodes the peroxiredoxin Tsa1p , which plays a critical role in scavenging endogenous ROS [7]. In the absence of Tsa1p, ROS levels rise, leading to oxidative stress and an increased likelihood of mutations. Defects in DNA repair systems, such as the base excision repair (BER) and nucleotide excision repair (NER) pathways, have also been shown to elevate ROS levels and thereby increase mutations rates and promotion of large-scale genome rearrangements.

SVs in *S. cerevisiae* have been observed in strains with a deletion of the TSA1 gene and with severely compromised BER and NER pathways, demonstrated by Contour-Clamped Homogenous Electric Field (CHEF) electrophoresis [7]. CHEF electrophoresis is a type of pulse-filed electrophoresis which can separate large DNA molecules. However, due to the limitations of this technique, including low resolution, certain genomic changes could not be detected and therefore remained characterized. Thus, studying genomic changes arising from these genetic variations that causes oxidative stress is of interest using higher-resolution techniques, such as OGM.

It has been shown that chromosome II is more susceptible to SVs caused by oxidative stress, compared to, for instance, chromosome III [7]. This was demonstrated using CHEF electrophoresis, a technique that more easily resolves smaller chromosomes like chromosome III, enabling a direct comparison. These findings suggest that chromosome II is particularly prone to oxidative stress-induced genomic instability.

Additionally, chromosome II contains a high density of retrotansposon sequences, particularly in the regions surrounding its centromere. These mobile genetic elements, especially when present in repeated copies, are know to act as hotspots for genome rearrangements [53]. This heightens the susceptibility to SVs.

3

Methodology

The methodology of the project included the culture of *S. cerevisiae*, extraction of genomic DNA, the use of OGM with a competitive binding labeling strategy, as well as the processing and analysis of the obtained data.

3.1 Cultivation of *S. cerevisiae*

The S. cerevisiae strain used in this project was BY4742, along with a mutant variant lacking the TSA1 gene ($\Delta tsa1$), which is known to be more susceptible to oxidative stress.

S. cerevisiae was cultivated using both liquid culture and solid agar plates. The methodologies for each approach are described in Sections 3.1.1 and 3.1.2, and were primarily followed during experiments aimed at establishing a cultivation scheme suitable for the objectives of this project. The resulting cultivation scheme is shown in Section 3.1.3 and was followed for experiments to asses changes to the genome due to oxidative stress.

3.1.1 Cultivation in liquid YPD media

A liquid culture of *S. cerevisiae* was inoculated by transferring a yeast colony from an agar plate into a sterile Falcon tube containing 10 ml of YPD media. Two replicates were prepared, along with a control containing only YPD media to confirm sterility during handling of yeast. The tubes were incubated overnight at 30°C in a shaker set at 200 rpm.

On the following day, a small volume of the overnight yeast culture was transferred to a sterile Erlenmeyer flask containing YPD media. For monitoring growth optical density measured at a wavelength of 600 (OD_{600}) was used. The goal was to adjust the starting culture to an OD_{600} of 0.1. For example, this could be achieved by diluting 100 ml of YPD media with 2 ml of yeast culture, followed by adjusting with additional yeast culture as needed to reach the target OD_{600}. Once the culture reached an OD_{600} of 0.1, the flask was incubated at 30°C with shaking at 200 rpm for 6-8 hours, until an OD_{600} of approximately 1.0 was reached. Then 100 ml of the culture was divided into three sterile Falcon tubes with an estimated total of 6×10^8

cells in each tube. The tubes were centrifuged at 5000 x *g* for 10 minutes at 4°C. Following centrifugation, the supernatants were discarded, and the cell pellets were frozen at -20°C for storage.

3.1.2 Cultivation on agar YPD plates

When cultivating *S. cerevisiae* on YPD agar plates, a single colony was picked using an inoculation loop and streaked onto a fresh YPD agar plate. The plate was incubated at 30°C for at least 3 days. After incubation, the newly formed colonies could be used for subsequent experiments.

3.1.3 Cultivation Scheme and Creation of Passages

Taking into account the two DNA extraction techniques evaluated in Section 4.1, a yeast cultivation scheme involving five serial passages was established. The cultivation scheme includes cultivation of two strains of BY4742 *S. cerevisiae*, wild-type (WT) and Δtsa1 (TSA), as illustrated in Figure 3.1. Five passages have been shown to be sufficient to observe genome rearrangements, as described by Degtyareva et al. [7], and hence why this number of passages was chosen. Here, one passage is defined as re-streaking one colony of yeast onto a fresh YPD agar plate and letting it grow at optimal conditions for 3 days. As indicated in the illustration of Figure 3.1, four different sample types are in the end created named by the type of *S. cerevisiae* strain and the amount of passages it has been subjected to: WT 0, WT 5, TSA 0 and TSA 5.

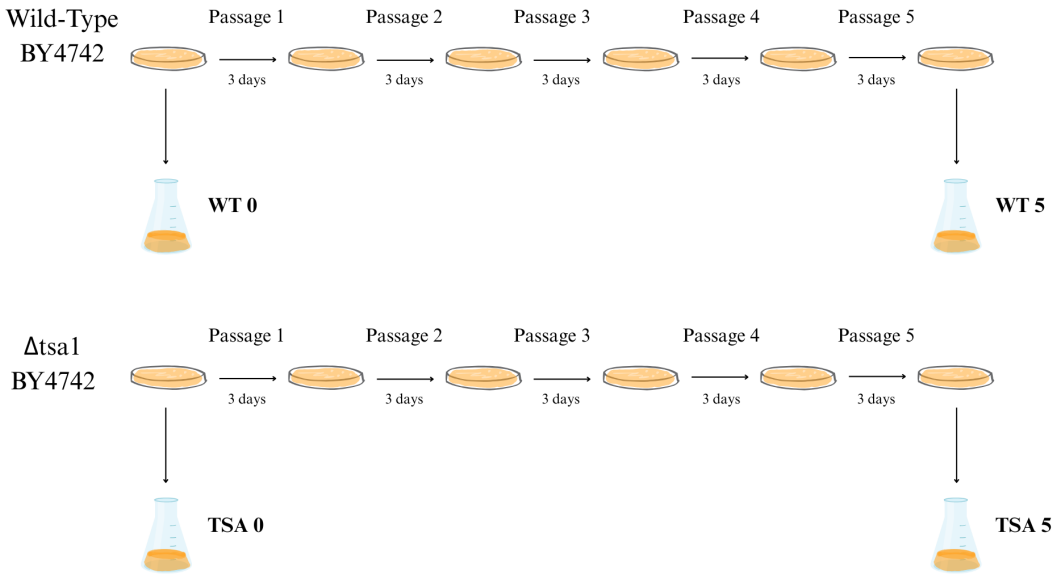

Figure 3.1: Cultivation and passage scheme for two strains of BY4742 *S. cerevisiae*, WT and Δtsa1, indicating the sample names WT 0, WT 5, TSA 0 and TSA 5.

Before creating Passage 1, colonies were picked to create liquid cultures with the purpose to increase the mass of yeast. This was done to ensure sufficient amount of yeast for genomic DNA extraction as well as getting a sufficient DNA concentration for down stream experiments. Liquid cultures were also created after passage 5 for the same due to the same reasoning as described previously.

For each liquid culture created, one colony was placed into a 100 ml shake flask containing 10 ml of YPD media. This corresponds to approximately 1.5×10^5 amount of cells. One extra liquid culture for each type of strain was created for each experiment to be able to monitor the growth using a spectrophotometer to measure OD_{600}. In addition to monitoring yeast growth, OD_{600} measurements were also used to ensure that the cells did not grow for more than five generations, thereby minimizing the risk of potential genomic changes.

When the yeast had grown for sufficient time, not surpassing the set limit of 5 generations, the content of each separate shake flasks were transferred to separate Falcon tubes for centrifugation. Each sample were then centrifuged for 10 minutes at 5000 x g to pellet the cells. Once the supernatant containing the YPD media was removed by decantation and pipetting, the pellets were frozen at -20°C for storage.

This was done for the four samples indicated in the cultivation scheme in Figure 3.1.

3.2 Genomic DNA Extraction

Extraction of genomic DNA from *S. cerevisiae* was performed either by embedding DNA in agarose plugs using CHEF Yeast Genomic DNA Plug Kit from BIO-RAD or through a strategy including separation on glass beads with Monarch® HMW DNA Extraction Kit for Tissue.

3.2.1 Preparation of Agarose Embedded DNA

The genomic DNA extraction for yeast cultivated on YPD plates using the CHEF Yeast Genomic DNA Plug Kit from BIO-RAD was performed according to the manufacturer's protocol, with modifications to accommodate yeast grown on solid media rather than liquid cultures.

To prepare agarose DNA plugs, 62.5 µl of Cell Suspension Buffer (CSB) was added to 1.5 ml Eppendorf tubes, one for each plug. Yeast colonies were transferred directly from the plate into the CSB using small inoculation loops. The tubes were then frozen at -20°C for at least 16 hours. After freezing, the tubes were thawed and placed on a 37°C heating block. The next step included addition of enzymes for cell wall degradation, which was either lyticase or Zymolyase Ultra and the volume used varied depending on experiment. Separately, 2% CleanCut Agarose soulution (CCA) was carefully melted in a microwave until fully clear and liquefied. 40 µl of CCA was then added to each yeast and CSB solution while carefully mixing by pipetting, keeping on a 50°C heating block. A total of 100 µl of the mixture was transferred to a plug mold and incubated at 4°C for approximately 20 minutes to allow the agarose to solidify. Once solidified, the plug was gently pushed out into a new 1.5 ml Eppendorf tube. Depending on the enzyme used and the volume, additional enzyme and 250 µl of lyticase buffer were added to the tube. The solutions were then incubated at 37°C for a time dependent on the experiment. After incubation, the solutions were removed and the agarose plugs were washed with MQ water. 250 µl of proteinase K reaction buffer together with 10 µl proteinase K stock were added to each plug and incubated at 50°C overnight.

After incubation, 1X Wash Buffer was prepared from 10X Wash buffer by diluting with MQ water, and each plug was washed 4 times for 1 hour. The finished agarose DNA plugs were then stored in the 1X Wash Buffer at 4°C.

3.2.2 Extraction of Agarose Embedded DNA

To extract DNA embedded in agarose plugs, 600 µl of 1X rCutSmart buffer was prepared for each plug. The plugs were washed twice with the buffer, each time for 15 minutes. When the plugs had been washed, 78 µl of MQ water and 20 µl of 10X rCutSmart buffer were added to the each plug and incubated for 10 minutes at

70°C to let the agarose melt. The plugs were additionally incubated for 10 minutes at 42°C to prepare for the following step where the enzyme agarase is added. After incubation, 2 µl of agarase was added to each sample and the solutions were mixed carefully by swirling around with a pipette tip. The agarase was left to work for approximately 3 hours while incubating at 42°C. During the incubation, the samples were checked every 30-60 minutes to ensure that the plugs were dissolving.

3.2.3 Modified Monarch DNA extraction

DNA extraction from yeast grown on YPD plates using the Monarch® HMW DNA Extraction Kit for Tissue was performed according to the protocol provided for yeast DNA extraction with modifications specifically for yeast grown on YPD plates and to obtain DNA molecules of 150 kbp or longer.

Yeast colonies were picked using inoculation loops and added to 1.5 ml Eppendorf tubes containing 62.5 µl CSB. The tubes were then placed in a -20°C freezer overnight or longer. After freezing, the tubes were thawed and placed on a 37°C heating block. 250 µl lyticase buffer, 40 µl MQ water and 11.5 µl of Zymolyase Ultra were added to each tube and incubated at 37°C for the amount of time that was evaluated for the specific experiment. After incubation, the samples were spun down for 3 minutes at 300 x g and the supernatants were removed. 150 µl of cold 0.5 M NaCl was added to each tube and pipetted up and down with wide-bore tips to resuspend the pellets. A master mix was prepared for each sample, containing 450 µl of HMW gDNA Tissue Lysis Buffer and 15 µl of proteinase K. Each sample received 450 µl of the master mix and were incubated at 56°C overnight. After the overnight incubation, 10 µl of RNase was added and the samples were mixed by inverting approximately 10 times before incubation for an additional 20 minutes at 56°C. The samples were then centrifuged for 3 minutes at 16 000 x g to get rid of cellular debris and the supernatants were transferred to new 1.5 ml Eppendorf tubes. 0.5 volumes of protein separation solution was added and mixed by inverting for 1 min. The samples were then centrifuged for 10 minutes at 16 000 x g, creating a phase separation where the DNA was present in the upper phases. The upper phases were then transferred with wide-bore tips to Monarch 2 ml Tubes. 2 DNA Capture Beads were added to each sample as well as 0.7 volumes of isopropanol. The samples were then placed on a vertical rotating mixer at 10 rpm for 5 minutes. Here, the DNA attached to the beads. The liquid was then removed by pipetting and 500 µl of gDNA Wash Buffer was added immediately and mixed by inverting 2-3 times. The washing step was repeated once and the solutions were removed. Bead Retainers were placed in Monarch Spin Collection Tubes, one for each sample, and the beads were poured into the retainers. To remove any residual solution from the beads, the tubes were pulse spinned in a bench top minicentrifuge. The Bead Retainers were then separated from the collection tubes and the beads poured into new Monarch 2 ml Tubes while the Bead Retainer were placed on new 1.5 ml Eppendorf tubes. 50 µl of Elution Buffer was added immediately after and the samples were placed on a 56°C heating block for a minimum of 5 minutes. During the incubation, the samples were checked by horizontally wiggling the tubes to make

sure that the beads were not stuck to the bottom. The beads together with the elute were then poured into the bead retainers and centrifuged for 30 seconds at 12 000 x g to ensure all elute had gone through the retainers. The Bead Retainers were then removed and the remaining elutes were pipetted up and down carefully 5 times with wide-bore tips. To ensure homogeneously in the samples before DNA quantification, the samples were either placed on a 37°C heating block for one hour or stored in room temperature overnight. The extracted DNA was stored at 4°C.

3.2.4 DNA Concentration Measurements with Qubit Fluorometer

To quantify the concentration of DNA obtained from each extraction, a Qubit Fluorometer with the ThermoFisher Qubit dsDNA Broad Range (BR) Assay Kit was used. The total volume of the measurement solution was consistently 200 µl, consisting of 5 µl of the DNA sample and 195 µl of Qubit Working Solution. The kit included two standards for preparing a standard curve, 10 µl of Standard #1 and Standard #2 were each mixed with 190 µl of Qubit Working Solution to achieve a total volume of 200 µl per standard. Before measuring the standards or samples in the Qubit Fluorometer, the solutions were vortexed for approximately 5 seconds. Ideally, three measurements were taken for each sample, one from the top, middle and bottom of the tube. The average DNA concentration of the three measurements were used for further calculations. However, in a similar way, if the DNA concentration was too low to be detected, ThermoFisher Qubit dsDNA High Sensitivity (HS) was used to verify.

3.3 Optical Genome Mapping

Optical genome mapping was performed using continuous labeling with competitive binding and stretching of DNA on a nanofluidic chip.

3.3.1 Staining of DNA

To perform OGM using competitive binding, the DNA molecules were stained with the intercalating fluorescent dye YOYO-1 and the AT-specific binding molecule netropsin.

Depending on the quantity DNA extracted and quantified using Qubit measurements, the total amount of DNA for the staining was determined to be between 2 and 8 µMbp (Molar base pair). The total amount of DNA included 70% extracted DNA and 30% λ-DNA. The quantities of YOYO-1 and netropsin added to the reaction were decided based on the desired molar ratio of 10:1:300 (base pairs DNA: YOYO-1: Netropsin). The staining reaction was performed in a total volume of 10 µl in 0.5X TBE buffer. An example of the components, concentrations and volumes used in a staining reaction is shown in Appendix A.1.

When all components had been added together, the solution was put on a 50°C heat-

ing block for 30 minutes. This was to allow for YOYO-1 and netropsin to bind and for the competitive binding to happen. Before imaging using fluorescence microscopy, each sample was diluted with 137 µl MQ water and 3 µl β-mercaptoethanol (BME). This dilution created an environment where negatively charged DNA molecules could self-repel, due to the salinity being reduced, promoting sufficient stretching.

3.3.2 Fluorescence Microscopy

For DNA stretching and imaging, a nanofluidic chip was used which consisted of 4 loading wells, two parallel microchannels connecting each pair of wells and 120 nanochannels connecting the microchannels, as shown in Figure 3.2.

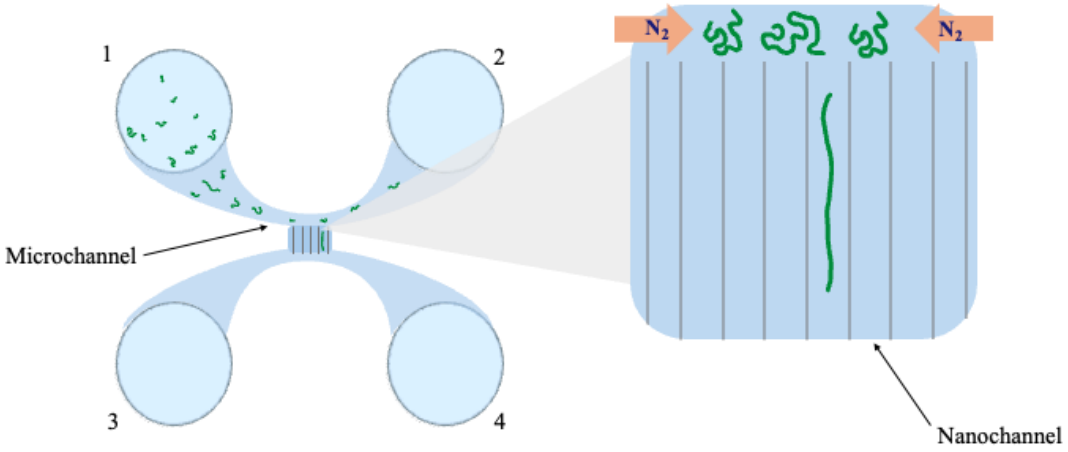

Figure 3.2: Illustration of the nanofluidic chip design. The four loading wells (labeled 1–4) are connected by two microchannels, which are further linked by an array of nanochannels. The illustration also demonstrates how DNA introduced into loading well 1 can be driven into the nanochannels by applying nitrogen pressure to wells 1 and 2.

The stained and diluted sample was added to the nanofluidic chip by loading 10 µl into any of the loading wells. The remaining three wells were filled with a buffer consisting of 0.033X TBE and 2% BME. By applying nitrogen gas pressure, the DNA molecules were guided out of the loading well and through the microchannel. Additional pressure was applied from both sides of the microchannel to facilitate the DNA to enter the nanochannels for stretching. The microchannels had the dimensions 50 µm (width) x 850 nm (height) and the nanochannels 150 nm (width) x 100 nm (height) x 500 µm (length). For imaging, a Zeiss Observer D.1 fluorescent microscope was used with a 63x oil immersion objective, Colibri 7 light source and a Photometrics Evolve EMCCD camera resulting in a pixel size of 254 nm/px.

For each DNA molecule imaged, 30 frames were captured to account for thermal fluctuation.

Prior to and between each sample being analyzed, all channels on the nanofluidic chip where flushed with 5% hypochlorite, followed by MQ water and buffer consisting of 0.033X TBE and 2% BME.

To ensure sufficient stretching of the DNA molecules, the lengths of λ-DNA molecules while stretched in the nanochannels were measured each imaging session. λ-DNA has a contour length of 16.5 µm, thus samples where λ-DNA molecules measuring less than approximately 11.5 µm (around 70% of the contour length) were considered to be insufficiently stretched, compromising the accuracy of the analysis. If the λ-DNA molecules were too short, the DNA sample was further diluted with 49 µl of MQ water and 1 µl BME.

3.4 Data processing

The processing of the data was performed through in-house developed MATLAB and R scripts from Westerlund's lab and MATLAB scripts that have been developed by Tobias Ambjörnsson's group at the Department of Astronomy and Theoretical Physics at the University of Lund. The data processing primarly involved the detection of individual DNA molecules from a series of fluorescence microscopy images, followed by the generation of kymographs. These kymographs were used to extract intensity profiles, here referred to as barcodes. The resulting barcodes were either aligned to theoretical reference barcodes derived from the *S. cerevisiae* genome, or analyzed for overlapping regions within the barcodes of the same dataset to enable *de novo* assembly of partial or complete chromosomes. Eventually, assembled genomic regions from the two different strains of *S. cerevisiae*, wild-type and Δtsa1, and DNA extracted from both before and after employing 5 passages, were compared to each other for revelation of SVs. As mention previously in Section 3.1.3, the four distinct samples are referred to as WT 0, WT 5, TSA 0 and TSA 5.

3.4.1 Molecule Detection

The raw imaging data acquired from the fluorescence microscope consisted of movies containing 30 frames and in .czi format. These files were imported into MATLAB and processed using a script called DNA Barcode Matchmaker (DBM). The parameters used for DBM is presented in Appendix B.1. As part of the DBM script, the files were converted into .tif format and kymographs were subsequently generated. Kymographs were produced by stacking sequential image frames, enabling the visualization of intensity variations along the length of individual DNA molecules over time. Additionally, intensity profiles could be obtained by plotting the time-average intensity from the kymographs along the length of the DNA molecules, obtaining what is referred to as the experimental barcodes. Before averaging the itensity over the molecule from the kymographs, here, the kymographs were first aligned, to account for fluctuation or movement of the DNA molecules when confined within

nanochannels. At this point, it was possible to perform quality control of the data by visually inspecting the generated kymographs and barcodes.

The DBM script also included a pipeline for λ-DNA detection. Here, the output consisted of intensity profiles and kymographs of the detected λ-DNA, as well as information such as the length, which was used for calculation of the DNA extension.

Due to occasional failures of the λ-DNA detection pipeline, manual measurements were performed of λ-DNA molecules using ImageJ. λ-DNA molecules exhibit a distinct and recognizable intensity profile, which made it possible to manually distinguish and sort them out by observing the generated kymographs. Measurements were carried out by manually tracing the length of these molecules, using a known pixel size of 254 nm/px.

3.4.2 Reference-Based Alignment

The next step in the data analysis involved reference-based alignment, performed using a MATLAB pipeline called Human Chromosome Alignment (HCA). The input for this pipeline consisted of the kymographs previously generated through the DBM pipeline. Within HCA, experimental barcodes were once again extracted by averaging the intensity across all aligned frames of each kymograph along the length of the corresponding DNA molecule.

To perform the alignment, theoretical barcodes were also generated to serve as references. These were derived from the complete genome sequence of *S. cerevisiae*, obtained from publicly available FASTA files from the Saccharomyces Genome Database (SGD) [37]. Using the HCA pipeline, these sequences were converted into theoretical intensity profiles that could be compared to the experimental barcodes. To ensure comparability between experimental and theoretical barcodes, the extension of DNA molecules in nm/bp, previously determined from the λ-DNA measurements, was applied to the theoretical barcodes.

During the alignment process, an additional parameter of stretching factor was applied to account for variability in DNA extension while also getting the best possible alignment. The stretching factor was set to $\pm 15\%$, corresponding to a range from 0.85 to 1.15, with a step size of 1%. This meant that each experimental and theoretical barcode was systematically scaled, both compressed and stretched, within this range.

The quality of each alignment was assessed using a correlation coefficient score (cc-score), calculated based on the Pearson correlation coefficient. The cc-score quantifies the linear relationship between the experimental and theoretical barcodes, ranging from -1 (perfect negative correlation), through 0 (no correlation), to 1 (perfect positive correlation). Here, it is desirable to have a cc-score as close to one as possible, meaning that the aligned barcode correlates well to the reference.

The output of the HCA pipeline included the cc-scores for all attempted alignments.

A length threshold could also be applied to prioritize longer experimental barcodes, which tend to provide more reliable alignment results.

For downstream analysis and visualization of the HCA results, custom R scripts developed in-house were used. For evaluating the coverage of the *S. cerevisiae* genome, experimental barcodes were aligned using a cc-score threshold of 0.5. Additionally, the difference between the best and second-best match was required to exceed 0.03 in cc-score to not be filtered away. A minimum length threshold of 120 kbp was also applied.

3.4.3 *De novo* assembly

For *de novo* analysis of the obtained data, firstly pair-wise comparisons were made between each barcode in every set of data. The pair-wise experiments were followed by parametrization, where parameters were obtained and later used for bargrouping of barcodes.

3.4.3.1 Pair-wise comparison experiments

Pair-wise comparison experiments was conducted without the use of theoretical barcodes, instead solely the experimental barcodes for each distinct sample.

To begin, data was loaded into the pair-wise comparison analysis pipeline in the form of kymographs, which were once again converted into experimental barcodes.

To improve alignment reliability and reduce artifacts, masking procedures were applied to the barcodes. First, the initial and terminal six pixles of each barcode were removed, as these regions often contain peaks of low intensity and that are not representative of the actual DNA sequence. Such peaks could inflate the cc-score and thereby lead to false-positive alignments.

In addition to masking of the ends, extreme intensity values or peaks, both high and low, were also removed. Any peaks that deviated more than 3.5 standard deviations from the mean intensity were masked. Furthermore, repetitive regions of experimental barcodes where also masked. If an experimental barcode was fully repetitive, it was excluded from further analysis. For barcodes containing partially repetitive regions, only the repetitive segments were masked.

A critical parameter in the pair-wise comparison analysis was the minimum length of the overlapping region required for a valid alignment. Overlap window thresholds of 150 px, 120 px, and 100 px were tested to evaluate their impact on assembly quality. The stretching factor was set to $\pm 10\%$ with a 1% step size.

During the pair-wise comparison analysis, each barcode was systematically compared to all others, initiating alignment at every possible pixel offset. As in previous steps, alignments were evaluated based on the cc-score derived from the Pearson correlation coefficient.

3.4.3.2 Parametrization

Parametrization is a procedure used to determine what constitutes a "good" cc-score within a given dataset. This process is important because the quality and characteristics, such as length and noise, may vary between datasets, influencing the cc-score interpretation.

The parametrization pipeline involves plotting the distribution of cc-score under the null hypothesis that no meaningful match exist between barcodes. Specifically, a β-extreme value probability density function is fitted to the cc-score data to approximate this null distribution.

To actually perform the parametrization, 100 experimental barcodes were picked out from each set of data (or fewer if the dataset was smaller). A segment or a cut-out was extracted from each of these barcodes. These cut-outs were then pairwise aligned and evaluated using both global and local cc-score. The resulting cc-scores are then fitted to the β-extreme value distribution, form which four fitting parameters are eventually extracted, two for the global cc-score distribution and two for the local cc-score distribution.

This parametrization procedure allows for calibration of cc-score thresholds, depending on how much false positives should be allowed, that are specific to the quality and characteristics of each dataset. By comparing experimental cc-scores to the empirically generated null distribution, the reliability and ability to interpret the resulting cc-scores could be improved.

3.4.3.3 Bargrouping of Barcodes

Bargrouping refers to the visualization and evaluation of barcodes that have been grouped together based on overlapping regions, with the aim of forming larger consensus structures. A bargroup may consist of as few of two barcodes, though larger groups were preferred for higher confidence and coverage. The structures obtained from the pairwise comparison experiments, along with their associated parameters from the parametrization procedure, were used as input for the bargrouping process.

Several additional parameters were specified during the execution of the bargrouping script, many of which required optimization. These parameters included the allowed degree of local stretching of barcodes, maximum distance from the true alignment position for grouping, as well as the p-value threshold used for false positive control. The specific parameter setting used for each sample type for the bargrouping experiments described in in Section 4 are provided in Appendix B.2.

Generated bargroups were visualized by plotting individual bargroups alongside a consensus barcode generated from their alignment. These bargroups were manually handpicked based on visual inspection. A high-quality bargroup was defined as one where barcodes aligned well and contributed to an extended consensus region. Bargroups were discarded if, for instance, they contained only short alignments and not contributing to an overall consensus extension, such as when the generated

bargroup's consensus length was no longer than the longest single bargroup in it. Each bargroup contained metadata such as which were the barcodes that contributed to the bargroup which allowed for traceability.

Once the high-quility bargroups had been selected, they were compared to the theoretical barcodes of *S. cerevisiae* using the HCA algorithm. Both global and local cc-scores were used to assess the alignments. To improve the coverage of the genome, barcodes that had not been included in bargroups could be added to the HCA, together with the bargroups. A minimum overlap threshold was applied to determine whether an experimental bargroup or barcode could be aligned to the theoretical barcodes.

Finally, consensus sequences representing entire chromosomes were assembled by combining both experimental bargroups together with supporting individual experimental barcodes.

3.4.4 Comparing Consensus for calling of SVs

The generated consensus barcodes from each sample were used to evaluate potential SVs in the genome. Specifically, the output of the consensus generation pipeline served as the basis for comparison across samples.

To ensure meaningful comparisons, the extension factor determined from the λ-DNA measurements of each sample were taken into account. Since the physical DNA extension of experimental barcodes differed between samples, normalization was required to be able to compare. To visualize and compare consensus barcodes within a single plot, the consensus with the largest extension factor was scaled down to match the consensus with the smallest extension factor.

To quantify similarities and differences between consensus barcodes, comparisons were performed over defined pixel lengths. For each pairwise comparison, two fixed window sizes were used: one for identifying the region with the lowest cc-score and another for detecting the region with the highest cc-score between the consensus barcodes.

4

Results and Discussion

The following chapter presents the results and discussion of the obtained data. It begins with evaluation of DNA extraction methods suitable for OGM. This is followed by reference-based alignment to evaluate the genome-wide coverage of *S. cerevisiae* using the generated experimental barcodes. The generation of bargroups is then discussed, along with subsequent reference-based alignments. Finally, the focus shifts to chromosome II for the identification of potential SVs.

4.1 Evaluating DNA extraction methods suitable for OGM

To perform ODM, the DNA extraction method must yield long and intact DNA molecules, preferably at least 150 kbp, which has been shown to be suitable for OGM. The evaluate this, two DNA extraction kits were tested: the CHEF Yeast Genomic DNA Plug Kit from BIO-RAD and Monarch® HMW DNA Extraction Kit for Tissue, in combination with the different compositions for cell wall degradation: lyticase and zymolyase. The evaluation of the DNA extraction methods were taken into account when establishing the cultivation scheme of *S. cerevisiae* presented in section 3.1.3 of the Methodology.

4.1.1 Adaption of protocol for Agarose Embedded DNA

The initial task involved evaluating the CHEF Yeast Genomic DNA Plug Kit from BIO-RAD on yeast grown on YPD plates, rather than in liquid media as specified by the protocol. The enzyme lyticase was used for cell wall digestion during this process. Lyticase and the protocol have worked in the past when extracting high molecular weight DNA from liquid cultivated *S. cerevisiae*, *Schizosaccharomyces pobe* and *Candida albicans*. The concentration of the extracted DNA, measured using Qubit fluorometer, is compared to the mass of yeast picked, which is shown in Figure 4.1.

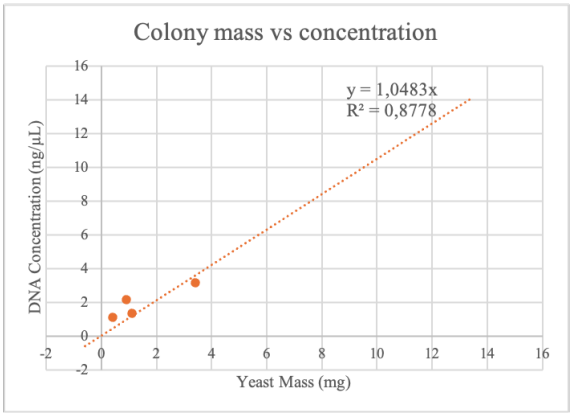

Figure 4.1: Results from a DNA extraction showing the DNA concentration (ng/µl) with corresponding mass of yeast (mg) using a linear fit.

A linear fit in Figure 4.1 shows that as the mass of yeast increases, the concentration of extracted DNA also increases. While the Qubit measurements confirm successful DNA extraction, the samples also contained cellular debris, including non-lysed yeast cells. This was not observed when extracting DNA from yeast grown in liquid culture, leading to the hypothesis that differences in cultivation methods contribute to the cells being harder to digest, likely due to a more robust cell wall. The contamination of cellular debris and intact cells disrupted downstream analysis by preventing DNA from entering the microchannels and nanochannels of the nanofluidic chip. This issue highlighted a limitation of the DNA extraction protocol.

In Figure 4.2, a separate experiment is shown where some modifications to the DNA extraction protocol had been implemented in an attempt to address the issues with insufficient cell wall degradation. Specifically, the incubation time with the enzyme lyticase was extended from 4 hours to 18 hours (overnight), allowing more time for the enzyme to act, in the hope of achieving sufficient degradation of the cell wall. The sample was stained with YOYO-1, placed on a coverslip and imaged. Strands of DNA were clearly distinguishable, along with bright green dots representing undigested yeast cells. The DNA appeared to be long and the sample was therefore further analyzed using the nanofluidic chip to evaluate whether the presence of yeast cells would cause clogging and disrupt imaging. Figure 4.3 shows the well on the nanofluidic chip where the same sample was loaded. A high concentration of yeast cells can be seen accumulating near the entrance of the microchannel while nitrogen gas is flushed through the system. The strands of DNA observed on the coverslip are less prominent here, overshadowed by the abundance of yeast cells. Further into the microchannels, by the entrances to the nanochannels, where the field of view is not disturbed by the brightness of the yeast cells, some strands of DNA can be seen. However, these strands are short, present in low quantities, and insufficient for the goals of this study, which require long DNA strands.

When comparing the DNA extraction results shown in Figures 4.2 and 4.3, where DNA was extracted directly from yeast grown on YPD plates, to samples prepared using the agarose embedded DNA method from yeast cultivated in liquid YPD me-

dia, clogging was no longer an issue. In the case of the liquid culture samples, long DNA strands were observed exiting the loading well and moving through the microchannel. These strands could then be successfully pushed into the nanochannels without obstruction.

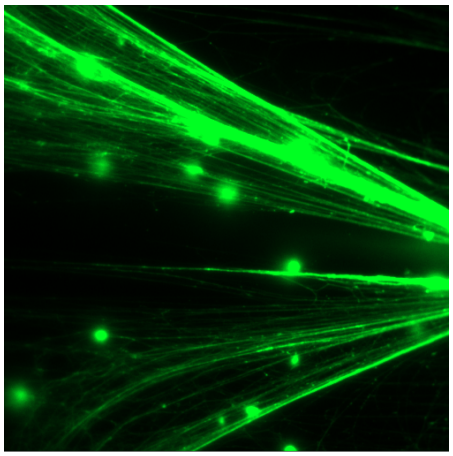

Figure 4.2: Image taken of a DNA extraction sample stained with YOYO-1 and placed on a coverslip, where the strands are DNA and the bright dots are intact yeast cells.

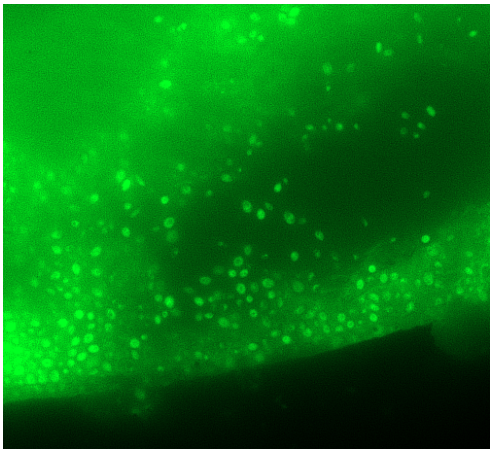

Figure 4.3: Image taken of a loading well on the nanofluidic chip where the same DNA sample as in Figure 4.2 has been loaded, showing intact yeast cells accumulating by the microchannel entrance.

Additionally, modifications to the original protocol, beyond focusing on the role of lyticase, were evaluated in an effort to obtain uncontaminated DNA samples. These included extending the incubation time with Proteinase K and increasing the number of washes with the washing solution. These strategies were tested to determine whether contaminants such as protein remnants or cell wall components were causing the disruptions, despite lyticase being effective. However, these changes did not alter the outcome, as intact yeast cells continued to clog the system and disrupt the imaging process. The most reasonable approach to addressing these issues remained focusing on the enzyme used to degrade the cell wall of *S. cerevisiae*. While Proteinase K is also an enzyme, its primary role is to degrade proteins. The presence of intact yeast cells indicated that improving lyticase's efficacy was a more appropriate strategy for enhancing the DNA extraction protocol and enabling successful OGM.

4.1.2 Zymolyase Ultra for Cell Wall Degradation

To address the insufficient degradation of the cell wall, Zymolyase Ultra from ZymoResearch was evaluated as a replacement for the lyticase provided in the CHEF Yeast Genomic DNA Plug Kit from BIO-RAD. According to ZymoResearch, Zymolyase Ultra is more than 50 times more effective than lyticase and should efficiently target the cell wall of *Saccharomyces* [54]. Zymolyase Ultra comprises four distinct enzymes, each with unique lytic actions targeting specific parts of the cell wall to allow for a more effective degradation. Based on the claims regarding the

efficiency by ZymoResearch, the initial protocol adaption for Zymolyase Ultra involved half the enzymatic units of lyticase recommended by the BIO-RAD protocol for each DNA agarose plug. The initial attempts with Zymolyase Ultra for cell wall digestion were not fully successful, as major challenges such as cloggage persisted when attempting to capture DNA in the nanochannels. However, microscopy revealed fewer intact yeast cells in the loading well on the chip, suggesting some improvement in cell wall digestion.

A sample was treated with 115 U of Zymolyase Ultra, which is double the enzymatic units specified by the BIO-RAD protocol for each DNA agarose plug when using lyticase, and incubated for 2 hours. The sample was then imaged on a coverslip, as shown in Figure 4.4. No yeast cells were observed in the image, indicating that incubation with Zymolyase Ultra was successful in degrading the cell wall.

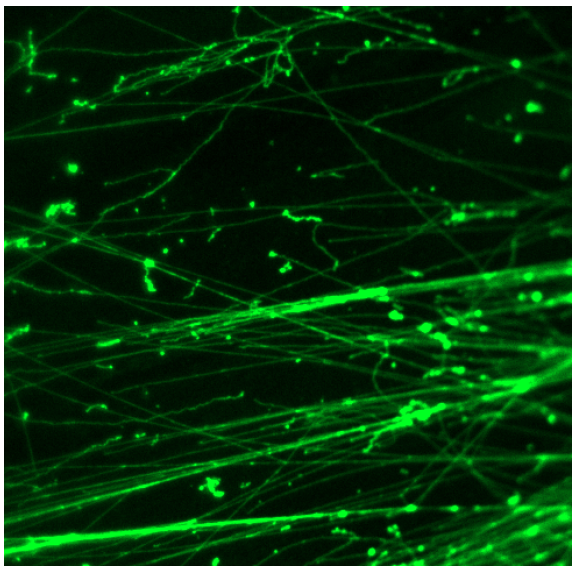

Figure 4.4: Image taken of a DNA extraction sample stained with YOYO-1 and placed on a coverslip, where strands of DNA can be seen and notably no yeast cells are visible.

While the sample shown in Figure 4.4 demonstrate improvement in the DNA extraction protocol for *S. cerevisiae* grown on YPD plates and potential for OGM, the high enzyme concentration used in this experiment introduced challenges. The DNA concentration in this sample was low, making imaging difficult and the DNA molecules too short for this purpose and suggesting zymolyase could promote some DNase activity. Out of the 28 molecules aligned in the Human Chromosome Alignment (HCA), only 3 molecules exceeded the threshold of 150 kbp, which is essential for this application. Another issue observed was the stickiness of the sample, which caused DNA strands to adhere to one another, to the microchannel walls and at the entrances to both microchannels and nanochannels. The stickiness was likely due to contamination, possibly from the high concentrations of Zymolyase Ultra used in the experiment, or from components that were not sufficiently broken down, remaining in the sample and disrupting the imaging process. This stickiness interfered

with imaging by creating bright spots, making it harder to visualize individual DNA molecules. Over time, this led to imaging being nearly impossible.

To address the problem of stickiness, the experiment was repeated with a prolonged incubation time of three days with Zymolyase Ultra and reduced amounts (14.375 U, 28.74 U and 57.5 U of Zymolyase Ultra). DNA was successfully extracted in all three cases, but stickiness persisted across all samples. Additionally, the sample treated with 57.5 U of Zymolyase Ultra showed the presence of yeast cells, further complicating the imaging. These observations of stickiness and contamination highlighted inconsistencies in the method, suggesting that additional adjustments are needed to optimize the protocol.

Data was extracted and processed for samples treated with 28.74 U and 57.5 U of Zymolyase Ultra. For the sample treated with 28.74 U of Zymolyase Ultra, 76 molecules were detected by the DBM across 76 images, 12 molecules met the criteria of having a length greater than 150 kbp and a cc-score above 0.5. In Figure 4.5, cc-scores are plotted against the length of the molecules that passed the criteria. Molecules with high cc-values are highlighted in green in Figure 4.6, where their matches to different chromosomes are shown. The highest-scoring molecule in this dataset aligns with chromosome XI. No molecules were detected as λ-DNA. As a result, the extension was assumed to match that of another sample imaged on the same day, with results presented later in Figures 4.7 and 4.8.

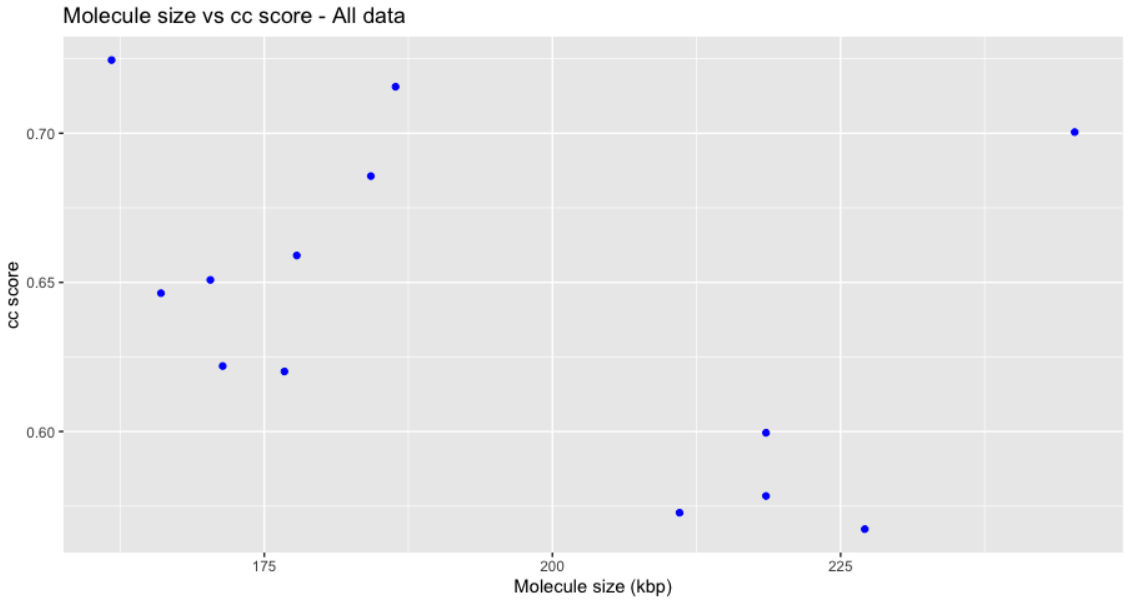

Figure 4.5: The cc-scores of all aligned molecules of a DNA extraction sample that had been treated with 28.74 U of Zymolyase Ultra for cell wall digestion, plotted against their corresponding length (kbp).

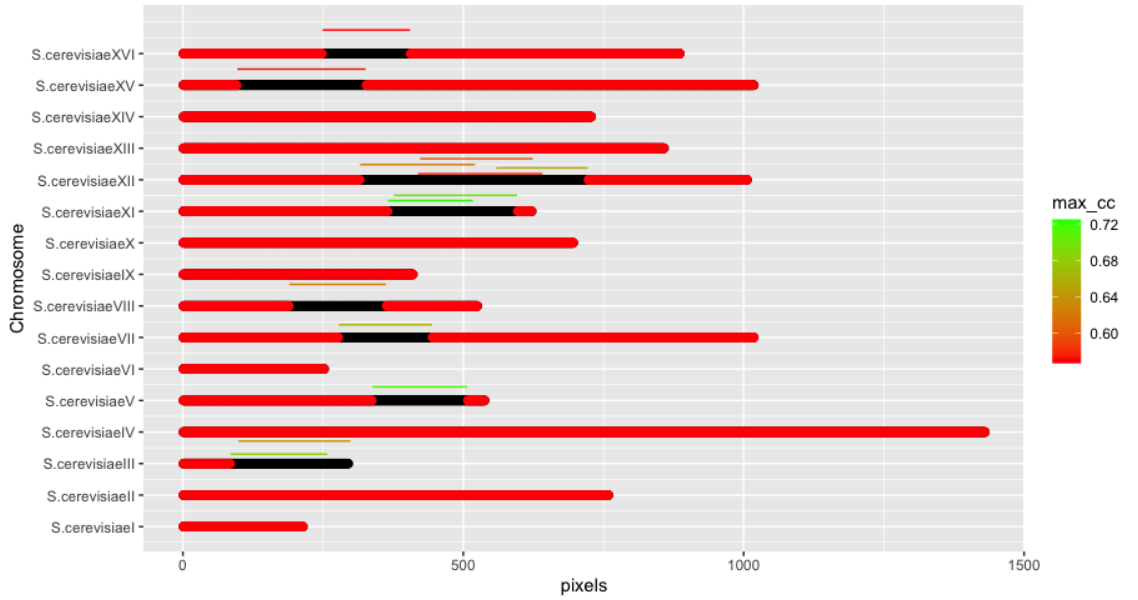

Figure 4.6: Coverage of all 16 chromosomes of *S. cerevisiae* of a DNA extraction sample treated with 28.74 U Zymolyase Ultra for cell wall digestion. Each chromosome is represented by black and red bars. Aligned DNA molecules are shown as narrower bars; matches to the chromosome are indicated in black. The corresponding cc-scores are shown by the gradient color bar on the right.

The results from the sample treated with 57.5 U of Zymolyase Ultra are shown in Figures 4.7 and 4.8. As illustrated in Figure 4.7, 7 molecules exceeded the length threshold of 150 kbp, had a cc-score above 0.5 and showed a difference greater than 0.03 between the best and second-best matches. A total of 62 molecules were detected by the DBM pipeline across 29 images. Additionally, 3 λ-DNA molecules were identified, and an extension rate of 0.196 nm/bp was applied for analysis. The extension was low and the sample should have been diluted to decrease the salt concentration and thereby increase the extension. Figure 4.8 highlights chromosome XI as the highest-scoring match in this dataset.

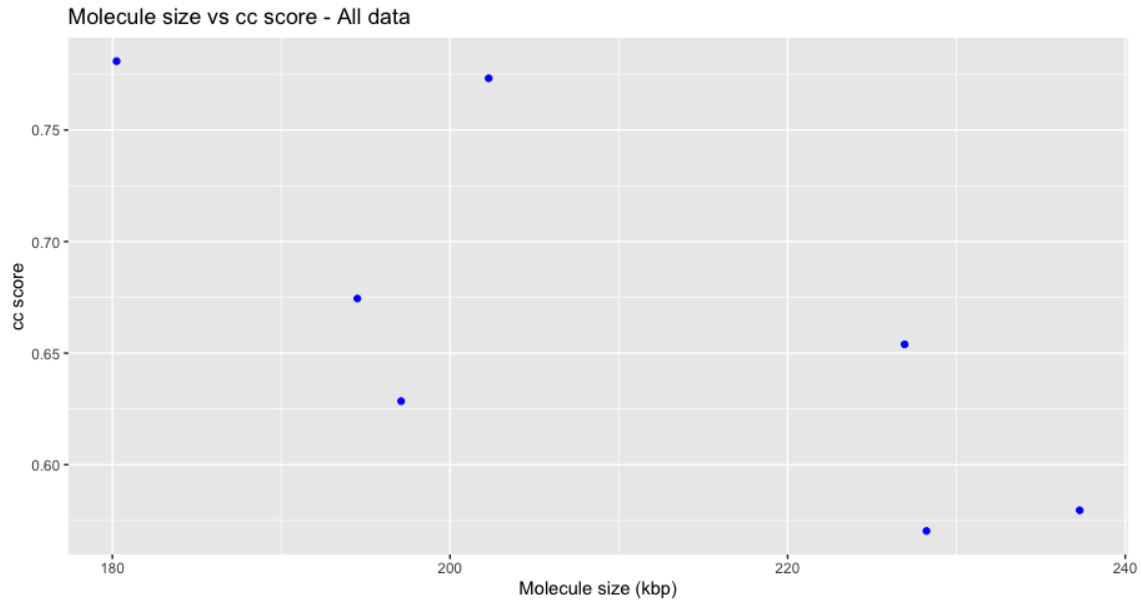

Figure 4.7: The cc-scores of all aligned molecules of a DNA extraction sample that had been treated with 57.5 U of Zymolyase Ultra for cell wall digestion, plotted against their corresponding length (kbp).

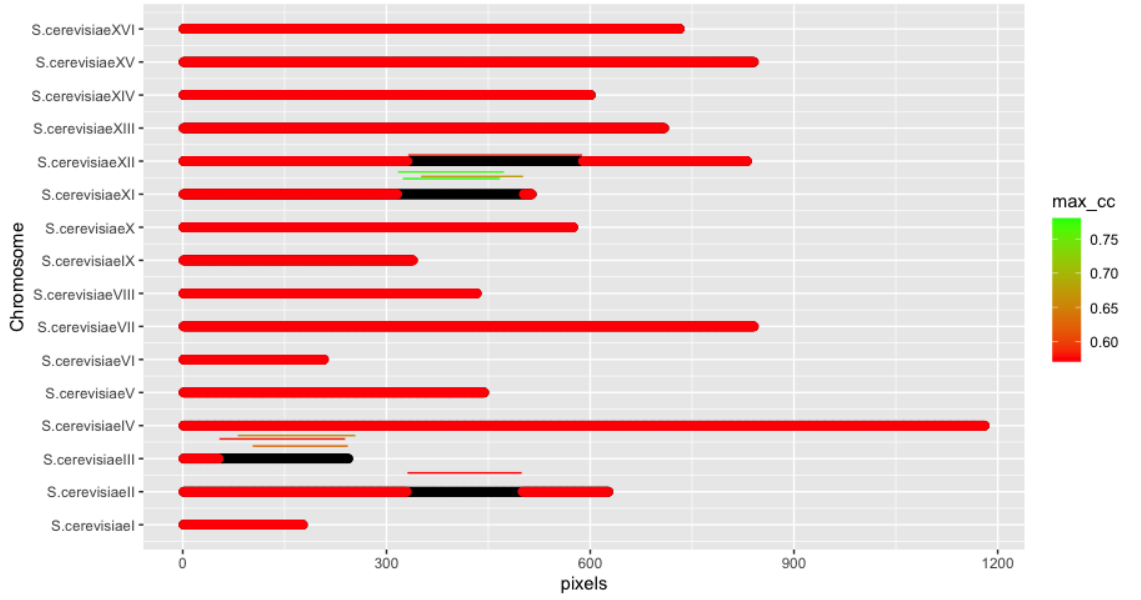

Figure 4.8: Coverage of all 16 chromosomes of *S. cerevisiae* of a DNA extraction sample treated with 57.5 U Zymolyase Ultra for cell wall digestion. Each chromosome is represented by black and red bars. Aligned DNA molecules are shown as narrower bars; matches to the chromosome are indicated in black. The corresponding cc-scores are shown by the gradient color bar on the right.

Although DNA was successfully extracted from *S. cerevisiae* cells cultivated on YPD plates using Zymolyase Ultra for cell wall degradation, the quality of the samples

brought challenges for microscopy. This highlighted the need for further protocol optimization to improve sample quality and ensure reliable analysis.

4.1.3 Monarch DNA extraction

As previously noted, the poor quality of the samples hindered imaging. To address this issue, strategies to improve sample cleanliness were explored. One common method, alcohol precipitation, was evaluated using ethanol. This technique is widely used to reduce DNA solubility, enabling its precipitation into a pellet while leaving other soluble components in the supernatant. However, since the goal was to separate DNA from yeast cells, which do not dissolve in solution, this method proved ineffective. The experiments demonstrated that alcohol precipitation could not successfully isolate DNA from yeast under these conditions.

Another approach of isolating DNA was performed using the Monarch® HMW DNA Extraction Kit for Tissue. This kit utilizes a method where DNA binds to borosilicate glass beads, facilitating separation. The process combines two strategies including increasing the salt concentration to neutralize the DNA's negative charge and adding isopropanol to reduce DNA solubility, enhancing it to bind to the silica beads. Ionic interactions between the silica and the DNA enable this separation.

Here, the same amount of Zymolyase Ultra which was considered most efficient in previous experiments (28.74 U) was used. Since this kit did not involve embedding the DNA in agarose plugs, it was expected that less incubation time with Zymolyase Ultra was going to be needed. This was because the enzyme would not have to diffuse through a layer of agarose to get to the yeast cells. Four different incubation times were evaluated, 2 hours, 4 hours, overnight (∼18 hours) and over the weekend (∼72 hours). The results indicated that 4 hours of incubation yielded the highest amount of DNA. Longer incubation periods, particularly over the weekend, may have reduced DNA yield due to potential DNase activity from the Zymolyase. On the other hand, shorter incubation times did not allow for sufficient time for the the enzymes to effectively degrade the cell walls.

Microscopy of the sample with 4 hours incubation time revealed a significant improvement over those samples extracted from agarose plugs. No yeast cells could be observed and although some stickiness remained, the overall quality was noticeably better. Here, the data captured included 90 images where 270 molecules were detected. Figure 4.9 presents the size distribution and the average size of the molecules that met the selection criteria of a length threshold of 150 kbp and a cc-score over 0.5. The analysis reveals that the average size of the molecules is just below 200 kbp, which is considered a good size for this application.

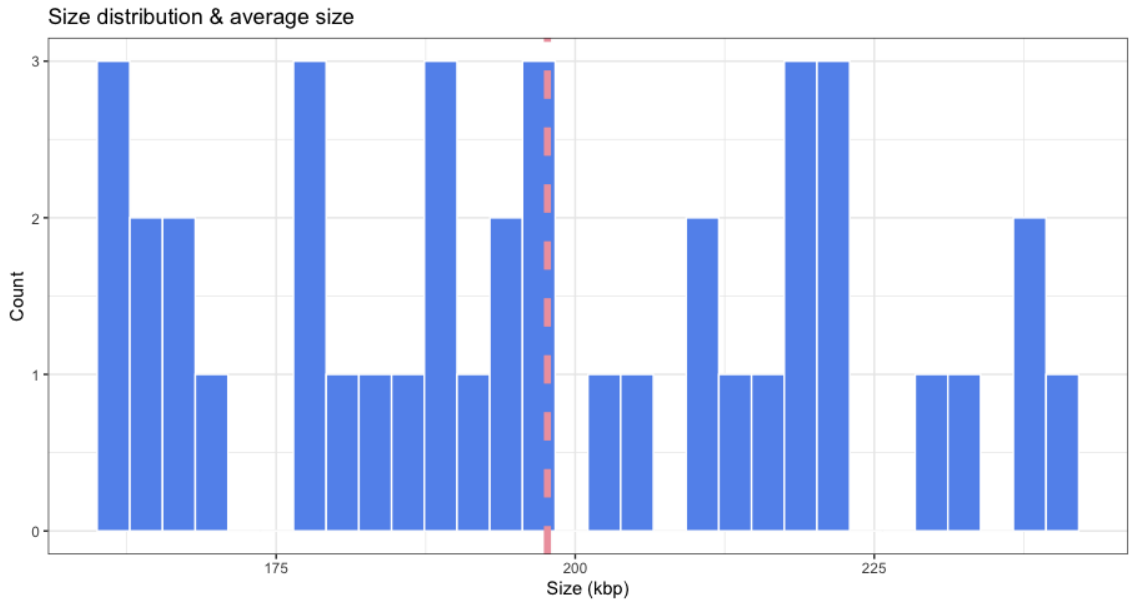

Figure 4.9: The size distribution of DNA molecules extracted using the Monarch® HMW DNA Extraction Kit for Tissue where the average size is represented as a red dashed line.

Figure 4.10 shows the cc-scores and the length of the molecules from the same data shown in the previous picture. In Figure 4.11, the matches to the chromosomes are instead presented and notably chromosome III seems to be completely covered.

The cc-scores of all aligned molecules of a DNA extraction sample that had been treated with 57.5 U of Zymolyase Ultra for cell wall digestion, plotted against their corresponding length (kbp).

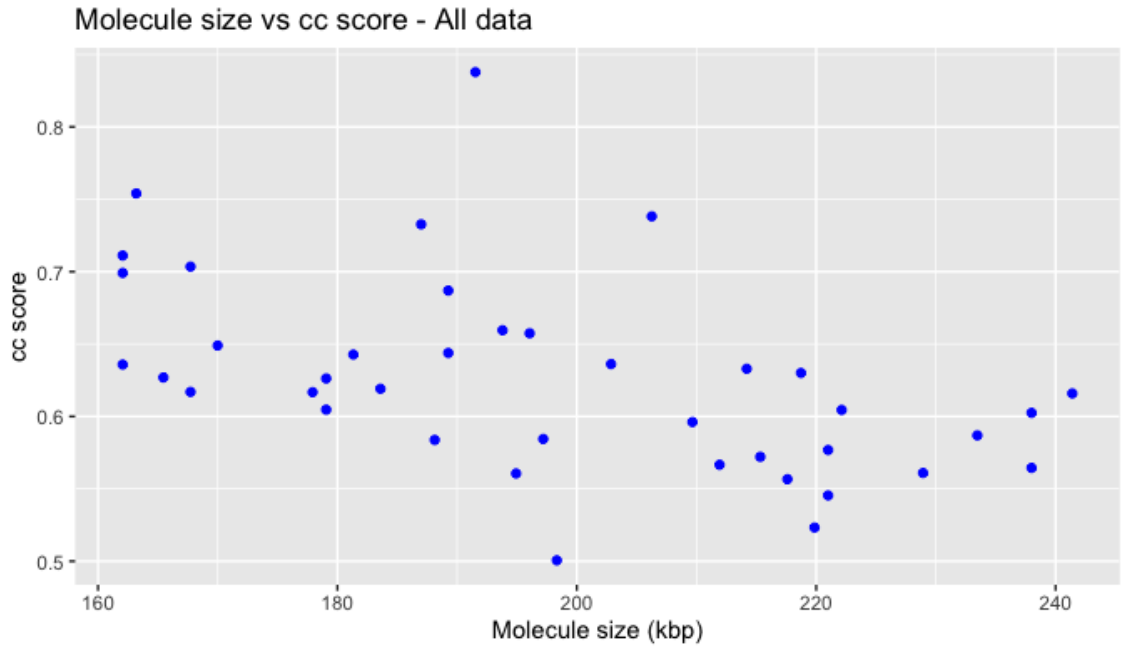

Figure 4.10: The cc-scores of all aligned DNA molecules of DNA extraction using the Monarch® HMW DNA Extraction Kit for Tissue, plotted against their corresponding length (kbp).

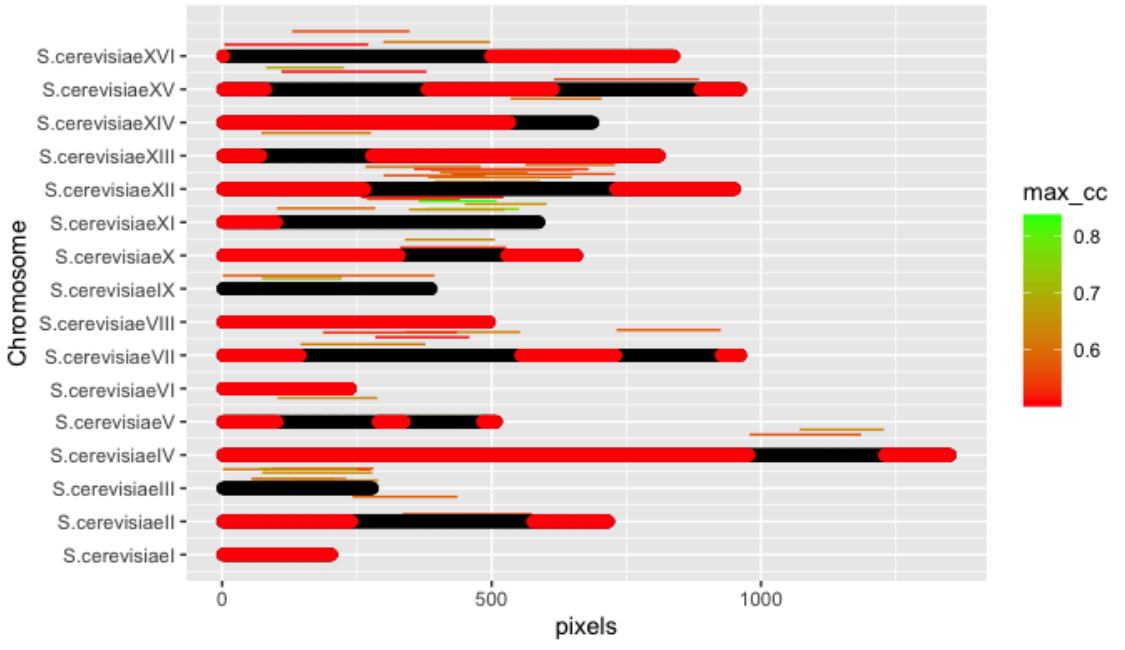

Figure 4.11: Coverage of all 16 chromosomes of *S. cerevisiae* of DNA extracted using Monarch® HMW DNA Extraction Kit for Tissue. Each chromosome is represented by black and red bars. Aligned DNA molecules are shown as narrower bars; matches to the chromosome are indicated in black. The corresponding cc-scores are shown by the gradient color bar on the right.

Figure 4.12 displays the number of molecules matched to each chromosome, represented by colored bars, alongside the expected distribution of matches indicated by black dots. One major deviation from the expected distribution is the large number of matches to chromosome III, which as shown in Figure 4.11, was fully covered.

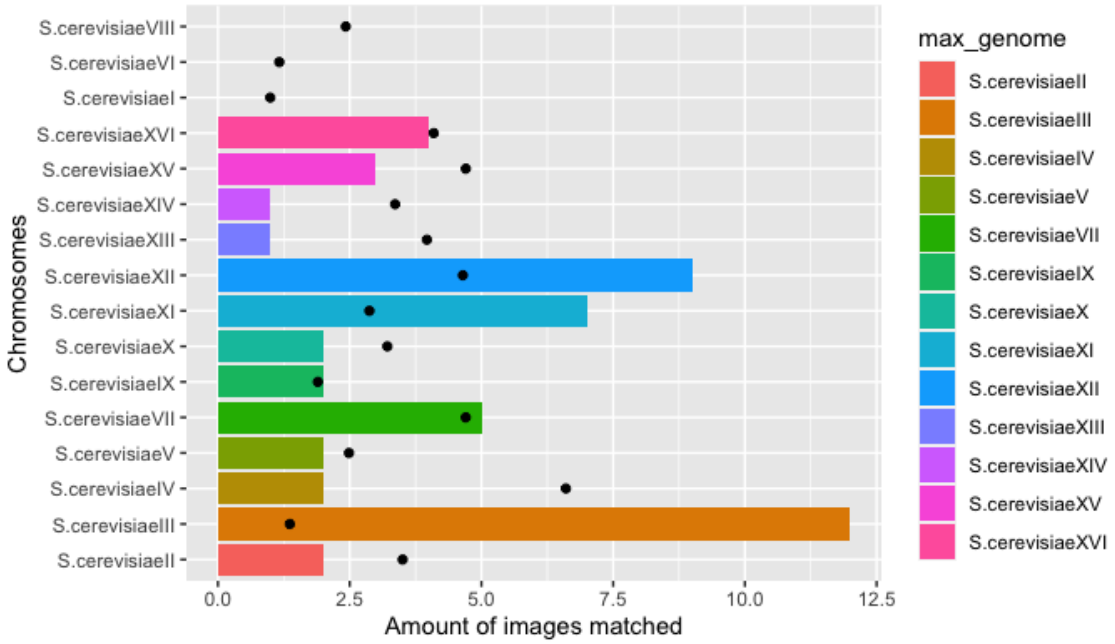

Figure 4.12: Coverage of the 16 reference chromosomes of *S. cerevisiae* by DNA extracted using the Monarch® HMW DNA Extraction Kit for Tissue, where the number of DNA molecules matched are represented by the colored bars and the expected distribution of matches are indicated by the black dots.

To improve the quality of the data, which may explain why chromosome III has by far the highest amount of matches, more λ-DNA could be captured. By capturing more λ-DNA, the extension factor can be determined more accurately. Additionally, while imaging, it is important to verify that the extensions are not too short.

A cultivation scheme for *S. cerevisiae* was established to study the development of SVs, employing the presented DNA extraction methods. The cultivation scheme can be found in Methodology 3.1.3.

4.2 Data collection

Following the cultivation scheme, DNA from the four different sample types were extracted. Initially, DNA extraction was attempted for all samples using the Monarch extraction method, as this approach is expected to yield DNA samples with less contamination compared to the DNA agarose plug method, thereby facilitating efficient imaging. However, successful extraction using the Monarch method was only achieved for the WT 0 sample. The success of WT 0 DNA extraction was likely

due to a coincidentally higher yeast biomass was obtained in the liquid culture. To reliably use the Monarch method further optimization would be required, particularly in determining the number of generations allowed to grow during liquid culture growth and in balancing this with the need for genetic homogeneity in the sample.

The remaining samples (WT 5, TSA 0, and TSA 5) were extracted using the DNA agarose plug method. Each sample, along with its corresponding DNA extraction method, is listed in Table 4.1.

Table 4.1: The four distinct samples listed together with their corresponding DNA extraction method, average λ-DNA length and average λ-DNA extension. Due to variation of DNA extension in sample TSA 0 and TSA 5, the data was split up into two sets, hence two values are given for average λ-DNA length and average extension.

Sample	**Extraction Method**	**Average λ-DNA length (µm)**	**Average λ-DNA extension (nm/bp)**
WT 0	Monarch	12.15	0.251
WT 5	Agarose Plug	9.39	0.194
TSA 0	Agarose Plug	9.43 & 11.86	0.194 & 0.244
TSA 5	Agarose Plug	8.63 & 10.88	0.178 & 0.224

Each imaging session using fluorescence microscope and nanofluidics device yielded specific values for λ-DNA length and extension, both per sample and session. These values are presented in Table 4.1 as averages across every session for each sample type. However, as also shown in Table 4.1, the TSA 0 and TSA 5 samples each have two values listed for both average length and extension. This is due to high variability in DNA extension between imaging sessions. To improve the accuracy in downstream data analysis, it was therefore decided to split these samples into two separate datasets. A complete list of all imaging sessions included in averages presented in Table 4.1 can be found in Appendix A.2.

Notably, the λ-DNA length in the WT 0 sample stands out when compared to the other samples. As mentioned in Methods section 3.3.2, a λ-DNA length greater than 11.6 µm is desirable to ensure sufficient extension for reliable barcode alignment. The WT 0 sample showed consistently higher extension values, and it is likely not a coincidence that this sample was also the only one extracted using the Monarch method. Lower contamination in this sample likely contributed to improved DNA extension, as impurities such as excess salt can negatively affect molecular stretching. This trend was also evident during imaging where DNA molecules from the WT 0 sample appeared bright and could be imaged at a laser intensity of only 10%. In contrast, the WT 5, TSA 0 and TSA 5 samples appeared less bright, suggesting that the YOYO-1 dye had bound not only to the DNA but also other contaminants. These samples required laser intensities of 40-50% to be able to distinguish DNA

molecules from the background noise.

4.2.1 Chromosome Coverage Estimation

To obtain an indication on how much of the *S. cerevisiae* genome in the four separate samples was covered by the obtained imaging data, the HCA pipeline was run. All detected molecules were matches to a reference containing all 16 chromosomes based on the quality criteria described in section 3.4.2. Table 4.2 summarizes all samples, including the total number of images acquired across all imaging sessions, the number of kymographs generated and the number of molecules that passed the previously described threshold. Additionally, Table 4.2 presents the estimated coverage of the genome, which is defined as how many times the aligned molecules could span the entire genome. The mean length of the aligned molecules which passed the criteria are also presented in the same table.

Table 4.2: The four samples listed together with the amount of images captured, amount of kymographs generated, number of molecules matched in the HCA, coverage of the chromosomes and the mean length of the molecules matched in kbp.

Sample	**Images**	**Kymographs**	**Molecules (HCA)**	**Coverage**	**Mean Length (kbp)**
WT 0	543	1581	502	9.07	220
WT 5	192	334	127	3.65	351
TSA 0	458	1098	331	9.12	336
TSA 5	452	786	305	10.95	438

In Figure 4.13, molecules from sample WT 0 are aligned to references of the 16 chromosomes of *S. cerevisiae*. Aligned regions are shown in black, while regions without aligned DNA molecules are marked in red. The cc-scores of the aligned molecules are represented by the color of each line: brighter green indicates a higher cc-score, while a darker red corresponds to a lower cc-score.

A notable feature of WT 0 is the mean length of the aligned DNA molecules, which is 220 kbp and significantly shorter than in the other samples. This reflects a limitation of the Monarch DNA extraction method, which tends to yield shorter DNA molecules compared to the DNA agarose plug method.

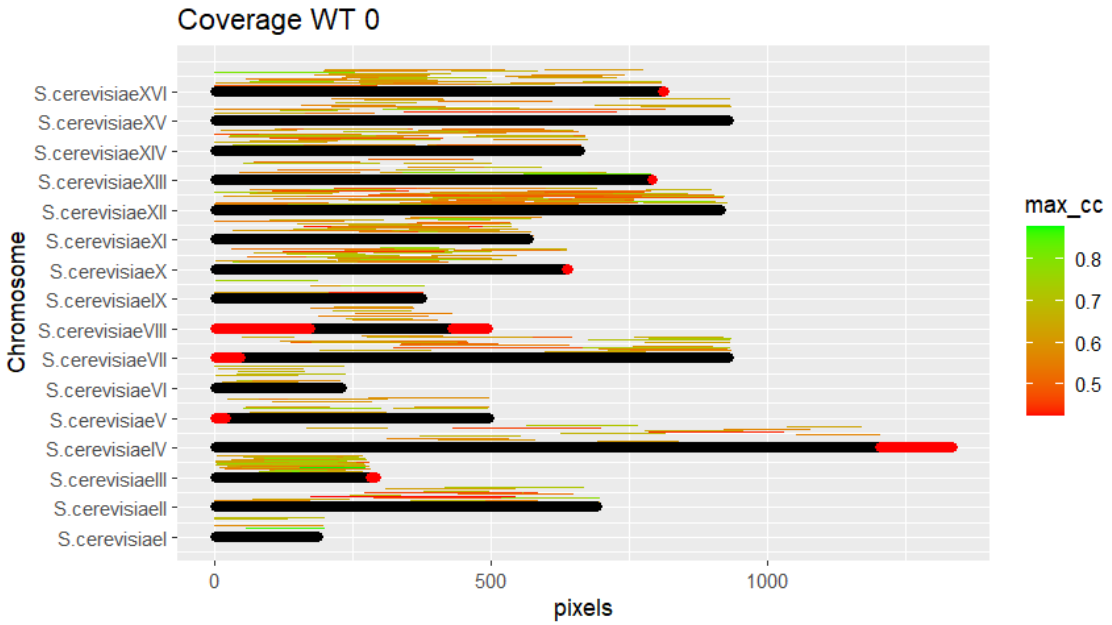

Figure 4.13: Coverage of all 16 chromosomes of sample WT 0. Each chromosome is represented by black and red bars. Aligned DNA molecules are shown as narrower bars; matches to the chromosome are indicated in black. The corresponding cc-scores are shown by the gradient color bar on the right.

The *S. cerevisiae* genome coverage of sample WT 5 is shown in Figure 4.14. Sample WT 5 exhibits lower coverage compared to the others, mainly due difficulties encountered during imaging of the sample. The overall quality was poor and contamination was evident, as the DNA molecules appeared dark and not sufficiently clear for detection by the DBM algorithm.

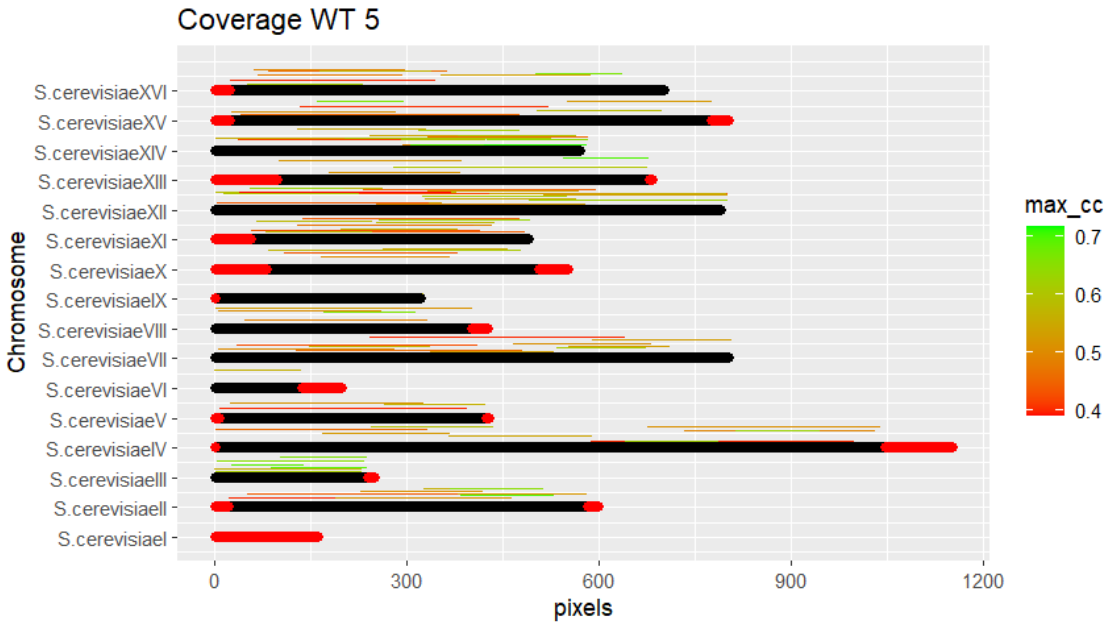

Figure 4.14: Coverage of all 16 chromosomes of sample WT 5. Each chromosome is represented by black and red bars. Aligned DNA molecules are shown as narrower bars; matches to the chromosome are indicated in black. The corresponding cc-scores are shown by the gradient color bar on the right.

A higher *S. cerevisiae* genome coverage can be distinguished from both of the TSA samples. In Figure 4.15, the aligned DNA molecules of sample TSA 0 is shown.

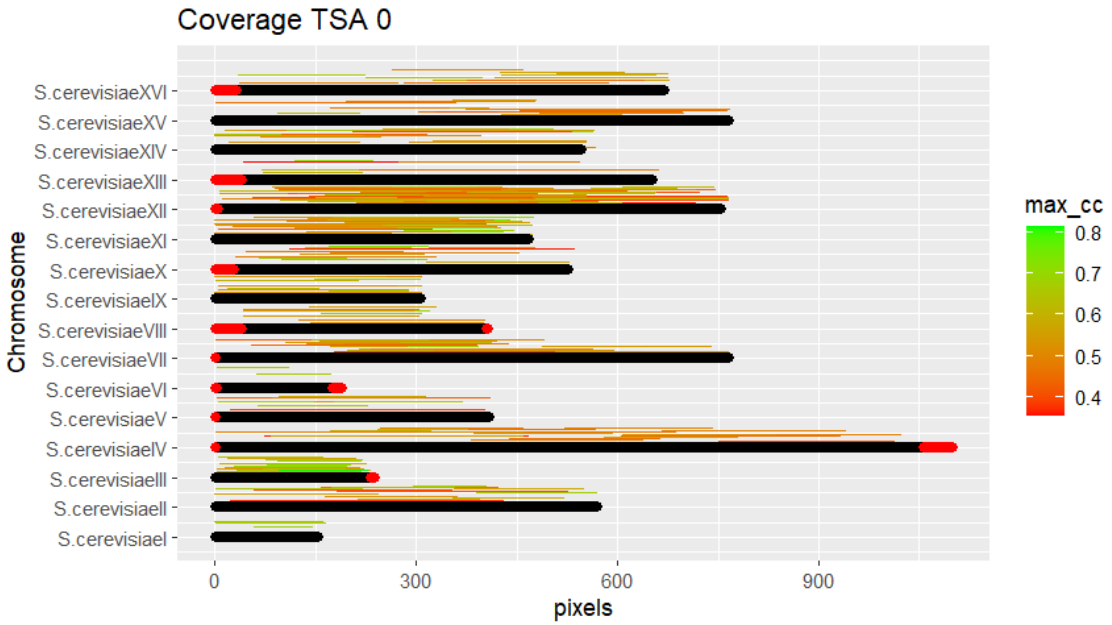

Figure 4.15: Coverage of all 16 chromosomes of sample TSA 0. Each chromosome is represented by black and red bars. Aligned DNA molecules are shown as narrower bars; matches to the chromosome are indicated in black. The corresponding cc-scores are shown by the gradient color bar on the right.

The highest coverage was obtained for sample TSA 5, which is illustrated in Figure 4.14.

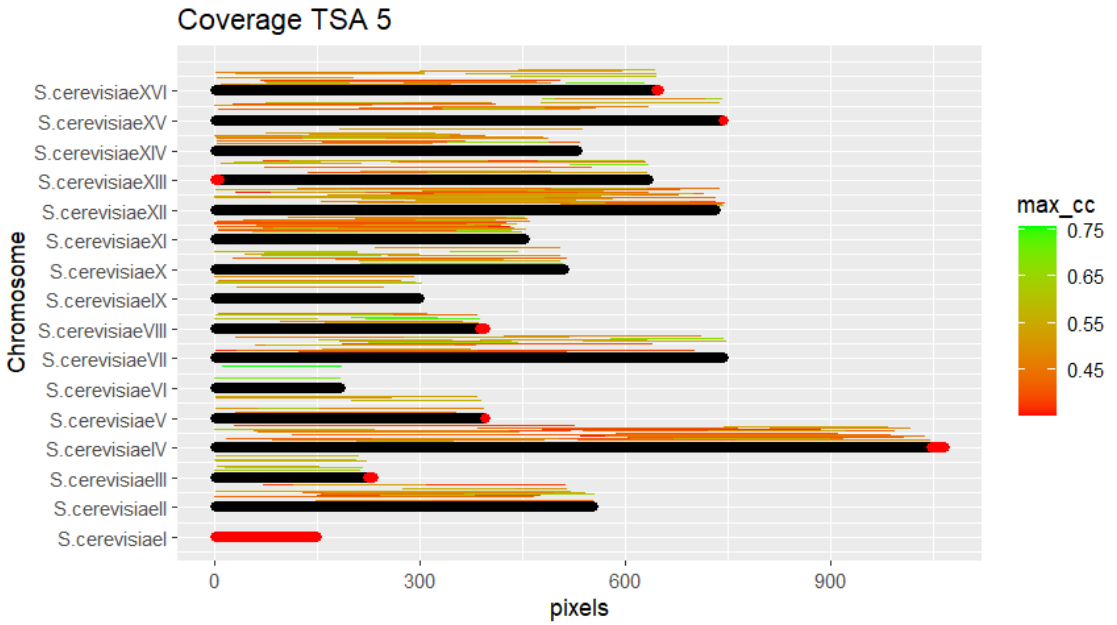

Figure 4.16: Coverage of all 16 chromosomes of sample TSA 5. Each chromosome is represented by black and red bars. Aligned DNA molecules are shown as narrower bars; matches to the chromosome are indicated in black. The corresponding cc-scores are shown by the gradient color bar on the right.

In two cases, specifically WT 5 and TSA 5, no molecules were aligned to chromosome I. However, this does not necessarily reflect the true absence of chromosome I sequences in the data of those samples. Chromosome I is the shortest chromosome in the *S. cerevisiae* genome, consisting of approximately 230 kbp. The HCA performed here used a global alignment strategy, meaning that each DNA molecule was aligned as a whole to the reference. In this approach, molecules that only partially match a chromosome or extend beyond its boundaries are excluded from the alignment. Since the allowed stretching was set to 15%, molecules stretched to the maximum (115% of their original length) must still be equal to or smaller than the size of the corresponding chromosome. Since longer molecules were prioritized during imaging, shorter fragments which would more likely match to chromosome I, may have been underrepresented or overlooked. Additionally, the reference genome used in the analysis may be incomplete, particularly near the chromosome ends. Fragments that contain the edges likely extends a couple of pixels out of the reference and therefore does not align. This is reflected across all four samples, where terminal regions of the chromosomes often appear red, indicating no aligned molecules in those areas.

In fact, when focusing on local alignment, which evaluate the correlation coefficient over a fixed number of consecutive pixels, matches to the theoretical barcode of chromosome I do appear. As shown in Figure 4.17, barcodes generated from the first dataset of TSA 5 align with the theoretical barcode of chromosome I. This contrasts with the overview shown in Figure 4.16, where chromosome I appears to

lack coverage. In Figure 4.17, the black segments of the barcodes indicate aligned regions based on a local cc-score calculated over 100 consecutive pixels. The corresponding local and global cc-score values are displayed in the bottom right corner of the figure.

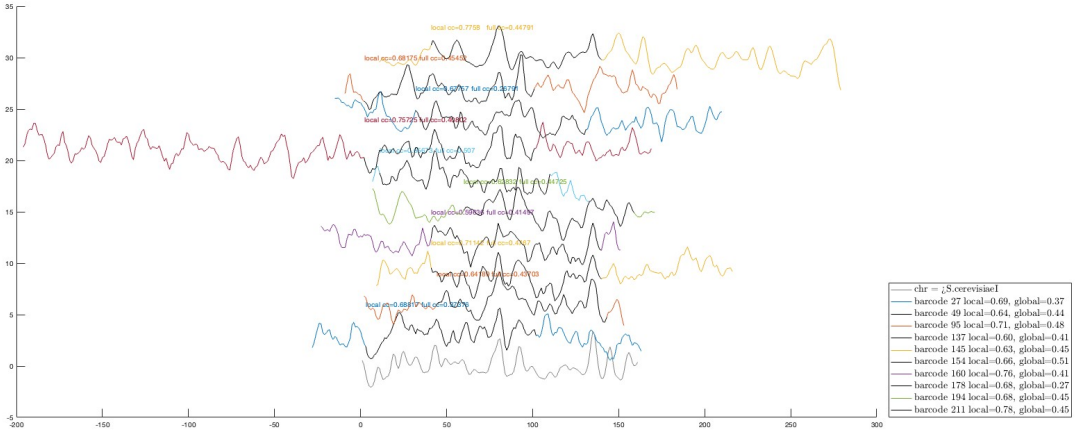

Figure 4.17: Barcodes from the first TSA 5 dataset aligned to chromosome I using a 100-pixel overlap window. Both global and local cc-scores are shown in the bottom right corner.

Overall, the genome coverage appears sufficient, except in the case of WT 5. Although a higher coverage, would be preferable. According to BioNano Genomics, a coverage of 80 times is recommended for *de novo* genome assembly [KÄLLA]. A limitation of this project was the need to image four distinct samples, which led to reduced coverage per sample due to time constraints.

4.3 Bargrouping and Reference-Based Alignment

Pair-wise comparison experiments were initially performed for all datasets: WT 0, WT 5, TSA 0 and TSA 5. Note that two datasets were analyzed for both TSA 0 and TSA 5 due to differences in DNA molecule extension during imaging. The pair-wise comparison experiments where used to generate bargroups. Generation of bargroups is an iterative process of grouping of barcodes that could belong to the same locus of the genome based on similarity of their intensity profiles. Here, no theoretical barcodes of the reference genome of *S. cerevisiae* was used.

In addition, the resulting bargroups together with ungrouped experimental barcodes were aligned to theoretical barcodes of the *S. cerevisae* genome. This alignment was performed to evaluate both genome coverage and the quality of the experimentally generated bargroups.

4.3.1 Bargrouping Experiments

Bargrouping involves grouping and visualizing barcodes with overlapping regions to form larger consensus structures. A bargroup may contain as few as two barcodes, though larger groups are preferred for increased reliability. The primary purpose of bargrouping the experimental barcodes was to extend their coverage length and achieve *de novo* assembly of whole chromosomes, and improve the alignment to theoretical barcodes. Bargroups generated from pair-wise comparisons with overlap windows of 150 pixels, 120 pixels and 100 pixels were evaluated. An overlap window of 120 pixels were later selected for continued analysis based on visual inspection of resulting bargroups across all samples.

As with previous alignments, the similarity between barcodes was evaluated using the Pearson correlation coefficient (cc-score), which depends not only on the quality of alignment but also on the length of the overlapping region. A trade-off exists between overlap length and the alignment reliability: shorter overlaps, such as 100 pixels, may produce artificially high cc-scores that do not necessarily reflect true similarity between two segments of barcodes. Furthermore, extreme intensity peaks in the barcodes disproportionally influence cc-scores, potentially skewing results. As described in the Methodology section 3.4.3.1, these effects were mitigated by masking intensity peaks that deviated more than 3.5 standard deviation from the mean, to enhance alignment reliability.

Another key parameter in bargroup formation was the p-value threshold used for false positive control. By simulating a null distribution using cc-scores from random barcode alignments, a p-value-based threshold was set (see Methodology section 3.4.3.2), allowing control over the false discovery rate. While a p-value threshold of 0.05 may be sufficient for one dataset, another dataset with more variable barcode quality might require a more permissive threshold, such as 0.1, to retain true positive alignments.

Figure 4.18 shows a bargroup composed of five experimental barcodes from the first TSA 5 dataset. Notably, this bargroup demonstrates consistent features across barcodes and a clear extension beyond the length of individual barcodes. Importantly, this group was only formed when the p-value threshold was set to 0.1. At a threshold of 0.05, the same bargroup was excluded. This example highlights the sensitivity of bargroup formation to the chosen false-positive threshold and underscores the need to tailor this parameter to the specific characteristics of each dataset.

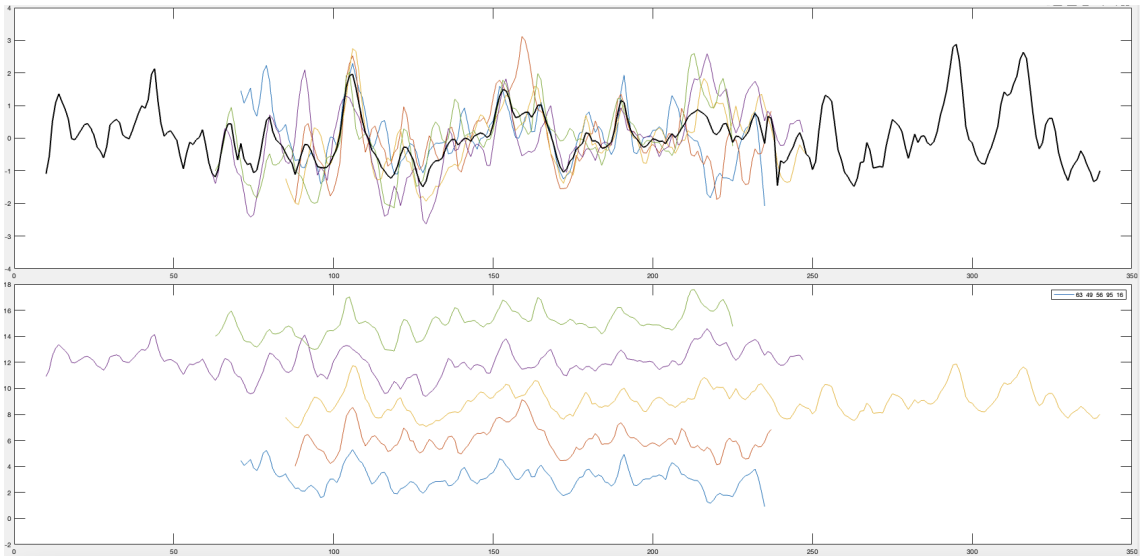

Figure 4.18: Bargroup created from the first dataset of TSA 5 with a p-value threshold for false positives of 0.1, containing five experimental barcodes.

For all distinct datasets, the generated bargroups were manually evaluated based on visual inspection of both alignment quality and extension length. Bargroups that were deemed satisfactory were selected for subsequent reference-based alignment to theoretical barcodes of the *S. cerevisae* genome. An overlap length of 120 pixels was determined to be optimal across all datasets, consistently producing the highest quality bargroups.

4.3.2 Reference-based Alignment

The bargroups selected based on the previously described criteria were aligned to theoretical barcodes of all *S. cerevisae* chromosomes to asses genome coverage. This alignment aimed to evaluate how well the chosen bargroups represented the genome. Both global and local cc-scores were computed for each alignment, and a specific overlap window (in pixels) was set for each case.

For instance, in Figure 4.19, the selected bargroups from the first TSA 0 dataset were aligned to chromosome V using a 200 pixel overlap window. The resulting alingment demonstrates a full coverage of the chromosome. Notably, bargroup 7 spans the entire length of the chromosome. Both global and local cc-scores for each aligned bargroup are displayed in the figure.

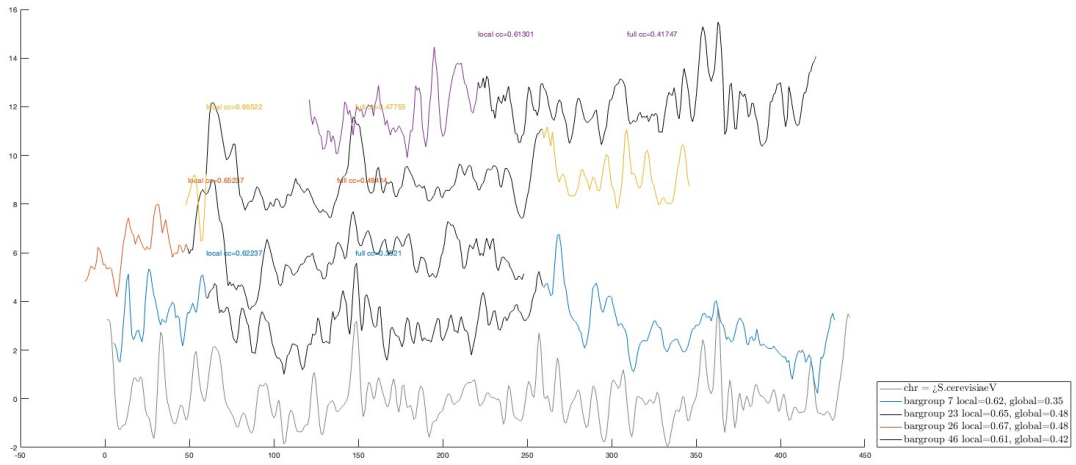

Figure 4.19: Bargroups from the first TSA 0 dataset aligned to chromosome V using a 200-pixel overlap window. Both global and local cc-scores are shown in the bottom right corner.

There were also instances where the selected bargroups alone were insufficient to achieve full chromosome coverage. In Figure 4.20, bargroups together with individual experimental barcodes from the first TSA 0 dataset were aligned to chromosome XI using a 200 pixel overlap window. It is evident from the alignment that the bargroups alone could not span the entire chromosome, prompting the inclusion of additional experimental barcodes. Notably, the alignment shows extension on both ends of the chromosome. Furthermore, the presence of ungrouped barcodes raises the question of why these were not incorporated during bargrouping. This highlights the distinction between alignment based on similarity to the theoretical barcode, as opposed to similarity between experimental barcodes. Both local and global cc-scores are displayed in the figure.

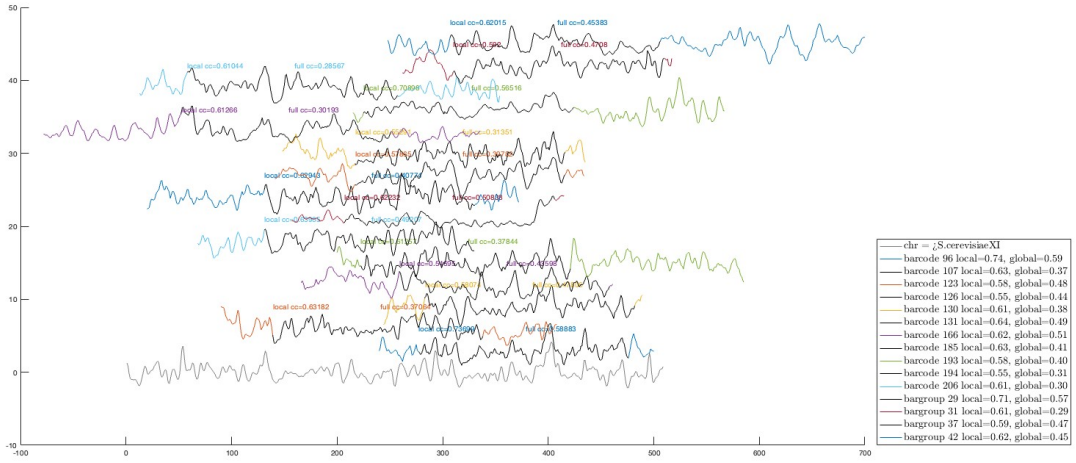

Figure 4.20: Bargroups and barcodes from the first TSA 0 dataset aligned to chromosome XI using a 200-pixel overlap window. Both global and local cc-scores are shown in the bottom right corner.

In Figure 4.21, bargroups and individual barcodes from the first TSA 5 dataset have been aligned to the theoretical barcode of chromosome II. Chromosome II is of particular interest in this project, as it will be examined further for potential structural changes to the genome, or SVs. In this case, the bargroups alone were insufficient to achieve full coverage of the chromosome, which is why additional experimental barcodes were included in the alignment.

The pixel overlap for the alignment presented in Figure 4.21 was set to 300 pixels, which may partly explain why the local cc-scores shown in the figure are generally lower compared to those in Figures 4.19 and 4.20. Another possible explanation for the lower local cc-scores is the presence of structural alterations in this chromosome, such as SVs, which would naturally result in poorer alignment. Such changes would be plausible, given that this sample comes from the Δtsa1 strain of *S. cerevisiae*, which has a compromised defense system against endogenous ROS, and has also undergone five passages through the full cultivation scheme. This possibility will be investigated further in the following section.

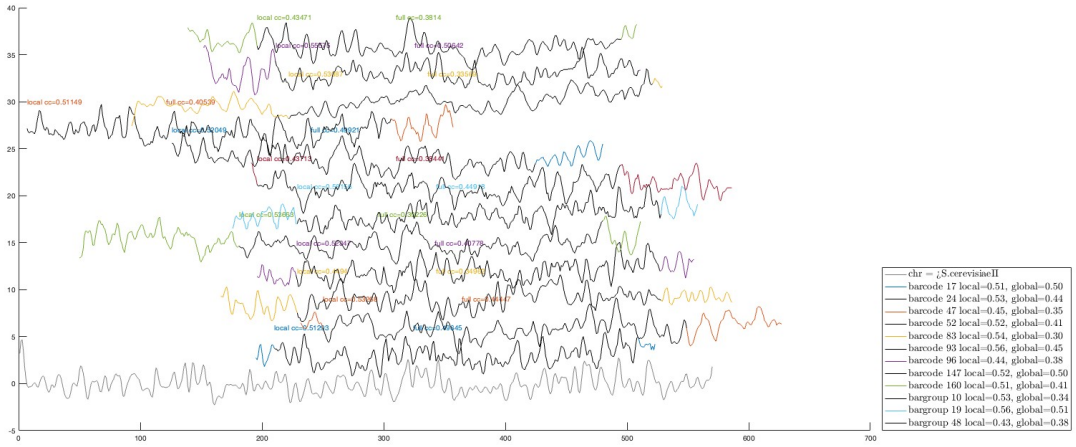

Figure 4.21: Bargroups and barcodes from the first TSA 5 dataset aligned to chromosome II using a 300-pixel overlap window. Both global and local cc-scores are shown in the bottom right corner.

4.4 SV calling focusing on Chromosome II

The following section focuses on chromosome II of *S. cerevisiae*. SVs have been previously reported in this chromosome, particularly in strains with compromised endogenous ROS scavenging systems, such as the Δtsa1 strain. This makes chromosome II especially relevant for further investigation.

Telomeric regions, referring to the ends of chromosomes, are known hotspots for genomic alterations and are particularly sensitive to structural changes. These areas are therefore of high interest when studying SVs.

In Figure 4.22, a consensus of experimental bargroups and barcodes from the first TSA 5 dataset have been made and aligned to the theoretical barcode of *S. cerevisiae* chromosome II. The blue line represents the alignment based on the best local cc-scores from the dataset of barcodes and bargroups, used to generate a consensus barcode. In contrast, the black line shows the consensus barcode generated from all barcodes and bargroups that align to chromosome II within the dataset. The approximate position of the centromere is also marked in the figure, as this region has been previously reported prone to genomic alteration [7].

While a sharp low intensity feature in both of the consensuses is observed at the site of the centromere, consistent with its AT-rich composition and conserved structure. The regions surrounding the centromere, particularly to the left, show more significant deviation between the consensuses and the theoretical barcode. This suggests a possible structural alteration in these flanking regions. However, it is difficult to definitively determine whether these differences originate from centromere changes or extend into subtelomeric or telomeric regions.

Additionally, Figure 4.22 shows an apparent extension beyond the theoretical end of the chromosome on the right-hand side. This is visible in the black line (the globally assembled consensus), which continues past the expected end of the reference. This could be indicative of structural changes such as insertions or duplications in the sample, or potentially incomplete reference information in this region. Therefore, to more accurately identify SVs, it may be more appropriate to compare the TSA 5 sample directly to the WT 0 sample, rather than to the theoretical barcode of chromosome II.

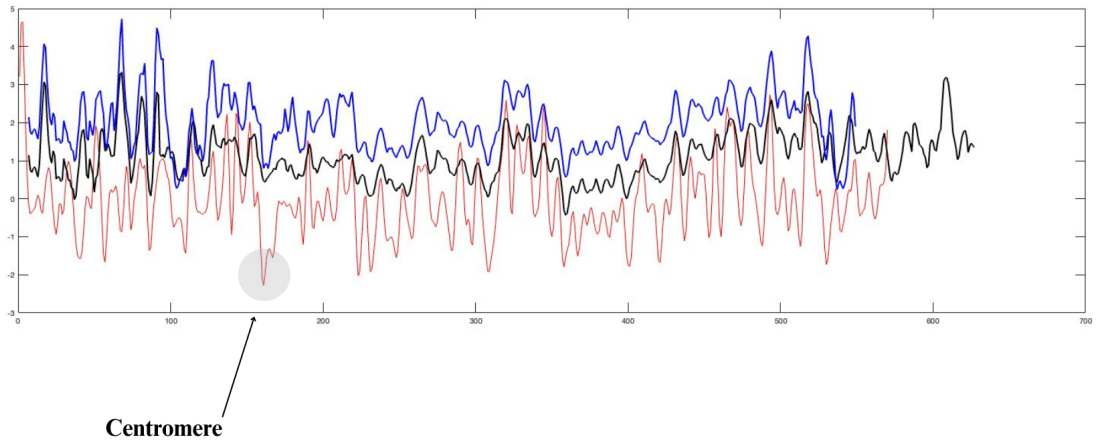

Figure 4.22: Experimental bargroups and barcodes from TSA 5 aligned against chromosome II theoretical barcode. The blue line represents the alignment of the local cc-score while the black line represents the global alignment.

The same experimental consensus of TSA 5 sample is shown in Figure 4.23, compared to the experimental consensus of the WT 0 sample. In this figure, the green segment highlight the 200 pixels where the local alignment of the two consensuses is strongest, with a corresponding local cc-score of 0.85. In contrast, the red segment marks the 150 pixels with the weakest alignment, having a local cc-score of -0.17. The global cc-score for the full alignment between the two consensus barcodes is 0.38.

Visual inspection of the green region clearly suggests that the two consensus barcodes share substantial similarity. However, it is important to not that the consensus barcodes were generated based on alignment to the theoretical *S. cerevisiae* reference genome, using the HCA algorithm. As a result, larger genomic alterations, such as rearrangements or insertions, may not be captured using this method, as only regions that align well to the reference genome are used to construct the consensuses.

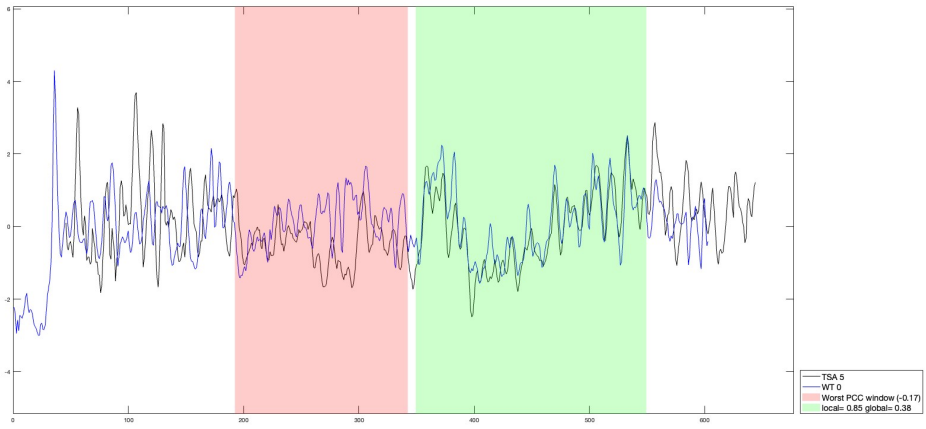

Figure 4.23: Experimental consensus barcodes of WT 0 and TSA 5. The 200-pixel region with the highest local cc-score is highlighted in green, while the 150-pixel region with the lowest local cc-score is highlighted in red. The x-axis represents the length in pixels.

A comparison between the first dataset of TSA 0 and the first dataset of TSA 5 is shown Figure 4.24. As before, the green segment highlights the 200 pixels with the highest local alignment, which in this case yields a cc-score of 0.77. The red segment marks the 150 consecutive pixels with the weakest alignment, with a cc-score of just 0.11. The global cc-score for the entire alignment of the two consensus barcodes is 0.43.

Similar to the comparison shown in Figure 4.23, where WT 0 was used instead of TSA 0, the best local alignment in Figure 4.24 is also centered around the middle of the consensus barcodes. This central region shows consistent peak patterns across samples suggesting it is a conserved region of chromosome II.

Notably in Figure 4.24, TSA 5 shows an extended signal in the left telemetric region compared to TSA 0. While this could indicate that TSA 5 has acquired additional genomic content in this area, it may also reflect insufficient data coverage for TSA 0. To draw stronger conclusions, additional DNA molecules form the TSA 0 sample would be needed to improve consensus reliability. The observation that the lowest local cc-score appears at the far right ens of the alignment supports however the hypothesis that telomeric regions are more prone to SVs or instability.

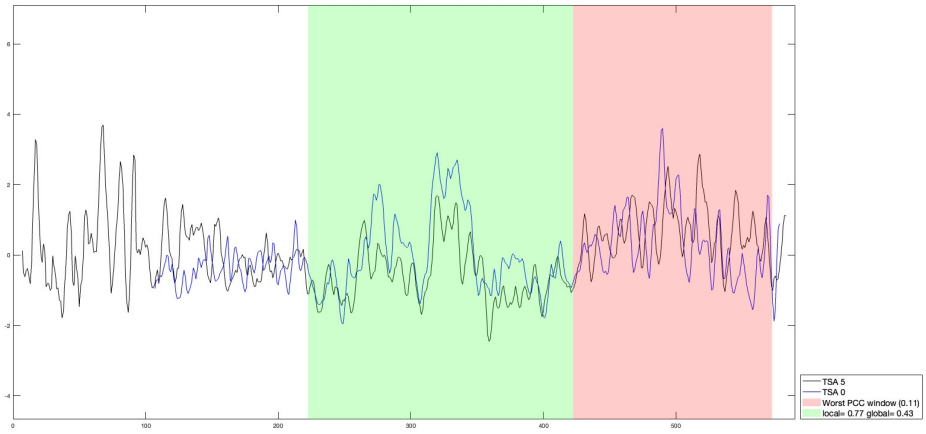

Figure 4.24: Experimental consensus barcodes of TSA 0 and TSA 5. The 200-pixel region with the highest local cc-score is highlighted in green, while the 150-pixel region with the lowest local cc-score is highlighted in red. The x-axis represents the length in pixels.

When instead focusing on the second dataset for both TSA 0 and TSA 5, where the DNA molecules exhibited longer extension, leading to higher resolution, their corresponding consensus barcodes are shown in Figure 4.25. The longer pixel scale here reflects either just an increased physical extension (in nm/bp) of the DNA during imaging, or improved coverage of chromosome II. Considering that the theoretical barcode of chromosome II is 813 kbp and the DNA extension in the case of Figure 4.25 is 0.224 nm/bp, the corresponding consensus theoretical barcode would be expected to span approximately 717 pixels in width. The 150 consecutive pixels with the highest local cc-score yields a value of 0.74, while 150 pixel segment with the lowest cc-score drops to -0.45. The overall global cc-score for the alignment is relatively low, at 0.16. Despite this, the alignment continues to follow the general trend observed in previous comparisons: strong agreement in the central region of the chromosome, with increasing divergence towards both ends. This again suggests potential SVs or data coverage issues in the telomeric regions.

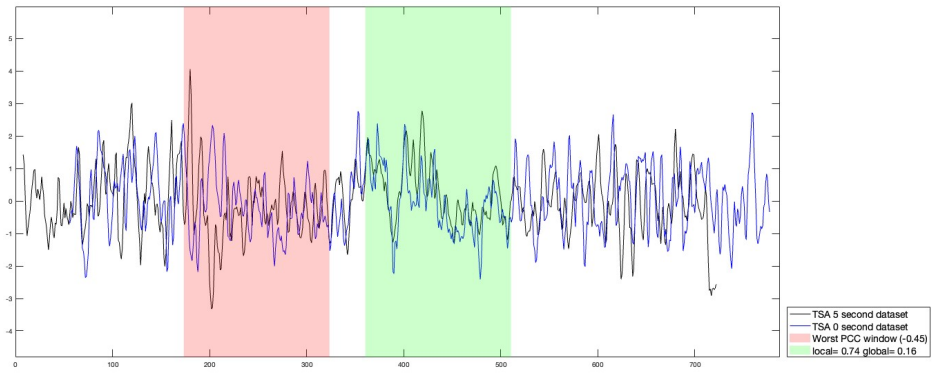

Figure 4.25: Experimental consensus barcodes of second dataset of TSA 0 and second dataset of TSA 5. The 150-pixel region with the highest local cc-score is highlighted in green, while the 150-pixel region with the lowest local cc-score is highlighted in red. The x-axis represents the length in pixels.

When performing comparisons between consensus barcodes, it is necessary to define the pixel length for the consecutive sequences used to evaluate both the lowest and highest local cc-score regions. This choice directly influences the outcome of the cc-scoring. For this reason, telomeric regions which often show visible extension without proper alignment, have been excluded from the lowest local cc-score display segments in two of the three previously shown consensus comparisons. However, when considering the global cc-score instead of local segments, a clearer picture of overall alignment quality emerges. In all three cases illustrated in Figures 4.23, 4.24 and 4.25, the global cc-scores are relatively low, indicating substantial differences across the full length of consensus barcodes. Notably, the lowest global cc-score is observed in the comparison between the consensus barcodes from the second datasets of TSA 0 and TSA 5.

Since consensus barcodes were generated from all experimental barcodes and bar-groups associated with each chromosome from every sample, some filtering out of those with low local cc-scores could have improved overall quality. However, this might reduce coverage, reinforcing the need for more data to increase confidence in the consensus barcodes and to reliably identify SVs. It is also possible that there exist differences within a DNA sample, meaning that it was not entirely homogeneous. This is particularly relevant during the final stage of the cultivation scheme, where the yeast was grown in liquid cultures solely to increase biomass and ensure cell-wall composition possible for DNA extraction. Although this step was limited to a maximum of five generations of growth, genomic heterogeneity may still have been introduced in the DNA samples.

5

Conclusion and Future Work

By evaluating two different methods for chromosomal DNA extraction, and testing two distinct enzyme compositions for cell wall degradation in *S. cerevisiae*, important insights were gained on how to best cultivate this organism for the purposes of this project. A cultivation scheme was established that involved multiple passages, i.e., re-streaking single colonies of YPD agar plates, to induce oxidative stress while also maintaining a relatively homogeneous genome within each sample.

Optical genome mapping was carried out using competitive binding-based staining and DNA stretching via nanofluidics, followed by fluorescence microscopy to image the DNA. This approach was applied to four distinct sample types: WT 0, WT 5, TSA 0 and TSA 5. Intensity profiles, referred to as barcodes, were generated for each sample and aligned to theoretical barcodes of *S. cerevisiae* to assess genome coverage. To obtain more reliable results and extract deeper insights, the analysis would have benefited from greater data volume and a larger coverage of the *S. cerevisiae* genome. At present, it remains unclear what level of coverage, meaning how many times experimental barcodes need to span the theoretical genome, is necessary to confidently reconstruct the correct barcode. During this project, the highest coverage achieved for a single sample was approximately 10x. For future work, the number needed should be systematically investigated and likely increased. In general, higher coverage would improve confidence and resolution in detecting structural variations.

In addition to coverage, the quality of the DNA samples had significant impact on the analysis. One major challenge during the nanofluidics experiments was achieving sufficient and consistent DNA extension. The extension, measured using reference λ-DNA, determines the resolution, meaning how many base pairs are represented per pixel. For meaningful comparisons across samples, DNA extension should ideally be similar between experiments. Extension quality can be influenced by factors such as sample contamination, suggesting that further optimizing of DNA extraction protocols would improve both resolution and reliability.

The observation made regarding chromosome II are particularly interesting, indicating potential structural changes made to the genome arising from oxidative stress. However, the limitation in coverage and resolution must be addressed to confidently interpret these findings. While there is evidence suggesting that SVs may have de-

veloped in the TSA 5 sample, the conclusion is limited due to low coverage. Since the analysis primary relied on alignment to theoretical barcodes, a more *de novo* strategy might uncover structural variants that through the current approach might have been discarded. Such an approach would strongly benefit from improved data depth.

This project focused on chromosome II due to prior reports of variation in this region and time constrains. However, future studies could broaden this scope. Larger chromosomes, such as chromosome IV or XV, are also of particular interest, as their size and the presence of repetitive regions have made them challenging to study using conventional methods like CHEF electrophoresis or NGS. Another interesting approach would be to expand the range of *S. cerevisiae* strains studied, particularly those with mutations leading to compromised DNA repair systems, comparing to the Δtsa1 mutant. This could help in understanding the mechanisms leading to structural variation.

Finally, while this project focused on *S. cerevisiae*, optical genome mapping hold strong potential for broader applications, including human genomics. As the technology improves, this method could play a role in studying structural variants associated with human disease.

Bibliography

1. Crick F. Central Dogma of Molecular Biology. Nature 1970; 227:561–3. DOI: 10.1038/227561a0
2. Ho SS, Urban AE, and Mills RE. Structural variation in the sequencing era. Nature Reviews Genetics 2020; 21:171–89. DOI: 10.1038/s41576-019-0180-9
3. Schütte J, Reusch J, Khandanpour C, and Eisfeld C. Structural variants as a basis for targeted therapies in hematological malignancies. Front Oncol 2019; 9:475303
4. Zhu Y, Brown HN, Zhang Y, Stevens RG, and Zheng T. Period3 structural variation: a circadian biomarker associated with breast cancer in young women. Cancer Epidemiol Biomarkers Prev 2005; 14:268–70. DOI: 10.1158/1055-9965.268.14.1
5. Müller V and Westerlund F. Optical DNA mapping in nanofluidic devices: principles and applications. Lab Chip 2017; 17(4):579–90. DOI: 10.1039/C6LC01439A
6. Parapouli M, Vasileiadis A, Afendra AS, and Hatziloukas E. Saccharomyces cerevisiae and its industrial applications. AIMS Microbiol 2020; 6:1–31. DOI: 10.3934/microbiol.2020001
7. Degtyareva NP, Chen L, Mieczkowski P, Petes TD, and Doetsch PW. Chronic oxidative DNA damage due to DNA repair defects causes chromosomal instability in Saccharomyces cerevisiae. Molecular and Cellular Biology 2008; 28:5432–45. DOI: 10.1128/MCB.00307-08
8. Brown T. Genomes. 2nd ed. Chapter 1: The Human Genome. Oxford: Wiley-Liss, 2002
9. Alberts B, Johnson A, Lewis J, and al. et. Molecular Biology of the Cell. 4th. The Structure and Function of DNA. New York: Garland Science, 2002
10. Lipfert J, Doniach S, Das R, and Herschlag D. Understanding nucleic acid-ion interactions. Annual Review of Biochemistry 2014; 83:813–41. DOI: 10.1146/annurev-biochem-060409-092720
11. Bencurova E, Akash A, Dobson RCJ, and Dandekar T. DNA storage—from natural biology to synthetic biology. Computational and Structural Biotechnology Journal 2023; 21:1227–35. DOI: 10.1016/j.csbj.2023.01.045
12. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, and Walter P. Molecular Biology of the Cell. 4th ed. New York: Garland Science, 2002. Chap. Chromosomal DNA and Its Packaging in the Chromatin Fiber. [Accessed on: 2025 Jun 24]

13. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G, and Malonia SK. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology 2023; 12:997. DOI: 10.3390/biology12070997
14. Yuan Y, Chung CYL, and Chan TF. Advances in optical mapping for genomic research. Computational and Structural Biotechnology Journal 2020; 18:2051–62. DOI: 10.1016/j.csbj.2020.07.018
15. Volpi EV and Bridger JM. FISH Glossary: An Overview of the Fluorescence in Situ Hybridization Technique. BioTechniques 2008; 45:385–409. DOI: 10.2144/000112811
16. Cui C, Shu W, and Li P. Fluorescence In Situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications. Frontiers in Cell and Developmental Biology 2016; 4:89. DOI: 10.3389/fcell.2016.00089
17. Yuan Y, Chung CYL, and Chan TF. Advances in optical mapping for genomic research. Computational and Structural Biotechnology Journal 2020; 18:2051–62. DOI: 10.1016/j.csbj.2020.07.018
18. Lam E, Hastie A, Lin C, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology 2012; 30:771–6. DOI: 10.1038/nbt.2303
19. Müller V, Dvirnas A, Andersson J, Singh V, KK S, Johansson P, Ebenstein Y, Ambjörnsson T, and Westerlund F. Enzyme-free optical DNA mapping of the human genome using competitive binding. Nucleic Acids Research 2019; 47(15):e89. DOI: 10.1093/nar/gkz489
20. Frykholm K, Müller V, KK S, Dorfman KD, and Westerlund F. DNA in nanochannels: theory and applications. Quarterly Reviews of Biophysics 2022; 55:e12. DOI: 10.1017/S0033583522000117
21. Becker A and Murialdo H. Bacteriophage lambda DNA: the beginning of the end. J Bacteriol 1990; 172:2819–24
22. Reisner W, Pedersen JN, and Austin RH. DNA confinement in nanochannels: physics and biological applications. Reports on Progress in Physics 2012; 75:106601. DOI: 10.1088/0034-4885/75/10/106601
23. Gruszka D, Jeffet J, Margalit S, Michaeli Y, and Ebenstein Y. Single-molecule optical genome mapping in nanochannels: multidisciplinarity at the nanoscale. Essays in Biochemistry 2021; 65:51–66. DOI: 10.1042/EBC20200021
24. Reisner W, Larsen NB, Silahtaroglu A, Kristensen A, Tommerup N, Tegenfeldt JO, and Flyvbjerg H. Single-molecule denaturation mapping of DNA in nanofluidic channels. Proceedings of the National Academy of Sciences of the United States of America 2010; 107:13294–9. DOI: 10.1073/pnas.1007081107
25. Chan S, Lam E, Saghbini M, Bocklandt S, Hastie A, Cao H, Holmlin E, and Borodkin M. Structural Variation Detection and Analysis Using Bionano Optical Mapping. *Copy Number Variants: Methods and Protocols*. Ed. by Bickhart DM. New York, NY: Springer New York, 2018 :193–203. DOI: 10.1007/978-1-4939-8666-8_16
26. Nilsson AN, Emilsson G, Nyberg LK, Noble C, Stadler LS, Fritzsche J, Moore ER, Tegenfeldt JO, Ambjörnsson T, and Westerlund F. Competitive binding-based optical DNA mapping for fast identification of bacteria–multi-ligand

transfer matrix theory and experimental applications on Escherichia coli. Nucleic Acids Research 2014; 42:e118. DOI: 10.1093/nar/gku556
27. Rye HS, Yue S, Wemmer DE, Quesada MA, Haugland RP, Mathies RA, and Glazer AN. Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: properties and applications. Nucleic Acids Res 1992; 20:2803–12. DOI: 10.1093/nar/20.11.2803
28. Kundukad B, Yan J, and Doyle PS. Effect of YOYO-1 on the mechanical properties of DNA. Soft Matter 2014; 10:9721–8
29. Berman HM, Neidle S, Zimmer C, and Thrum H. Netropsin, a DNA-binding oligopeptide: structural and binding studies. Biochim Biophys Acta 1979; 561:124–31
30. Nyberg LK, Persson F, Berg J, Bergström J, Fransson E, Olsson L, Persson M, Stålnacke A, Wigenius J, and Tegenfeldt JO. A single-step competitive binding assay for mapping of single DNA molecules. Biochem Biophys Res Commun 2012; 417:404–8
31. Smith AC, Neveling K, and Kanagal-Shamanna R. Optical genome mapping for structural variation analysis in hematologic malignancies. American Journal of Hematology 2022; 97:975–82. DOI: 10.1002/ajh.26587
32. Feuk L, Carson AR, and Scherer SW. Structural variation in the human genome. Nature Reviews Genetics 2006; 7:85–97. DOI: 10.1038/nrg1767
33. Currall BB, Chiangmai C, Talkowski ME, and Morton CC. Mechanisms for Structural Variation in the Human Genome. Current Genetic Medicine Reports 2013; 1:81–90. DOI: 10.1007/s40142-013-0012-8
34. Duina AA, Miller ME, and Keeney JB. Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. Genetics 2014; 197:33–48. DOI: 10.1534/genetics.114.163188
35. Goffeau A et al. Life with 6000 Genes. Science 1996; 274:546–67. DOI: 10.1126/science.274.5287.546
36. Winston F, Dollard C, and Ricupero-Hovasse SL. Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast 1995; 11:53–5. DOI: 10.1002/yea.320110107
37. Engel SR, Wong ED, Nash RS, Aleksander SB, Alexander M, Shu S, Weng S, and Cherry JM. The Saccharomyces Genome Database in 2022: curated annotations for all known genes of the budding yeast Saccharomyces cerevisiae. Nucleic Acids Research 2022; 50:D956–D961. DOI: 10.1093/nar/gkab1053
38. Yue J, Li J, Aigrain L, Hallin J, Persson K, Oliver K, Bergström A, Coupland P, Warringer J, Lagomarsino M, et al. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. Nature Genetics 2017; 49:913–24. DOI: 10.1038/ng.3847
39. Lesage G and Bussey H. Cell Wall Assembly in Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews 2006; 70:317–43. DOI: 10.1128/MMBR.00038-05
40. Klis FM, Boorsma A, and De Groot PWJ. Cell wall construction in Saccharomyces cerevisiae. Yeast 2006; 23:185–202. DOI: 10.1002/yea.1349
41. Orlean P. Architecture and Biosynthesis of the Saccharomyces cerevisiae Cell Wall. Genetics 2012 Nov; 192:775–818. DOI: 10.1534/genetics.112.144485

42. Kawai S and Murata K. Transformation of Saccharomyces cerevisiae: Spheroplast Method. *Genetic Transformation Systems in Fungi, Volume 1*. Ed. by Berg M van den and Maruthachalam K. Fungal Biology. Cham: Springer, 2015. DOI: 10.1007/978-3-319-10142-2_5
43. Scott JH and Schekman R. Lyticase: Endoglucanase and Protease Activities That Act Together in Yeast Cell Lysis. J Bacteriol 1980; 142:414–23. DOI: 10.1128/jb.142.2.414-423.1980
44. Rodríguez-Peña JM, Díez-Muñiz S, Bermejo C, Nombela C, and Arroyo J. Activation of the yeast cell wall integrity MAPK pathway by zymolyase depends on protease and glucanase activities and requires the mucin-like protein Hkr1 but not Msb2. FEBS Lett 2013; 587:3675–80. DOI: 10.1016/j.febslet.2013.09.030
45. Dymond JS. Chapter Twelve - Saccharomyces cerevisiae Growth Media. *Methods in Enzymology*. Ed. by Lorsch J. Vol. 533. Academic Press, 2013 :191–204. DOI: 10.1016/B978-0-12-420067-8.00012-X
46. Reynolds TB, Jansen A, Peng X, and Fink GR. Mat Formation in Saccharomyces cerevisiae Requires Nutrient and pH Gradients. Eukaryotic Cell 2008; 7:865–74. DOI: 10.1128/ec.00310-06
47. Barrios-González J. Solid-state fermentation: Physiology of solid medium, its molecular basis and applications. Process Biochem 2012; 47:175–85. DOI: 10.1016/j.procbio.2011.11.016
48. Lievense JC and Lim HC. The Growth and Dynamics of *Saccharomyces cerevisiae*. *Annual Reports on Fermentation Processes*. Ed. by Tsao GT. Vol. 5. Elsevier, 1982 :211–62. DOI: 10.1016/B978-0-12-040305-9.50010-8
49. Gray BF and Kirwan NA. Growth rates of yeast colonies on solid media. Biophysical Chemistry 1974; 1:204–13. DOI: 10.1016/0301-4622(74)80006-2
50. Sies H and Jones DP. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. Nature Reviews Molecular Cell Biology 2020; 21:363–83. DOI: 10.1038/s41580-020-0230-3
51. Jamieson DJ. Oxidative stress responses of the yeast Saccharomyces cerevisiae. Yeast 1998; 14:1511–27. DOI: 10.1002/(SICI)1097-0061(199812)14:16<1511::AID-YEA356>3.0.CO;2-S
52. Hofmann B, Hecht HJ, and Flohé L. Peroxiredoxins. Biological Chemistry 2002; 383:347–64. DOI: 10.1515/BC.2002.040
53. Mieczkowski PA, Lemoine FJ, and Petes TD. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae. DNA Repair 2006; 5:1010–20. DOI: 10.1016/j.dnarep.2006.05.027
54. Zymo Research. Zymolyase Ultra Datasheet. Accessed: 2025-01-28. 2024. Available from: https://files.zymoresearch.com/datasheets/ds2713_zymolyase_ultra_datasheet.pdf

A

Appendix 1

A.1 Example of a staining reaction for competitive binding

For the following example of a staining reaction for competitive binding, extracted DNA concentration had been decided with Qubit measurement to 5.20 ng/µl, meaning 8.00 µMbp DNA. The total amount of DNA for the staining was decided to be 5 µMbp, where 70% consisted of extracted DNA sample and 30% λ-DNA. In Table A.1, the components of the staining reaction are presented together with the respective final concentrations and volumes. The total final volume was 10 µl.

Table A.1: Components, final concentrations and volumes used for a staining reaction.

Component	**Final Concentration (µM)**	**Volume (µl)**
YOYO-1	0.5	0.5
Netropsin	150	1.5
λ-DNA	1.5	3.0
Extracted DNA	3.5	4.4
0.5X TBE	-	0.63

A.2 λ-DNA lengths

All experiments that yielded usable data for analysis are listed in Table A.2. The table includes the sample type, date of the experiment, the number of λ-DNA molecules detected togehter with their average size.

Table A.2: Sample type, date of experiment and average λ-DNA size for all experiments which were used for the data processing.

Sample Type	**Date**	**Average size (µm)**
WT 0	250311	12.80
WT 0	250325	10.50
WT 0	250326	12.98
WT 0	250416	12.94
WT 0	250417	11.80
WT 0	250425	11.86
WT 5	250404	9.07
WT 5	250422	9.81
TSA 0	250314	8.65
TSA 0	250324	9.01
TSA 0	250328	9.23
TSA 0	250331	10.02
TSA 0	250403	10.20
TSA 0	250424	11.74
TSA 0	250424	11.97
TSA 5	250331	8.44
TSA 5	250407	8.40
TSA 5	250409	8.80
TSA 5	250410	8.70
TSA 5	250410	8.80
TSA 5	250425	10.88

B

Appendix 2

B.1 Parameters for DBM

The parameters used for the molecule detection and generation of kymographs using the MATLAB script called DNA Barcode Matchmaker (DBM) are presented in Figure B.1. The pixel size (nmPerPixel) was set to 254 nm/px which is a number based on the fluorescent microscope used.

Figure B.1: The parameters used for running the MATLAB script DBM for molecule detection and generation of kymographs.

B.2 Parameters for Bargrouping

The parameters used for bargrouping which yielded the bargroups used for continued analysis are listed in Table B.1. Note that the TSA 0 and TSA 5 samples are each divided into two datasets. Here, local stretch refers to the degree of local stretching allowed when assigning a barcode to a bargroup. Distance allowed indicates the maximum deviation (in pixels) from the true alignment position that is tolerated for inclusion in a bargroup. The p-value represents the threshold used to control for false positives during bargroup assignment.

Table B.1: The parameters used for bargrouping of barcodes, indicating the dataset, local stretch and distance (in pixels) allowed when assigning to bargroup. The p-value for false positive control is also presented.

Dataset	**WT 0**	**WT 5**	**TSA 0 #1**	**TSA 0 #2**	**TSA 5 #1**	**TSA 5 #2**
Local stretch	0.05	0.05	0.05	0.05	0.05	0.05
Distance allowed	20	20	20	20	20	20
P-value	0.05	0.05	0.05	0.05	0.1	0.05

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY