Diplomarbeit

Interaction of Werner gene and SNEV

ausgeführt am

Institut für angewandte Mikrobiologie
Universität für Bodenkultur Wien

von

Hannes Rußmayer

Wien, im September 2010

Danksagung

Ich möchte mich herzlich bei allen Personen bedanken, die mich bei dieser Diplomarbeit und im vorausgegangenen Studium unterstützt haben.
Mein Dank gilt meiner Familie, die stets ein offenes Ohr für mich hatten.

Ganz besonderer Dank gilt **Dr. Johannes Grillari**, **Dr. Michael Sauer** **Dr. Hans Marx** und **Dr. Michael Breitenbach** für die wissenschaftliche Betreuung und Unterstützung bei der Durchführung meiner Diplomarbeit.

Weiters möchte ich mich noch bei **Dr. Markus Ralser** und **Dr. Bernd Timmermann** bedanken, die mich während meines Aufenthaltes am Max Planck Institut für Molekulare Genetik in Berlin betreut haben.

Abstract

The accumulation of somatic mutation is a major cause of aging and age-related diseases. In addition to this accumulation the efficiency and fidelity of the DNA repair mechanism declines with progression of aging. During the lifecycle of a cell it faces various types of DNA damages. One of the most harmful DNA damages is DNA interstrand cross links (ICLs). The exact repair process of ICLs is not known yet, but it involves different DNA repair pathways. So there is a big effort to understand the ICL repair pathway. This diploma thesis deals with two specific genes, which have a role in ICL repair. The complexity of multicellular organism often makes it hard to understand cellular processes. Therefore *Saccharomyces cerevisiae* was used as a single cell eukaryote model organism.
The first gene is Prp19/Pso4, which plays a dual role in the cell. It has an essential function in pre-mRNA splicing and a role in processing of DNA lesions especially ICLs. In humans the Prp19/Pso4 has two different names, SNEV (Senescence evasion factor) and hNMP200 (human nuclear matrix protein). Prp19/Pso4 is an essential gene in yeasts. *pso4-1* mutants show a pleiotropic phenotype. The second gene Sgs1 belongs to the RecQ-helicase family. RecQ-helicases are highly conserved DNA helicases that are involved in various processes like as DNA repair, DNA replication and recombination. The human homologue of Sgs1 is the WRN-gene. A mutation within this gene causes the Werner Syndrome. Werner syndrome is an autosomal recessive disorder and belongs to the group of pre-mature aging disorder. These pre-mature aging diseases open a chance to better understand the aging process itself.
Another chapter of this diploma thesis was the analysis of next generation sequencing data. Two different next generation sequencing technologies were used to identify specific mutations in yeast strains. 454 Life Sciences/Roche sequencing technology was used to identify a dominant mutation in K6001-B7, which causes oxidants resistance. The phenotype causing mutation was identified by subtractive bioinformatic analysis and finally was located in the Tsa1 gene. Tsa1p encodes a peroxiredoxin.
This same approach was used to identify mutations in two independent developed revertant yeast strains. The revertants were generated by incubation of a *pso4-1* strain for 2-3 days at 30°C. As mentioned above the *pso4-1* strain has a temperature sensitive phenotype. The revertant yeast strains show a revision of this temperature sensitivity. The Illumina/Solexa sequencing technology was used to sequence the two revertants and the parental *pso4-1* strain. The final outcome was that 1 putative variant was found in the revertant 1 and 10 variations were found in revertant 2. These variations have to be checked by PCR and test for restoring the revertant phenotype in the *pso4-1* strain.
The combination of next generation sequencing technologies and functional genomics may facilitate solving of future genetic problems.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Yeast as a model organism in aging research

Yeasts are known for years as a powerful tool in biotechnological processes. Especially the baker´s yeast *Saccharomyces cerevisiae* is a well known eukaryote and often used in fermentative processes. The ability to ferment sugar under aerobic and anaerobic conditions to ethanol and carbondioxid is one of the most important fermentation in food industry. *Saccharomyces cerevisiae* belongs to the group of fungi more precisely to the higher fungi (Eumycetes). Yeast strains can be maintained as haploid cells or diploid cells. For Reproduction yeasts use two different ways. On the one hand there is the vegetative growth through mitotic cell division. On the other hand diploid cells sporulat under special starving conditions. Sporulation produces four haploid spores via a meiotic cell division. Diploid and haploid cells can grow vegetatively. To restore a diploid cell two haploid cells of different mating types can conjugate. The lifecyle of yeast is shown in Fig. 1.1.

S. cerevisiae is one of the most thoroughly researched eukaryotic microorganisms. The genome was first sequenced in 1996 and comprises 16 well characterized chromosomes. The genome size is 12.5 Mb and contains about 6000 genes. Approximately 1200 of these genes are essential for living. Most genes encode proteins and only a few yeast genes possess introns (Goffeau et al., 1996).

The second important role of *Saccharomyces cerevisiae* is the use as a model organism to study genetics and cellular mechanism. Why is yeast a good model organism in genetics. Cellular and genetic mechanisms are conserved between yeast and human. A large amount of cell cycle proteins and proteins involved in DNA damage are similar in yeast of humans. Especially in aging research yeast is often used as a model organism to understand the complex genetic mechanism involved in the aging process (Breitenbach et al., 2003).

Interestingly about 120 human disease genes have counterparts in yeast. This suggests that such diseases disrupt basic cellular mechanism, for example DNA repair mechanism(Sharma et al., 2006). A yeast mutant that has lost the functional orthologue of a humane disease gene can be screened for thousand of drugs to identify compounds that restore normal function to the yeast cell.

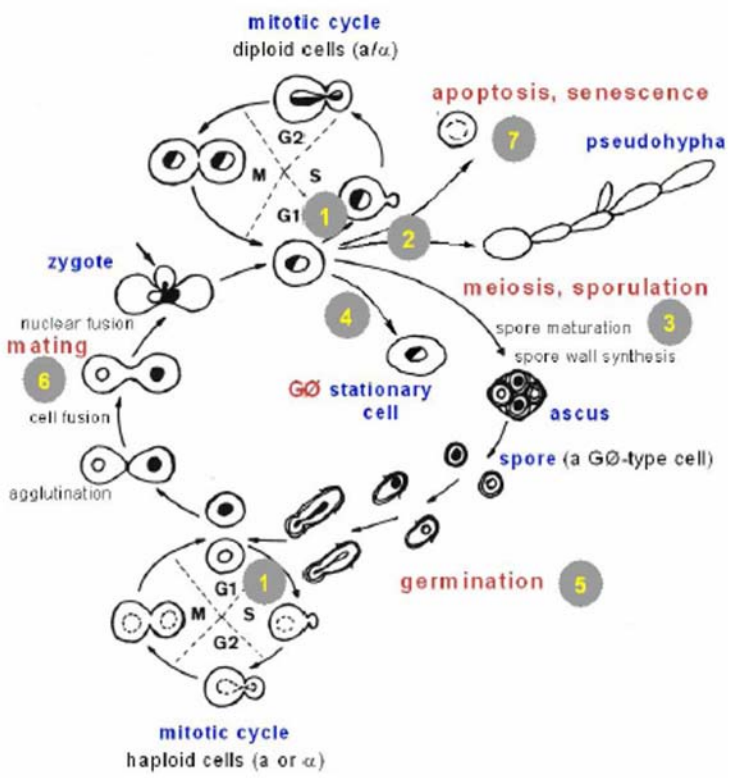


Figure 1.1: Cell Cycle of *Saccharomyces cerevisiae* : 1. mitotic cycle of haploid/diploid cells, 2.Formation of pseudohypha, 3.Meiosis and Sporulation, 4. Entering stationary phase, 5.Germination, 6.Mating and 7. Apoptosis (©Michael Breitenbach)

Several advantages makes yeast also attractive as an model organism for ageing research:

- Yeast is at the same time cell and organism. Ageing of an organism and of a cell can be investigated in one study.
- Yeast grows rapidly, it takes only about 90 minutes for one cell division. Due to this short generation time material for physiological, biochemical and molecular analysis is produced in a short time
- Yeast life spans are short, only a few days
- Yeast mutants are easily produced. The big advantage of haploid yeast mutants is that there is no second interfering allele.
- Methods for lifespan determination are well established.
- Yeast databases are accessible online, which makes functional genomics analysis easier.

Two different aging processes are known in yeasts. On the one hand there is the process of postmitotic ageing, which indicates the behavior of cells in stationary phase over a certain period of time without adding any new nutrients. Also in humans brains postmitotic aging can be observed. In this case neurons cannot be compensated and the function of the brain depends on the lasting neurons. (postmitotic neurons) On the other hand there is an aging process called replicative aging or mother cell specific aging.(Fig. 1.2) Mother cell specific means that an asymmetric cell division happens and only the mother cell ages. The daughter cell resets the cell divisions clock and has the potential of a full lifespan. After a certain number of cell divisions the mother cell will reach a state of senescence, at which no further cell division can be performed. Only the very last daughter cells undergo no asymmetric division and show a shorter lifespan. The replicative lifespan of a mother cell also depends on the environmental conditions.

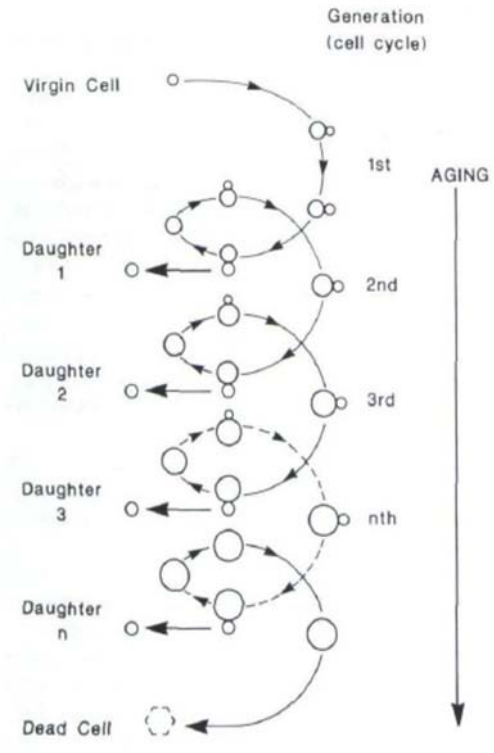


Figure 1.2: Mother cell specific aging (Jazwinski S. M., 1990)

1.2 Reactive oxygen species as a reason for aging

Aging is a decline of function and goes along with a higher mortality risk. It is a complex process and involves a series of different pathways. The complexity of the aging process led to the development of more than one theory for the aging process. Some of these theories only work in multicellular organism, yeast or other eukaryotes.

- Autoimmune Theory
- Death Hormone Theory
- Caloric restriction
- Asymmetric inheritance of damage
- Telomere Theory
- The Somatic Mutation Theory
- The Free Radical Theory

For this work the free radical theory was the most interesting one. The theory implies that oxidative stress and also the generation of free radicals within the cell is involved in the aging process. This theory was postulated by Denham Harman in 1956. Free radicals or radicals are atoms or molecules with an unpaired electron pair. Due to this unpaired electrons almost all radicals are highly chemically reactive. Denham based his theory on observation of irradiated living organisms and saw a shortened lifespan. So he stated that formed radicals cause this shorter lifespan. The most common source for radicals in biological systems is oxygen. Oxygen is an essential molecule in aerobic organism for energy production and cellular metabolism. However, the use of oxygen as a terminal electron acceptor normally converts it to H_2O, but in some cases oxygen is converted to superoxide anion O_2^- and H_20_2. These two molecules belong to the group of reactive oxygen species (ROS). ROS are chemically reactive molecules and contain oxygen. A higher ROS level causes damages of cells in many different ways (DNA damage, oxidation of unsaturated fatty acids in lipids and oxidation of amino acids in proteins). Also, Harman described this in his theory that ROS harm cells in different ways and the accumulation of these damages enhance the aging process. Up to 90% of oxygen consumed from the cell are used in mitochondria. The level of ROS within this organelle has to be higher than in the rest of the cells. The number of oxidative lesions occurring in the mitochondria is ten times higher than in the nucleus and additionally accumulated more quickly with age (Ames et al., 1993). The higher damage rate leads to a less efficient production of ATP, which will disrupt the energy household. Recent studies also showed that mitochondria from older organisms produce far more ROS than younger ones. Different studies also showed a functional decline of mitochondrial function in aging and age-related diseases (Johnson F.B. et al., 1999).

Because of the hazardous potential of ROS, cells developed different ways to protect itself against ROS. The protection involves a battery of enzymes, like the **superoxide dismutases**, **catalases**, **lactoperoxidases**, **glutathione peroxidases** and **peroxiredoxins**.The superoxide dismutases (SOD) catalyses the dismutation of superoxides. The oxidized form of the enzyme reacts with a superoxidion and produces oxygen and the reduced form of the enzyme. The reduced form now reacts with a second superoxidion and 2 protons to hydrogen peroxide. Subsequently hydrogen peroxide is converted in water and oxygen via the enzyme catalase. SOD´s can be divided into three groups, because of the metal iron in their catalytic region: copper/zinc-, manganese-, iron-SOD.
The glutathione peroxidase reduces lipid hydro peroxides to their corresponding alcohols and also splits hydrogen peroxide to water and oxygen. One of the most important compounds to defend rising ROS levels is glutathione itself. Glutathione is a tripe tide containing L-glutamicacid, L-cysteine, glycine. It contains an unusual peptide bond between the gamma-carboxly group of glutamicacid and the amino group of cysteine. The thiol group of glutathione is highly reactive and provides reducing equivalents, which

is a good target for ROS. ROS react with the reduced form of glutathione (GSH) and convert it to the oxidized form. This oxidized form immediately reacts with a second one via disulfide bond formation to from GSSG. GSSG is later reduced with glutathione reductase (Sies, 1999).
The peroxiredoxins are another group of enzymes involved in oxidative stress defense. Peroxiredoxins are thiol proteins with a molecular mass of 17 to 24 kDa. They reduce a broad spectrum of peroxides like H_20_2 and lipid hydroperoxide. Mammalian cells comprise different peroxiredoxins, but the enzymes share the same basic mechanism. A redox-active cysteine in the active center is oxidized to sulfenic acid by the substrate. The mammalian peroxiredoxins are divided in three groups: typical 2-Cys Prxs, atypical 2-Cys Prxs, and 1-Cys Prxs (Kim et al., 1988).
Yeasts also have peroxiredoxins called thiol-specific antioxidant (TSA). It was first discovered as a protein that protects glutamin synthetase against oxidative stress (Kim et al., 1988). A few years later TSA was identified as a member of the peroxiredoxin family via substitution of TSA with alkyl hydroperoxide reductase (AhpC) that is an bacterial peroxiredoxin from *Salmonella typhimurium* (Chae HZ. et al. ,1993). It was also found that thioredoxin was the physiological reductant of TSA.

1.3 DNA repair mechanism

Genetic variation is one of the most important parts in evolution, but survival of the population requires genetic stability. To maintain the genetic stability both an accurate DNA replication mechanism and DNA repair mechanism are necessary. During the lifecycle of a cell it is exposed to several DNA damaging agents leading to changes in DNA. These changes have different reasons, like exposure to radiation of various kinds, exposure to substances in the environment, metabolic by-products. This leads to a tallish number of lesions per day. Due to the remarkable efficiency of DNA repair in cells only a few of these lesions accumulated as a mutation in the DNA. The importance of a functional DNA repair system is evidenced by the fact that various human diseases are linked with a decrease in DNA repair.
The activation of a specific DNA repair mechanism depends on the type of DNA damage and also on the cell cycle phase of the cell. DNA damage can change the conformation of the helix. The alternation is recognized by the DNA repair mechanisms. The next step is the aggregation of specific repair molecules at or near the site of damage to ensure a fast an accurate repair of the damage. The type of molecules which are involved in the repair processes depend on the type of damage.

If only one strand of the DNA helix is damaged three major pathways exists to repair this lesion. The first one is the mismatch repair pathway (MMR).
MMR is an evolutionarily conserved process that repairs mismatches from DNA repli-

cation or recombination. MMR involves the recognition of mismatched bases and also short insertion and deletions loops. The two proteins MutS and MutR are responsible for this recognition steps. MutS binds specifically to a mismatched base pair and the MutR protein scans the damaged strand for a nick. Once MutR found a nick it triggers the degradation of the strand all the way back and through the mismatch. Then the resynthesis of the excised segment starts using the old strand as a template. The most important step is the recognition of the damaged strand. In *E.coli* the strand distinction mechanism depends on the methylation state of the strand. The newly synthesized strand has no methylation. The template strand is methylated this give an opportunity for the mismatch repair proteins to distinguish between correct and incorrect nucleotide. In eukaryotes this recognition process independent of DNA methylation. Here the recognition process works via single strand breaks signaling in the newly synthesized strand (Li, 2008).

The second repair pathway is the base excision repair.(BER) The BER mechanism is active through the entire cell cycle and gets actived by alkylated, oxidized, deaminated C and A and bases with opened rings. There are three different categories of enzymes involved in BER: glycosylases, endonuclase, and polymerase. Depending on the chemical modification of the nucleotide different DNA glycosylases are activated (Krokan et al., 1997). The first step in the BER is the removal of the damaged bases by cleavage of the N-glycosylic bond between the base and the deoxyribose moietes. Once the base is removed the AP site (apurinic, apyrmidinic) is removed by an AP - endonuclease or an AP-lyase. This enzyme cleaves the DNA 5´ or 3´to the AP-site. The remaining deoxyribosephosphat is excised with a phosphodiestarase. The gap is filled by a DNA-Polymerase β and gets ligated by a DNA-Ligase.(Seeberg et al., 1995) There are 2 subpathways of BER. One of the pathwaya is the above mentioned pathway were only a single nucleotide is incorporated during the repair process. The second one is the long patch repair pathway. In this case more than one nucleotide is incorporated during the repair process. Long patch (LP) involes several additional proteins, including flap endonuclease (FEN-1), poly(ADPribosepolymerase) 1 (PARP-1) and Werner protein (WRN). WRN exonuclease and helicase activity together with Polymerase β process BER intermediates, containing a 3´-mismatched substrate (Harrigan et al., 2006).
The third pathway which belongs to the group of single strand damage repair is the nucleotide excision repair (NER). NER repairs DNA damages, which incorporate more than one nucleotide and cause larger changes in the structure of the DNA double helix. Lesion repaired by NER can vary from small DNA distortions that are produced by photoproducts, to larger adduct formed by chemical reactions with endogenous or environmental compounds. Larger lesions are especially formed by reaction with carzinogen benzopyrene. Two different sub pathways can be distinguished, but only the initial recognition step is different. On the one hand there is the transcription coupled repair (TCR-NER). In this case the RNA-Polymerase acts as the initial sensor for damages. For non-transcribed regions the global genome repair NER is used.(GG-NER) Specific

damage recognition factors screen the genome for aberrations.(Fig. 1.3) After the initial recognition step the following cascade of reaction like verifying the damage and removing parts of the damage containing strands are the same in TCR-NER and GG-NER. The proteins involved in the initial recognition step in TC-NER are called CSB protein. CSB proteins are believed to recruit additional compounds to the stalled RNA-Polymerase II, like RPA, TFIIH, p300, CSA-DDB1 and other NER core proteins. CSB/CSA-protein may facilitate the access to the damage site and also help by the addition of factors to the stalled transcription complex. GG-NER depends on the activity of a specific complex XPC-Rad23. The next step in the cascade is the unwinding of the DNA helix. A DNA-complex is formed comprising XPC-Rad23, TFIIH, XPA and RPA. The whole unwinding process is performed by XP-helicase, which consist of the XPB and XPD subunit of the TFIIH. Then the segment, containing the DNA damage, is removed. This is done by an incision on the 3´- and 5´-site of the lesion. The incision is made asymmetrically around the lesion starting with the incision on the 3´-site done by the XPG-protein. The second incision on the 5´-site of the lesions is induced by the XPF-ERCC1 complex resulting in a single strand of 20-30 nucleotides. RPA is needed for the protection of the single strand against nucleases. The gap filling step is done by DNA-Polymerase δ and ϵ together with PCNA. The last step in NER is the ligation of the strands. A disorder connect with the NER is Xeroderma Pigmentosa. XP is an autosomal recessive disorder. Patients with XP are susceptible for damages caused by ultraviolet light. Typical symptoms are sunburns, freckles and blistering after a small dose of light, solar keratoses, scaly skin and higher risk of skin cancer. XP is associated with a variety of mutated genes, like XPA, ERCC3 (XPB), XPC, ERCC2 (XPD), DDB2 (XPE), ERCC4 (XPF), ERCC5 (XPG), ERCC1, and POLH (XP-V). Mutation of XPA and XPC are found in 50 % of all XP patients (Van Hoffen A. et al., 2003).

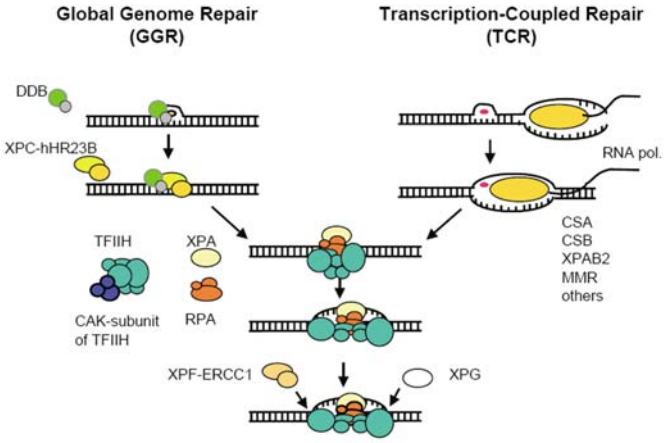


Figure 1.3: Overview Nucleotide Base Excision repair. The first step in TCR-NER and GG-NER are different. The processing of the damage is equal in both pathways (Van Hoffen A. et al., 2003)

This was a brief summary of repair mechanism for single strand damage, but what happens when a double strand break (DSB) occurs. This kind of DNA damage is one of the most harmful. DSB are repaired by two different pathways: non-homologous end joining (NHEJ) and homologous recombination (HR). These two pathways are highly conserved throughout eukaryotes. Only the relative importance differs from one organism to an-

other. For example *Saccharomyces cerevisiae* or *Schizosaccharomyces pompe* use mainly HR to repair DSB. In mammals NHEJ is the first choice in repairing such damages. Also the cell cycle stage plays a role in choosing the repair mechanism. In the G_0 and G_1 phase, where no duplicate of the DNA is available, NHEJ is often use. In the S-Phase the homologous repair pathway is used for repairing DSB.
The NHEJ pathway doesn´t need a second undamaged DNA molecule. The repair pathway comprises different proteins, but the central role plays the Ku protein. Ku is a heterodimer consisting of two subunits Ku70 and Ku80. The Ku protein binds to the DNA in a non-sequence dependent manner. The binding of Ku is determined by the DNA DSB. Ku binds to the DNA ends and brings both ends closer together. The yeast homologue for the human Ku heterdimer is the Hdf1/Hdf2 complex. The Ku heterodimer recruits an additional factor called DNA-dependent protein kinase catalytic subunit. (DNA-PKcs) DNA-PKcs itself has affinity for DNA ends and gets activated by single stranded DNA region derived from DSB. (Jackson, 2002) After binding of DNA-PKcs Ku recruits two other factors XRCC4 and Ligase IV. This two additional factors cause the religation of the lesion. Sometimes the direct ligation of both ends is not possible. In this case the MRE11-RAD50-NBS1 complex, which has an exo- and endonucleases and helicase activities, start with the processing of the DNA ends before ligation. The yeast homologue for the ligation complex XRCC4/Ligase IV is the Dn14-Lif1 complex. Lif1 appears to act as an adaptor between Ku and Dn14.(Jackson, 2002) The NHEJ pathway leads generally to small deletion of DNA and also bigger chromosomal re- arrangements are possible.
The fidelity is much higher for HR than NEHJ. HR is precise, because it uses the genetically identical sister chromatide. HR is not only used for repair of DNA damage. The process of homologous recombination is also often used during meiosis to join DSB. In this case the generation of a DSB has a beneficial impact on the genetic variety. HR in *Saccharomyces cerevisiae* involves a set of genes - RAD50, RAD51, RAD52, RAD54, RAD55, RAD57, RAD59, MRE11 and XRS2 (Sung and Klein, 2006). The first step is a resection in 5´ to 3´direction by RAD50/MRE11/XRS2 to produce on both DNA ends 3´single-stranded DNA tails. RAD51p binds to the single stranded DNA. This binding is influenced by a battery of other proteins like Rad52p, RPA, and Rad54p. The Rad51p nucleoprotein filament interacts with the undamaged DNA molecule and screens it for homologous region. After localization of homologous regions strand exchange events are induced. The damaged DNA molecule enters the intact DNA helix and displaces one strand as a D-loop. Now the 3´end of the damaged DNA is extended by a DNA-polymerase using the undamaged DNA strand as template. The ends are ligated by DNA-ligase I. The last step is the resolution of the holliday junction. The holliday junction is produced by migration of the crossing over during the recombination process. To solve the holliday junction cleavage and religation of the DNA strands are needed to get two intact DNA helices. Homologous recombination repair is a complex mechanism. Most of the proteins involed in HR belong to the Rad52-epistatis group. Recent studies showed that there are two different sub pathways for HR.(Fig. 1.4)

The first steps are equal for both, but after the invasion of the 3´-overhang they split up in double strand break repair (DSBR) pathway and synthesis-dependent strand annealing (SDSA). The above explained mechanism refers to the DSBR. Alternatively the SDSA pathway could be used. Here the invading strand is displaced after repair synthesis and anneals with the single strand DNA of the original strand. The process is finished by gap filling by DNA synthesis and ligation. During the SDSA process no crossing over occurs. If we look at the DSBR there is the possibility to from a crossing over (Sung and Klein, 2006). Also the WRN protein is involved in both NHEJ and HR. There is a hypothesis which says that WRN may plays a role in controlling the DSB repair by suppressing NHEJ and facilitate recruitment of error-free HR machinery. (Comai and Li et al., 2004) Concerning normal ageing a similar process may regulates these pathways during the ageing processes the mode of repair will shift from HR to NHEJ (Gorbunova and Seluanov, 2004). This would lead to an accumulation of DNA damages and gross genomic rearrangements and finally a decline of function.

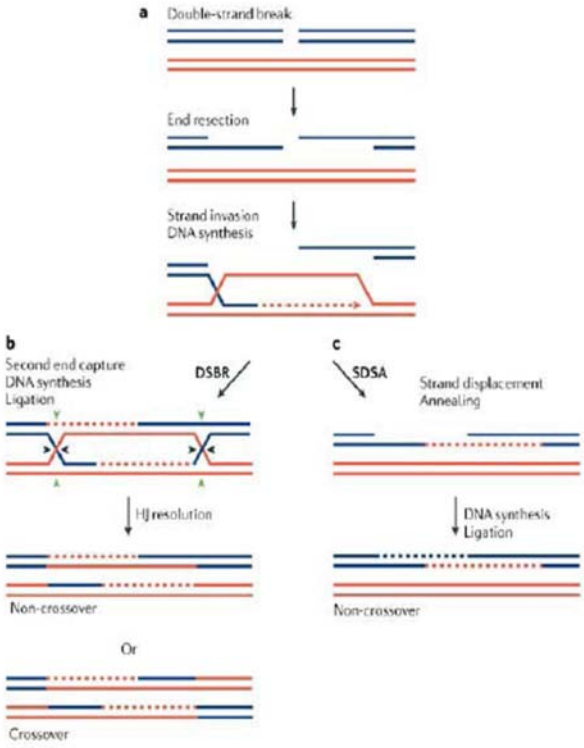


Figure 1.4: The two subpathways of homologous recombination DNA double strand repair(Sung and Klein, 2006)

1.4 RecQ-helicase Family

The RecQ protein family is group of DNA helicases that are highly conserved throughout eukaryotes and prokaryotes. For example *Escherichia coli* (RecQ) , *Saccharomyces cerevisiae* (Sgs1) and *Schizosaccharomyces pompe* (Rqh1) contain one RecQ-helicase . Humans and mice contain five RecQ-homologues (RecQ1, BLM, WRN, RecQ4, RecQ5). All members of the ReQ-helicase family possess a highly conserved helicase domain, but there are two additional conserved regions (Sharma et al., 2006):

RQC domain (RecQ C-terminal)
The RQC motif is located C-terminal to the helicase domain. This region contain a Zn^{2+} binding domain and has a role in binding to DNA and other proteins.

HRDC domain (helicase and RNaseD C-terminal)
Mutations in the HRDC alter the binding ability of the enzymes. HRDC domains in Sgs1 and RecQ may have a function in substrate recognition.

The ReQ-helicase family plays a role in maintenance of the genome stability. They are involved in essential DNA metabolic pathways like replication, recombination or DNA repair.

For example ReQ helicase are used for:

- resolving aberrant structures during the DNA replication
- replication restarts at collapsed replication forks
- replication recovery at stalled replications forks by promoting fork regression and homologous recombination
- resolution of recombinogenic structures
- telomere maintenance via recombination (ALT)
- resolution of the holliday junction during the meiotic segregation

The disruption of a specific human RecQ helicase is often associated with diseases. Almost all of them belong to a group of diseases called human premature aging disorder. Patients with such diseases show most of the clinical signs and symptoms associated with normal aging process.

Following diseases belong to the group of human pre-mature aging diseases:

Bloom syndrome
The Bloom syndrome or also know as BloomTorreMachacek syndrome is an autosomal recessive disorder. This disorder was first described by the dermatologist Dr. David Bloom in 1954. The cause of this disease is a mutation in the Bloom gene, which belongs to the group of human ReCQ-helicases. Mutations in BLM gene cause the loss of protein helicase function or the loss of the whole protein. The phenotypic characteristics for Bloom syndrome are skin color changes, Butterfly-shaped facial rash, fertility problems, increased susceptibility to cancer and leukemia. Patients with the Bloom Syndrome die with a median age of 27 years often related to cancer (Monnat, 2010).

Figure 1.5: Patient with Bloom Syndrome (Left), Patient with Cockayne Syndrome (Right)

Cockayne syndrome

The Cockayne syndrome was first described by Edward Alfred Cockayne and is an autosomal recessive inherited disorder. People with this disease are sensitive to sunlight, have short stature and have the appearance of premature aging. Interestingly, unlike all diseases linked with DNA repair disorders, Cockayne syndrome is not related to cancer. The early congenital form is apparent since birth (Type II). The classical form of this syndrome appears with an age of 1 year. This disease is caused by a disruption of a DNA repair mechanism called transcription coupled NER. Two genes in Cockayne syndrome, CSA and CSB, are identified. The products of these two genes are involved in the transcription coupled NER (Van Hoffen A. et al., 2003).

Rothmund-Thomson syndrome (RTS)

RTS is a rare condition that affects many parts of the bodies. Rothmund-Thomson syndrome is characterized by sparse hair, eyebrows, and eyelashes, slow growth and small stature, abnormalities of the teeth and nails, gastrointestinal problems in infancy and chronic diarrhea and vomiting. People with Rothmund-Thomson syndrome have an increased risk of developing cancer, particularly a form of bone cancer called osteosarcoma. The cause of this disorder is a mutation in the RecQ4. A mutation of this gene leads to a truncated version or complete loss of this protein (Monnat, 2010).

Werner Syndrome

Werner syndrome is an autosomal recessive genetic disorder. A mutation in the WRN gene causes this disorder. The WRN gene encodes a protein, which belongs to the group of RecQ helicase. This shows again the close relation between genomic instability and disruption of RecQ-helicases. Werner syndrome was first described by Otto Werner in 1904. The typical symptoms appear in adolescence or early adulthood and results in the appearance of old age by 30-40 years of age and are short stature, wrinkled skin, baldness,

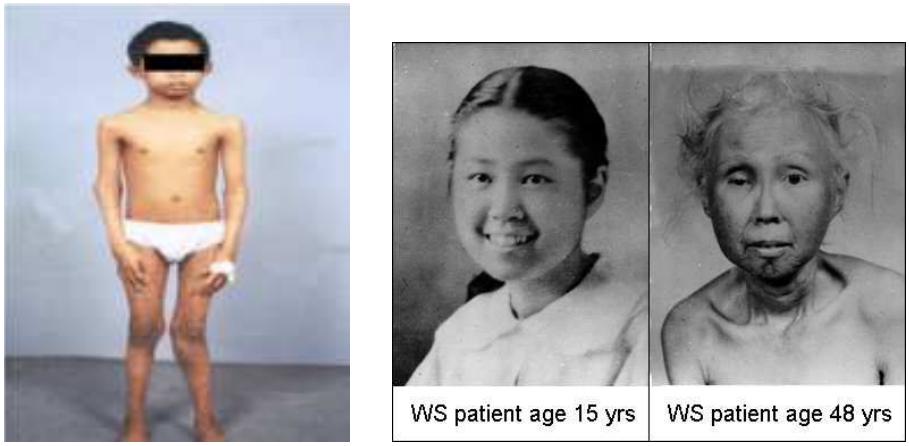


Figure 1.6: Patient with Rothmund Thomson Syndrome (Left), Patient with Werner Syndrome (Right)

cataracts, muscular atrophy and a higher risk for diabetis melitus. The cellular effects of Werner syndrome are genomic instability (chromosomal rearrangments), shortened telomerase, hyperrecombination, shortened replicative lifespan. Due to the mutation in the WRN gene truncated version of the protein or almost no protein is produced. Normally the WRN protein is localized in the nucleus. For the truncated version of WRN it is not possible to get in the nucleus and this seems to be critical for the phatogenesis of Werner syndrome (Moser M. et al., 1999).

WRN is involved in different cellular processes:

Roles in DNA Replication
WRN was found to interact with RPA. RPA is needed for DNA replication and enables the WRN helicase to unwind DNA helix up to 1000 bp. WRN has different specific functions in the replication process, like resolving replication fork blocks or restarting the process. It also has been found that WRN helicase activity helps the DNA-polymerase to synthesize through complex structures in the DNA (hairpins, trinucleotide repeats, tetraplex structures) (Opresko et al., 2003).

Functions in DNA repair pathways
WS cells show larger chromosomal rearrangements and also bigger deletions due to decrease in fidelity of repair processes, especially DSB repair. It is suggested that WRN interacts with factors of the NHEJ and HR repair pathway. Concerning the NHEJ pathway WRN interacts with the Ku-heterdimer and DNA-PKcs. The interaction of WRN with Ku strongly stimulates the WRN exonuclease activity, but not only in elevating the processivity, also in extending the substrates that can be processed (Opresko et al., 2003). Also in HR repair WRN interacts with typical factors of the repair pathway. The main task for WRN in this case is to resolve recombinational intermediates. A third

repair pathway that involves WRN is one of the most important pathways the BER. As stated above the long patch BER incorporates more than one nucleotide producing a 5´dRP containing flap. FEN-1 performs the critical step in this pathway by removing this flap. It is suggested that WRN stimulated the activity of FEN-1 and also unwinds DNA intermediates (Opresko et al., 2003).

Telomere Maintenance
WS-cells show a telomere dysfunction, which are overcome by expression of exogenous telomerase (Opresko et al., 2003). This shows an role of WRN in the maintenance of telomeres. Deficiency in WRN alters the telomere integrity and structure of telomeres. Recent studies suggested that WRN- and also BLM-helicase participate in the alternative lengthening of telomeres (ALT) pathway. ALT is an alternative pathway to maintain telomere length. WRN also interacts with the capping proteins. Capping is important at the telomeres, because without capping the cell will consider the telomerase as DSB breaks. This would lead to gross chromosomal rearrangements due to wrongly activated repair mechanism.

1.4.1 Sgs1 in S. cerevisiae as a member of the ReQ Family

The Sgs1 gene in *Saccharomyces cerevisiae* encodes a DNA Helicase of the ReQ-like family. Sgs1p is the yeast homologue to the above mentioned human RecQ-helicases, like WRN and BLM. Sgs1 (slow growth supressor1) is localized at chromosome XIII with a size of about 4343 bp. It was first characterized by a mutation that caused a suppression of the slow growth phenotype of top3 (topoisomerase III) mutants (Gangloff et al., 1994). Sgs1 is the RecQ-helicase in *Saccharomyces cerevisiae* and yeast deficient in Sgs1 show also phenotypes, which are seen in WS cells. A deletion of Sgs1 in budding yeast causes mitotic hyperrecombination, sensitivity to hydroxyurea and a defect in the intra-S phase DNA damage checkpoint, but no telomere shorting (Cohen and Sinclair, 2001). Hydroxyurea leads to stalling of replication forks, which are restarted by RecQ-helicases. Also a shortened replicative lifespan and abnormal chromosomal structures are seen in Δsgs1 cells. The shorter replicative lifespan was found in both WS cells and Δsgs1 cells. The increased recombination happens also at the repeated ribosomal DNA (rDNA) locus (Sinclair and Guarente, 1997). The formation of extrachromosomal rDNA circles goes much faster in Sgs1 deficient cells. The accumulation of ERC in yeast cell is directly linked to aging in yeast. ERC are formed during the life´s cycle of a yeast cell and accumulate in the mother cell. ERC determine the end of the replicative lifespan in yeast. In older yeast cell with a high number of ERC a enlargement of the nucleoli and fragmentation are observed (Sinclair and Guarente, 1997). Sgs1p is involved in a variety of process and has many interactions with other proteins in yeast cells. The single ReQ-helicase in budding yeast seems to be less specialized and takes over almost all function that in humans are divided to the five RecQ helicases.

1.5 Prp19/Pso4 and the role in DNA repair

As the standard name Prp19/Pso4 tells us this gene was identified by two different cellular processes. First it was identified as a splicing factor associated with the spliceosome and named Prp19 (pre-mRNA processing 19). Other studies identified it as Pso4, due to the study of the *pso4-1* mutant strain. *pso4-1* is a psoralen sensitive mutant. The mutant yeast strain shows a pleiotropic phenotype containing temperature sensitivity, block of sporulation, affected induced mutagenesis, mutagen and radiation sensitivity and accumulation of pre mRNA. (Grey et al., 1996) The *pso4-1* mutant was originally named xs9 isolated from Benathen and Beam (Grey et al., 1996) and is slightly X-ray sensitive and then renamed to *pso4-1* (Henriques and Moustacchi, 1980). Prp19/Pso4 has been found as the orthologue of SNEV (Senescence evasion factor) (Grillari et al., 2005). In humans SNEV has also been identified as hNMP200 (human nuclear matrix protein). Prp19/Pso4 consists of several domains, like U-box with an E3-Ligase domain, self interaction domain (SID), low complexity region (LCR), a coiled coil region (CC) and four WD40 segments on the C-terminus, which form a globular domain that later build a typical beta-propeller structure held together by a central stalk formed by the coiled-coil domains (Fig. 1.7) (Ohi et al., 2005).

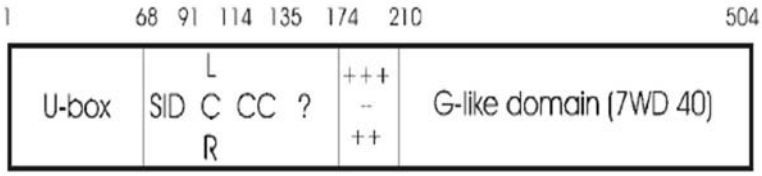


Figure 1.7: Scheme of the hPrp19 domains, Numbers indicate the amino acids(Grillari et al., 2005)

Prp19 is a core protein of the nineteen complex (NTC) that is essential for splicing of pre-mRNA. It interacts with itself and produces a tetrameric complex via the self interaction domain. This tetrameric complex is a structural core component of the NTC. The NTC is an non-snRNA containing protein complex and comprises at least 9 other proteins additionally to Prp19 (Ohi et al., 2005). This protein complex seems to be conserved between eukaryotes, because mammalian cells have a functional equivalent complex called Prp19/CDC5L. The NTC complex binds to the assembling splicosome after the dissociation of U4 snRNA and stabilizes the U5 and U6 snRNAs for the first step in the splicing processes. The first step is a cut of the pre-mRNA at the 5´splicing site. The complex remains at the spliceosome during the second splicing step at the 3´ splicing site (Ohi et al., 2005).

SNEV/Prp19 also contains a U-Box showing an E3-Ligase activity. The substrates of this E3-Ligase are still unknown. It was proven that SNEV interact with a β subunit of the proteasome, PSMB4. (Löscher M. et al 2005) Beside this proteolytic activity SNEV/Prp19

is also involved in the cellular process of senescence and DNA repair. An overexpression of SNEV in human umbilical endothelial cells causes a prolonging of the replicative lifespan of these cells (Voglauer et al., 2006). Another aspect is a link between SNEV and breast cancer. In breast cancer cells express a higher level of SNEV compared with normal cells. Within the different samples of breast cancer cells big deviation between the expression levels of SNEV can be seen. Cells with a higher expression level of SNEV lived longer after diagnosis and also showed a lower number of metastases (Voglauer et al., 2006). All these facts together support a role of SNEV in repair mechanism. Interestingly, a study suggests that SNEV and additional proteins are involved in interstrand crosslink repair (ICL) (Nianxiang et al., 2005). Because of the involvement of SNEV/hPrp19/hPso4 in ICL repair and the functional relation between SNEV and Prp19 reinforce the participation of SNEV in DNA repair processes.

1.6 Next generation sequencing technologies

The definition of DNA sequencing says that it is a process for determining the type and order of the nucleotides in a DNA template. Since the discovery of the structure of the DNA structure there was a big interest to understand biological systems and there for the availability of genomic information is the first step. Also for biotechnological and environmental processes a good characterization of microorganism is of high interest. Due to this motivations DNA sequencing technologies were developed. In 1977 Frederick Sanger and Allan Maxam and Walter Gilbert developed independently methods for DNA sequencing. Until today Sanger sequencing is the method of choice to get an exact sequencing result. Sanger sequencing use dideoxynucleotides (ddNTP´s) as DNA chain terminators. The first sequencing machine was introduced in 1987. At the beginning of the 1990´s only a handful of groups worldwide were able to sequence DNA up to 100.000 bp, but under high costs. To sequence greater genome, like the human genome, the cost and also the performance of the sequencing processes have to be improved. Technical variations of chain termination sequencing were done to improve the sequencing process. For example dye termination sequencing was one of the improvements. In this case the four nucleotides are labeled with a fluorescent dye, each of them with a different wavelength of fluorescence and emission. Also the use of capillary electrophoresis made it easier to sequence larger stretches of DNA. The improvements in DNA sequencing technologies it was essential to filter the generated data to get the correct sequence of the DNA template. So it was time to develop an accurate algorithm to do a correct base calling. The sequencing technologies up to this point were limited to fragment with sizes from 300-1000 bp, because of the inefficient separation of the produced chain differing in only one nucleotide. The capacity of the sequencing machines gets higher and so the very last capillary sequencer produces 6 Mb per 24 hours. Also the sample preparation was a very time intensive step. The next step in development was the introduction of

next generation sequencing methods. So the first next generation sequencer (GS20) was released in 2005 by 454 Life Science/Roche. GS20 had an average read length of 100-250 bp, round about 200.000 reads and produced 20 Mb of sequencing data per run. 454 Life Science steadily improved the technology and so today the GS FLX Titanium sequencer has an average read length of 350-450 bp, 1 million of reads and produces 400-500 Mb of sequencing data per run (10 hours) In the last view years not only 454 Life Science/Roche worked on the field of next generation sequencing.

Following next generation sequencing technologies are available today:

- 454 Life Science/Roche GS FLX
- Illumina Solex Genome Analyzer GAII
- Applied Biosystems Solid^{TM}
- Helicos Heliscope^{TM}
- Pacific Biosciencies SMRT

Helicos Heliscope^{TM} comes recently available and Pacific Biosciencies SMRT will likely launch commercially in 2010. Helicos and Pacific Biosystems Instruments are so called "single molecule" sequencer and do not require an amplification of DNA fragments for the sequencing process. The process of single molecule sequencing could be stated as third generation sequencing.

Applications of next generation sequencing:

- **Research Applications**
  - De-Novo genome sequencing
  - Whole genome resequencing
  - Transcriptome sequencing
  - small RNA sequencing
- **Clinical Applications**
  - variations detection (SNP discovery and identification)
  - Sequencing of interesting clinical samples

1.6.1 Roche/454 FLX Pyrosequencer

The 454 FLX Pyrosequencer was the first next generation sequencer on market. For sequencing the thousands of DNA fragments an alternative sequencing technology called pyrosequencing is used. In pryrosequencing, each incorporation of a nucleotide by DNA polymerase leads to a release of pyrophosphate. This pyrophosphate is later converted in a series of enzymatic steps to ATP and in the finally step this ATP is used from the firefly enzyme luciferase to produce light. The amount of released light is proportional to the number of nucleotides incorporated. The first step in the sequencing process is the library preparation. Here the genome gets fragmented by nebulization and 454 specific adaptors are ligated. The adaptors are important for the steps that follow, including purification of the ligated DNA-fragments using biotin label B adaptor. The dsDNA fragments are purified using Streptavidin beads. The single stranded fragments (ssDNA) with the adaptors that are released from the streptavidin beads are quantified for the emPCR reaction by titration. To generate the library no cloning and colony picking is needed, which is a time saving improvement. The next step in the sequencing process is the emulsion PCR (emPCR). Therefore the library fragments are mixed with agarose beads. These beads carry on their surface olignucleotides that are complementary to the specific adaptors on each fragment. Only one read is per bead. Each of these complexes is isolated in an oil/water micelle that contains PCR reagents. Due to the emPCR millions of copies of the single reads are made. After the emPCR the beads are filled in a picotiter plate. The size of each well of the picotiter plate is big enough to carry one single bead. Enzyme containing beads are added to the microtiter plate that later are used for the sequencing process. During the sequencing process not all four nucleotide at the same time compete for incorporation. Only one nucleotide is cycled over the picotiter plate per time. The picotiter plate is placed into the GLX pyrosequencer opposite to a CCD camera. Now the sequencing process starts. The first four nucleotides (TCGA) on the adaptor beside the sequencing primer are used for calibration of the light emission. This ensures accurate base calling. The big problem concerning pyrosequencing are big scratch (>6) of one single nucleotide also called homopolymers. These areas are prone to base insertion or deletion errors during base calling. The average read length of this next generation sequencer using titanium technology is 350 - 500 bp. Approximately 1 million of reads produce 500 - 600 Mb per run. One run takes about 9 hours.

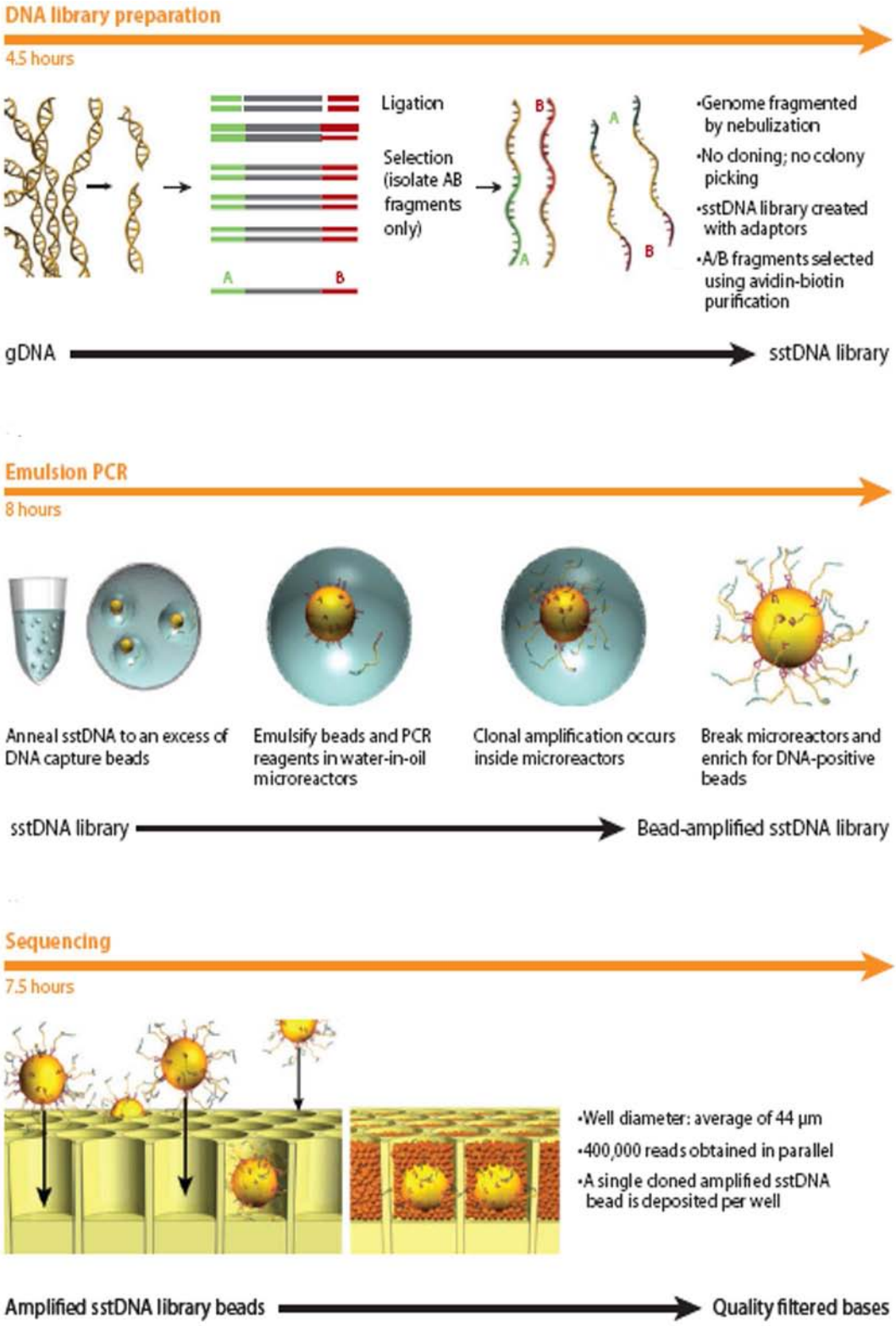


Figure 1.8: Sequencing workflow for the 454 Life Sciences/Roche technology (Mardis, 2008)

1.6.2 Illumina Genome Analyzer

The Illumina sequencing process starts with a fragmentation of genomic DNA and adaptor ligation. The adaptor ligated DNA fragments are washed over the surface of a flow cell and bind to it. This works via hybridization between specific oligonucleotides on the surface of the flow cell and the Illumina specific adaptor ligated to the DNA fragments. The next step is the amplification of the single stranded DNA fragment. This is done be bridge amplification and the process is called cluster formation. Cluster formation is done in an automated device named Cluster station. One cluster represents one DNA-fragment. The flow cell is an 8-channel sealed glass device. Normally one lane is used for one library. So it is possible to sequence 8 different samples at the same time. The Illumina sequencer uses the sequencing by synthesis approach. That means that all four nucleotides added simultaneously and compete for incorporation into the primed DNA fragment. Each of the four nucleotides is labeled with a unique fluorescent dye and additionally the 3´-OH group is chemically blocked. After the incorporation the imaging step follows. This is done by laser excitation and the emitted light with a specific wavelength is detected. After the imaging step the blocked 3´-OH group is now removed and the second sequencing cycle starts. The number of cycles is determined by the user. The sequencing process leads to an average read length of 25 - 100 bp and 80,000,000 reads per run. A run time is up to 8 days and produces 4.8 Gb per run. The advantage of the Illumina sequencing technology is the high accuracy of the sequencing process. Furthermore there is no problem with homopolymers.

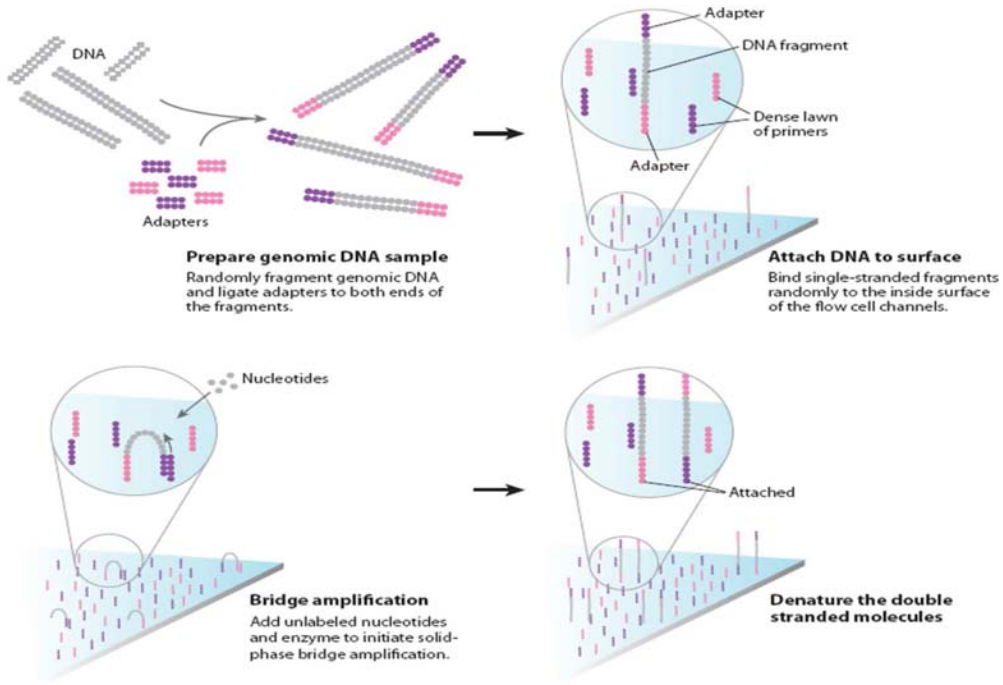


Figure 1.9: Sequencing workflow for the 454 Life Sciences/Roche technology (Part 1)(Mardis, 2008)

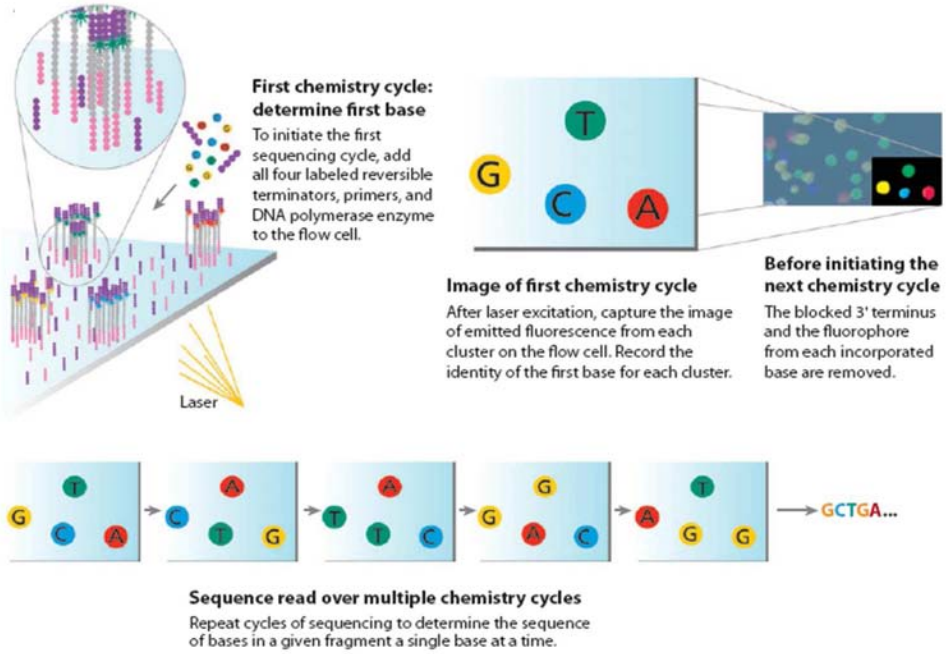


Figure 1.10: Sequencing workflow for the 454 Life Sciences/Roche technology (Part 2) (Mardis, 2008)

1.6.3 Applied Biosystems SOLiD^{TM} Sequencer

The first step using the SOLID platform is the generation of an adaptor ligated library. The ligated DNA fragments are immobilized on magnetic beads. The amplification of the single fragments is done by emPCR. SOLID uses a unique sequencing by ligation approach. After emPCR the beads are moved on a glass slide, which can be divided in several unique sections to sequence more than one sample per run. The sequencing process starts with the annealing of a primer at the adapter sequence and then DNA-ligase is provided along with specific fluorescence labeled 8mers. Two bases in the 8mers define the fluorescence dye. A set of four labeled dinucleotide probes compete for the ligation to the template. The ligation step is followed by an imaging process and then the cleavage of bases form the ligated 8mers. Multiple cycles were done to sequence the template DNA. After a series of cycles the extended product is removed and the process is reseted with a new primer shifted per one position. A whole sequencing process comprises 5 of this primer resets. So every base in a single read is captured two times. The 2 base encoding of the AB SOLID systems facilitates the distinction between base errors and true single nucleotide polymorphism and insertions or deletions. The average read length is about 25 bp and produces 20 Gb per 7 days sequencing run.

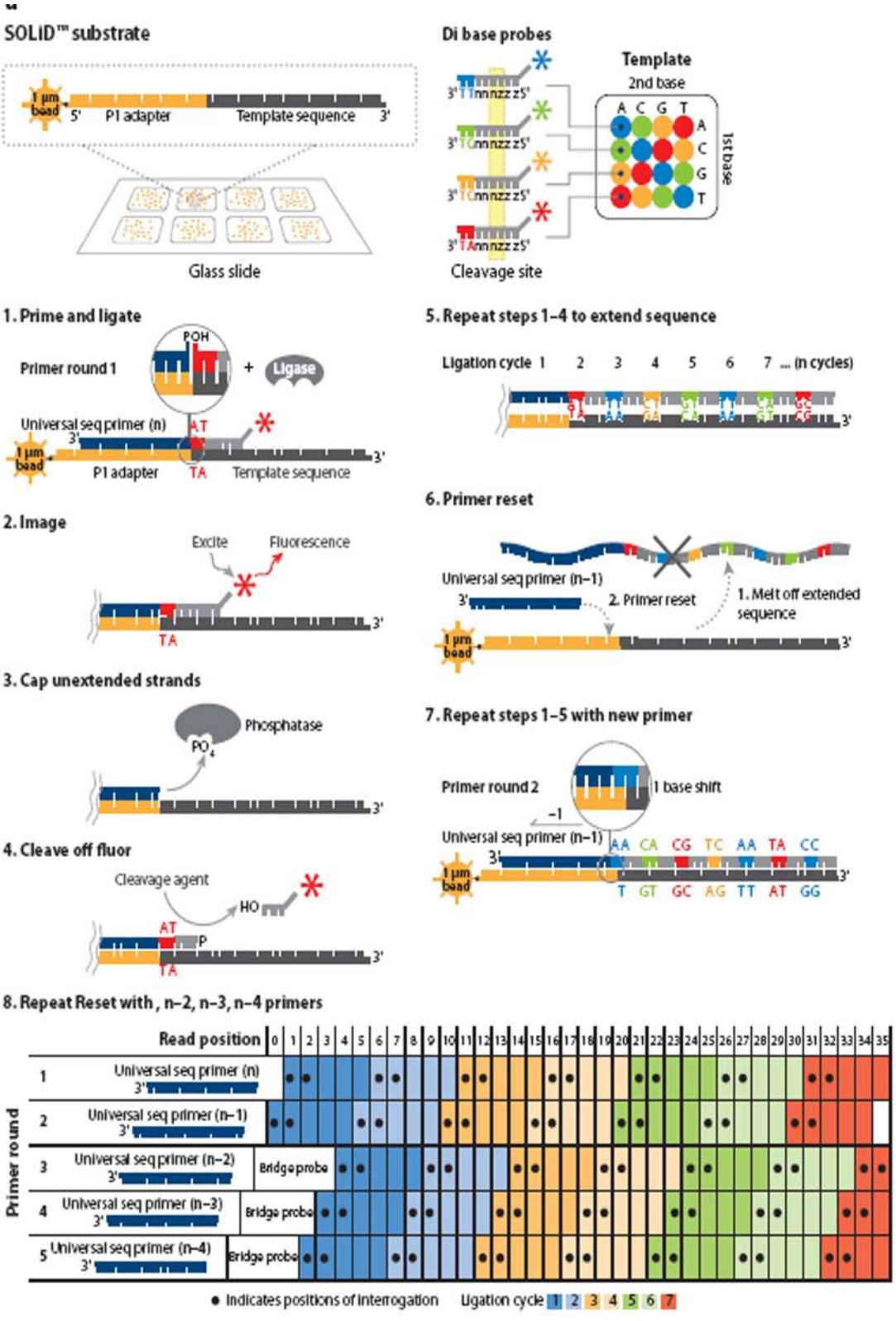


Figure 1.11: Sequencing workflow for the Applied Biosystems SOLiD^{TM} sequencing technology (Mardis, 2008)

1.6.4 Next generation sequencing data analysis

Depending on the used technology different software packages were provided by the company. For example 454 life science/Roche provide the GS De Novo Assembler,GS Reference Mapper and other software tools to further process the data. There are a lot of different software tools also commercially available.
The first steps in sequencing data analysis are the assembly and mapping of the sequencing data. If the sequencing process was only a whole genome resequencing of an existing genome, like yeast or human, a mapping against the reference genome is the first step. Instead, when there is no reference genome you have to do a de novo assembly to compare the produced consensus sequence with other genomes. A very interesting point during the analysis of whole genome resequencing data is the detection of variation between the reference genome and the resequenced genome. The variations normally are split in two groups the single nucleotide variations (SNV´s) and in insertion/deletions (INDEL´s). Again it depends on the sequencing platform, which software is used for the variation calling. If there is no software tool available neither from the company nor from other source the programming language PERL is used for establishing new scripts. PERL is a general-purpose, interpreted, dynamic programming language. It was developed from Larry Wall in 1987 as a UNIX scripting language to make report processing easier. Due to the fact that biological data produced by the above mentioned technologies proliferates fast it is important to integrate bioinformatic tools in sequence data analysis.

2 Experimental Procedures

2.1 Strains,Plasmids and Media

2.1.1 E. coli Strains

Strain	Genotype
TOP10	F- mcrA D(mrr-hsdRMS-mcrBC) F80lacZDM15 DlacX74 recA1 araD139 D(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG

2.1.2 Yeast Strains

Strain	Genotype
BY4741	MATa his3Δ1;leu2Δ0;met15Δ0;ura3Δ0
CEN PK	MATa/MATα;ura3-52/ura3-52;trp1-289/trp1-289;leu2-3_112/leu2-3_112 his3Δ1/his3Δ1;MAL2-8C/MAL2-8C;SUC2/SUC2
MG5128	MATa/MATα; ura3Δ52/ura3Δ0;ade2Δ0/ADE;can1Δ0/CAN;pso4-1/pso4-1
JSY171 (SUB592)	MATa lys2810;leu23,-112;ura352;his3-D200;trp11[am];ubi1-D1::TRP1 ubi2-D2::ura3;ubi3-Dub-2;ubi4-D2::LEU2;[pUB221] [pUB100]

2.1.3 Media

YPD	2% Glucose 2% Peptone 1% Yeast Extract
YPD-Geniticin	2% Glucose 2% Peptone 1% Yeast Extract 200 mg/L Geniticin
YPG	3% Glycerol 2% Peptone 1% Yeast Extract
SC	2% Glucose 0.17% Yest Nitrogen Base 0.5% Ammonium sulfate per liter: 5 ml 1% Tyrosine 4 ml 1% Uracile 1 ml 1% Adenine 10 ml Complete Dropout
5- FOA(5-Fluoorotic Acid)	2% Glucose 0.17% Yest Nitrogen Base 0.5% Ammonium sulfate 0.1% 5-Fluoorotic Acid 0.05% Uracile
LB	1% Tryptone 0.5% Yeast Extract 0.5% NaCl
LB-Amp	1% Tryptone 0.5% Yeast Extract 0.5% NaCl 100 mg/L Ampicillin

To prepare plates 2% Agar was added to the media.

Complete Dropout	0.2% arginine 0.1% histidine 0.6% isoleucine 0.6% leucine 0.4% lysine 0.1% methionine 0.6% phenylalanine 0.5% threonine 0.4% tryptophane
1% Tyrosine (100ml)	1 g Tyrosine 10 ml 1 M NaOH

2.1.4 Plasmids

This plasmid was kindly provided by Michael Breitenbach.

pUG6-pso4-1
The pso4-1 gene was cloned in the pUG6-vector containing a kanMX cassette that is flanked by two loxP sites. The kanMX cassette is downstream of the pso4-1 gene.

2.1.5 Primer

Name	T_m	Sequence
pUG6-sequencing	59.4°C	5' CGATTCGATACTAACGCCGC 3'
pUG6-seq-reverse	59.4°C	5' CAATACGCAAACCGCCTCTC 3'
pso4-1-integr-sense	55.3°C	5' AGAAAGCCAACTAGGGAACA 3'
pso4-1-integr-antisense	72.9°C	5' AATGTTACTACTATTACACAGGTTTATTTA GAAAGTACAACAGCTGAAGCTTCGTACGC 3'
prp19-start	63°C	5' TATATACATATGCTTTGTGCTATTAGTGGG 3'
prp19-stop	67°C	5' TAACTCGAGTTAGGGTGTCAATGCAAC 3'

2.1.6 Chemicals

Hydrogen Peroxid 30% (w/w)
H_20_2 / Molecular Weight: 34.01 g/mol / CAS-Number: 7722-84-1 / (Sigma Aldrich: 31642)

Methyl methanesulfonate
Methanesulfonic acid methyl ester / $CH_3SO_3CH_2CH_3$
Molecular Weight: 110.13 g/mol CAS-Number: 62-50-0/(Sigma Aldrich: M0880)

Diethyl maleate
Maleic acid diethyl ester / $C_2H_5OCOCH{=}CHCOOC_2H_5$
Molecular Weight: 172.18 g/mol CAS-Number: 141-05-9 / (Sigma Aldrich: D97703)

2.2 DNA Preparation

2.2.1 Preparation of plasmid DNA by alkaline lysis (Quaigen Miniprep Kit)

2.2.1.1 Material

Buffer P1: 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 μg/ml RNaseA
Buffer P2: 200 mM NaOH, 1% SDS
Buffer N3: 4.2 M Gu-HCl, 0.9 M potassium acetate (pH 4.8)
Buffer PE: 10 mM Tris-HCl (pH 7.5), 80% ethanol
Buffer EB: 10 mM TrisCl (pH 8.5)

2.2.1.2 Method

Plasmid Purification is performed with Qiagen Miniprep Kit and the according protocol. Inoculate a single bacterial colony into 2 ml of LB-Medium containing the desired antibiotic and incubate overnight at 37°C with vigorous shaking. Centrifuge 2 ml overnight culture for 1 min at high speed (13.000 rpm) in a microcentrifuge. Remove the supernatant (with a disposable tip attached to a vacuum pump) as much as possible. Resuspend the bacterial pellet with 250 μl Buffer P1. Add 250 Buffer P2 and gentle invert the microcentrifuge tube 4-6 times. Add 350 μl Buffer N3 and invert the tube again 4-6 times. Centrifuge the solution 15 minutes at high speed (13.000 rpm). A compact white pellet will form, which contains the chromosomal DNA, cell debris. Apply the supernatant to the QIAprep spin column by decanting or pipetting. Centrifuged 30-60s, at high speed (13.000 rpm). Discard the flow-through and add 600 μl Buffer PE. Centrifugate again 1 minute at high speed (13.000 rpm). Discard the flow-through and centrifuge again 1 minute at high speed. Place the QIAprep column in a new 1.5 ml microcentrifuge tube. To elute DNA, add 50 μl Buffer EB or water to the center of each QIAprep spin column, let it sit for 1 min, and centrifuge for 1 min.

2.2.2 Preparation of genomic DNA from Yeast

2.2.2.1 Material

EDTA-Solution: 500mM
Ethanol: 100%
Sorbitol buffer: 1 M sorbitol, 100 mM EDTA, 14 mM ß-mercaptoethanol

For the purification of genomic yeast DNA the Qiagen DNeasy© Blood Tissue Kit was used. Inoculate a single yeast colony into 2 ml appropriated yeast medium and grow on 28°C on a shaker. Growth time depend on the different strains and medium. The liquid culture should contain a maximum of $5x10^7$ cells. The cells were harvested by centrifugation for 10 minutes at 7500 rpm. Resuspend the pellet in 600 μl sorbitol buffer. The sorbitol buffer has to be freshly prepared. After Resuspension add 200 units (20 μl) lyticase and incubate at 30°C for 30 minutes. The lysis time and yield can vary from strain to strain and depend on also on the cell number. Centrifuged 10 minutes at 3000rpm to pellet the sheroblasts. Add 180 μl of ATL buffer and resuspend the pellet. Add 20 μl proteinase K and mix thoroughly by vortexing. After mixing incubate at 56°C until the spheroplasts are completely lysed. Incubation for 45 minutes was done on a Thermomixer to disperse the sample. After lysis of the spheroblasts add 4 μl RNase and incubate at 37°C for 1 hour. A master mix of Buffer AL and ethanol was prepared. After addition of 400 μl Master Mix the solution was mixed immediately. Transfer the mixture into the DNeasy Mini spin column, which is placed in a 2 ml collection tube. Centrifuged, 6000g for one minute. The DNeasy Mini spin column was placed in a new collection tube and 500 μl of buffer AW1 were added. Centrifuge again 6000 g for 1 minute. Place the DNeasy Mini spin column in a new collection tube and add 500 μl buffer AW2. Centrifuge at high speed (20.000 g) for 3 minutes to dry the membrane. It is important that all of the ethanol is dried away, because it may interfere with subsequent reactions. DNeasy Mini spin column is placed in a clean 1.5 ml micro centrifuge tube and 200 μl buffer AE were directly added on the DNeasy membrane. Incubate for 1 minute at room temperature. Afterwards centrifuge for 1 minute. For a higher DNA yield this step repeat this step.

2.2.3 Agarose gel electrophoresis

This method was used for separation and identification of DNA fragments. Agarose gels were prepared by mixing 8 g 50xTAE (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH of 50X TAE: 8.4), 4g Agarose and 400 ml AD in an Schott flask. To melt the agarose the mixture was heated in a microwave. After melting the mixture has to cool down and 40 μl Cybersafe (10,000x stock in DMSO) Stock were added. Then the gel was poured into chambers. For the electrophoresis 1xTAE buffer was used. A 6x loading dye (10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol,

60 mM EDTA, Fermentas) was used to monitor the moving front and to ensure that the DNA sinks evenly into the wells.

2.2.4 Agarose gel extraction

For the purification of DNA from agarose gels the Metabion mi-Gel Extraction Kit was used. Visualization of the DNA under UV-light and excise the bands with a scalpel and place the gel in an clean micro centrifuge tube. Add 300 μl gel Extraction Buffer and incubate the mixture at 65°C until all of the gels are liquid. The mixture should be mixed every 2-3 minutes. Transfer the solution in a spin column and centrifuge 1 minute at 12.000 g. Discard the flow-through and add 500 μl Column Wash Buffer and centrifuge again at 12.000 g for 1 minute. Repeat this step and discard the flow-through. It is important to remove the column wash buffer. This is done by another centrifugation step. Eluate the DNA in 10-30 μl AD water by centrifugation at 12.000 g. The volume used for eluation depends on the later use of the DNA.

2.2.5 DNA manipulation

2.2.5.1 DNA restriction

DNA-Template	5μl
Digestion buffer(10x)	2μl
Enzyme (10u/μl)	0.5 μl
H20	12.5 μl

Mix on ice and incubate at 37°C. For a single enzyme digestion a unique buffer was used. In case of a double digestion a buffer with the highest activity of both enzymes was used.

2.3 Transformation of yeast and bacteria

2.3.1 Transformation of yeast

2.3.1.1 Materials

50 % Polyethylenglycol
100 mM Lithiumacetat
1 M Lithiumacetat
ssDNA(salmon sperm DNA)-Stock 2 mg/ml

For the transformation of yeast the lithium acetate method was used.

2.3.1.2 Method

Inoculate 10ml appropriated medium with a yeast single colony and incubate overnight at 28°C on the shaker. Measure the $OD_{600}$the preculture. For the main culture inoculate 50 ml of media with a OD_{600} of 0.2 - 0.4. Incubate for another 4 hours at 28°C on the shaker. The cells should double twice before transformation to ensure higher transformation efficiency. The culture used for transformation should be in an exponential growth phase (OD_{600} 0.7 - 1.0). After incubation measure the OD_{600} of the main culture if the OD_{600} is between 0.7 - 1.0 pour the culture in a 50 ml Greiner tube. Harvest the cells by centrifugation at 3000 g for 5 minutes and discard the supernatant. Wash the cells with 25 ml of sterile water. Centrifuge again for 5 minutes at 3000g. Resuspend the cell pellet with 3 ml 100mM LiAc solution. Incubate the mixture for 15 minutes at 28°C on the shaker. In the meantime the transformation mix can be prepared.

Transformation Mix:
240 μl 50% PEG
36 μl 1M LiAc
50 μl 2 mg/ml ssDNA Stock
34 μl plasmid DNA

To use different amounts of plasmid DNA add water instead of plasmid DNA. For every transformation 300 μl of chemical competent yeast cells were aliquoted in a 1.5 microcentrifuge tube. Centrifuge for 5 minutes at 3000 g and discard the supernatant. Add the transformation mix to the cell pellet und resuspend the cells. Incubate the mixture for 30 minutes at 30°C on a thermomixer. Heat shock at 42°C for 40 minutes in the water bath. Invert the tubes every 10 minutes. Centrifuge 4500g for 5 minutes. Resuspend the pellet in 1 ml sterile water and resuspend the cells. 100 μl of this culture were platted on the appropriated media. Incubate the cells for 2-3 days at 28°C.

2.3.2 Transformation of electrocompetent E.coli

2.3.2.1 Materials

LB-Medium
SOC-Medium

2.3.2.2 Method

Set the electroporation apparatus to 2.5 kV, the capacitance to 25 μF and the pulse controller to 1000 Ohm. Add plasmid DNA to the electrocompetent *E. coli*. Transfer the plasmid DNA/cell mixture in an ice-chilled clean electroporation cuvette. Place the cuvette in the sample chamber and pulse. Add 1 ml SOC-Medium and mix the solution. Transfer the SOC/cells mixture in a microcentrifuge tube and incubate at 37°C for 30 minutes. After the incubation time 100 μl of the cell suspension were plated on agar plates containing the appropriated antibiotic. Different dilutions of cells were platted to ensure that single colonies can be isolated from the plates. Incubate the plates at 37°C overnight.

2.4 PCR

For this work the Phusion©-polymerase was used for PCR. This polymerase is a genetically modified fusion protein with proof reading function. The Phusion© PCR contained 50 μl of PCR mixture:

Volume	Component
10μl	5x Phusion Puffer
1μl	DNA-Template
1μl	10mM dNTP´s
2.5 μl	Forward Primer
2.5 μl	Backward Primer
0.5 μl	Phusion DNA Polymerase
32.5 μl	H_2O

Following PCR-program was used:

Temperature	**Time**	**Cyles**
98°C	30 sec	1
98°C	10 sec	25
51°C	30 sec	
72°C	1 min	
72°C	10 min	1

2.5 Spotting tests

SC-glucose agar plates were made and contain different concentrations of the desired chemicals. Therefore the SC-Agar was autoclaved and later aliquoted into 50 ml *Greiner* tubes. It is important that the SC-agar is as cold as possible by the addition of the chemicals. After addition of the chemicals the agar was poured in petri dishes. Inoculate 10 ml SC-Medium with a yeast single colony. Incubate the culture overnight at 28°C on the shaker. Measure the $OD_{600}$the overnight and harvest the cells by centrifugation at 3000g for 5 minutes. Wash the cells with sterile $H_2O$and dilute it to an OD 3, OD 1, OD 0.3, OD 0.1, and OD 0.01. 5 μl aliquots were spotted onto the SC-Glucose agar plates. Incubate the plates for 3 days at 28°C.

3 Results

3.1 Testings for the temperature phenotype

In the beginning of my work I received following strains. All these strains were kindly provided by Dr. Michael Breitenbach.

- BY a
- REV1α
- pso4-1α
- 708-3A
- 708-3B
- 708-3C
- 708-3D
- 708 dipl.

Rev1 is a revertant and was produced by streaking out the parental temperaturesensitive strain *pso4-1α* and incubate it for 2-3 days at 37°C. The tetrade 708 was produced by crossing Rev1 with the wildtype strain BY4741. This crossing was necessary to show that the new mutation in Rev1 is not in the *pso4-1* mutated gene. If the mutation in Rev1 is in the *pso4-1* gene all the four spores will grow at 37°C. If the mutation is in another gene at least one of the spores will show no growth at 37°C. Additional experiments showed that the second mutation is recessive. The experiments were done by Prof. Michael Breitenbach. There is also a second revertant strain, which shows the same phenotype. The two strain developed independent from each other. The phenotyp causing mutation is not in the same gene if the two strains were compared. For Rev2 it have to be tested if the mutation is recessive or dominant. The strains were tested for their temperature sensitivity phenotype on 23°C, 30°C, 37°C and on YPD (C-Source: Glucose) and YPG (C-Source: Glycerol) agar. The following table shows the result of the temperature sensitivity tests.(Tab.3.1) The temperature sensitivity tests showed a coldsensitive phenotype of Rev1 after 2 days incubation. Interestingly the parental strain *pso4-1α* didn´t show this phenotype.(Fig. 3.1) The outcome of this experiment was that the revertant really showed growth at 37°C and can overcome the temperature sensitive phenotype. Now it was interesting which gene carries the mutation which causes this new phenotype. Therefore the revertant and the parental strain *pso4-1* were sent for deep sequencing to the Max Planck Institute for Molecular Genetics in Berlin. The result of the deep sequencing will be shown later in this work.

	YPD 23°C	YPD 30°C	YPD 37°C	YPG 23°C	YPG 30°C	YPG 37°C
BY a	+	+	+	+	+	+
REV1α	-	+	+	-	+	-
pso4-1 α	+	+	-	-	+	-
708-3A	+	+	-	+	+	-
708-3B	+	+	+	+	+	+
708-3C	+	+	+	+	+	-
708-3D	+	+	+	+	+	+
708 dipl.	+	+	+	+	+	+

Table 3.1: This table shows a summary of the results from the temperature sensitivity tests (after 2 days incubation). Legend: + growth, - no growth

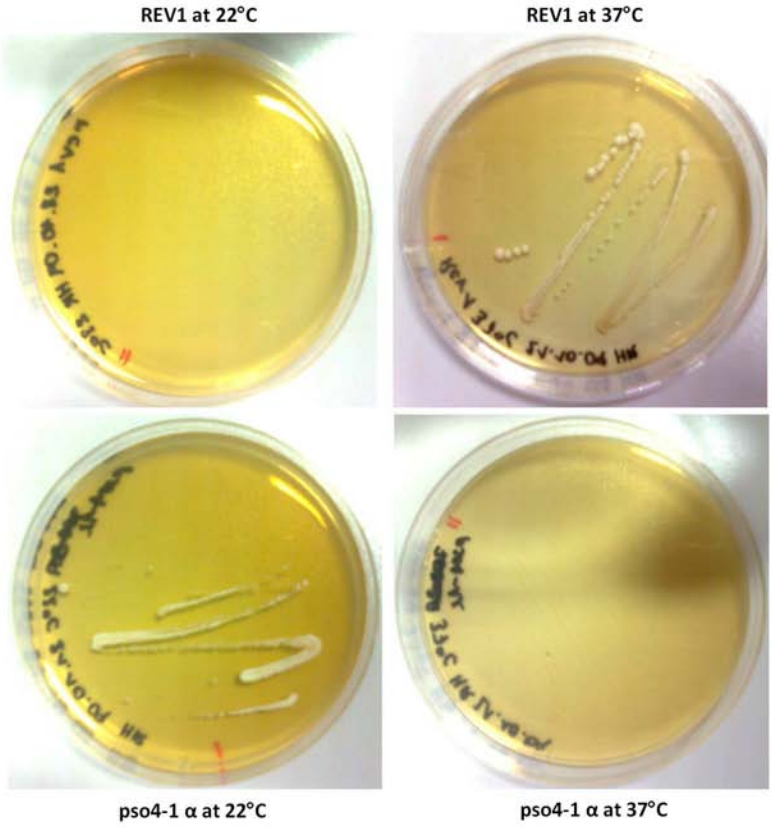


Figure 3.1: Testing of temperature sensitivity of *pso4-1* α and REV1

3.2 Sensitivity test of the mutant strains by spotting tests

All yeast strains which were used for the spotting test were kindly provided by Dr. Michael Breitenbach. The strains were produced by crossing of MG5128 MATa with BY4741α Δsgs1.

Following strains were used for spotting tests:

Spore	Genotype
4A	MATa pso4-1/SGS1$^+$
4B	MATα *PSO*4$^+$ Δsgs1
4C	MATα *pso4-1*Δsgs1
4D	MATα *PSO*4$^+$ SGS1$^+$
17A	MATα *pso4-1*Δsgs1
17B	MATα *pso4-1*Δsgs1
17C	MATa PSO4$^+$ SGS1$^+$
17D	MATa PSO4$^+$ SGS1$^+$
Rev1	-
pso4-1 a	-
pso4-1 α	-

3.2.1 Testing for H_20_2 sensitivity

Hydrogen peroxide is one of strongest reactive oxygen species (ROS). Due to the fact that oxygen is the terminal electron acceptor in the respiratory chain hydrogen peroxide and superoxide are also produced under normal conditions. ROS can damage a variety of cellular processes, including proteins and unsaturated lipids. Especially the product from lipid peroxidation, production of alkoxy/peroxyl radicals and aldehyde compounds, which are products of radical lipid reactions triggered by ROS, are toxic. Detection of ROS is an essential step in the oxidative stress response. The presence of hydrogen peroxide in yeast activates the general ESR cluster genes and the specific H_20_2 stimulon, containing most of the antioxidant enzymes, such as thioredoxins peroxidase (peroxiredoxin), thioredoxin reductase, superoxide dismutase and cytochrome c peroxidase. To test the yeast strains for their behavior against oxidative stress different H_20_2 concentrations were applied. The strains were tested on SC-plates containing following H_20_2 concentrations: 0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM, 1 mM, 1.5 mM and 2 mM. The concentrated H_20_2 was diluted with sterile water to the desired concentrations. The concentration of 0.6 mM H_20_2 was choosen, because here the best differences were observed.

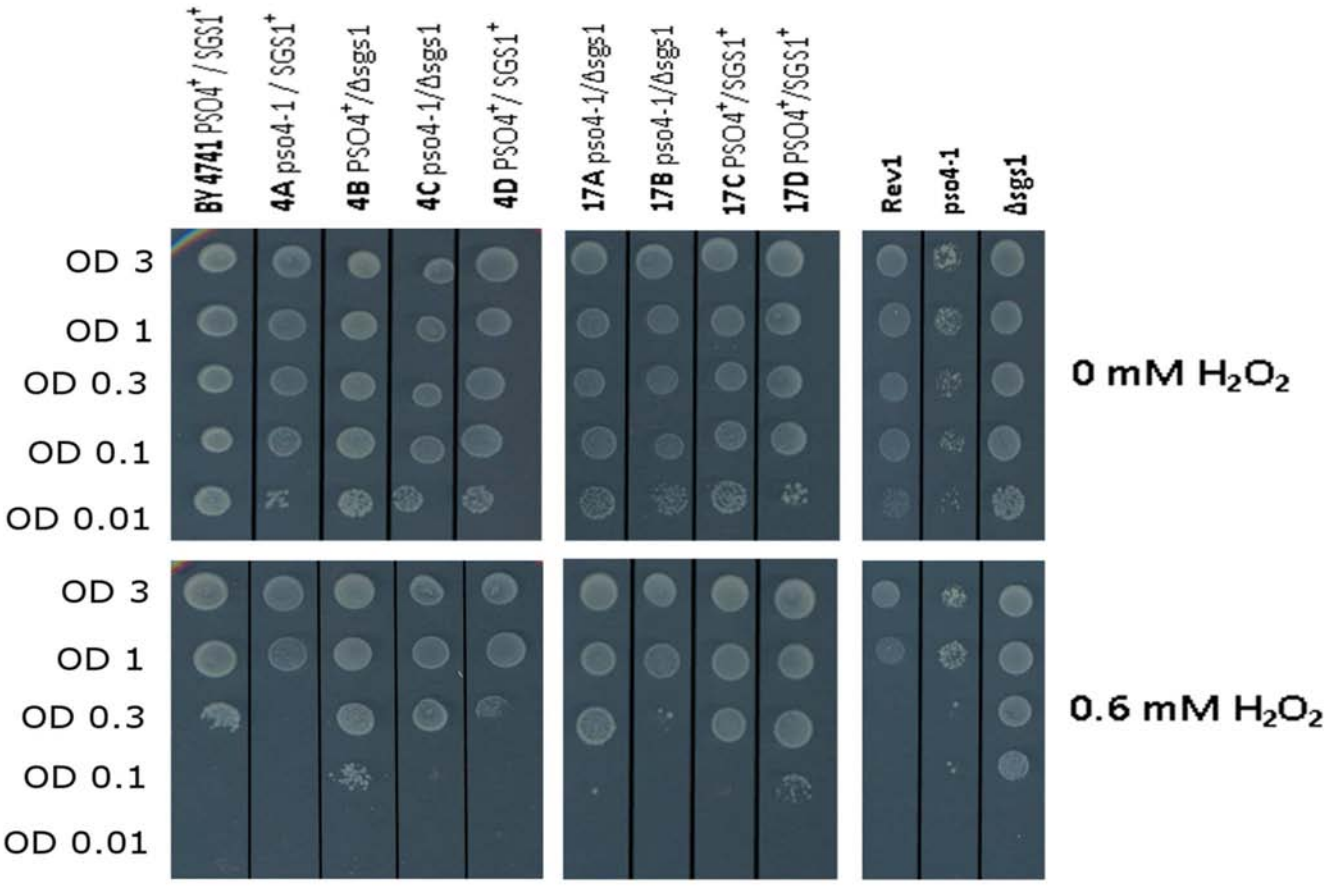


Figure 3.2: Testing of sensitivity against H_20_2: The two tetrades 4, 17, single mutants, wildtype and revertant were tested.

The results of the spotting test are not really clear, because the *pso4-1* mutants and the wildtype strains show almost the same sensitivity to H_20_2. Maybe a reason for this result could be the use of a non isogenic wildtype. Also for the *pso4-1* strain and rev1 strain no clear result can be seen, because of the retarded growth on the plates. 17A and 17B, which carry a pso4-1 mutated gene, show almost the same sensitivity to hydrogen peroxide like the wildtype BY 4741 and the spores 17C, 17B (both $Pso4^+/Sgs1^+$). Also the other spores 4A and 4C, carrying the *pso4-1* mutated gene, show nearly the same sensitivity like the wildtype.

3.2.2 Testing for DEM sensitivity

Diethlymaleate(DEM) conjugates with glutathione this is know from different studies.(Boyland E. and L.F. Chasseaud, 1968) Glutathione is the major endogenous antioxidants produced by the cell, participating directly in the neutralization of free radicals and ROS. Therefore DEM raises the level of ROS in the cell and so it was an ideal substance to test the yeast strains for oxidative stress. Strains were test on SC-Plates with following concentrations of DEM: 0.5 mM, 1.0 mM, 1.5 mM, 2.0 mM, 2.5 mM, 3 mM, 3.5 mM DEM is insoluble in H_2O, but is soluble in 96% EtOH. The concentration of 2 mM DEM was choosen, because here the best differences were observed. The spores 17B,

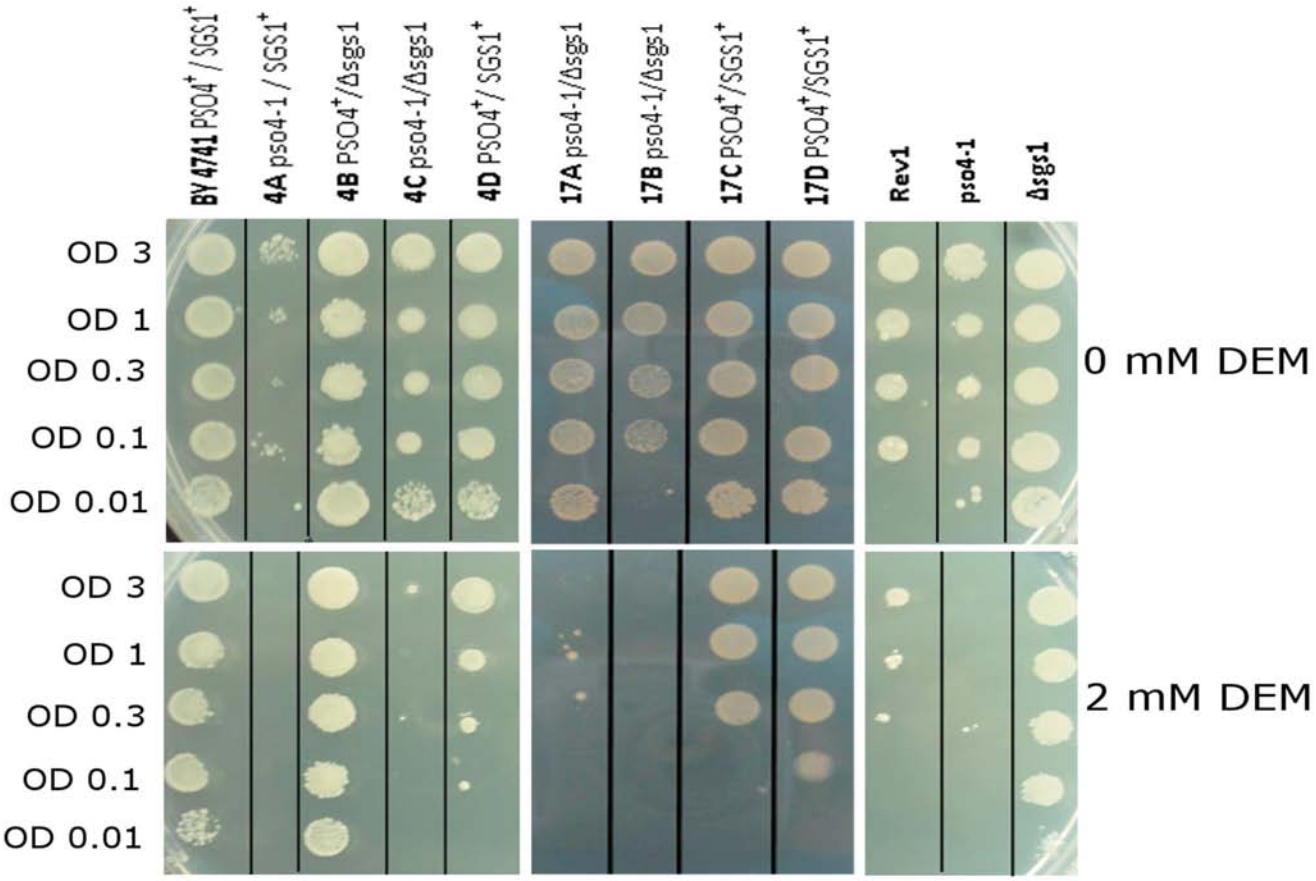


Figure 3.3: Test for DEM sensitivity. Again the two tetrades, single mutants, wildtype and revertante were tested

4A showed nearly no growth at the control plate. So this spore can´t be used for the analysis of genetic interaction between *pso4-1* and Sgs1. The results show a sensitivity of *pso4-1* to higher ROS levels. The Δsgs1 strain shows almost no effect on higher levels of ROS due to glutathione depletion. In principal all strain containing the *pso4-1* mutated gene showed a higher sensitivity to ROS. The reason for this phenotype is not really clear. There are more explanations for it. The first one is that Prp19/Pso4 has two or

more functional domains and only one of these is fully active at 30°C. The first domain is involved in pre-mRNA splicing. Due to the fact that no accumulation of pre-mRNA at 30°C in *pso4-1*-mutants occur this domain has to be fully functional. Only the function in DNA repair is perturbed (Grey et al., 1996). The second hypothesis deals with the fact that the phenotype can be caused by incorrect splicing of pre-mRNA. Yeast has 238 intron containing genes, but only one of these is linked with DNA repair. This gene is RAD14 and causes incisions of damaged DNA, but pso4-1 mutants are able to incise DNA (Grey et al., 1996).

3.2.3 Testing for MMS sensitivity

MMS (methyl methanesulfonate) is a DNA alkylation agent and has been used for many years as an DNA damaging agent to induce mutagenesis. MMS modifies guanine to 7 - methylguanine and adenine to 3 - methyladenine. This modified nucleotides cause base mispairing and replication blocks (Lundin C. et al., 2005). DNA Damages likes this are repaired using the base excision repair (BER) pathway and DNA alyktransferases. Studies showed that MMS treatment leads to a fragmentation of the DNA due to formation of double strand breaks (DSB). Also mutants which are deficient in homologous recombination (HR) showed a higher sensitivity to MMS (Lundin C. et al., 2005). Recent studies suggest that not the MMS treatment itself introduces the DSB more the processing of the damage cause the DSB. The reason for this formation of DSB is the stalking of the replication fork due to the alkylated nucleotides. HR is a massive tool to overcome the strand breaks after repair of stalked replications forks. This also emphasizes the higher sensitivity of HR deficient mutants. Concerning the yeast strains also the Sgs1p is from interest, because also this protein is involved in repair of stalled replication forks. Following concentrations of MMS were used: 0.0025%, 0.005%, 0.01%, 0.015%, 0.02%, 0.025%, 0.03%, 0.035%.The concentration of 0.01% mM MMS was choosen, because here the best differences were observed. The spores 17B, 4A showed nearly no growth at the control plate. So this spore can´t be used for the analysis of genetic interaction between *pso4-1* and Sgs1. At the plate with 0.1% MMS revertants can be seen at the spot of 17B. The other double mutants show an increased sensitivity to MMS compared with *pso4-1*. Also the Δsgs1 strain shows a slightly higher sensitivity to MMS then *pso4-1*. This additive effect, which can be seen for the double mutants, indicates a role of *pso4-1* in a Sgs1 independent repair pathway of MMS damage. The lower sensitivity of *pso4-1* against MMS could be explained by the function of Sgs1 overcoming stalled replication forks. MMS cause stalled replication forks. This fact may explain observed result.

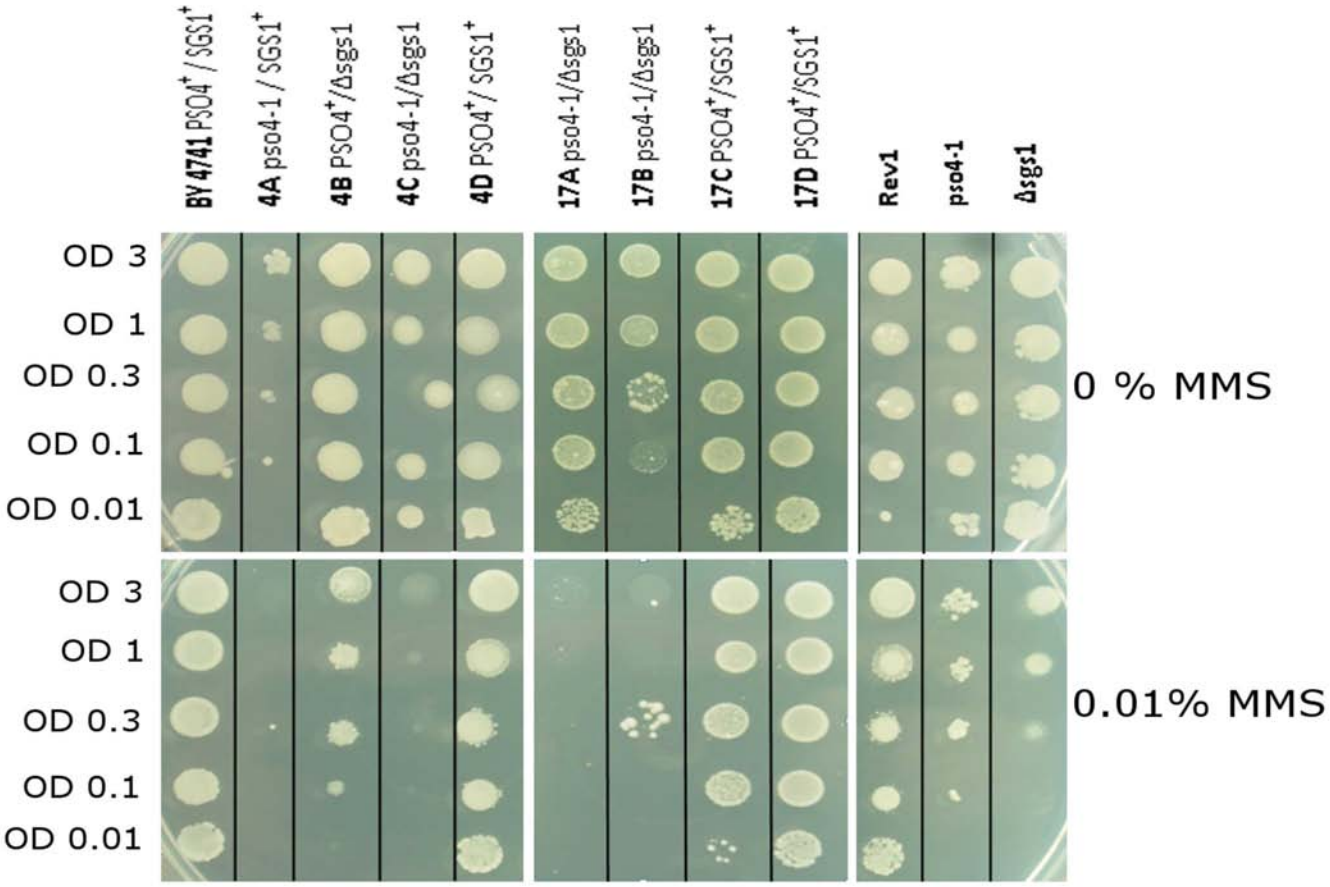


Figure 3.4: Test on MMS. Two tetrads, single mutants, wildtype and revertant were tested.

3.3 E3-Ligase Domain in Prp19/Pso4 and targets of ubiquitination

As mentioned in the introduction Prp19/Pso4 contains different domains. One of these domains is an E3-Ligase which is needed in the process of ubiquitination. Ubiquitination is an important process within a cell and has not only a function in protein degradation. ubiquitination is also used for signaling in DNA repair mechanism. So it is very interesting to know the putative targets of Prp19/Pso4 E3-Ligase domain. The temperature sensitive phenotype of the *pso4-1* mutated gene is a proper tool to determine the targets of ubiquitination. The first task was to construct a yeast strain which possesses the *pso4-1* gene and a His tagged ubiquitin. For the constructions we had different approaches which will be described later in this section.

3.3.1 Construction of a strain containing *pso4-1* and His-tagged ubiquitin

For the construction we used the yeast strains SUB592, which was kindly, provided from Dr.Daniel Finley. This strain contains two different plasmids. One plasmids was an high copy Plasmid with an URA3-selectionmarker and expresses a 6xHis tagged Ubiquitin. The second plasmid contains an expression cassette for expressing the tail of the Ubi1 gene and posseses a HIS selectionmarker. Furthermore there is a knockout of all ubiquitin genes in this strain. The first point was to characterize this yeast strain. Therefore we determined the selection markers and showed that the plasmid, containing the 6xHis tagged ubiquitin, is really necessary for survival of the yeast strain. Normally a knockout of all ubiquitin genes is lethal.

Testings of the selection markers

To test the selection markers minimal agar with different amino acid compositon was produced. We tested for following markers: histidine(HIS), tyrptophane(TRP), leucine (LEU), uracil (URA) and lysine (LYS). The yeast strain SUB592 was scratch from a minimal agar plate and resuspended in sterile water. Then the cell suspension was plated at the different minimal agar plates. The agar plates were incubated at 28°C for 2-3 days. The Table 3.2 shows the composition of each agar for test.

The result of the selection marker test was that only the addition of lysine is essential for the growth of the yeast strain. The strain contains two different plasmids. One plasmid with the 6xHis tagged Ubiquitin posseses a URA selctionmarker. The second plasmid expressing the Ubi1 tail contains a HIS selectionmarker. Compared with the genotype an the knowledge of the selectionmarker of the two plasmids this result was expected.

	YNB	HIS	LEU	LYS	TRP	URA
Control	+	+	+	+	+	+
1	+	-	+	+	+	+
2	+	+	-	+	+	+
3	+	+	+	-	+	+
4	+	+	+	+	-	+
5	+	+	+	+	+	-

Table 3.2: The amino acid composition of the different minimal media used for the selection marker tests. + Component added, - Component not added

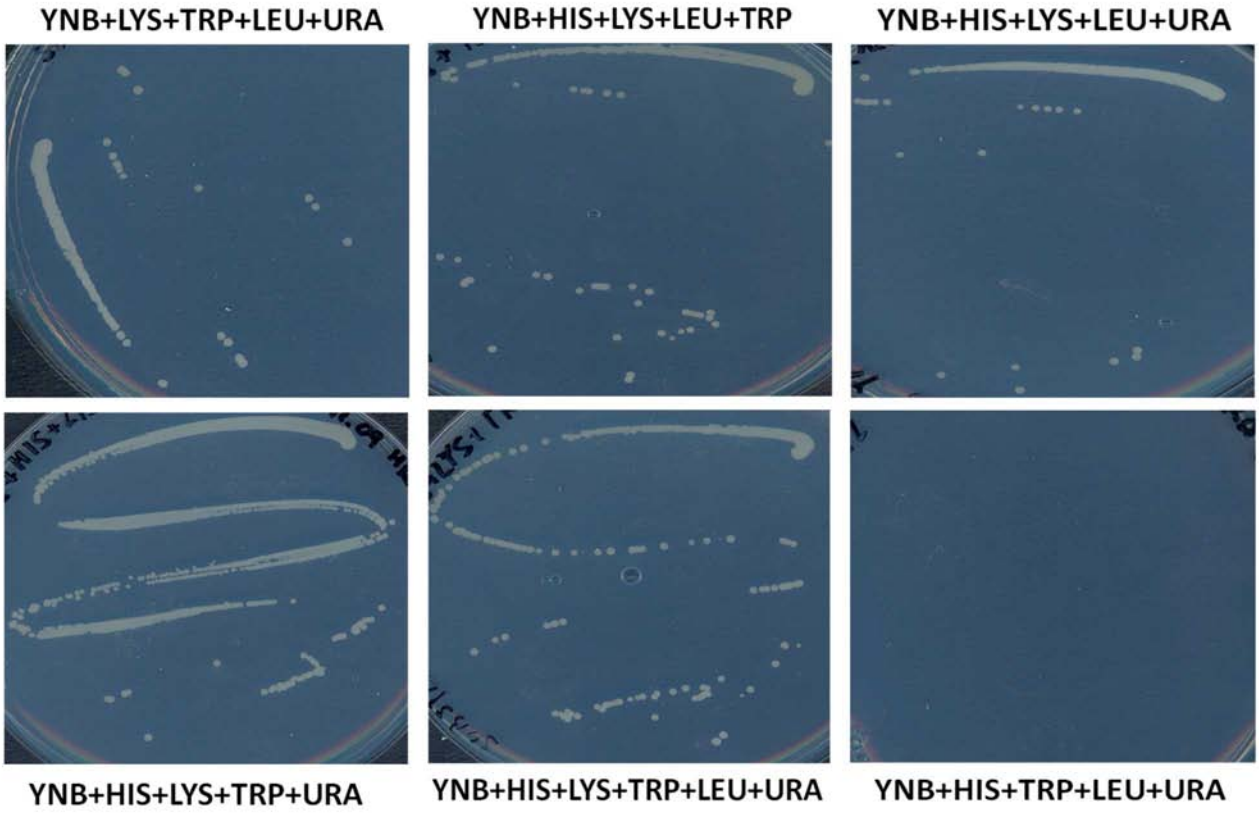


Figure 3.5: Result of the selection marker test for SUB592

Test for plasmid loss

To show that the URA-Plasmid, which contains the expression cassette for the 6 x His-Tag ubiquitin is essential for survival of the yeast strain, a test with minimal agar plates containing 5´-FOA (5´-Fluororoticacid) was performed. If the URA-plasmid is really essential no growth on the 5´FOA plates is expected. To be sure that the test works a control strain was used. This CEN PK strain contains a URA-plasmid which is not essential for growth. The different strains were plated and then incubated at 30°C

for 2-3 days. For SUB592 on the 5´-FOA plates no growth was observed. This shows that the plasmid containing the 6 x His-Tag ubiquitin is essential for the growth. The control strain CEN PK showed growth on the 5´FOA plates. This also indicates that the test worked.

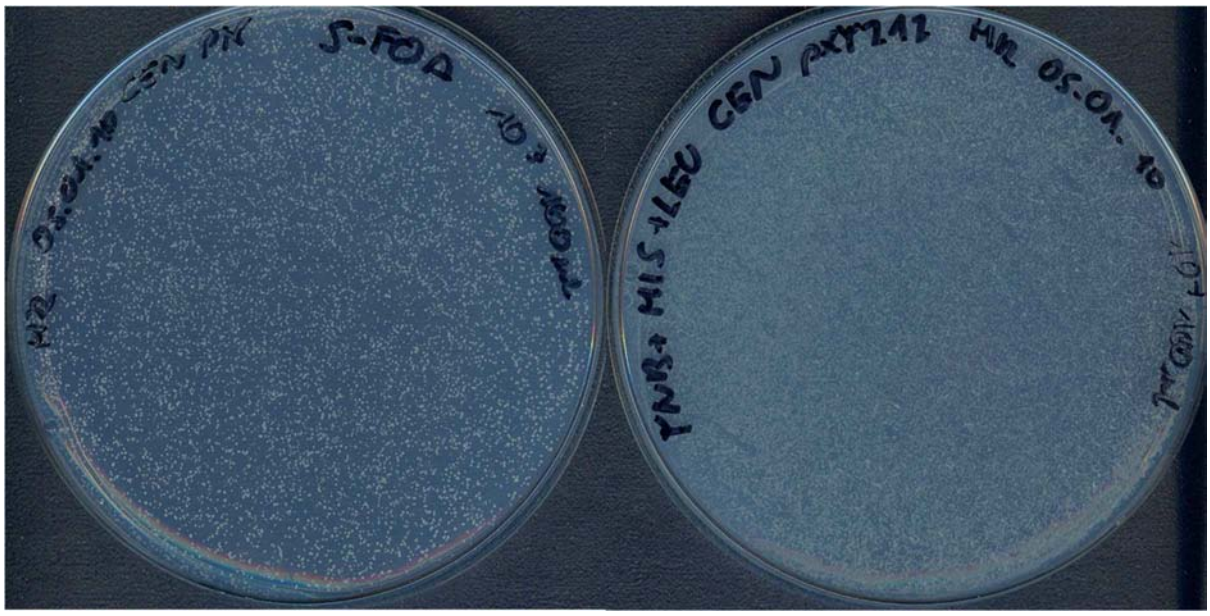


Figure 3.6: Reference strain CEN PK: On the left side is the 5-FOA plate on the right side is the control plate

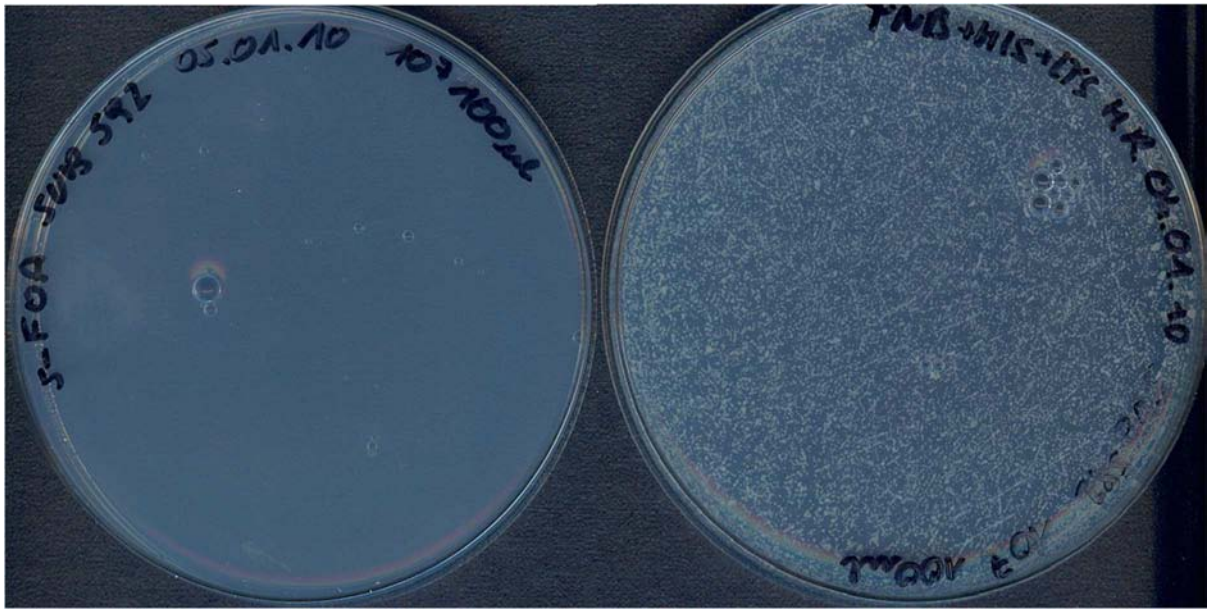


Figure 3.7: SUB592: On the left side the 5´-FOA plate and on the right side the control plate

The result of this test was that the plasmid in SUB592 is essential for the growth of this strain.

After these tests the construction of the SUB592 strain containing the *pso4-1* mutated gene started. The next section of this work will describe which approaches were used and the result of the experiments.

3.3.2 Crossing of *pso4-1*α with SUB592

The work was done by Prof. Michael Breitenbach. This approach didn´t succeed because of the bad sporulation efficiency of pso4-1 strains.

3.3.3 Introduction of the *pso4-1* in SUB592 via transformation

Another approach to introduce the *pso4-1* gene was the transformation of SUB592 with a plasmid containing the *pso4-1*. The plamid additionally contain a kanMX within two loxP site for selection. The wildtype $Pso4^{+}$ gene in SUB592 will be replaced by the mutated *pso4-1* gene via homologous recombination. The *pso4-1* gene together with the kanMX cassette and loxP sites was transformed into SUB592. The kanMX is essential for the selection of positive transformants and YPD plates containing G418. After the integration of the *pso4-1* gene in the genome and selection of the positive transformants it is possible to remove the kanMX cassette by a second transformation with a plasmid which contains Cre-lox Recombinase. The proteins will be expressed after induction with galactose. This plasmid was kindly provided from Dr. Michael Breitenbach. To get the desired fragment, which later was used for transformation a Phusion© PCR was done. The Phusion© PCR produced a 3189 bp large fragment, containing the *pso4-1* gene and the kanMX cassette within the loxP sites. This fragment later was transformed into SUB592 and the transfromants were crossed out on YPD-G418. After 3 days incubation at 30°C positive candidates could be found. To test if the *pso4-1* integrates at the right locus the temperature sensitive phenotype was tested. Therefore the positive transformants were stroked out on minimal agar and YPD agar. These plates were incubated for 2-3 days at 30°C and 37°C. The outcome of the test was that there were no temperature sensitive clones. This result indicates that the *pso4-1* didn´t integrate at the $Pso4^{+}$ wildtype locus. Because of the changed genetic background it was possible that there is also a shift of the temperature sensitivity to higher temperatures. To be really sure about the integration of pso4-1 gene at the Prp19/Pso4 gene locus in SUB592 we isolated genomic DNA and made a Phusion© PCR with the primers Prp19-Start and Prp19-Stop. The Phusion© PCR produced a 1810 bp fragment and later was digested with the restriction enzyme Afl II. The point mutation in *pso4-1* destroys the restriction site for Afl II. To differentiate between $Pso4^{+}$ and *pso4-1* this was an easy way to test. For each clone a digested and undigested fragment were applied to agarose electrophoresis. As you can see in figure 3.8 the digested band is smaller than the undigested band. This indicates the cut of the fragment by AflII in a band with a size of 1482bp and a smaller band with

a size of 382bp. The final outcome of the restriction analysis was that the *pso4-1* gene didn´t integrated at the Prp19/Pso4 gene locus.

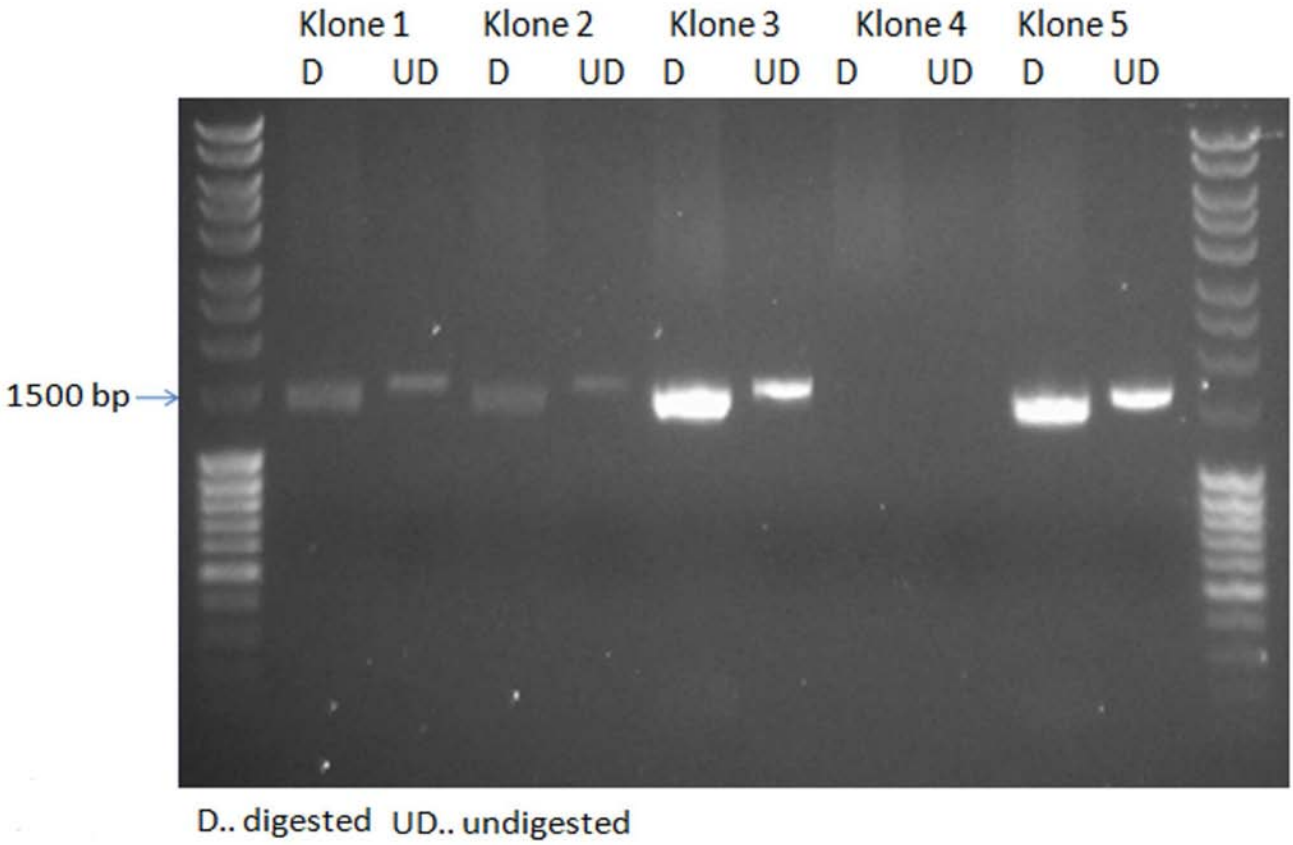


Figure 3.8: Restriction digestion of genomic DNA from SUB 592 by Afl2

3.4 Determination Mutations with different next generation sequencing technologies

This part of my diploma thesis comprises two projects. Two different next generation sequencing technologies (454 Roche and Illumina Sequencing) were used for whole genome resequencing of yeast genom. The objective of both projects was to find a mutation, which cause a specific phenotype in yeast strain. For processing the sequencing data PERL was used. PERL is the most widely used programming language in Bioinformatics. This novel combination of next generation sequencing technologies with genomics may facilitate solving difficult biological problems.

3.4.1 454 Roche Sequencing data analysis

3.4.1.1 Yeast strains K6001 and B7

The W303 derived yeast strain K6001 provides a model system for the study of replicative life span in *S. cerevisiae*. Normally the determination of replicative lifespan in yeast is a very time consuming experiment K6001 allows an enrichment of yeast mother cells and an easy determination of the replicative lifespan of these cells. To have a link between aging and oxidative stress, K6001 cells where used for a random mutagenesis experiment. With an exponentially grown K6001 culture a random mutagenesis experiment with ethyl methansulfonate (EMS) was done. Afterwards the cells were grown under non selective conditions to fix the mutation. The cells were plated on agar containing oxidant cumene hydroperoxide (CHP). The isolated stress resistant mutants were tested for their lifespan and K6001-B7 was used for whole genome resequencing to determine the putative mutation, which cause the specific dominant phenotype.

3.4.1.2 454 Roche Sequencing

A whole genome resequencing with the 454 Roche/Life Sciences sequencing technology was done. The genomic DNA of both yeast strains K6001 and K6001-B7 were isolated and processed for the sequencing process. The output of the sequencing process is a sff-file, which is a binary file and it contains all the sequenced reads. For the mapping process the GS Reference Mapper Version 2.3 was used. This software was provided by Roche. The reference genome was the Saccharomyces reference genome S288C. The mapping produces a consensus sequence, which align against the reference genome and it computes also variations found in the reads, relative to the reference genome. Also some statistical analyses are done.

During the mapping algorithms the software performs different processes:

- If the read aligns at multiple positions the best alignment is chosen.
- Identify putative differences of the reads to the reference genome. The software distinguishes between nucleotide variations (SNP) or structural variations. The list of variations is further processed to get high confidence nucleotide variations and structural variations.

The GS Reference Mapper Version 2.3 produces following output information:

- large contigs and corresponding quality parameters
- single nucleotide variations and structural variations
- for each position of the mapped genome the corresponding base and quality parameter (depth, consensus accuracy)
- alignment of each read to the reference genome and a classification of reads

For this work mainly the files with the high confidence single nucleotide variations were from interest. To do further analyses the files were processed with PERL. PERL is a programming language, which is often used in bioinformatics to handle big data files.

3.4.1.3 Descriptive analysis of the sequencing Data

3.4.1.4 General description of the sequencing Data

The sequencing process produced at 662 MB high quality data set. The average read length for this run was 499 bp. The genome of K6001 is covered with a depth of 19.5 and K6001-B7 with 19.2. For the merged genome 7660 uniform single nucleotide variations were found and 7499 of these showed more than 10 x coverage. This indicates that every 1600 bp a nucleotide exchange happened. This is a quiet high rate and shows a high evolutionary divergence between S288C and K6001. After merging both genomes the depth was 38.8. In this case depth means the number of reads mapped at this position. The descriptive analyses were all done with PERL. The analysis were done for the single genome K6001 and K6001-B7 and with the merged genome of K6001 and K6001-B7 (Timmermann et al., 2010).
All the results from the descriptive analyses can be seen in Table 3.3.

	K6001	B7	K6001/B7
Average depth	19.5	19.2	38.8
Number of reads	641056	684525	1325581
Mapped reads	584131/91.12%	612892/89.54%	1197517/90.34%
Unique reads	495304/77.26%	504414/73.69%	999718/75,42%
Unmapped reads	56679/8.84%	71385/10.43%	128064/9,66%
Mapped bases	274532646/84.43%	277494049/82.35%	552056738/83.38%
Unique mapped bases	235918676/72.56%	232050907/68.86%	467999626/70,68%

Table 3.3: Summary of the descriptive analyses of the whole genome resequencing from K6001 and K6001-B7

3.4.1.5 Genome Coverage, Coverage Coding/Noncoding regions

For K6001 96,98% of the whole reference genome was covered and for K6001-B7 96,88% coverage of the reference genome was calculated. To calculate the coverage of the whole genome and also the coverage of the coding and noncoding region we constructed a coverage file. This coverage file contains all covered bases of the yeast genome and additionally the depth of each position. Depth is defined as the number of unique mapped reads at this position. Using this coverage file the covered bases were counted and compared with the size of the reference genome. For the estimation of the coverage coding/noncoding region additionally the annotation file was used. The annotation file contains the starting and end position of the coding sequencing. Using a Perl script the coding sequence positions were counted and compared with coverage file of each strain. So the coverage of the coding/noncoding region in K6001 was 96,97%/96,98% and for K6001-B7 96,89%/96,83%. Another interesting point is the coverage uniformity. The coverage uniformity is a visualization of the distribution of the coverage over the 16 different chromosomes. Because the yeast genome has a genome with 12 million bases it is not possible to visualization every position. Therefore the chromosomes were split into 1000 bp units. For each unit the unique average depth was calculated and visualized using R. The coverage uniformity is an interesting point, because it is possible to see non covered regions or putative regions with sequencing errors. Especially at repeated sequences (centromere and telomere) the software often has problems with the alignment.

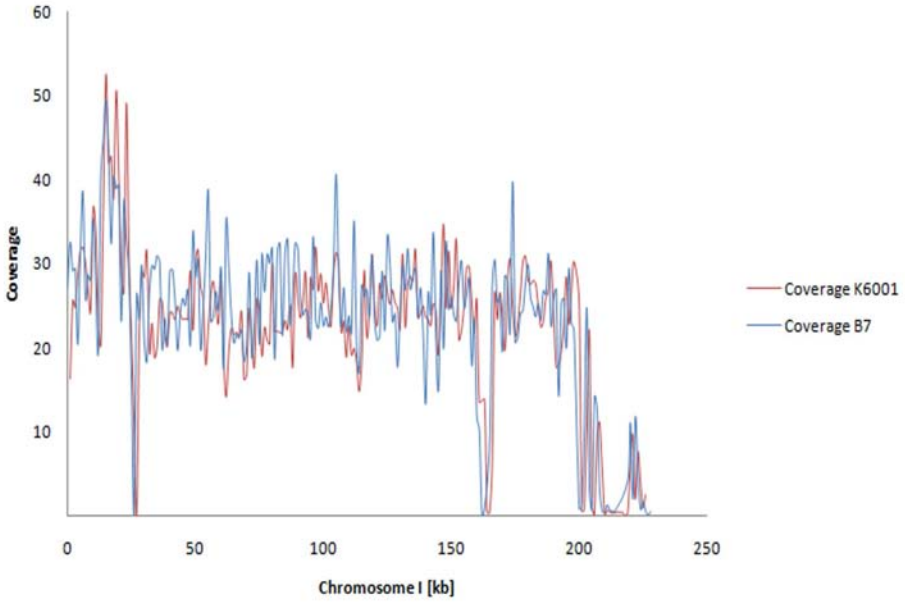


Figure 3.9: Coverage Uniformity Chromosome I: This diagram shows a homogeneous distribution of the chromosome and also between both strains.

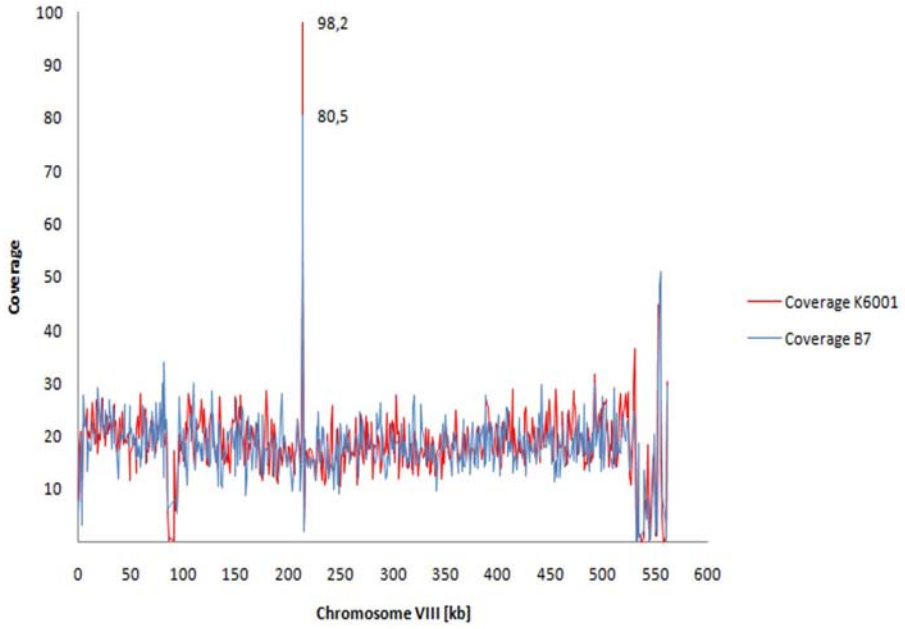


Figure 3.10: Coverage Uniformity Chromosome VIII: This diagram shows again a homogenous distribution of coverage over the chromosome for both strains. The high peak at about 450 bp suggests a region with repeated sequences.

3.4.1.6 Unmapped Reads

Due to the high number of unmapped reads it was interesting why these reads didn´t map for the S288C reference genome. The number of unmapped reads for K6001 was 56679 and for K6001-B7 it was 71385. To further investigate the unmapped reads we first used the UNIX grep command to extract the accession numbers of the unmapped reads. The accession number is unique for every read. The accession number was essential extraction of the unmapped reads out of the sff-file. The sff-file contains the information about all reads, which were produced during the sequencing process. The sequences of the unmapped reads wer saved in an new sff-file. The last step was to run an assembly of this reads for each strain and compare the assembled consensus sequence with the yeast genome. Different explanations for the unmapped reads are possible. Contaminations with foreign DNA during the extraction of DNA, genomic rearrangements in the sequenced genome (Insertion, Deletion) are possible explanations. For the mapping the mt genome of yeast was first used, because in the first mapping process with all reads, only the chromosomal genome was used as reference genome. We found that almost all assembled reads map to the mitochondrial genome. To quantify how many unmapped reads in both genomes map to the mitochondrial genome a mapping process with the unmapped reads sff-file and the mt Genome of yeast as reference genome was done. The mapping showed that 52215 of 56679 reads in K6001 and 66875 of 71385 in K6001-B7 reads map to the mitochondrial genome. So we expect that almost all unmapped reads come from mitochondrial DNA. In the end we covered the reference genome S288C with 99.26 % for K6001 and 99.30 % for K6001-B7.

Mapping against the mitochondrial yeast genome:

	K6001	**B7**
Unmapped reads	56679	71385
Mapped Reads	52215 (93,89%)	66875 (93,68%)
Overall Coverage Genome	99.26%	99.30%

Table 3.4: Summary of the analysis of the unmapped reads in K6001 and K6001-B7

3.4.1.7 Identification of the K6001-B7 mutation by whole genome resequencing

The approach was to identify the mutation via whole genome resequencing of K6001 and K6001-B7. The resequencing process identified 11230 high confidence differences between K6001 and the reference genome S288C. K6001-B7 showed 11397 differences. High confidence differences (HCdiffs) mean that the variation shows up at least 3 times and on both strands. The HCdiff-file contains single nucleotide variations and small Indels (Insertions and Deletions) up to 50 bp. The first step was to reduce the list of candidates. Therefore a cut off filter of 95 % was applied on the HCdiffs of K6001-B7.

All non-uniform SNV´s and Indels with a calling rate lower than 95% were eliminated. For a real SNV or Indel in a yeast genome you expect a calling rate of round about 100%, because K6001 and K6001-B7 are haploid yeast strains. The list of HCdiffs showed also SNV´s and Indels with lower calling rates (0-85%). Due to the elimination of this HCdiffs we reduced the list of candidates to 7213 uniform SNV´s and Indels. After the removal of this dubious SNV´s and Indels, the variations which were equal in K6001 and B7 had to be eliminated. This was done by a simple Perl script. The final outcome after the comparison of both strains was that there were 152 possible candidates left. The remaining 152 HCdiffs of K6001-B7 were controlled by hand using the information of the 454Alignment-file that contains the covered positions of the yeast genome, the consensus base and the reference base for each position. So every remaining HCdiff of K6001-B7 was compared with the corresponding consensus base of K6001. If the exchanged nucleotide in B7 and the consensus base of K6001 were equal the candidate was removed. Finally we ended up with a list of 13 putative candidates, which have to be tested via PCR and sequencing. 7 of these candidates were located in protein coding regions and 4 of this candidates showed residual changes. Those four genes CLC1, DON1, HOT1 and TSA1 were amplified by PCR. The next step was to check via Sanger sequencing that the mutation are real. After the verification of a real mutation all four genes where subcloned in an expression vector and the phenotype was tested by complementation tests (Timmermann et al., 2010).

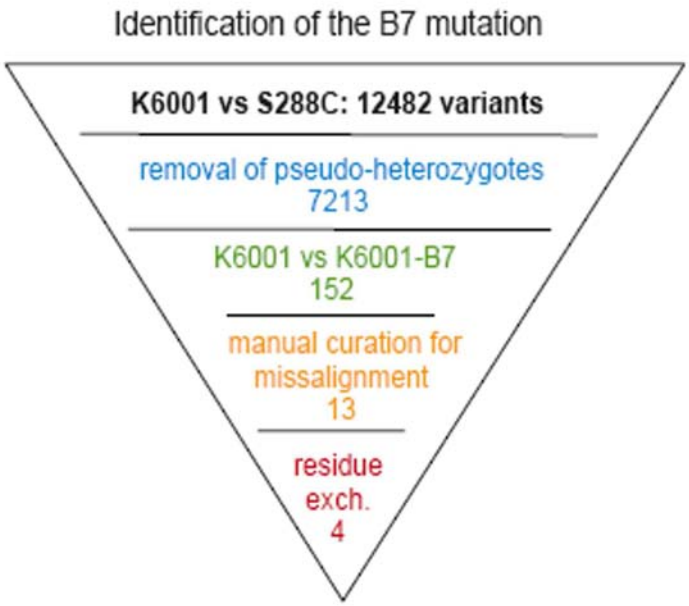


Figure 3.11: This scheme shows the workflow during for the identification of the K6001-B7-Mutation.(Timmermann et al., 2010)

position		S288C	K6001-B7	coverage	amino acid	gene	
>chrI	27082	G	A	21x	T/T	YAL063C	*FLO9*
>chrI	198629	GGAAT	AATAAA	9x			
>chrIV	1010474	T	C	14x	L/S	YDR273W	*DON1*
>chrIV	1345552	T	-	6x			
>chrVII	323609	AT	G	7x			
>chrVII	832778	-	GCCTGC	18x	- /AC	YGR167W	*CLC1*
>chrVII	984323	-	A	8x			
>chrIX	151512	C	T	16x			
>chrXIII	220682	C	T	16x	P/L	YML028W	*TSA1*
>chrXIII	606980	T	A	13x	V/D	YMR172W	*HOT1*
>chrXIII	914824	-	C	15x			
>chrXIV	98110	T	C	3x	L/L	YNL284C-B	
>chrXVI	843169	C	T	17x			

Figure 3.12: List of candidates after manual control (bottom). The red coloured candidates were used for complementation tests.(Timmermann et al., 2010)

3.4.2 Illumina Sequencing Data analysis

As mentioned before the revertant shows no temperature sensitive phenotype anymore and it is not known which mutations cause this phenotype. To find the mutation, which cause the reversion of the phenotype is an important point. Knowing the cause of the revision of the temperature sensitive phenotype gain more information about the pathway in which the two genes interact with each other. To identify this mutations the first step was a whole genome resequencing with Illumina/Solexa Sequencing Technology. The sequencing work was done at the Max Planck Institute for Molecular Genetic in Berlin. Three different yeast strains were sequenced. Two of this were revertant strains (Rev1, Rev2) and the third one was the parental strain *pso4-1*.

3.4.3 Illumina Solexa Sequencing Technology

The basic procedure is the same as in all other next generation sequencing technologies. Differences lie in the generation of the DNA libraries and the sequencing process itself. The big advantage of the next generation technologies is the simultaneously sequencing of millions of different reads. After the sequencing all reads has to align to a reference genome and putative differences can be located. Different software tools for aligning Illumina data are available. In this work the Burrows-Wheeler Alignment Tool (BWA) was used. To call the SNP´s and Indels (<50bp) and for further manipulation of the sequencing data Samtools and PERL were used.

3.4.4 Descriptive Analysis of the sequencing data

3.4.4.1 General Description of the sequencing data

The sequencing process produced an average of 3.4 GB sequencing data. The genome of all three Strains is covered with an average depth of 250. A paired end run with 2 x 100 bp reads was performed. For the mapping process we used again the S288C reference genome excluding the mitochondrial genome. In Table 3 is a summary of all descriptive analysis. Again Perl was used to process and analyze the three data files.

	Pso4 - Control	**Revertant 1**	**Revertant 2**
Number of reads	83140912	69945868	84457344
Unique mapped reads	63137438/75.94%	53734142/76.82%	67502508/79.92%
Multiple mapped reads	10636300/12.79%	9001218/12.87%	8969214/10.62%
Unmapped reads	9367174/11.26%	7210508/10.30%	7985622/9,46%
Number of bases	4179537869	3529815271	4258076606
Mapped bases	3707508739/88.70%	3166879853/89.71%	3855480665/90.55%
Unique mapped bases	3172581021/75.90%	2712449414/76.84%	3403517575/79.93%
Unmapped bases	472029130/11.30%	362935364/0.29%	3403517575/9.45%

Table 3.5: Summary of the descriptive analysis of the Illumina sequencing for PSO4-Control, Rev1 and Rev2

3.4.4.2 Genome Coverage, Coverage Coding/Noncoding regions

To calculate the Genome Coverage and Coverage Coding/Noncoding region we had to generate a coverage file. For the construction of this file we used the unique aligned reads and split them into single base. Due to the mapping process we knew the starting position of the read in the reference genome. To calculate the covered position we determined the length of the read and add it up to the starting position. Every single position which is covered is saved in the coverage file. The same Perl script determined the depth of each position. With a single Perl script we got information of the covered position and also the depth of this position. This information was used to calculate the Coverage Genome, Coverage coding/noncoding and the coverage uniformity. The *pso4-1* yeast strain showed a genome coverage of 96,56% and the coverage of coding/noncoding regions was 96.56% /96.53%. For the revertant 1 the genome coverage was 96.53% and the coverage coding/noncoding region was 96.54% / 96.50%. Revertant 2 strain had a genome coverage of 96.62% and the coverage of coding/noncoding was 96.60% / 96,69%. To visualize the coverage uniformity we used a second Perl script which split the covered position in 1000 bp units (1 kb bins), because it is not possible to visualized each single position of the yeast genome. The depth of all bases within this 1000 bp was added up and the total depth and the bins are visualized using R.

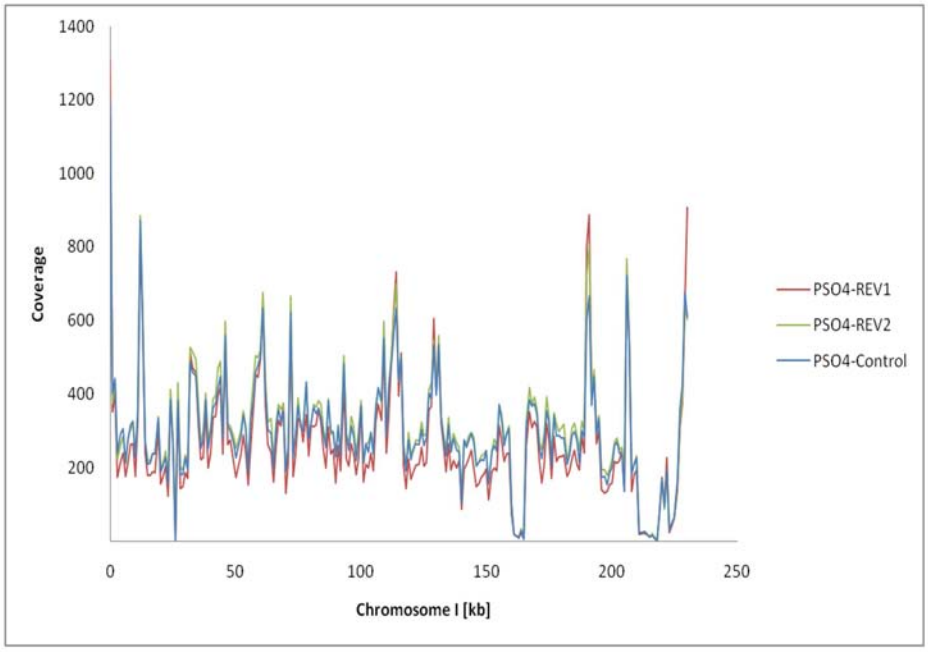


Figure 3.13: Coverage Uniformity Chromosome I: Rev2 and PSO4-Control show a higher coverage than the Rev1. Interestingly all three sequencing process cover the same regions of the chromosome.

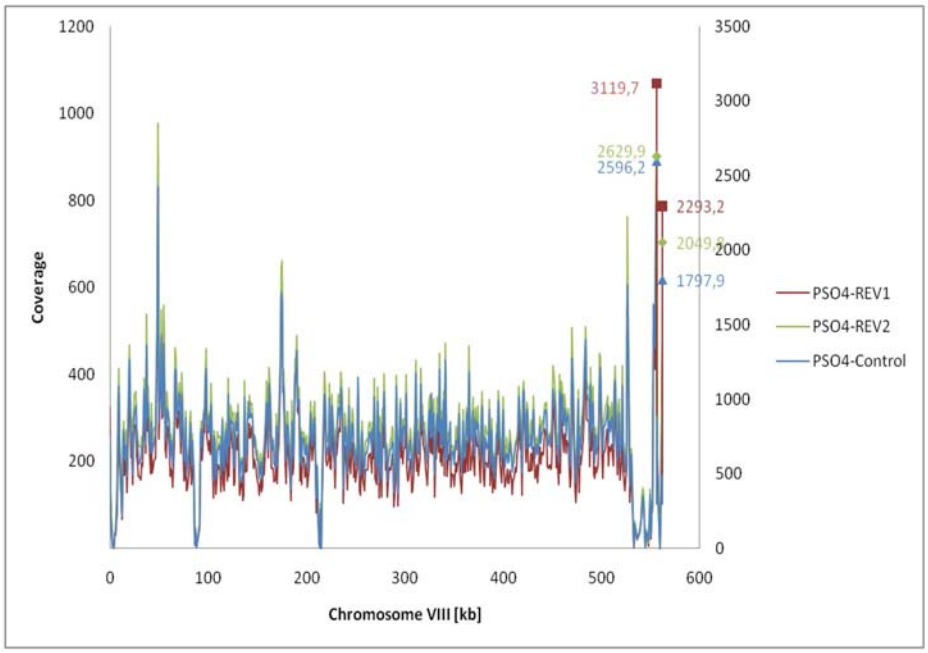


Figure 3.14: Coverage Chromosome VIII: The diagram shows two really high peaks. The reason for those peaks a repeated sequences.

3.4.4.3 Variations Statistics

At this point in the data analysis it was interesting to see, if there is a hot spot of variations to the reference genome or is the distribution of the variation nearly the same for all chromosome. Using random mutagenesis for production of muted yeast strains normally you expect a random distribution. In this case no random mutagenesis was done. Therefore the variations were split up in SNP´s and Indels. The number of each group related to the chromosome size for each chromosome and the whole genome were visualized in fig. 3.14 for Rev1 and in fig. 3.16 for Rev2. The diagram shows the frequency of variations on the chromosomes. The data shows no real hot spot with regard to variation frequency. A small number of chromosomes, like chromosome 1, 3, 7, show a higher frequency than the rest. The fig. 3.15 and 3.17 show the frequency of the variation on the different chromosomes in comparison with the whole genome.

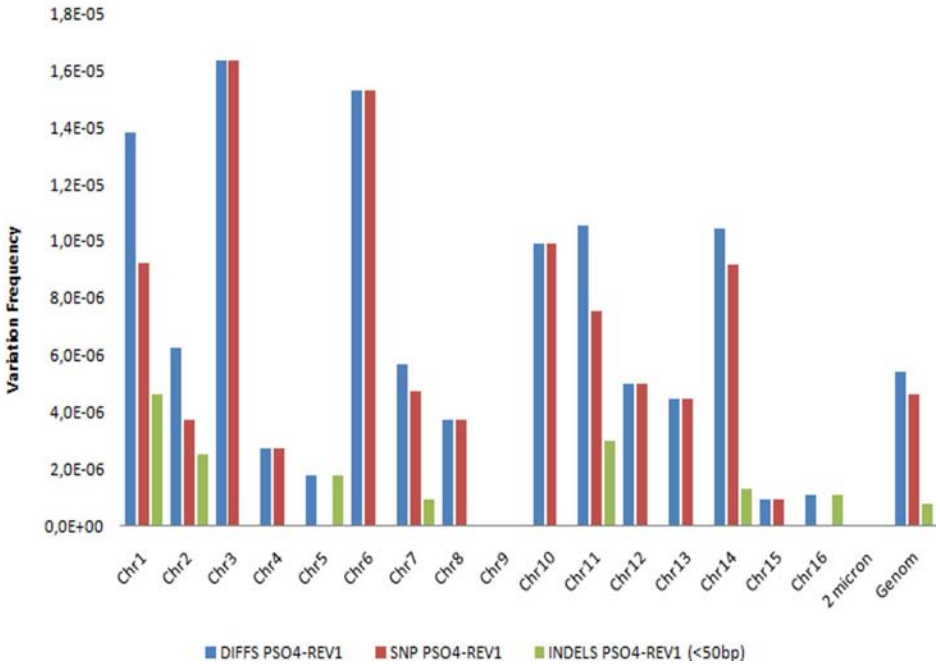


Figure 3.15: This diagram shows the frequency of variations on the different chromosomes in Rev1. For this analysis the remaining variations after the comparison with Pso4-Control were used.

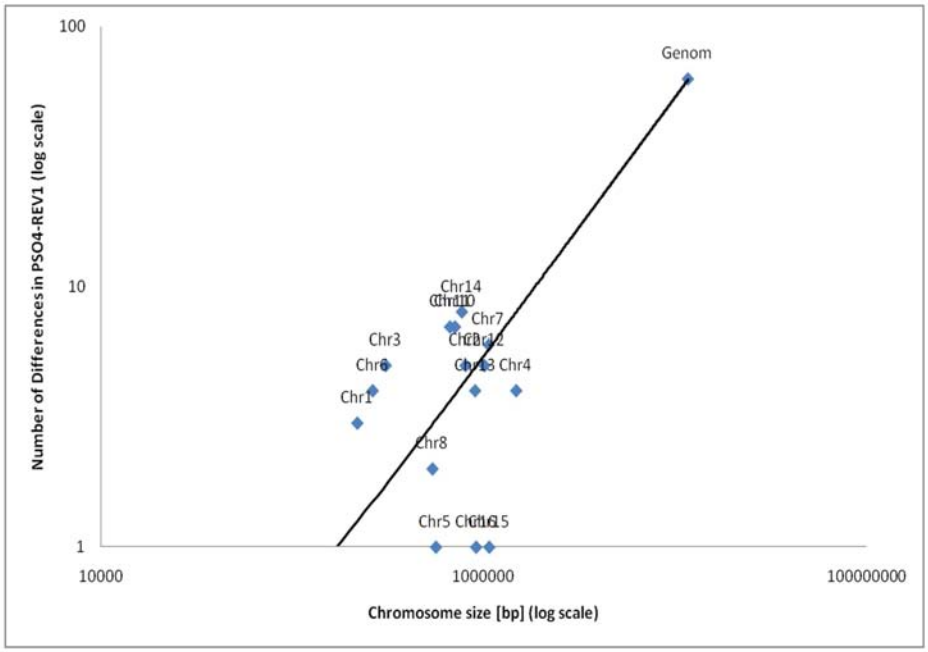


Figure 3.16: Here the same variations were used for the analysis. This figure shows the 16 chromosomes and the frequency of variations for each chromosome in comparison with the frequency for the whole genome.

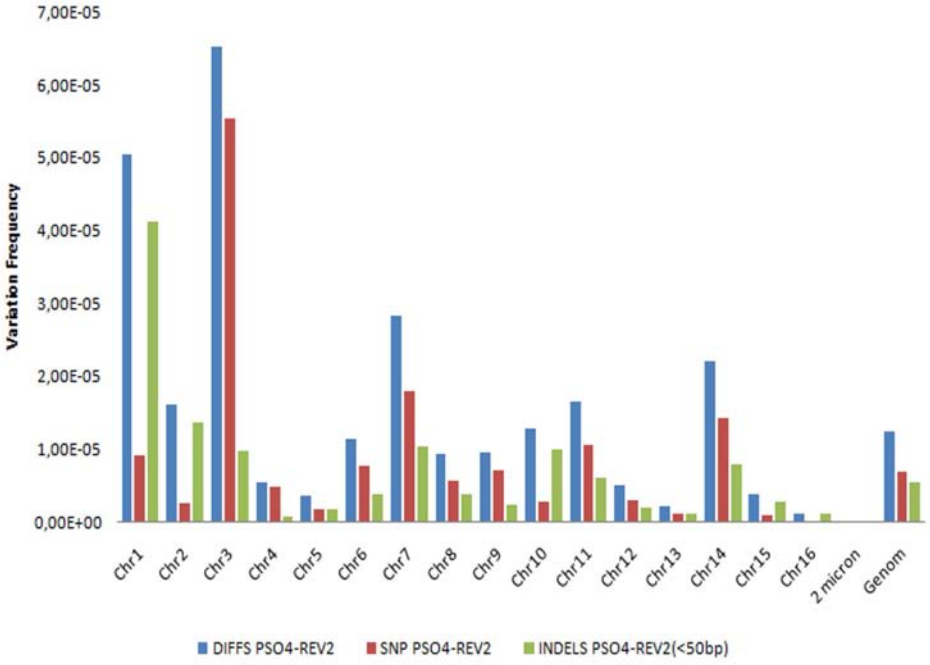


Figure 3.17: The diagram shows the frequency of variations over the whole genome and the 16 chromosomes for Rev2. Again the remaining variations after the comparison with Pso4-Control were used.

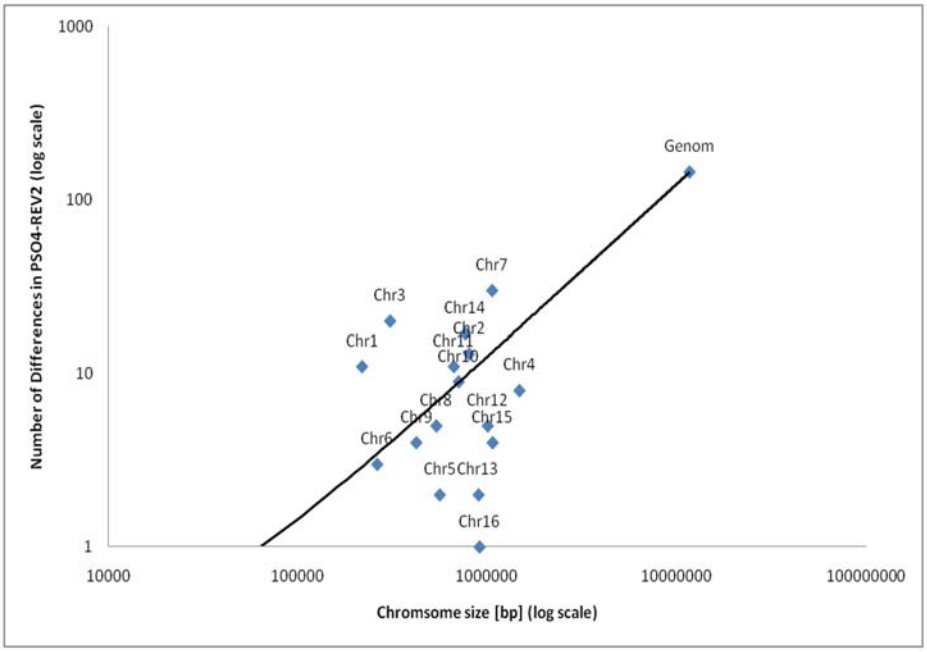


Figure 3.18: Frequency of variations over the whole genome and all chromosomes in Rev2.

3.4.4.4 Identification of the mutations in Rev 1 and Rev 2

For the identification of the putative mutations we used the same approach as in the 454 Roche project. The first step was the SNP´s and Indels (<50bp) calling. The result of the variation calling were 12159 for the Pso4 Control strain, 12051 for revertant 1 and 12585 for revertant 2 SNP´s and Indels (>50bp). In a next step the variations were split in SNP´s and Indels. So we got for Pso4-Control 10784 SNP´s and 1411 Indels. For the Rev1 we found 10749 SNP´s and 1302 Indels and for the Rev2 10575 SNPS´s and 2010 Indels. First we looked at the SNP´s, because they were easier to handle. To reduce the number of SNP´s we applied a cut off filter of 85 %. We used a lower cut off, because of the high coverage of the sequenced genome. Furthermore we calculated the variation frequency, which was important to apply the cut off filter. To calculate the variations frequency a Perl script was used. After removing the SNP´s with a lower variation frequency than 85% we got 10579 for Rev1 and 10431 for Rev2. To eliminate those SNP´s which are equal for Pso4-Control and the two reverant strains we compared the strains and filter out the SNP´s only occur in revertant 1 and revertant 2. After the comparison 54 SNP´s for revertant 1 and 81 SNP'S for revertant 2 were left. The final step was to look by hand at all remaining SNP´s, if there is really a difference to the control strain. After this manual control we got one SNP for revertant 1 and 4 SNP´s for revertant 2. We did the same procedure for the Indel´s the only thing that was different is the calculation of the variation frequency. We used the number of reads which support one allele to determine the variation frequency of the SNP´s. We calculated for each allele the frequency and used the allele with the highest frequency, which also has to be over 85%. After applying

the 85% cut off filter we reduced the number of Indels to 144 for revertant 1 and 212 for revertant 2. After comparison with the pso4 control strain 9 Indels for revertant 1 and for revertant 2 64 were left. The manual control of the remaining Indels eliminated all Indels for revertant 1 and for revertant 2 only 6 Indels were left. The next step is to check if the remaining mutations a real. This is done be isolation the genomic DNA and amplification of the different gene locus with PCR.

Position		S288C	Rev1	Gene
chrXV	411612	C	A	YOR043W/WHI2

Table 3.6: List of putative candidates after manual check for Rev1

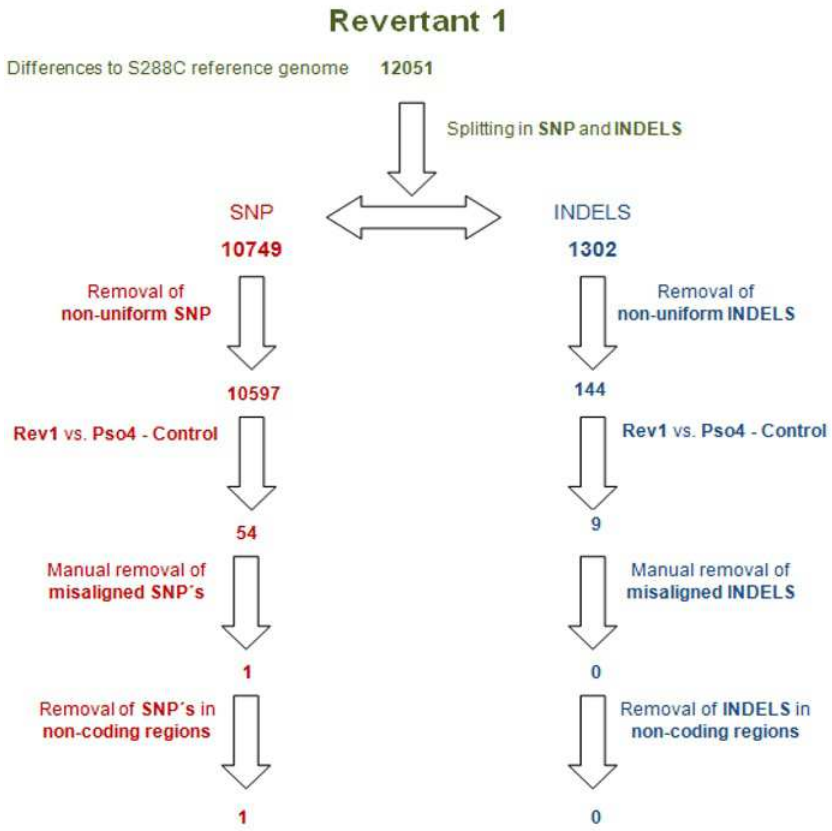


Figure 3.19: Workflow for Identification of Rev1 mutation

Position		S288C	Rev2	Gene
chrVII	8596	A	C	YGL259W
chrXIV	750702	C	T	YNR065C
chrXIV	750761	C	T	YNR065C
chrXIV	751286	G	A	YNR065C
chrII	92917	*	-CCG/*	YBL068W
chrVII	21893	*	G-/G-	YGL255W
chrXI	199359	*	T+/T+	YKL129C
chrXV	220192	*	T+/T+	YOL058W
chrXIII	53771	*	GAACTG+/GAAC	YML109W
chrVI	100379	*	CAGGTG-/CAGG-	YFL019C

Table 3.7: List of putative candidates after manual check for Rev2

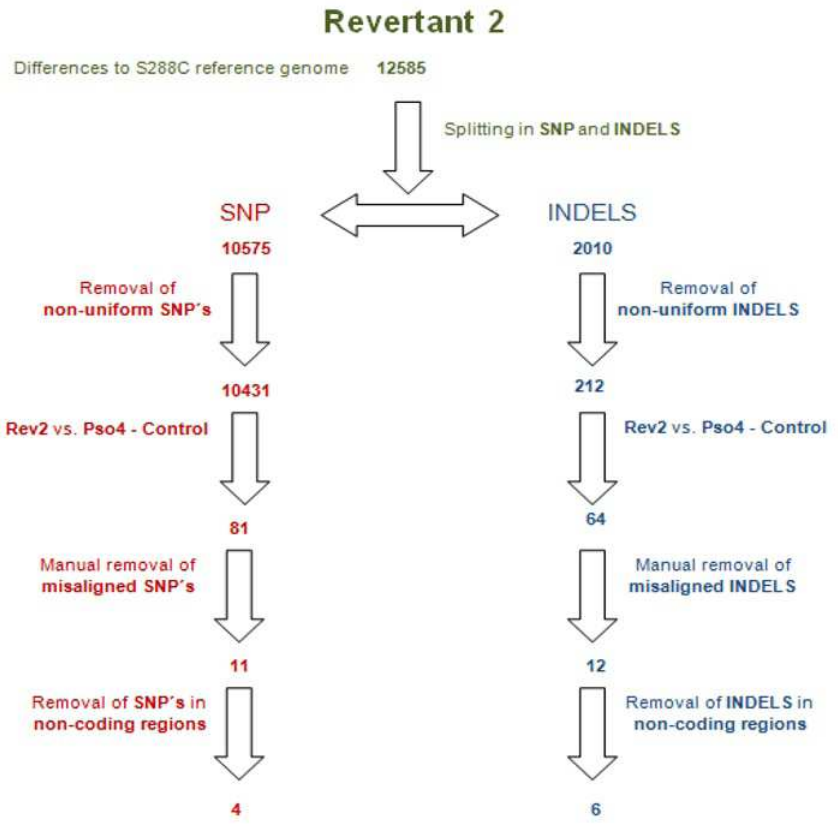


Figure 3.20: Workflow for Identification of Rev2 mutation

4 Discussion

Sensitivity test of the mutant strains by spotting tests

The first part in this work was the verification of spotting test results that we got from the University of Salzburg. Spotting tests were done using different chemicals reagents, like MMS, DEM, H_20_2. The strains which should be tested were kindly provided by Dr. Michael Breitenbach. The tested yeast strains comprise double mutants (*pso4-1*, Δsgs1) and single mutants. Single mutant strains either contain *pso4-1* or Δsgs1. Prp19/Pso4 is an essential gene in yeast and was identified by two different approaches. Prp19/Pso4 is a splicosome associated protein and helps by the activation of the splicosome. Prp19/Pso4 was identified via the *pso4-1* strain, which shows a psoralen sensitive phenotype. The *pso4-1* mutant shows a pleiotropic phenotype, which comprises temperature sensitivity, mutagen and radiation sensitivity, disturbed induced mutagenesis, block of sporulation.(Brendel and Henriques, 2001) This finding created a novel link between DNA repair and pre-mRNA processing. Sgs1 is the only member of the RecQ helicase family in yeasts. RecQ-helicases are involved in several processes for example DNA replication, recombination and DNA repair.(Sharma et al., 2006) Mutation in RecQ helicases are often linked with human diseases.(Werner Syndrome, Bloom Syndrome) Now it was interesting to check if there is interaction between Prp19/Pso4 and Sgs1 in DNA repair, especially in repair of interstrand cross links (ICL). The testing with MMS (methyl methanoe sulfonate), which is a alkylating agents and causes double strand breaks, showed that the double mutants a have increased sensitivity compared with *pso4-1*. The single mutant Δsgs1 is slightly less sensitive to MMS then the double mutants. The interesting thing is that the revertant shows almost no sensitivity to MMS. Because of the additive effect between the sensitivity of single mutants compared with double mutants it may indicate that Prp19/Pso4 plays a role in a Sgs1 independent repair pathway. Recent studies in in vitro cells also suggested that Prp19/Pso4 plays in role the repair of double strand breaks induced by interstrand crosslinks(Nianxiang et al., 2005). The results of the spotting test on diethylmaleate(DEM) showed that *pso4-1* is sensitive to higher ROS level. Interestingly the Δsgs1 strains showed almost no negative growth effects on DEM. The double mutants or single mutants (only *pso4-1*) showed a high sensitivity to DEM, which suggests that the presence of a functional Sgs1 has no influence on the sensitivity of *pso4-1* mutants to DEM. The revertant also shows a DEM sensitive phenotype, because of the fact that the revertant still has the *pso4-1* mutated gene. The reason for the *pso4-1* sensitivity to higher ROS levels is unclear. Studies stated that the spliceosomal function of Prp19 mutants at 30°C is close to the wildtype

functions, but the function in DNA repair is perturbed. This suggests that Prp19/Pso4 possesses two different domains and one of this is disrupted(Grey et al., 1996). Coming back to the results this is maybe a reason for the sensitivity to higher ROS levels. A second explanation could be also a disruption of the spliceosomal function, but this is rather unlikely because of the fact that at 30°C there is no pre-mRNA accumulated. *Saccharomyces cerevisiae* possesses 238 genes containing introns and only one of these is related to DNA repair. This gene is Rad14, which causes incision of damaged DNA, but *pso4-1* mutants are able to incise DNA(Grey et al., 1996).
A easy way to test if the Rad14 gene is really involed in the creation of the pso4-1 sensitivity to ROS could be a knock out of this gene in wildtype yeast strains and do again a spotting on DEM or H_20_2. A higher ROS sensitivity of the knock out strain compared to the wildtype strain would indicate that a functional Rad14 gene is necessary for repair of damages caused by ROS. The results of the H_20_2 spotting are not that clear. The double mutants and also the single mutants for *pso4-1* showed sensitivity to hydrogen peroxide.*pso4-1* mutants show a higher sensitivity to H_20_2 compared to mutants. The problem with the results form the H_20_2 spotting is that also the wildtype (BY) and spore 4D showed nearly the same sensitivity to H_20_2. This fact makes it hard to find a clear result. One possible reason for the observed results is maybe a non isogenic wildtype. Another reason can be a problem by the preparation of the plates used for spotting. The H_20_2 was maybe not the same at every position on the plate. The main goal of the spotting tests was to test the interacting of Prp19/Pso4 and Sgs1 in ICL repair. The chemicals used for spotting cause different stress condition. H_20_2 and DEM cause oxidative stress. MMS lead to double strand breaks, which are only an intermediate during the interstrand crosslink repair pathway. The results of the spotting test are good, but testing on more different chemicals would be a good chance to better understand the interaction Prp19/Pso4 and Sgs1 in repair mechanism. To have a really estimation of the interaction of Prp19/Pso4 and Sgs1 in ICL repair it is necessary to test with other chemicals. One chemical could be psoralen. Psoralen itself produces no ICL. The activation of psoralen with UVA causes this harmful DNA damages. A spotting on psoralen containing plates maybe will help by the elucidating of the interaction of Prp19/Pso4 and Sgs1 in ICL repair.

Targets of the Prp19/Pso4 E3-Ligase domain

The second part of this work was to identify substrates of the E3-ligase domain in the Prp19/Pso4 gene has. Therefore we got a *Saccharomyces cerevisae* strain (SUB592) from Dr. Daniel Finely. This yeast strain possesses a knockout for all ubiquitin genes in yeast and comprises two plasmids. One of the plasmids contains a expression cassette for an His-tagged ubiquitin. The other plasmid (pUB39) contains the Ubi1 tail. (Spence et al., 2000) So the ubiquitinylated substrates of Prp19/Pso4 E3-ligase domain in Pso4 can be easily purified and western plotted. Also the temperature sensitive phenotype of *pso4-1* was a big advantage, because this was a possible way to determine the substrate specific for the E3-ligase domain. At an incubation temperature of 30°C the function of the *pso4-1* gene is normal like every wildtype gene and also ubiquitination works. Incubation of the *pso4-1* strain at 37°C the function of the *pso4-1* gene is disrupted and no ubiquitination happens. Concerning both situations it is possible to extract the substrates of the E3-ligase. The first step to test this theory was to establish an SUB592 yeast strain with a *pso4-1* mutated gene. We had different approaches to do that. At first a crossing of the *pso4-1* strain with SUB592 was done. Through analysis of the tetrads the strain containing *pso4-1* mutation could be isolated. This approach didn´t worked because of the bad sporluation efficiency of *pso4-1* strain.

Another approach was to introduce the *pso4-1* gene via transformation in SUB592. Therefore a plasmid containing the *pso4-1* gene and a kanMX was used(Hager, 2009). The kanMX cassette was on the 3´end of the *pso4-1* gene. The kanMX cassette was needed for selection of the positive candidates because of the temperature sensitive phenotype no selection would be possible. Through amplifying a fragment containing the *pso4-1* gene and the kanMX by PCR and later transformation via LiAc method should insert the *pso4-1* gene on the Prp19/Pso4 wild type locus via homologous recombination. The result was that we got positive transformants and we checked then for the temperature sensitive phenotype. First we checked it by incubation of the transformants at 30°C and 37°C. If the *pso4-1* gene was successfully introduced at the wild type locus the SUB592 also should show the temperature sensitive phenotype. No transformants were found that showed this temperature sensitive phenotype. Furthermore to really be sure that the *pso4-1* gene didn´t insert at the wild type locus, genomic DNA was isolated and the Prp19/Pso4 locus was amplified via PCR. This fragment was then digested with the restriction enzyme Afl II. The wild type gene possesses an AflII restriction site. In the *pso4-1* mutated gene this restriction site is destroyed. Because of this fact you can distinguish between wild type gene and mutated gene. The digestion results in a 1482 bp fragment and another fragment with 328 bp. The final outcome of the digestion analysis was that the *pso4-1* gene didn´t insert at the wild type locus. Possible reason could be the use of a wrong transformation protocol, because of the applied heat shock in the LiAc protocol the *pso4-1* mutated cells could not survive. The transformation was repeated two times with the same result. Another approach would be a transform of the plasmid, containing the 6xHis tagged ubiquitin, into a *pso4-1* strain.

Determination of a specific mutation via next generation sequencing data

The last part of my diploma thesis deals with the analysis of next generation sequencing data using bioinformatic tools. The detection of real mutation via next generation sequencing data is a new approach. The normal approach is to isolated the genomic DNA of the mutant strain and clone it into plasmid vectors. The different plasmids are later transformed in the parental yeast strain. Then a selection step is necessary to check for positive candidates. From these candidates the plasmids are isolated and the inserted sequence has to be sequenced. After the sequencing a determination of the involved genes is possible. For the determination of the mutation it is also important to know if the mutation is dominant or recessive. There are more factors which complicated the way of finding the mutation.

The way used in this thesis to find a specific mutation was new and uses next generation sequencing data as starting point. So a reduction of the possible candidates before any time intensive laboratory work was possible. For the reduction of the candidates a simple Perl script was necessary to remove those variations, which are equal in the reference genome and the sequenced parental and mutant strand. The next step is a manual check of the remaining variations. This step was important to remove those variations from the mutant strain, which show the same nucleotide sequence as in parental strain. After the manual check the work in the lab start. The remaining candidates have to be checked by sequencing for a real mutation. Further complementation tests have to be done to find the right gene with the specific mutation.

In the first project the two yeast strains K6001 and K6001-B7 were sequenced by 454 life Science/Roche sequencing. K6001 is a yeast aging model strain and was the parental strain(Jarolim et al., 2004). The mutant K6001-B7 was isolated after treatment with EMS. The mutant strain phenotype was a higher oxidative stress resistance. To determine the gene which carries the mutation causing this phenotype was found by a subtractive whole genome resequencing approach. In this project a whole genome resequencing of both genome K6001 and K6001-B7 was done using 454 life Science/Roche next generation sequencing technology. The average read length was 499 bp and we covered the K6001 genome with a 19.2 fold average and the K6001-B7 genome with 19.5 fold coverage. The merged genome showed 38.8 fold coverage. Another interesting point was the coverage uniformity. At this point it was interesting to see, which regions of the 16 different chromosomes were covered and also the level of coverage. The outcome was that regions with no or high coverage were in both strains at the same position on the chromosomes. These results showed that no random failure happened in the sequencing process. The mapping of both genomes against the reference genome S288C showed that in K6001 96.98% and in K6001-B7 96.88% of the reference genome was covered. The variation detection using again S288C reference genome detected 11230 high confidence variations for K6001 and 11397 for K600-B7. Now it was time to reduce these variations by applying a cut off filter of 95%. After removing all non-uniform variations from

K6001-B7 we landed at a number of 7213 uniform high confidence variations to the reference genome. After removing all equal variation between K6001 and K6001-B7 and a manual control we had only 13 candidates left. 7 of these candidates were located in protein coding regions and 4 of the candidates showed residual changes. The four genes are CLC1, DON1, HOT1 and TSA1. They were amplified by PCR and checked for mutations with Sanger sequencing. The genes were then subcloned in a single-copy yeast expression vector and transformed into K6001 and the transformants were test for their phenotype on CHP (cumene hydroperoxid). The experiments showed that a mutation in TSA1 causes this specific phenotype of K6001-B7. TSA1p (thiol specific antioxidant) belongs to the group of peroxiredoxins and reduces H_20_2in presence of thioredoxins. For further details look at the attached publication in the appendix.

The second project had the same aim, but uses different yeast strains and another next generation sequencing platform. Also here a whole genome resequencing of two revertant strains (Rev1, Rev2) and the parental *pos4-1* strain was done using Illumina/Solexa next generation sequencing. The *pso4-1* strain shows a temperature sensitive phenotype at 37°C. The two revertants, which developed independent, didn´t show this temperature sensitive phenotype. They were developed by striking out the *pso4-1* strain out on plates and incubate them for a few days at 37°C. Further experiments showed that the mutations isn´t in the *pso4-1* gene. Also the temperature causing mutation in both strains is not in the same gene. Why it is interesting to know the mutated gene? /Prp19/Pso4 is an essential gene in *Saccharomyces cerevisiae* and a knock out is lethal for the strain. Because of this fact it would be of high interest to identify a new interaction partner of Prp19/Pso4, which reverse the phenotype of *pso4-1* temperature sensitivity. The same analyses were performed as in the first project. The mapping against the reference genome of S288C showed that Rev1 covered 96.53%, Rev2 96.62% and *pso4-1* 96.56% of the reference genome. The reference genome was also covered with a 200 fold average coverage. The coverage uniformity also showed a homogeneous coverage distribution for all three strains. The next step was the variation calling and the result of was 12159 variations for Pso4-Control strain, 12051 for Rev1 and 12585 variations for Rev2. Again we applied a cut off filter, but this time at 85 % variation frequency. The reason for the lower cut off filter was the high coverage of the genome. Due to the high coverage it is possible that a real SNP comes up with a lower variation frequency as expected. A possbile reason for the lower variation frequency is sequencing errors. It is also possible that a real SNP or mutation has a lower frequency than 85%. Also the way of filtering the Indels may comprise possible errors. The Samtools variation caller shows four different possible alleles, which were detected during the calling. We only used the two most possible alleles for the analysis. Also the way how the alleles were shown in the alignment viewer was sometimes real complicated. We decided to discard those Indels, which show up in a sequence were more than 3 Indels per 10 bp occur. If we didn´t find the phenotype causing mutation with this first

attempt the above mentioned points could be improved.(lowering the cut off filter). After comparing with Pso4-Control strain and a manually control afterwards the final number of variations in Rev1 was only one variation. For Rev2 we ended up with 10 variations. (4 SNP´s and 6 Indels) This SNP and Indels are all in a protein coding region. The next step in the process of determining the mutation in the two revertant is to check if the remaining variants are real mutations. This can be done by isolation of genomic DNA of all three strains and amplifying the candidate genes and check the sequence via Sanger sequencing. Those mutations which are real can be used for complementation tests. The next step depends on the character of the mutation. This means that it is crucial to know if the mutation is dominant or recessive. In case of an dominant mutation the gene is amplified from the genomic DNA of the revertant strain and cloned in a yeast expression vector. The plasmid is transformed in the parental strain. In case of a complementation the parental strain should show the same phenotype as the revertant strain. If the mutation is recessive the gen is amplified from the parental strain and transformed in the revertant strain. The dominant allele of the parental strain will then restore the old phenotype of the parental strain in the revertant. We yet know that the mutation in revertant 1 is recessive. For revertant strain 2 this analysis has to be done.
If we find also in this project using Illumina Sequencing Technology this would be a great step forward to apply this method for further analysis in functional genomics. The next generation sequencing technology develops very fast and has a great potential not only in estimation the nucleotide sequence of a genome. This technology can help by different problems in functional genomics. It is possible to sequence every genome, but sometimes it is a hard way to get from the raw data to the nucleotide sequence of the genome. So there is a lot of work to do in developing software which can handle and process such big data sets for a better understanding. This would open new fields for the next generation sequencing technologies.

A Appendix

A.1 Abbreviations

ROS	Reactive oxygen species
TSA	Thiol-specific antionxidant
NER	Nucleotide Excision Repair
BER	Base Excision Repair
MMR	Mismatch repair pathway
NHEJ	Non-homologous end joining
HR	Homologous repair
DSBR	Double strand break repair
SDSA	Synthesis-dependent repair
ALT	Alternative lengthening of teleomers
emPCR	Emulsion PCR
SNV	Single nucleotide variation
Indel	Insertion or Deletion
MMS	Methyl methane sulfonate
DEM	Diethyl maleate

A.2 Resluts Spotting Tests

A.2.1 H_20_2-Spotting

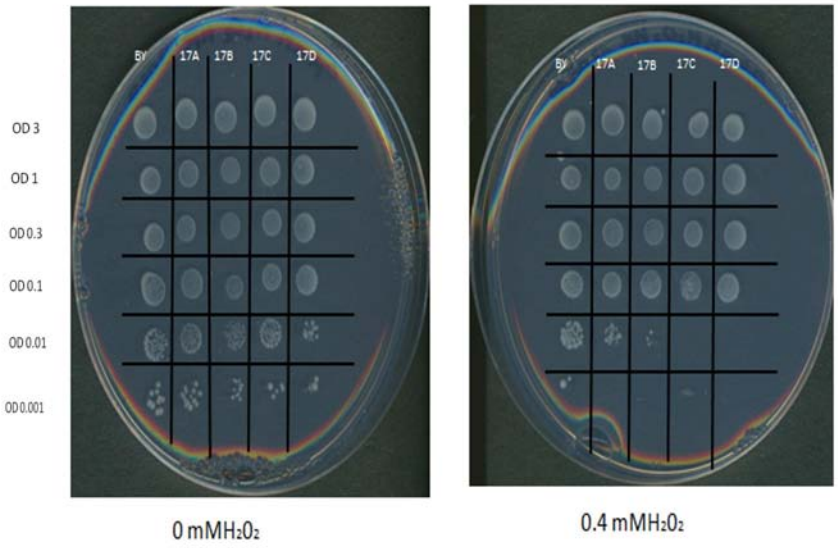


Figure A.1: Spotting Tetrade 17(0mM and 0.4mM)

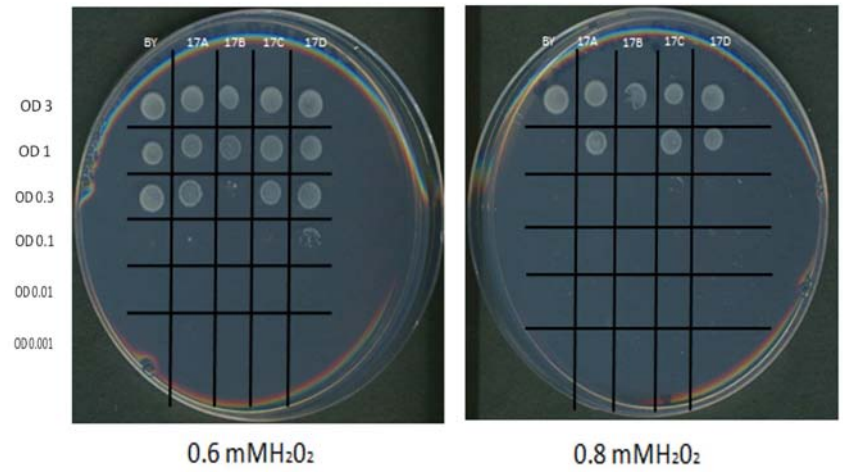


Figure A.2: H_20_2Spotting Tetrade 17(0.6mM and 0.8mM)

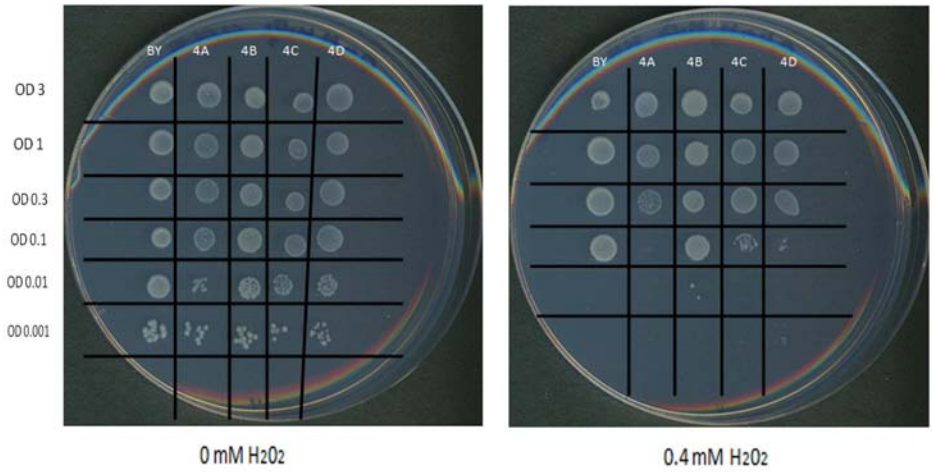


Figure A.3: H_20_2Spotting Tetrade 4(0mM and 0.4mM)

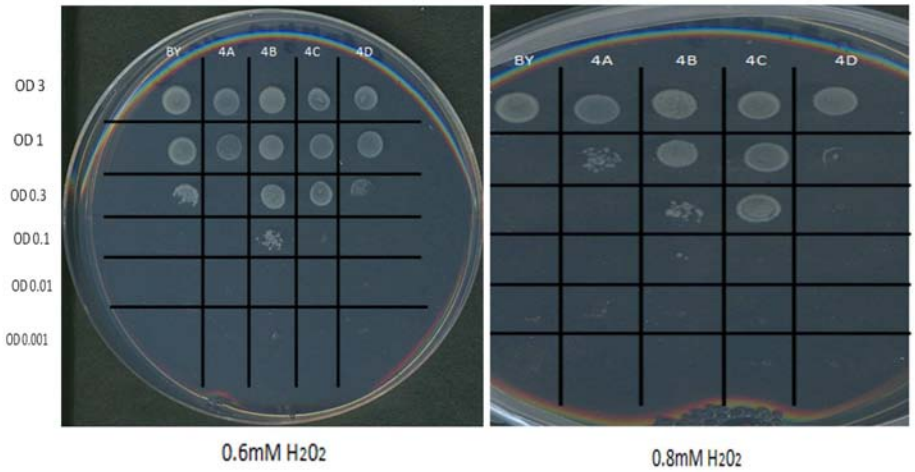


Figure A.4: H_20_2Spotting Tetrade 4(0.6mM and 0.8mM)

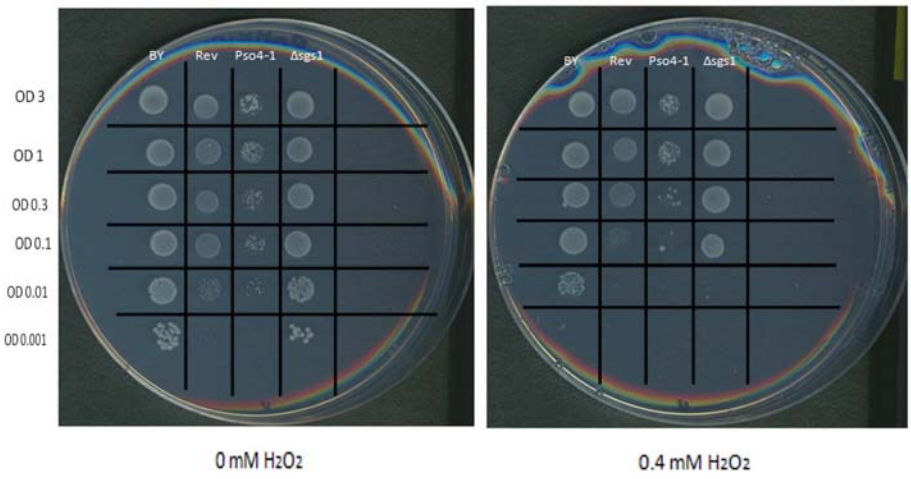


Figure A.5: H_20_2-Spotting pso4-1,Rev and $\Delta sgs1(0mMand0.4mM)$

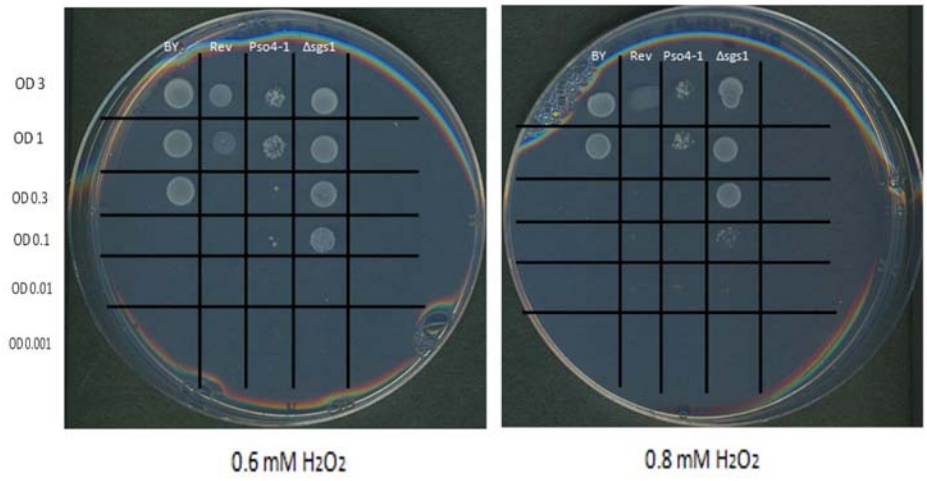


Figure A.6: H_20_2Spotting pso4-1,Rev and $\Delta sgs1(0.6mMand0.8mM)$

A.2.2 DEM-Spotting

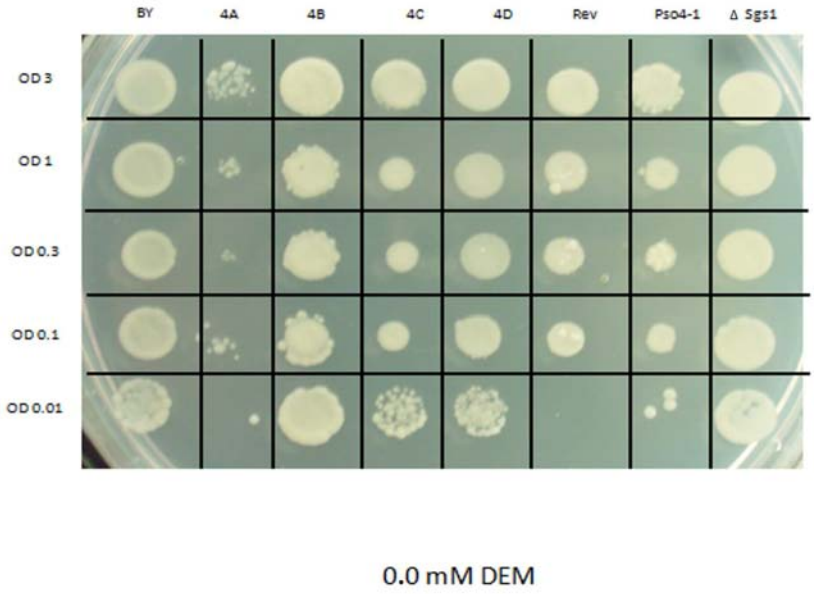


Figure A.7: DEM Spotting Tetrade 4, pso4-1 and $\Delta sgs1(0mMDEM)$

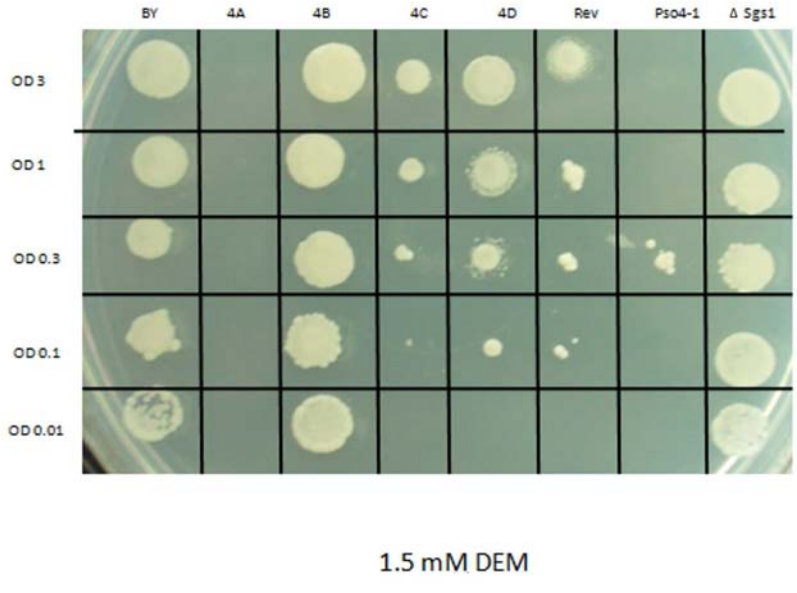


Figure A.8: DEM Spotting Tetrade 4, pso4-1 and $\Delta sgs1(1.5mMDEM)$

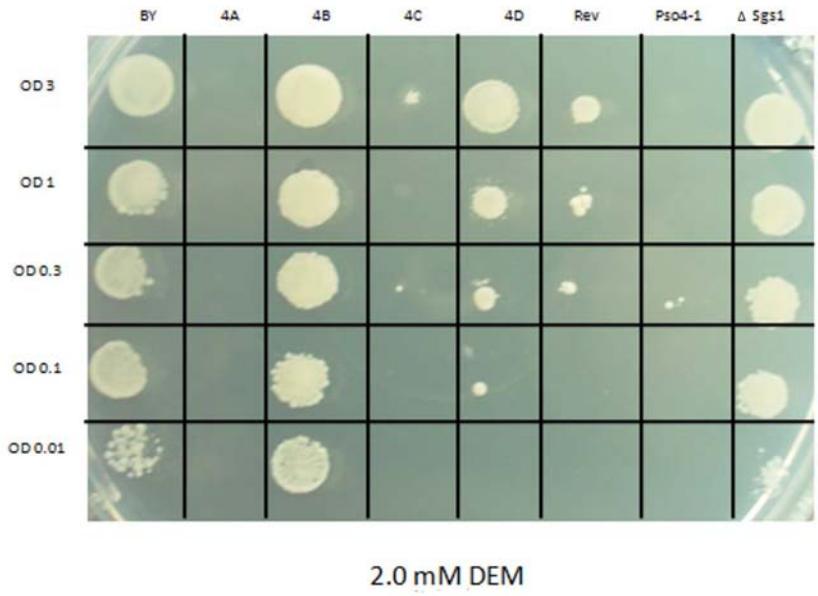


Figure A.9: DEM Spotting Tetrade 4, pso4-1 and $\Delta sgs1(2.0mMDEM)$

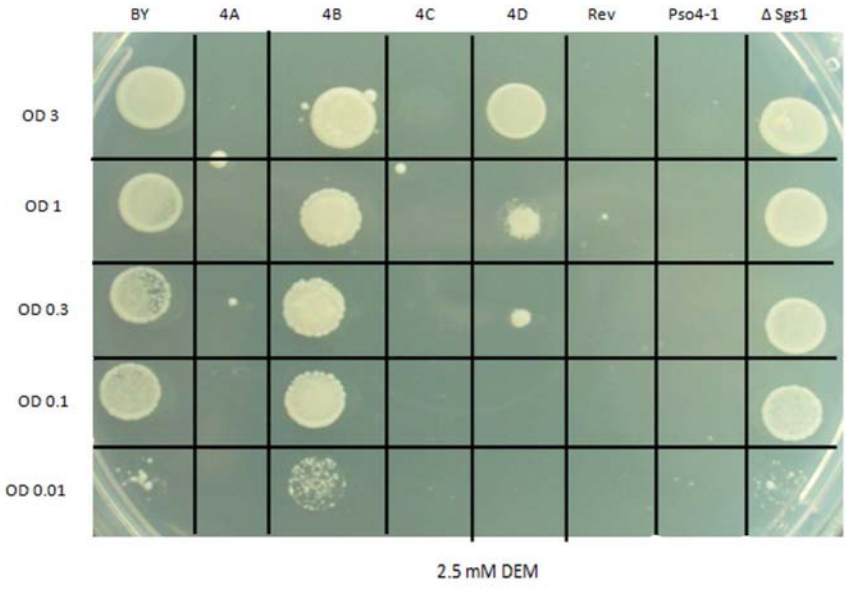


Figure A.10: DEM Spotting Tetrade 4, pso4-1 and $\Delta sgs1(2.5mMDEM)$

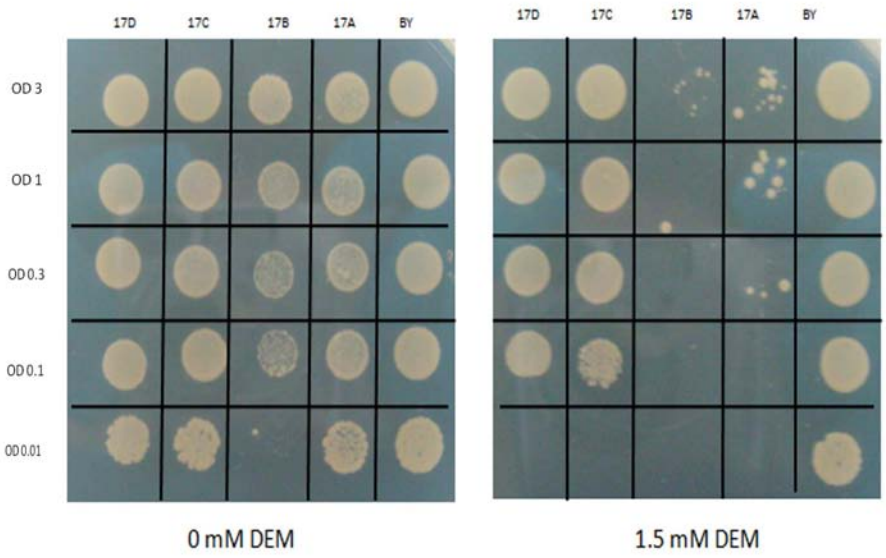


Figure A.11: DEM Spotting Tetrade 17 (0mM and 1.5 mM)

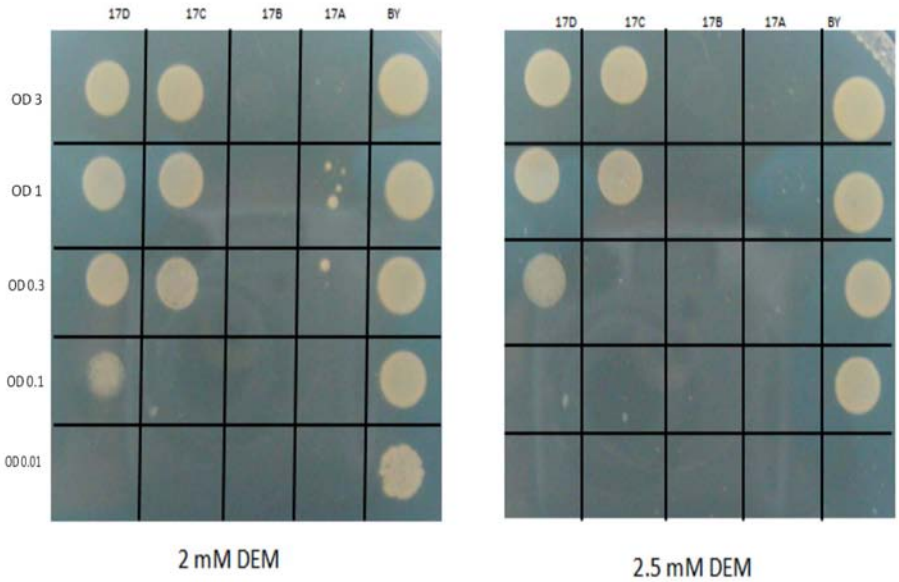


Figure A.12: DEM Spotting Tetrade 17 (2mM and 2.5 mM)

A.2.3 MMS-Spotting

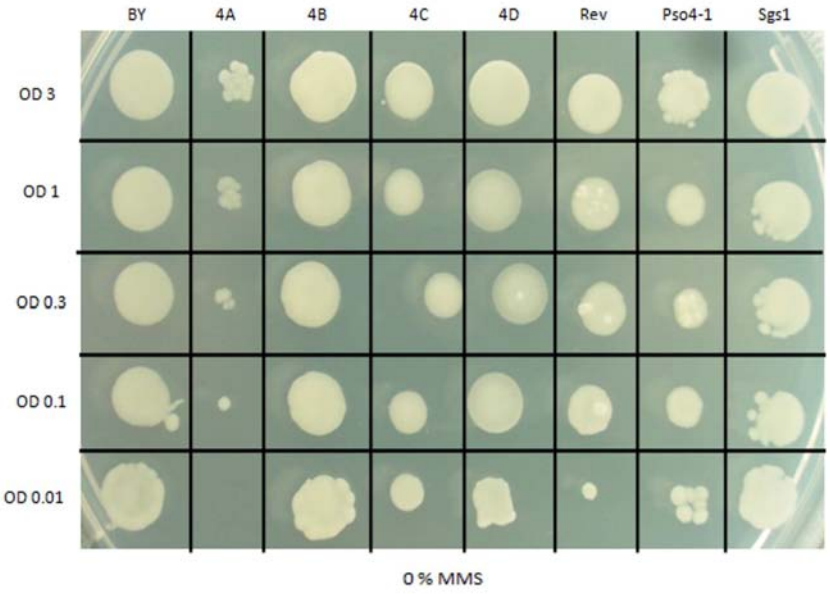


Figure A.13: MMS Spotting Tetrade 4, pso4-1 and $\Delta sgs1(0\% MMS)$

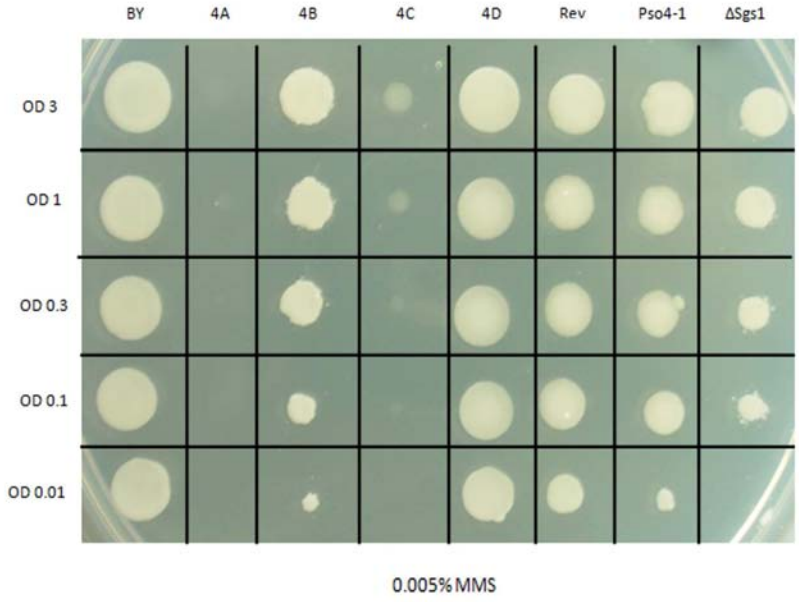


Figure A.14: MMS Spotting Tetrade 4, pso4-1 and $\Delta sgs1(0.005\% MMS)$

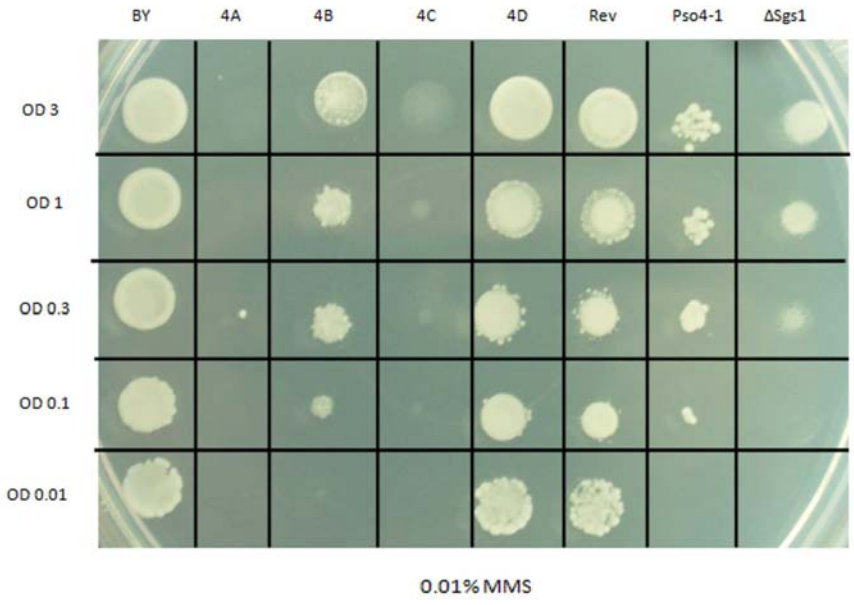


Figure A.15: MMS Spotting Tetrade 4, pso4-1 and $\Delta sgs1(0.01\% MMS)$

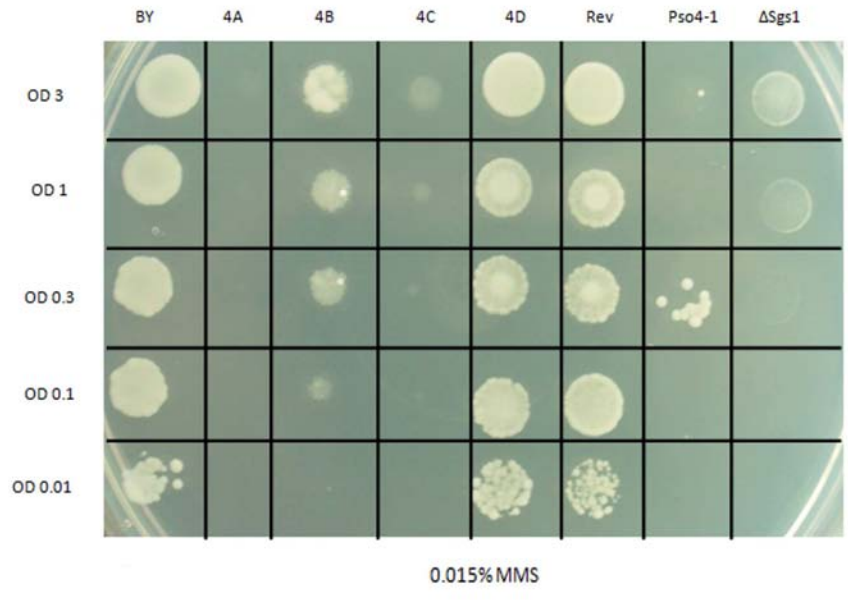


Figure A.16: MMS Spotting Tetrade 4, pso4-1 and $\Delta sgs1(0.015\% MMS)$

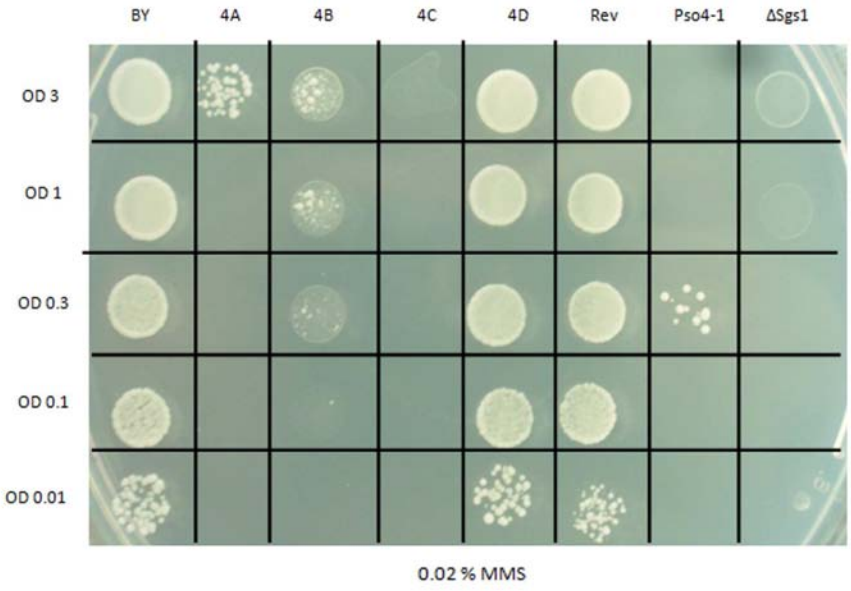


Figure A.17: MMS Spotting Tetrade 4, pso4-1 and $\Delta sgs1(0.02\% MMS)$

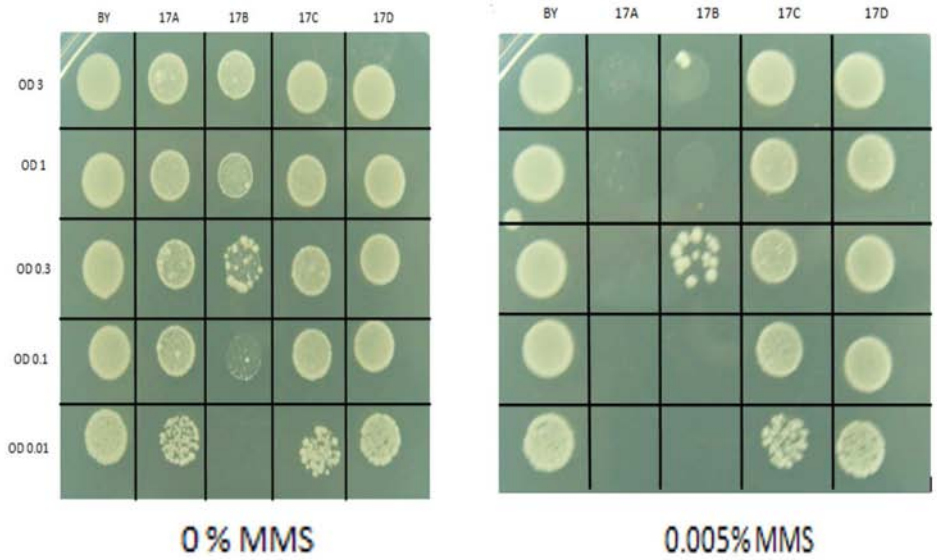


Figure A.18: MMS Spotting Tetrade 17 (0% MMS and 0.005% MMS)

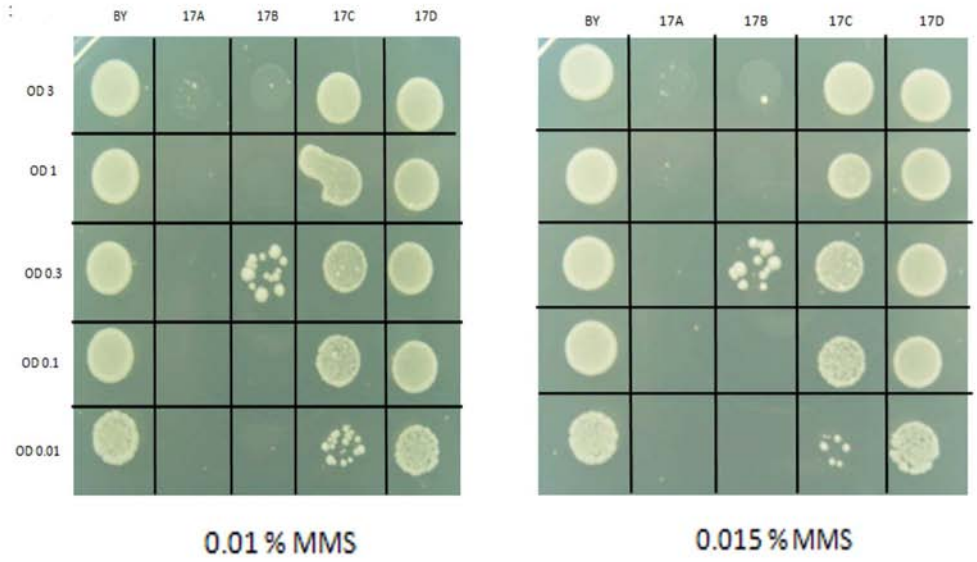


Figure A.19: MMS Spotting Tetrade 17 (0.01% MMS and 0.015% MMS)

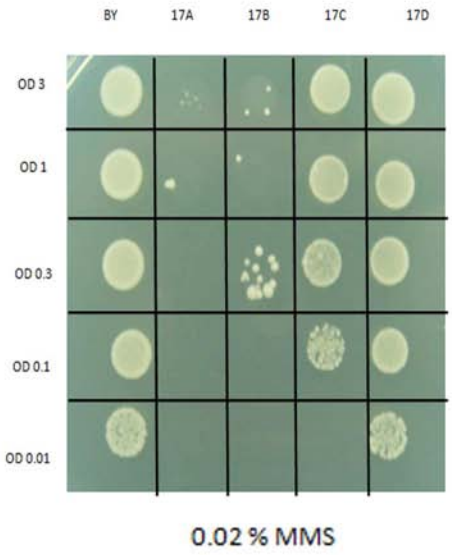


Figure A.20: MMS Spotting Tetrade 17 (0.02% MMS)

A.3 454 Life/Sciences Roche Coverage Uniformities

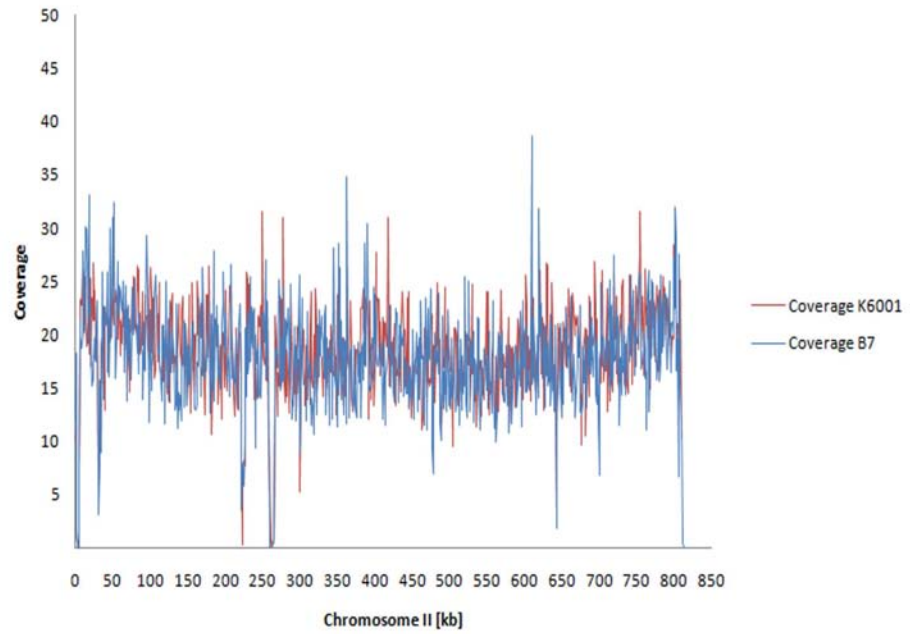


Figure A.21: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 2

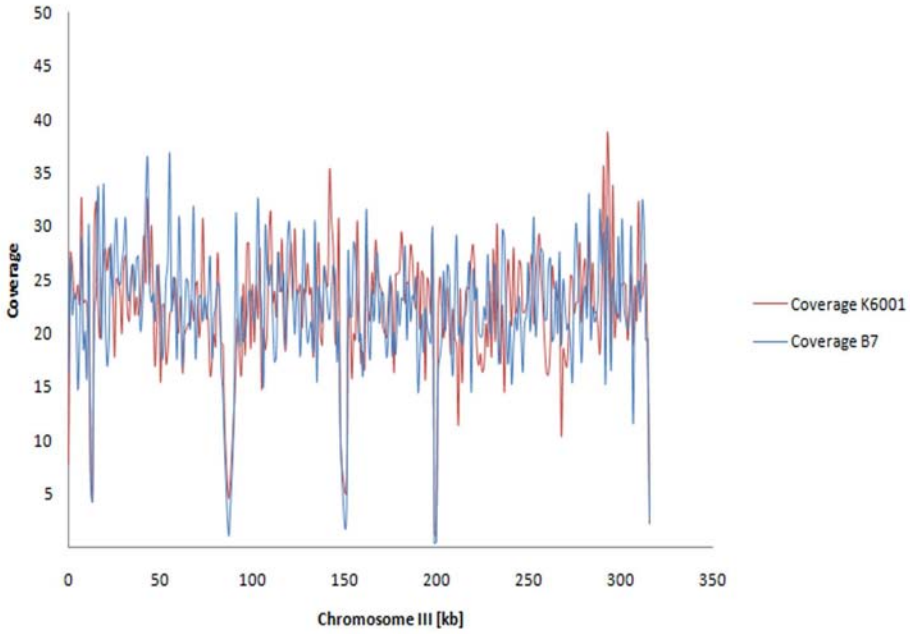


Figure A.22: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 3

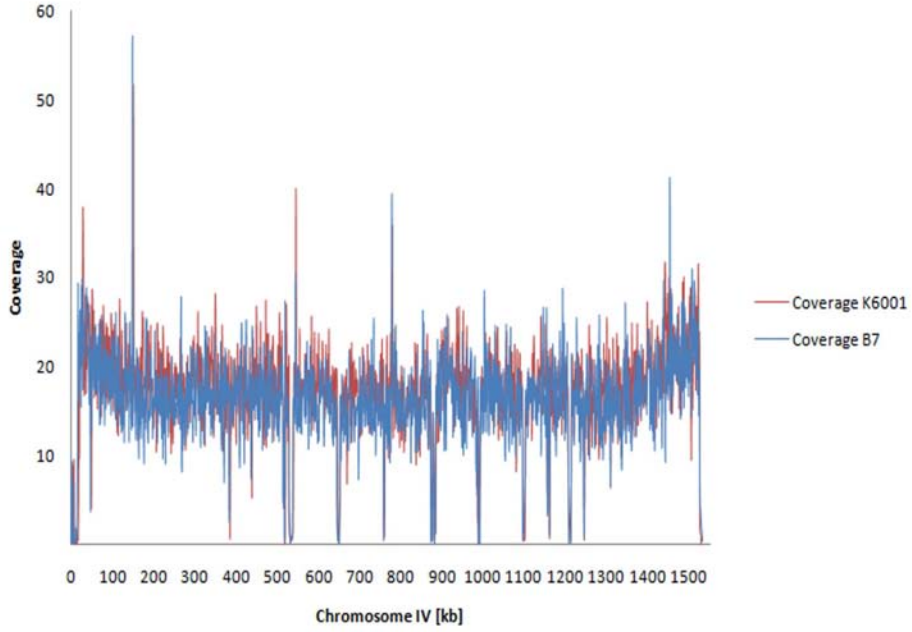


Figure A.23: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 4

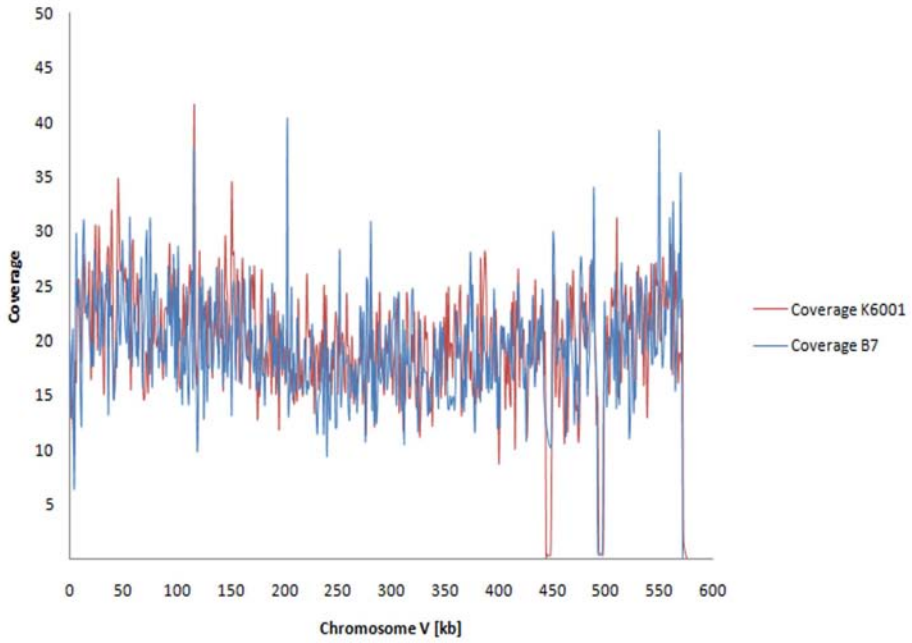


Figure A.24: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 5

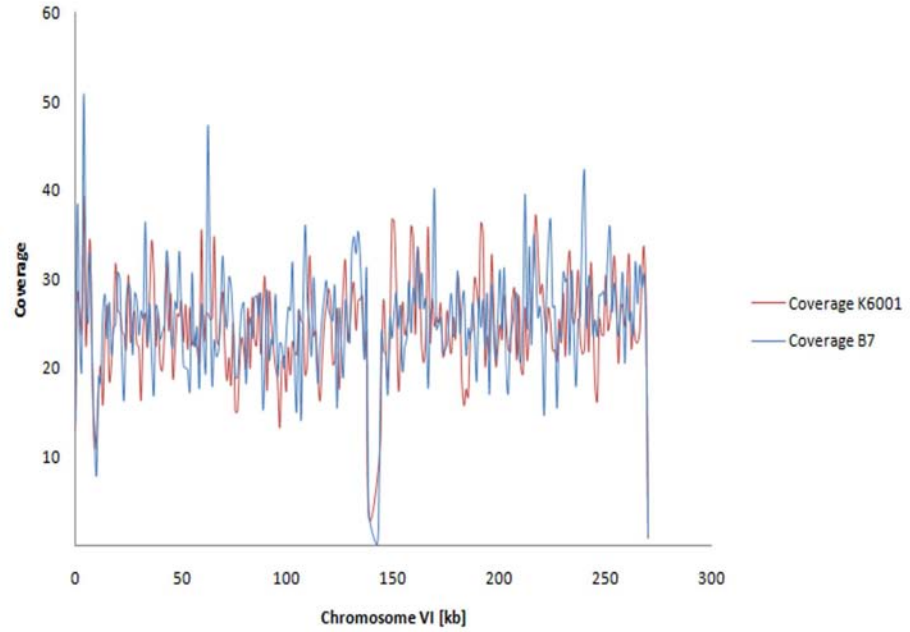


Figure A.25: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 6

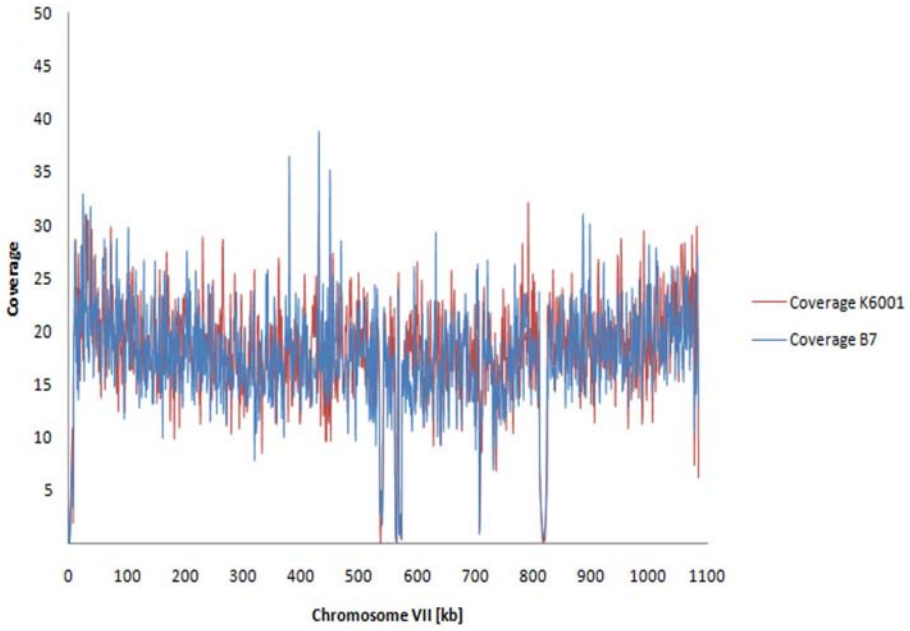


Figure A.26: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 7

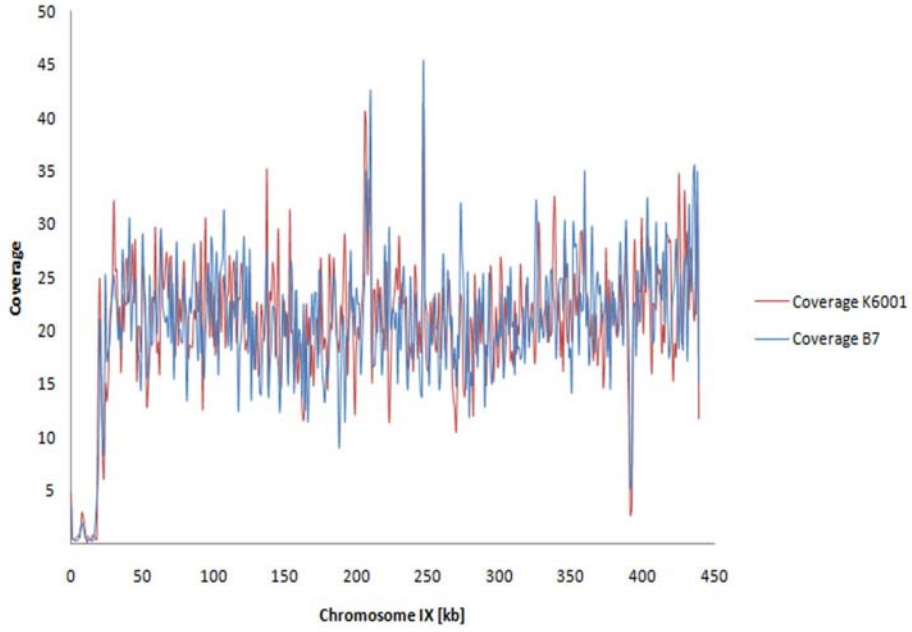


Figure A.27: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 9

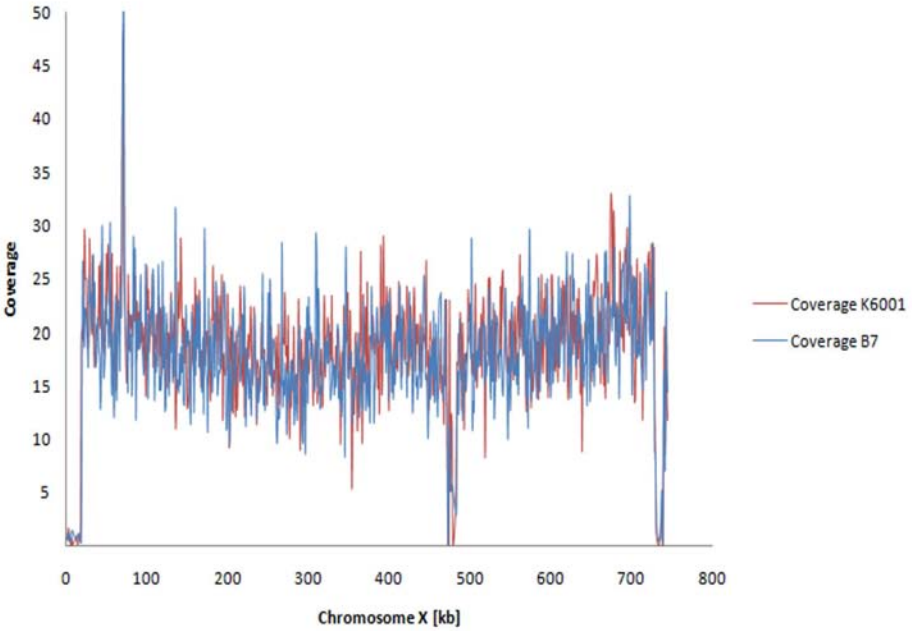


Figure A.28: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 10

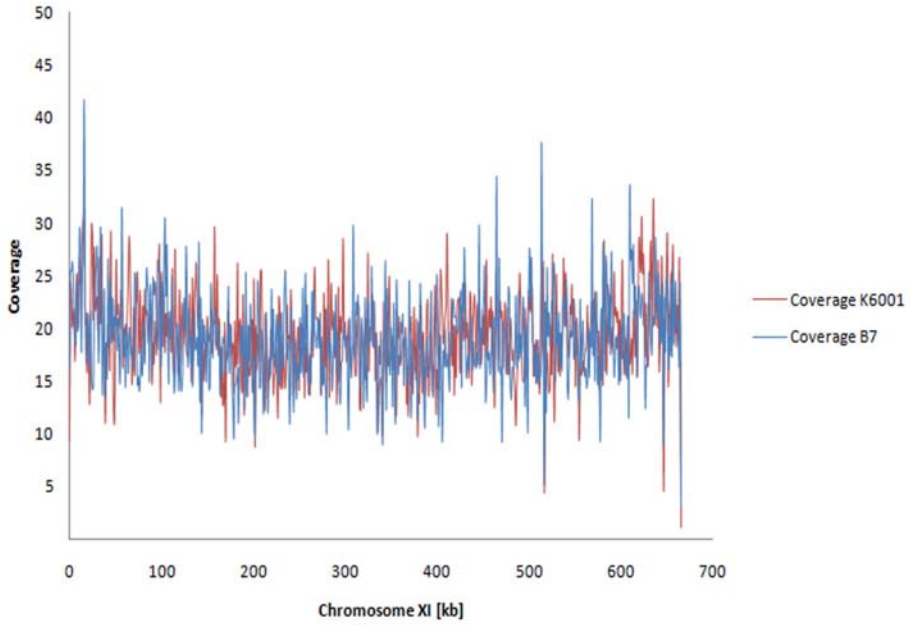


Figure A.29: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 11

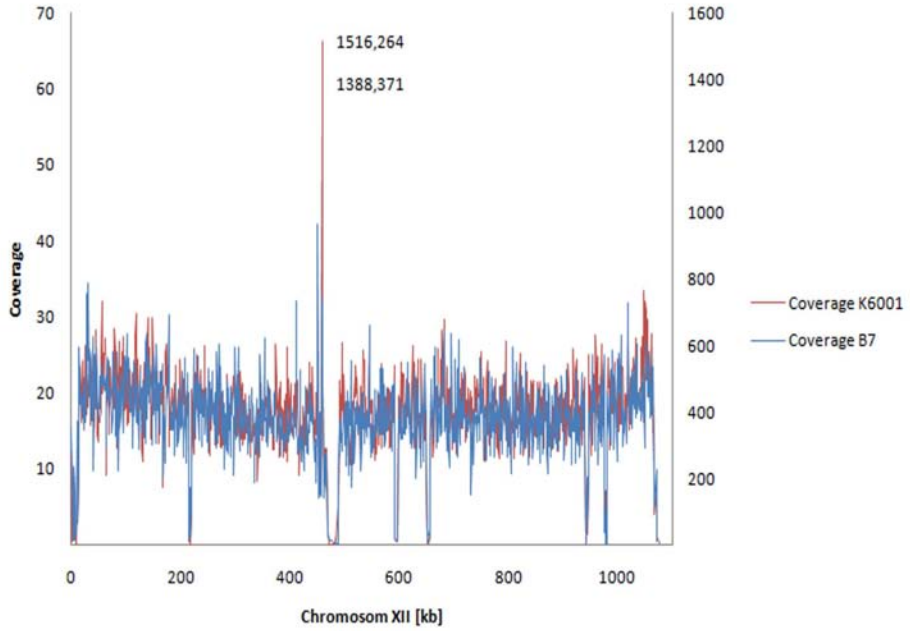


Figure A.30: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 12

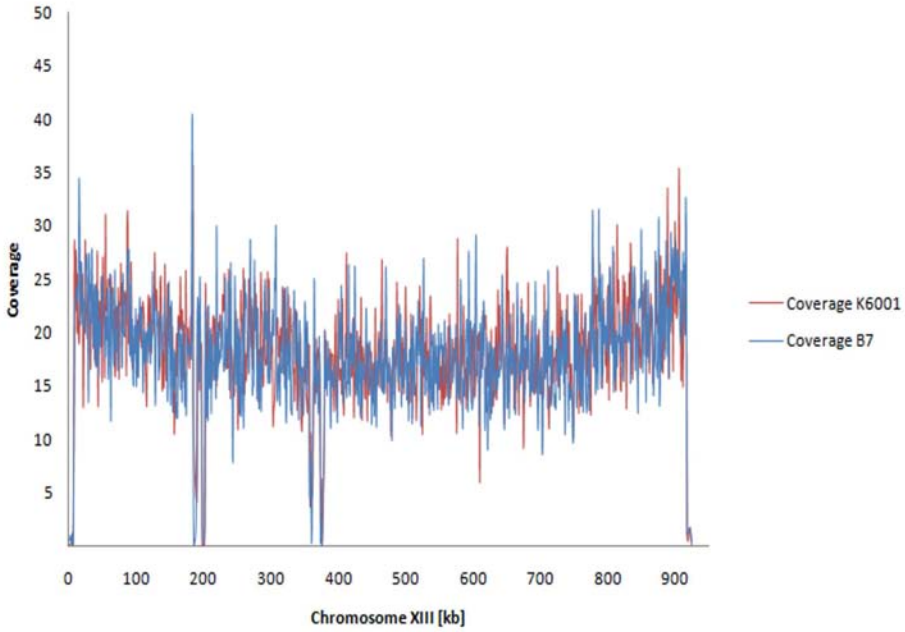


Figure A.31: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 13

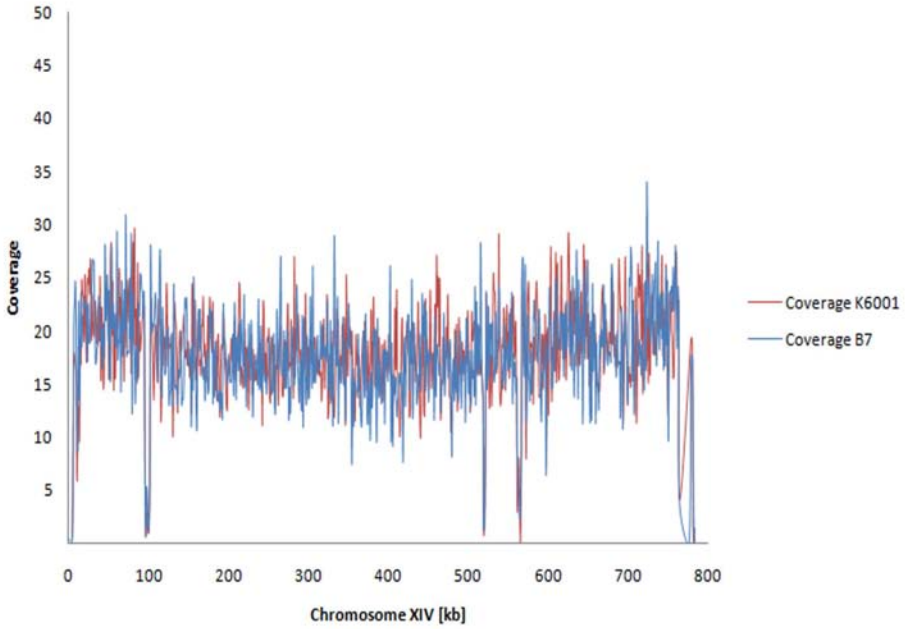


Figure A.32: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 14

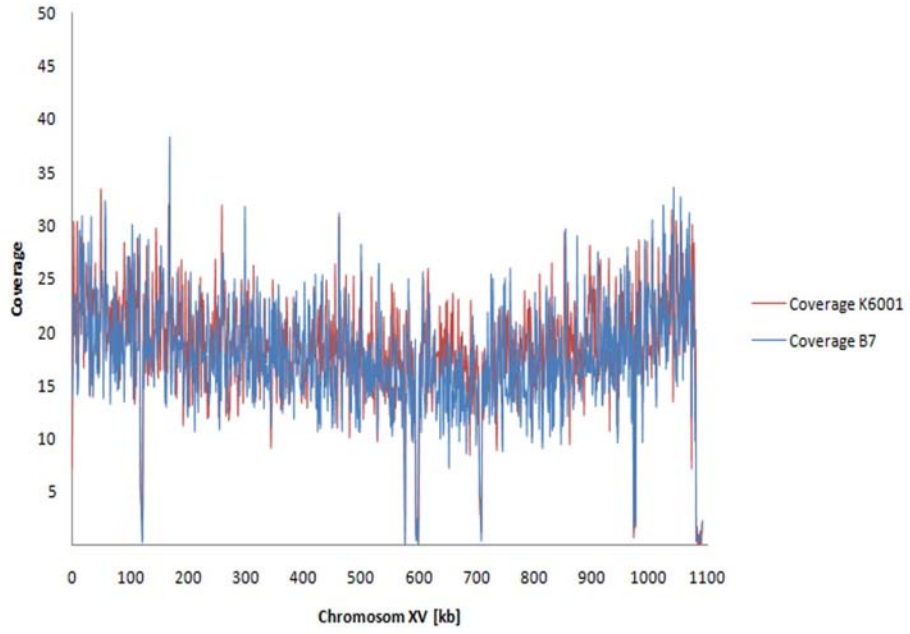


Figure A.33: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 15

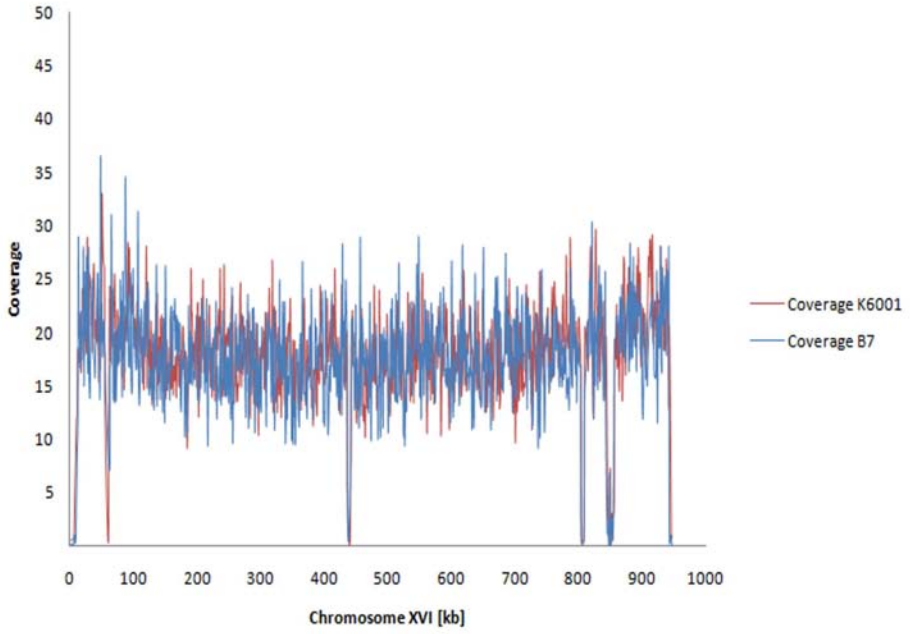


Figure A.34: 454 Life/Sciences Roche Sequencing Coverage Uniformity Chromosome 16

A.4 Illumina/Solexa Sequencing Coverage Uniformities

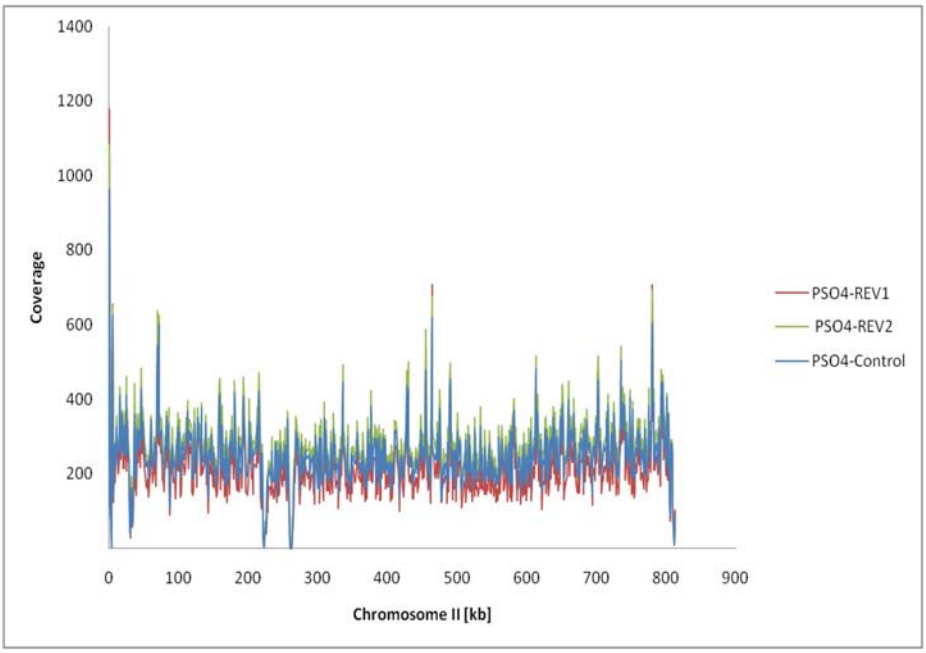


Figure A.35: Illumina/Solexa Sequencing Coverage Chromosome 2

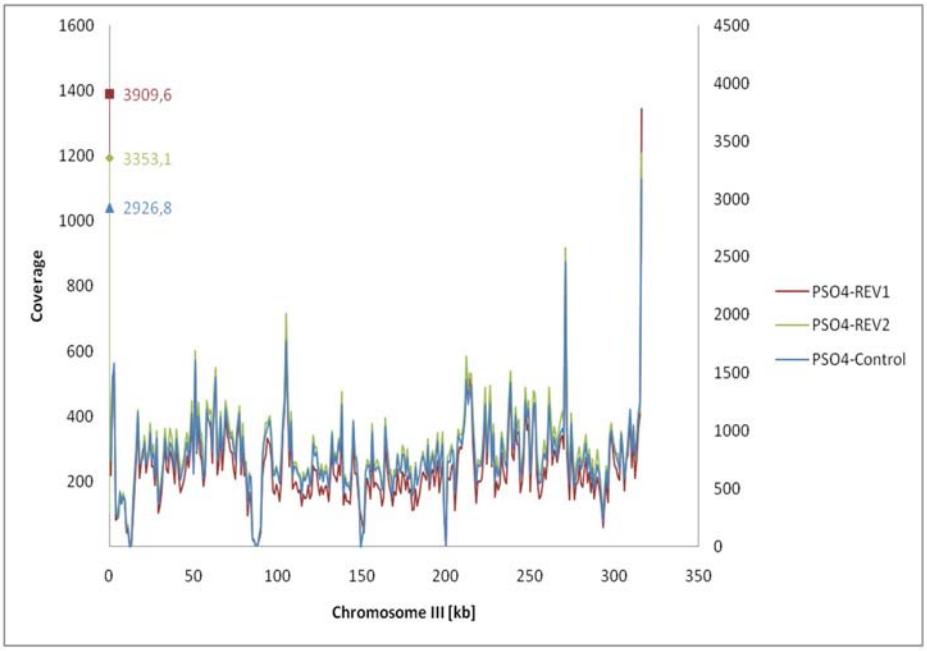


Figure A.36: Illumina/Solexa Sequencing Coverage Chromosome 3

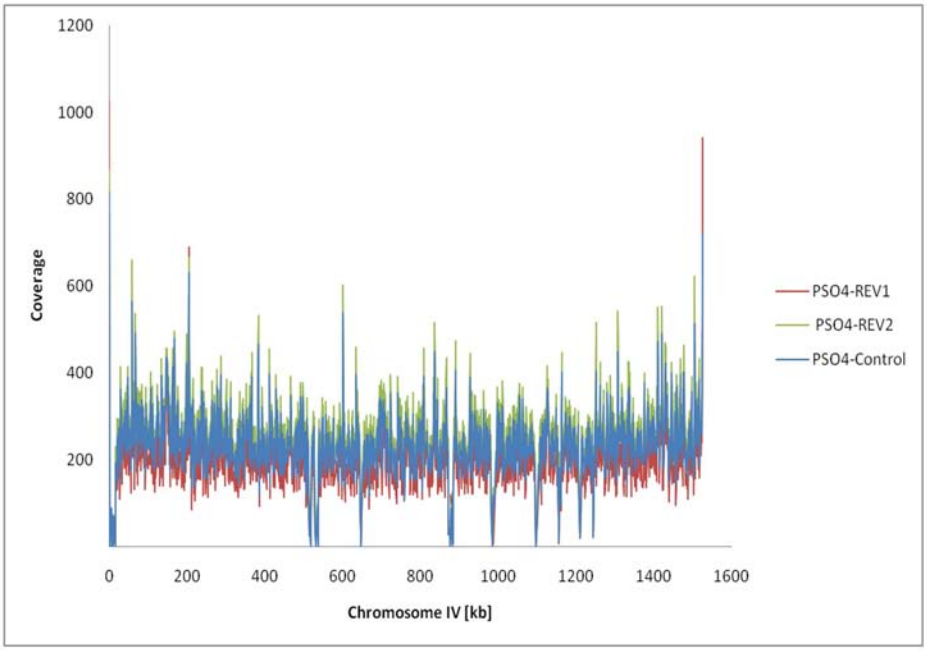


Figure A.37: Illumina/Solexa Sequencing Coverage Chromosome 4

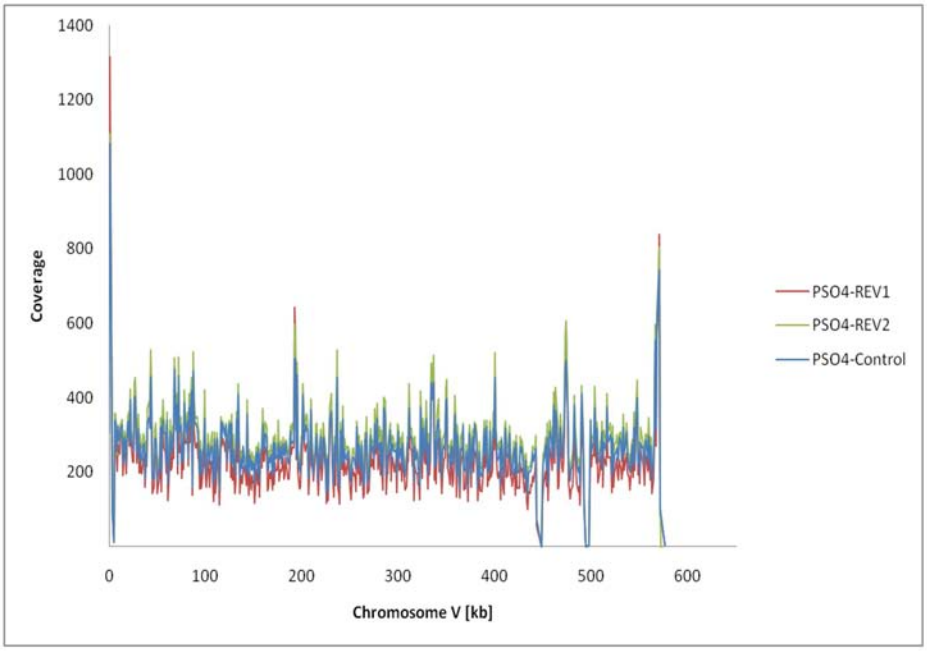


Figure A.38: Illumina/Solexa Sequencing Coverage Chromosome 5

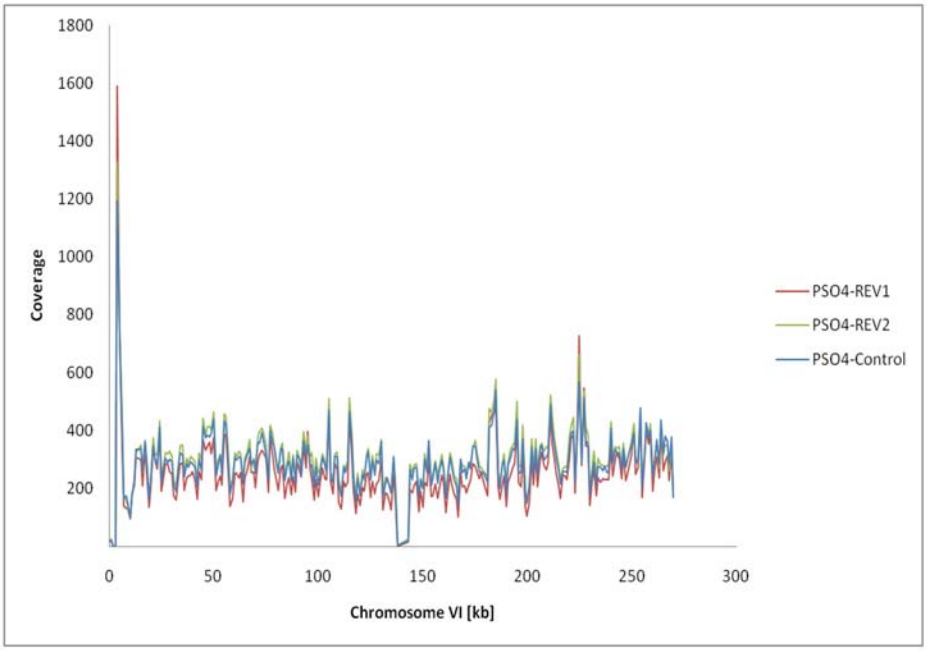


Figure A.39: Illumina/Solexa Sequencing Coverage Chromosome 6

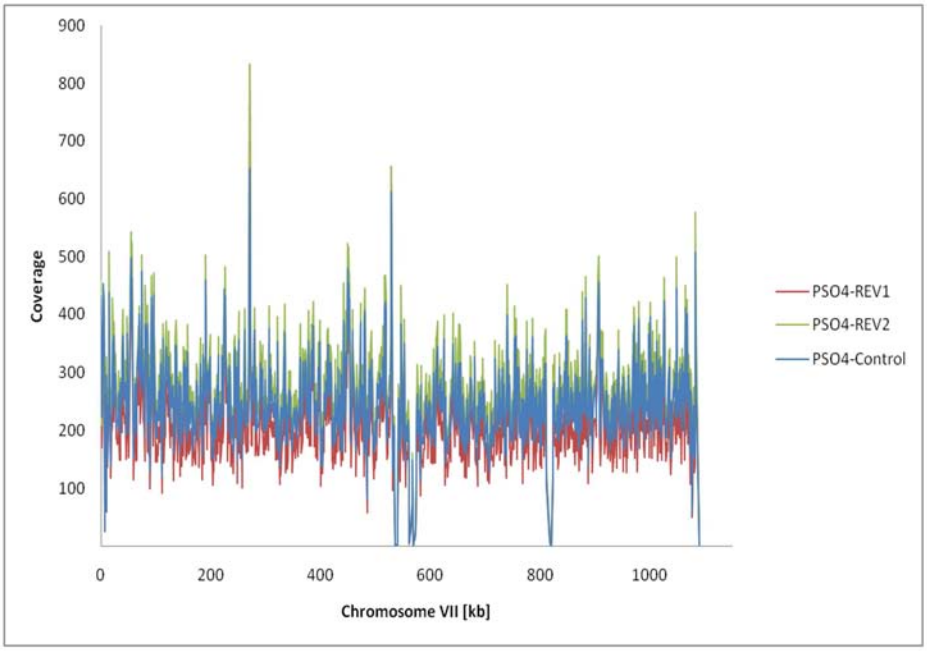


Figure A.40: Illumina/Solexa Sequencing Coverage Chromosome 7

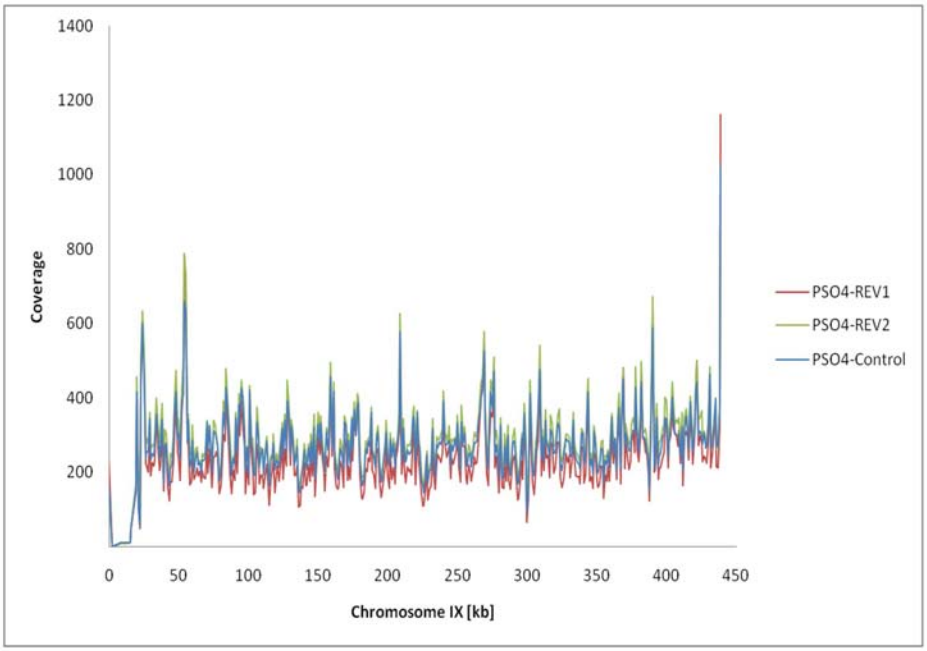


Figure A.41: Illumina/Solexa Sequencing Coverage Chromosome 9

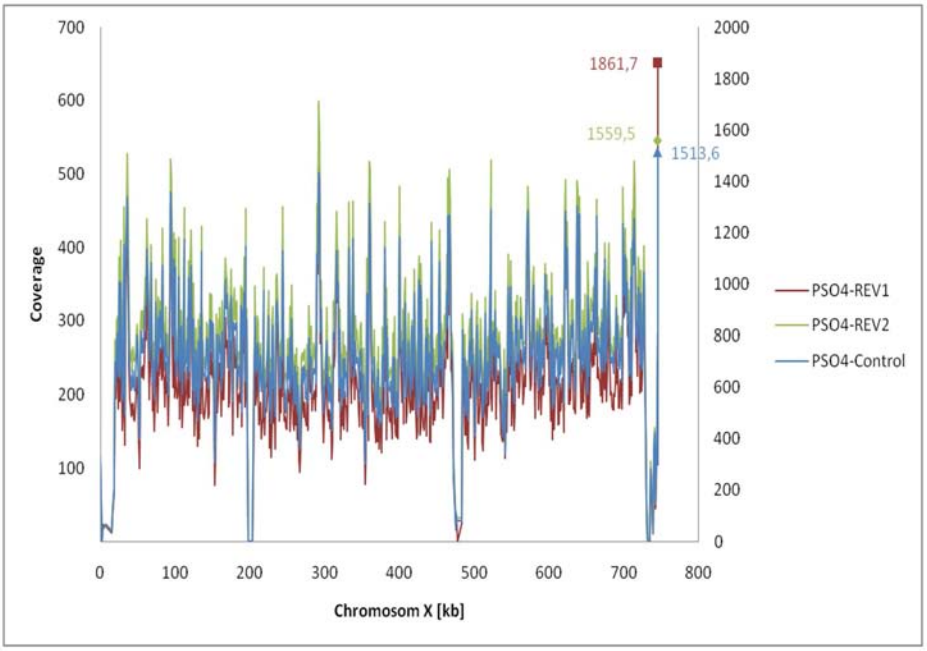


Figure A.42: Illumina/Solexa Sequencing Coverage Chromosome 10

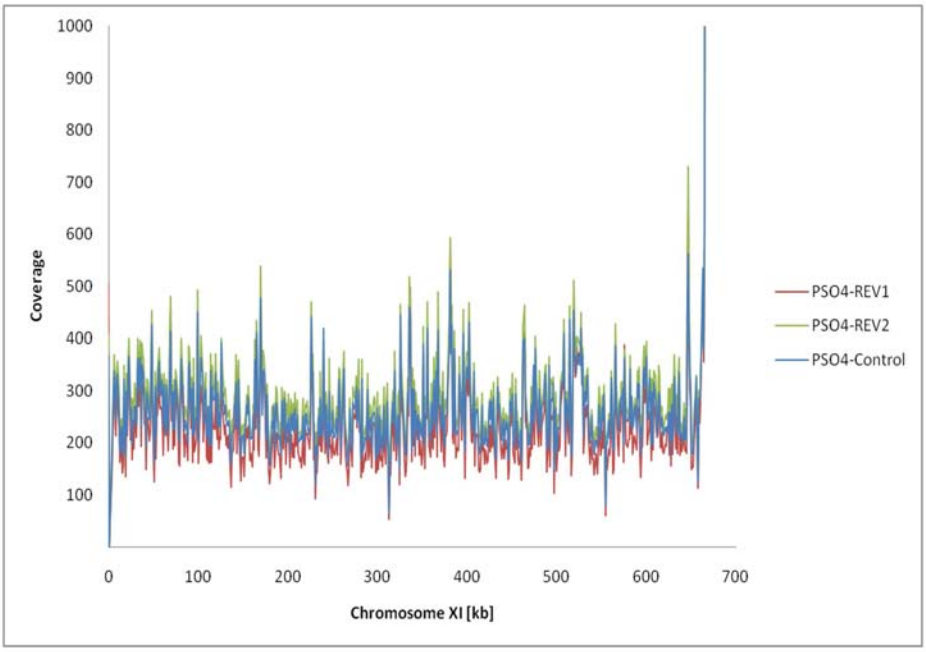


Figure A.43: Illumina/Solexa Sequencing Coverage Chromosome 11

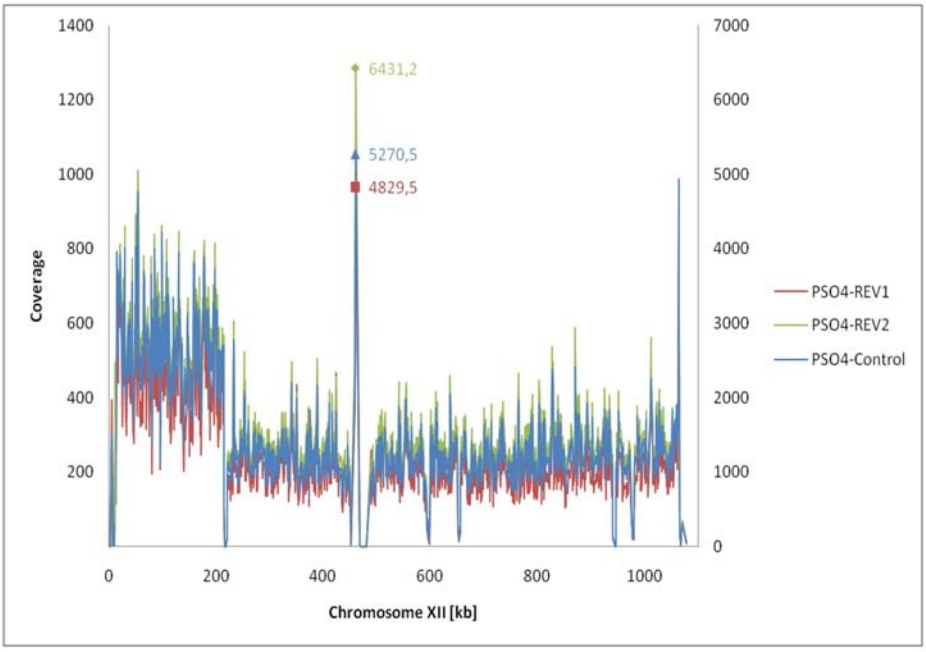


Figure A.44: Illumina/Solexa Sequencing Coverage Chromosome 12

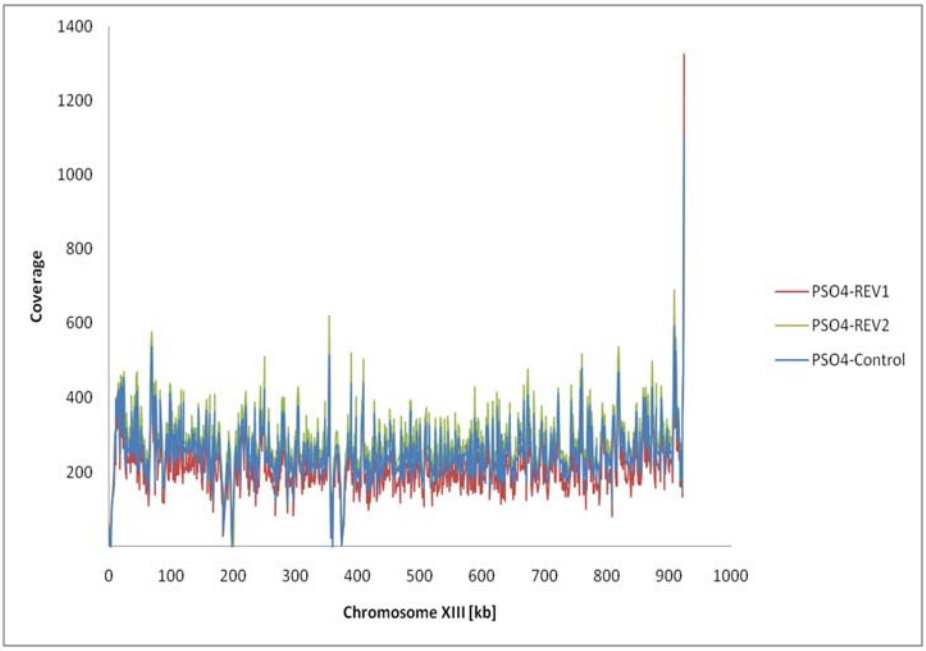


Figure A.45: Illumina/Solexa Sequencing Coverage Chromosome 13

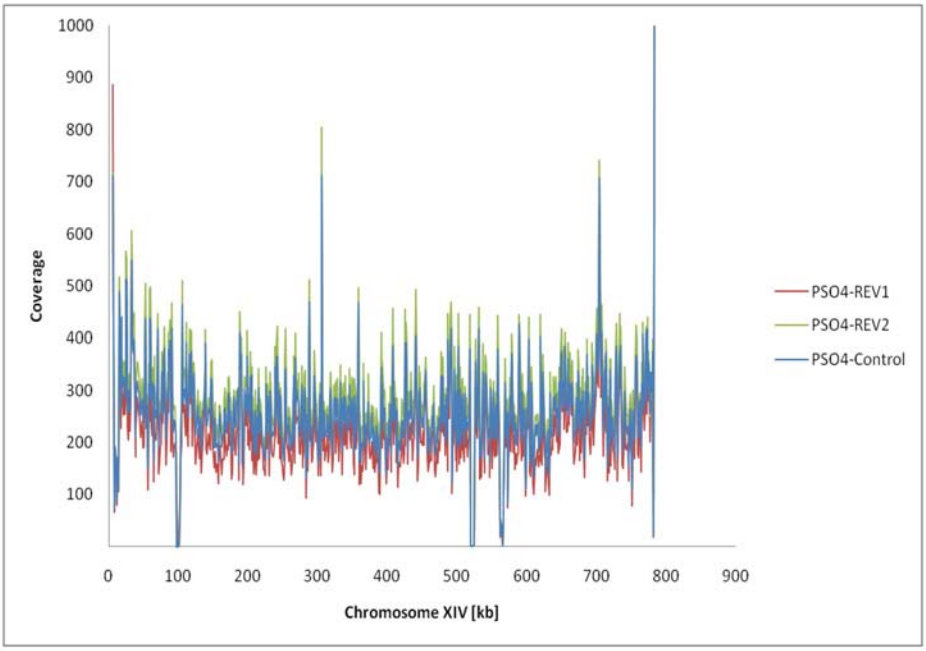


Figure A.46: Illumina/Solexa Sequencing Coverage Chromosome 14

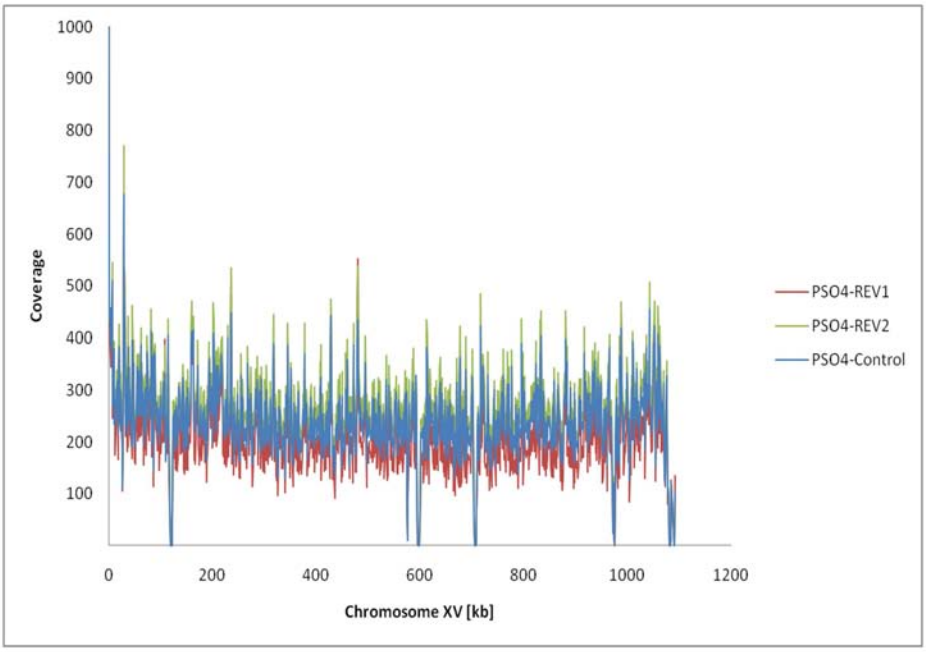


Figure A.47: Illumina/Solexa Sequencing Coverage Chromosome 15

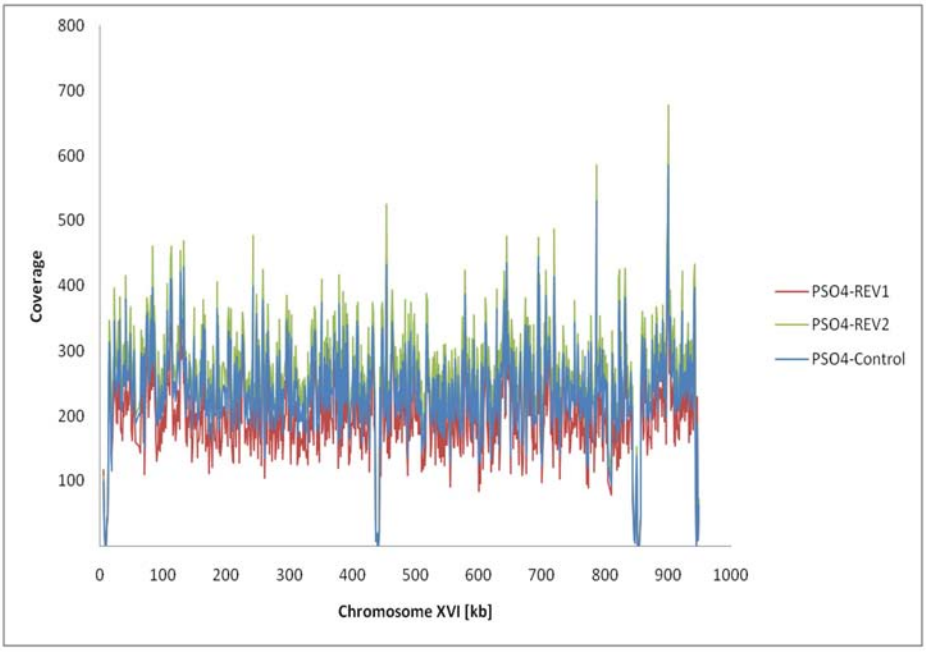


Figure A.48: Illumina/Solexa Sequencing Coverage Chromosome 16

A.5 Perl Scripts

A.5.1 Filtering of unequal differences

```
#!/perl/bin/perl -w

open (DATEI, "B7.txt") || die("Programmabbruch:
Konnte die Datei b7.txt nicht öffnen");
open (DATEI1, "k6001.txt") || die("Programmabbruch:
Konnte die Datei k6001.txt nicht öffnen");
open (OUT1, ">gleichinbeidenstaemmen.txt") || die("Programmabbruch:
Konnte die Datei gleichinbeidenstaemmen.txt nicht öffnen");
open (OUT2, ">differentenb7k6001.txt") || die("Programmabbruch:
Konnte die Datei differentenb7k6001.txt nicht öffnen");

my(%b7,%k6001,$chr,$chr2,$pos2,$nuc2,$pos,$nuc,$chr1,$pos1,$nuc1,%allepos
,$chr3,$pos3,@l);
while (<DATEI>) {
chomp;
@l=split("\t",$_);
$l[0] =~ /^>(\S+)/;
$chr2 = $1;
$pos2=$l[1];
$nuc2=$l[3];
$b7{$chr2}{$pos2}= $nuc2;
$allepos{$chr2}{$pos2}= 1;
}
for my $chromosomen1 (keys %b7) {
print "$chromosomen1\n";
 for my $position1 (keys %{$b7{$chromosomen1}}){
print "$position1\n";
}
}
my @k;
while (<DATEI1>) {
chomp;
@k=split("\t",$_);
my $wert2 = $k[1];
$k[0] =~ /^>(\S+)/;
$chr1 = $1;
```

```
$pos1=$k[1];
$nuc1=$k[3];
$k6001{$chr1}{$pos1}= $nuc1;
$allepos{$chr1}{$pos1}= 1;
}
close (DATEI);
close (DATEI1);
for my $chromosomen (keys %k6001) {
print "$chromosomen\n";
 for my $position (keys %{$k6001{$chromosomen}}){
print "$position\n";
}
}
for my $chromosomen3 (  keys %allepos) {
print "$chromosomen3\n";
 for my $position3 (  keys %{$allepos{$chromosomen3}}){
print "$position3\n";
}
}
foreach $chr ( keys %allepos){

     foreach $pos ( keys %{$allepos{$chr}}){

    if(exists $b7{$chr}{$pos} && exists $k6001{$chr}{$pos}){

        if($b7{$chr}{$pos} eq $k6001{$chr}{$pos}){
        print OUT1 "$chr\t$pos\t uninterssant\n";
        }else{
        print OUT2 " $chr\t$pos\t interessant\n";
        }

    }elsif(exists $b7{$chr}{$pos}){
        print OUT2 " $chr\t$pos\t schwierig,weil keine Kontrolle\n";
    }
     }
}
close (DATEI);
close (DATEI1);
close (OUT1);
close (OUT2);
```

A.5.2 Generation of Coverage file (Illumina)

```
#!/perl/bin/perl -w

use strict;

open (DATEI, "uniquereads.txt") || die("Programmabbruch:
Konnte die Datei uniquereads.txt nicht öffnen");
open (OUT1, ">coveragegenome.txt") || die("Programmabbruch:
Konnte die Datei coveragegenome.txt nicht öffnen");

my(@l,%position,$end,$start,$pos, $nucleotide,$posit,$chromosom,$chr,$nuc,$anzahl);

while (<DATEI>) {
chomp;
my @l;
@l=split("\t",$_);
$chr=$l[2];
$start=$l[3];
$anzahl=(length($l[9])-1);
$end= $start + $anzahl;
for($pos=$start;$pos<=$end;$pos++){
$position{$chr}{$pos} += 1;
}
}
foreach my $chromosom (keys %position) {
foreach my $posit(sort {$a <=> $b} keys %{$position{$chromosom}}){
print OUT1 "$chromosom\t$posit\t$position{$chromosom}{$posit}\n";
}
}
close (DATEI);
close (OUT1);
```

A.5.3 Splitting up covered positions in 1000 bp bins

```
#!/usr/bin/perl -w


open (DATEI, "eingangsfile.txt") || die("Programmabbruch:
Konnte die Datei eingangsfile.txt nicht oeffnen");
open (OUT, ">bins.txt") || die("Programmabbruch:
Konnte die Datei bins.txt nicht");

my(%hash,$pos,$chr,$binid,);


while(<DATEI>) {

@l= split("\t",$_);
$chr=$l[0];
$pos=$l[1];
$cov=$l[2];
$binid=int($pos/1000);
$hash{$chr}{$binid} += $cov;
}

foreach $chr (keys%hash){
foreach $bin (sort {$a <=> $b}keys%{$hash{$chr}}){
print OUT "$chr\t$bin\t$hash{$chr}{$bin}\n";


close (DATEI);
close (OUT);
```

A.5.4 Calculation of Variation frequency (Illumina)

```
#!/perl/bin/perl -w

use strict;

open (DATEI, "eingangsfile.text") || die("Programmabbruch:
Konnte die Datei eingangsfile.txt nicht öffnen");
open (OUT1, ">varfreq.text") || die("Programmabbruch:
Konnte die Datei varfreq.txt nicht öffnen");

my(@l,$cons,$count,$text,$textneu,$varfreq,$kleincons,$prozent);

while (<DATEI>) {
chomp;
@l=split("\t",$_);
$cons = $l[3];
$kleincons=lc($l[3]);
print"$kleincons\n";
print"$cons\n";
$text = "$l[8]";
$text =~ tr/[\$][\^]//d;
print "$text\n";
eval"\$text =~ tr/$cons$kleincons/1/";
$count=$text=~ tr/1/1/;
print"$count\n";
$varfreq = ($count / length($text))*100;
print OUT1 "$l[0]\t$l[1]\t$l[2]\t$l[3]\t$l[4]\t$l[5]\t$l[6]\t$l[7]\t$varfreq\n";
}
```

Bibliography

Ames, B. N., Shigenaga, M. K., , & Hagen, T. M. (1993). Oxidants, antioxidants, and the degenerative diseases of aging. *Proc. Natl. Acad. Sci. USA*, 90:7915–7922.

Benathen, A. & Beam, C. A. (1977). The genetic control of X-ray resistance of budding yeast cells. *Radiation Research*, 69:91–116.

Boyland, E. & Chasseaud, L. (1968). Enzymes catalysing conjugation of glutathione with alpha-beta-unsaturated carbonyl compounds. *Biochemistry Journal*, 109:651–661.

Breitenbach, M., Madeo, F., Laun, P., Heeren, G., Jarolim, S., Fröhlich, K.-U., Wissing, S., & Pichova, A. (2003). Yeast as a model for ageging and apoptosis. *Topics in Current Genetics*, 3.

Brendel, M. & Henriques, J. A. P. (2001). The *pso* mutants of *Saccharomyces cerevisiae* comprise two groups: One deficient in DNA repair and another with altered mutagen metabolism. *Mutation Research*, 489:79–96.

Cheng, W.-H., Kusumoto, R., Opresko, P. L., Sui, X., Huang, S., Nicolette, M. L., Paull, T. T., Campisi, J., Seidman, M., & Bohr, V. A. (2006). Collaboration of Werner syndrome protein and BRCA1 in cellular respones to DNA interstrand cross-links. *Nucleic Acids Research*, 34(9):2751–2760.

Cohen, H. & Sinclair, D. A. (2001). Recombination-mediated lengthening of terminal telomeric repeats requires the Sgs1 DNA helicase. *PNAS*, 98(6):3174–3179.

Dharmendra, K. S., Byungchan, A., & Bohr, V. A. (2009). Roles of RECQ helicases in recombination based DNA repar, genomic instability and aging. *Biogerontology*, 10:235–252.

Dhillon, K. K., Sidorova, J., Saintigny, Y., Poot, M., Gollahon, K., Rabinovitch, P. S., & Raymond J. Monnat, J. (2007). Functional role of the Werner syndrome RecQ helicase in human fibroblasts. *Aging Cell*, 6:53–61.

Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L., & Rothstein, R. (1994). The yeast type i topoisomerase top3 interacts with sgs1, a DNA helicase homolog: A potential eukaryotic reverse gyrase. *Molecular and Cellular Biology*, 14:8391–8398.

Goffeau, A., Barell, B. G., H.Bussey, Davis, R. W., Dujon, B., Galibert, H. F. F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science*, 274:546–567.

Gorbunova, V. & Seluanov, A. (2004). Making ends meet in old age: DSB repair and aging. *Mechanisms of Ageing and Development*, 126:621–628.

Grey, M., Düsterhöft, A., Henriques, J. A. P., & Brendel, M. (1996). Allelism of *PSO4* and *PRP19* links pre-mrna processing with recombination and error-prone DNA repair in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 24(20):4009–4014.

Grillari, J., Ajuh, P., Stadler, G., Löscher, M., Voglauer, R., Ernst, W., Chusainow, J., Eisenhaber, F., Pokar, M., Fortschegger, K., Grey, M., Lamond, A. I., & Katinger, H. (2005). SNEV is an evolutionarily conserved splicing factor whose oligomerisation is necessary for spliceosome assembly. *Nucleic Acid Research*, 33(21):6868–6883.

Grillari, J., Kartinger, H., & Voglauer, R. (2007). Contributions of DNA interstrand cross-links to aging of cells and organisms. *Nucleic Acids Research*, 35(22):7566–7576.

Hager, M. (2009). Aspects of the roles of oxidative stress and DNA repair in the aging process of *Saccharomyces cerevisiae*.

Harman, D. (1956). Role of free Radicals in Aging and Disease. *ANNALS NEW YORK ACADEMY OF SCIENCES*.

Harrigan, J. A., Wilson, D. M., Prasad, R., Opresko, P. L., Beck, G., May, A., Wilson, S. H., & Bohr, V. A. (2006). The Werner syndrome protein operates in base excision repair and cooperates with DNA polymerase b. *Nucleic Acids Research*, 34(2):745–754.

Henriques, J. A. P. & Moustacchi, E. (1980). Isolation and characterization of *pso* mutants sensitive to photo-addition of psoralen derivatives in *Saccharomyces cerevisiae*. *Genetics*, 95:273–288.

Jackson, S. P. (2002). Sensing and repairing DNA double-strand breaks. *Carcinogenesis*, 23(5):687–696.

Jarolim, S., Millen, J., Heeren, G., Laun, P., Goldfarb, D., & Breitenbach, M. (2004). A novel assay for replicative lifespan in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, 5:169–177.

Johnson, F. B., Sinclair, D. A., & Guarente, L. (1999). Molecular Biology of Aging. *Cell*, 96:291–302.

Krokan, H. E., Standal, R., & Slupphaug, G. (1997). DNA glycosylases in the base excision repair of DNA. *Biochemistry Journal*, 325:1–16.

Lehoczky, P., McHugh, P. J., & Chovanec, M. (2007). DNA interstrand cross-link repair in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev.*, 31:109–133.

Li, G.-M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Research*, 18:85–98.

Löscher, M., Fortschegger, K., Ritter, G., Wostry, M., Voglauer, R., Schmid, J. A., Watters, S., Rivett, A. J., Ajuh, P., Lamond, A. I., Katinger, H., & Grillari, J. (2005). Interaction of U-box E3 ligase SNEV with PSMB4, the ß7 subunit of the 20 S proteasome. *Biochemistry Journal*, 388:593–603.

Lundin, C., North, M., Erixon, K., Walters, K., Jenssen, D., Goldman, A., & Helleday, T. (2005). Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. *Nucleic Acids Research*, 33:3799–3811.

Mardis, E. R. (2008). Next-generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.*, 9:387–402.

Monnat, R. J. (2010). Human RECQ-helicases: Roles in DNA metabolism, mutagenesis and cancer biology. *Cancer Biology*.

Nianxiang, Z., Ramandeep, K., Xiaoyan, L., Xi, S., Li, L., & Legerski, R. J. (2005). The Pso4 mRNA splicing and DNA repair complex interacts with WRN for processing of DNA interstrand cross-links. *The Journal of biological Chemistry*, 280(49):40559–40567.

Ohi, M. D., Kooi, C. W. V., Rosenberg, J. A., Ren, L., Hirsch, J. P., Chazin, W. J., Walz, T., & Gould, K. L. (2005). Structural and functional analysis of essential pre-mRNA splicing factor Prp19p. *Molecular and Cellular Biology*, 25(1):451–460.

Opresko, P. L., Cheng, W.-H., von Kobbe, C., Harrigan, J. A., & Bohr, V. A. (2003). Werner syndrome and the function of the Werner protein; what they can teach us about the molecular aging process. *Carcinogenesis*, 24(5):791 802.

Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., & Gygi, S. P. (2003). A proteonomic approach to understanding protein ubiquitination. *Nature Biotechnolgy*, 21(8):921–926.

Seeberg, E., Eide, L., & Bjoras (1995). The base excisions repair. *TIBS*, 20:391–397.

Sharma, S., Doherty, K. M., & Rober M. Brosh, J. (2006). Mechanism of RecQ helicases in pathways of DNA metabolism and maintenance of genetic stability. *Biochemistry Journal*, 398:319–337.

Sies, H. (1999). Glutathione and its role in cellular functions. *Free Radicals and Medicine*, 27:916–921.

Sinclair, D. A. & Guarente, L. (1997). Extrachromosomal rDNA circles-A cause of Aging in Yeast. *Cell*, 91:1033–1042.

Spence, J., Gali, R. R., Dittmar, G., Sherman, F., Karin, M., & Finley, D. (2000). Cell Cycleregulated Modification of the Ribosome by a Variant Multiubiquitin Chain. *Cell*, 102:67–76.

Sung, P. & Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature Reviews/Molecular Cell Biology*, 7:739–750.

Timmermann, B., Jarolim, S., Rußmayer, H., Kerick, M., Michel, S., Krüger, A., Bluemlein, K., Laun, P., Grillari, J., Lehrach, H., & Ralser, M. (2010). A new dominant peroxiredoxin allele identified by whole-genome re-sequencing of random mutagenized yeast causes oxidant-resistance and premature aging. *Aging*, 2:375–386.

Voglauer, R., Chang, M., Dampier, B., Baumann, M. W. K., Hofer, D., Sterovsky, T., Schreiber, M., Kartinger, H., & Grillari, J. (2006). SNEV overexpression extends the life span of human endothelial cells. *Exp Cell Res*, 312:746–759.

von Hoffen, A., Balajee, A., von Zeeland, A. A., & Mullenders, L. H. (2003). Nucleotide excision repair and its interplay with transcription. *Toxicology*, 193:79–90.

Research Paper

A new dominant peroxiredoxin allele identified by whole-genome re-sequencing of random mutagenized yeast causes oxidant-resistance and premature aging

Bernd Timmermann[1]*, Stefanie Jarolim[2]*, Hannes Rußmayer[1,4], Martin Kerick[3], Steve Michel[3], Antje Krüger[3], Katharina Bluemlein[3], Peter Laun[2], Johannes Grillari[4], Hans Lehrach[3], Michael Breitenbach [2] and Markus Ralser[3]

[1] Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[2] Department of Cell Biology, University of Salzburg, 5020 Salzburg, Austria
[3] Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[4] Institute of Applied Microbiology, University of Natural Resources and Applied Life Sciences, 1180 Vienna, Austria
** These authors contributed equally to this work*

Key words: *Aging, whole genome resequencing, redox homeostasis, peroxiredoxin*
Received: *07/28/10; accepted: 08/12/10; published on line: 08/13/10*
Corresponding author: *Markus Ralser, PhD;* ***E-mail:*** *ralser@molgen.mpg.de*

Abstract: The combination of functional genomics with next generation sequencing facilitates new experimental strategies for addressing complex biological phenomena. Here, we report the identification of a gain-of-function allele of peroxiredoxin (thioredoxin peroxidase, Tsa1p) via whole-genome re-sequencing of a dominant *Saccharomyces cerevisiae* mutant obtained by chemical mutagenesis. Yeast strain K6001, a screening system for lifespan phenotypes, was treated with ethyl methanesulfonate (EMS). We isolated an oxidative stress-resistant mutant (B7) which transmitted this phenotype in a background-independent, monogenic and dominant way. By massive parallel pyrosequencing, we generated an 38.8 fold whole-genome coverage of the strains, which differed in 12,482 positions from the reference (S288c) genome. Via a subtraction strategy, we could narrow this number to 13 total and 4 missense nucleotide variations that were specific for the mutant. Via expression in wild type backgrounds, we show that one of these mutations, exchanging a residue in the peroxiredoxin Tsa1p, was responsible for the mutant phenotype causing background-independent dominant oxidative stress-resistance. These effects were not provoked by altered Tsa1p levels, nor could they be simulated by deletion, haploinsufficiency or over-expression of the wild-type allele. Furthermore, via both a mother-enrichment technique and a micromanipulation assay, we found a robust premature aging phenotype of this oxidant-resistant strain. Thus, *TSA1-B7* encodes for a novel dominant form of peroxiredoxin, and establishes a new connection between oxidative stress and aging. In addition, this study shows that the re-sequencing of entire genomes is becoming a promising alternative for the identification of functional alleles in approaches of classic molecular genetics.

INTRODUCTION

The free radical theory of aging implies that oxidative stress, and the generation of free radicals, are causally involved in the process of aging [1-3]. This theory is supported by many observations, including that yeast mother cells retain oxidatively damaged macromolecules, whereas the daughter cells are formed from a juvenile set of proteins [4], or inherit functional enzymes whereas the damaged species are retained with

the mother [5]. However, other discoveries challenge a causal implication of free radicals in different aging phenotypes [6]. Although most long-living mutants of *S. cerevisiae* and *C. elegans* tolerate high doses of oxidants, not all oxidant-resistant mutants are long living. As an example, both deletion of the metabolic regulator *AFO1* [7] and reduced activity of the metabolic enzyme *TPI1* [8] increase oxidant tolerances of yeast, but Δ*afo1* cells are massively long-living whereas TPI-deficient cells have a strong premature aging phenotype [7, 8]. In drosophila, mitochondrial ROS production correlates with aging but conversely, is not sufficient to alter lifespan [9]. All these observations are further complicated by the fact that mutants which are long-living under one environment/condition do not necessarily show this phenotype under other circumstances, i.e. yeast mutants with prolonged survival at 4°C are not enriched for mutants that are long-living at a higher temperatures [10]. In addition, there is a close relation of growth rate, metabolic activity and aging [11]. Since metabolism is itself a primary source for free radicals within the cell, it is difficult to distinguish between consequence and causality of oxidative damage during the aging process [6, 11]. Thus, although oxidative stress and free radicals are important players, their exact role during aging and the complex interplay of the involved genetic and biochemical components has yet to be clarified.

Systematic functional genetics/genomics is powerful in the identification of genetic components of biological processes and their interactions. With the introduction of systematic genetic libraries, such as an entire knock-out library of the yeast *S. cerevisiae* one decade ago [12], a new era of functional genetics began. Screening of systematic libraries allows circumventing the most time-consuming and limiting step of experimental genetics, which is the identification of the functional mutation. Screens with the systematic libraries identified many components that influence yeast aging phenotypes [10, 13, 14]. However, all pre-made libraries have the disadvantage of being limited in the number of alleles they contain. In practical terms, genome-wide knock-out, over-expression and copy number variation libraries can be generated, but nothing such as genome-wide 'point-mutation' libraries which would allow the isolation of alleles with new functionality. Therefore, although highly efficient, screens with systematic genetic libraries miss functional connections that can be identified by the isolation of new alleles in random-mutagenesis screenings. Thus, there is demand for experimental strategies that increase the efficiency of these classic approaches.

Here, we provide a case where whole-genome re-sequencing led to the identification of a new and dominant peroxiredoxin allele that causes oxidative stress resistance and premature aging. The W303-derived yeast strain K6001 [15], a model system that allows the enrichment of yeast mother cells [16], was used in a random mutagenesis experiment to isolate a mutant that exhibited strong resistance to the chemical oxidant cumene hydroperoxide (CHP). The mutant was dominant and segregated this phenotype in a regular way and independent of genetic background, observations which pointed to a monogenic trait. Using whole-genome re-sequencing of both strains by the Roche/454 system, and comparison of their genomes, we identified the mutation as a single nucleotide exchange in the gene, *TSA1*, coding for thioredoxin peroxidase (peroxiredoxin).

RESULTS

Random mutagenesis and isolation of an oxidant-resistant K6001 mutant

Cell division of *S. cerevisiae* is asymmetric; mother and daughter cells are clearly recognizable by their size. Mother cells can complete a limited number of cell cycles. At the end of this so-called *replicative* lifespan, the cells persist for some further time in a non-dividing state, and finally die, predominantly by apoptosis [17]. The yeast strain K6001 has been used to study this phenotype, because it facilitates the selective enrichment of yeast mother cells. On glucose media, daughters of K6001 do not express the essential *CDC6* gene, and are therefore unable to complete their cell cycle. Thus, the biomass which a K6001 culture achieves in glucose media, is a function of the number of daughters per mother and indicates its average replicative lifespan [16].

To identify new genetic connections between aging and oxidative stress, we used K6001 in a random mutagenesis experiment. An exponential K6001 culture was mutagenized with ethyl methansulfonate (EMS) and grown for two generations under non-selective conditions to enable mutation fixation. Mutagenized cells were then plated onto agar that contained the synthetic oxidant cumene hydroperoxide (CHP) at a concentration lethal for the wild-type strain. For every million of cells plated, an average of three colonies gained resistance and grew on the CHP-containing media. Obtained stress resistant mutants were tested for their replicative lifespan. We are presenting here one mutant (K6001-B7) that displayed resistance/sensitivity to multiple oxidants and clear regular 2:2 segregation in genetic crosses. The isogenic mating partner, K6001α, was obtained by mating type switching of K6001 induced by a plasmid carrying the wild type HO gene.

The mutant was mated to K6001α and the diploids were sporulated. Tetrads gained by sporulation were tested for resistance to CHP, t-butyl hydroperoxide (t-BHT), diamide, and hydrogen peroxide. Figure 1 summarizes the oxidant phenotypes of K6001-B7, the diploid obtained in the outcross just described, and the haploid progeny derived thereof (one typical tetrad). The mutant strain K6001-B7, the diploids obtained by outcrossing it to wild type, and the haploid progeny all showed the same phenotype: strong resistance to CHP, increased resistance to t-BHT, sensitivity to hydrogen peroxide (Figure 1A), and normal growth in the presence of diamide, the trait was inherited in a monogenic way (2:2 segregation in tetrads; Figure 1B) and the mutant allele was dominant even in crosses with a non-related wild type strain, BY4742 (Figure 1C lower panel).

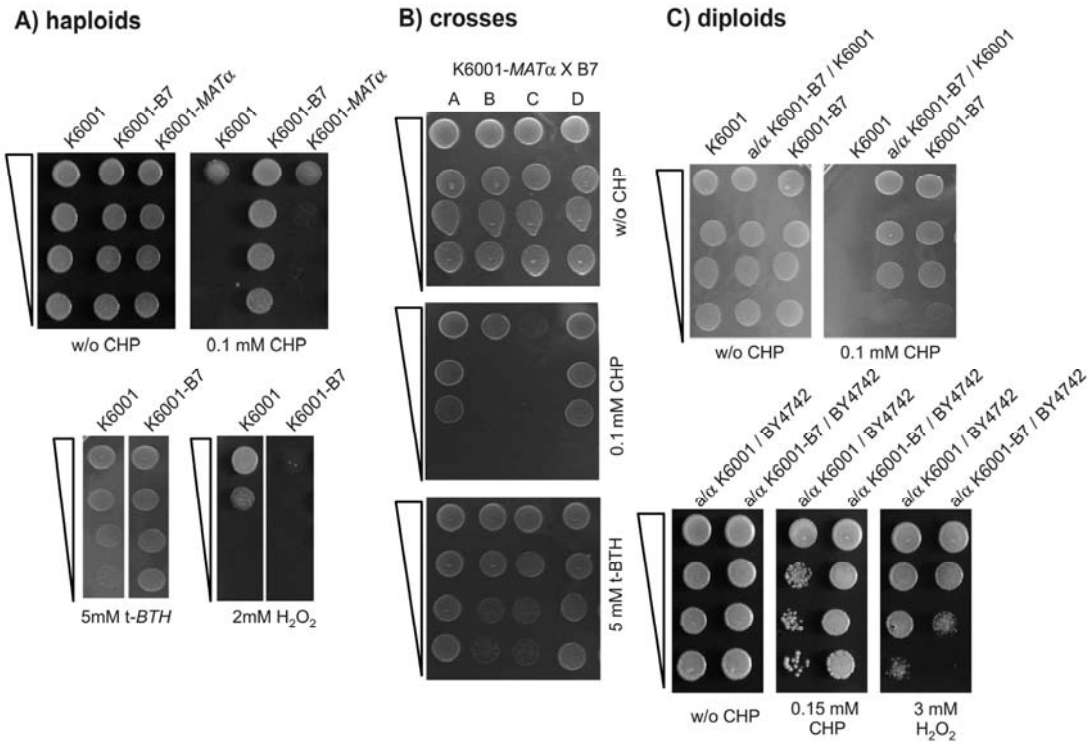
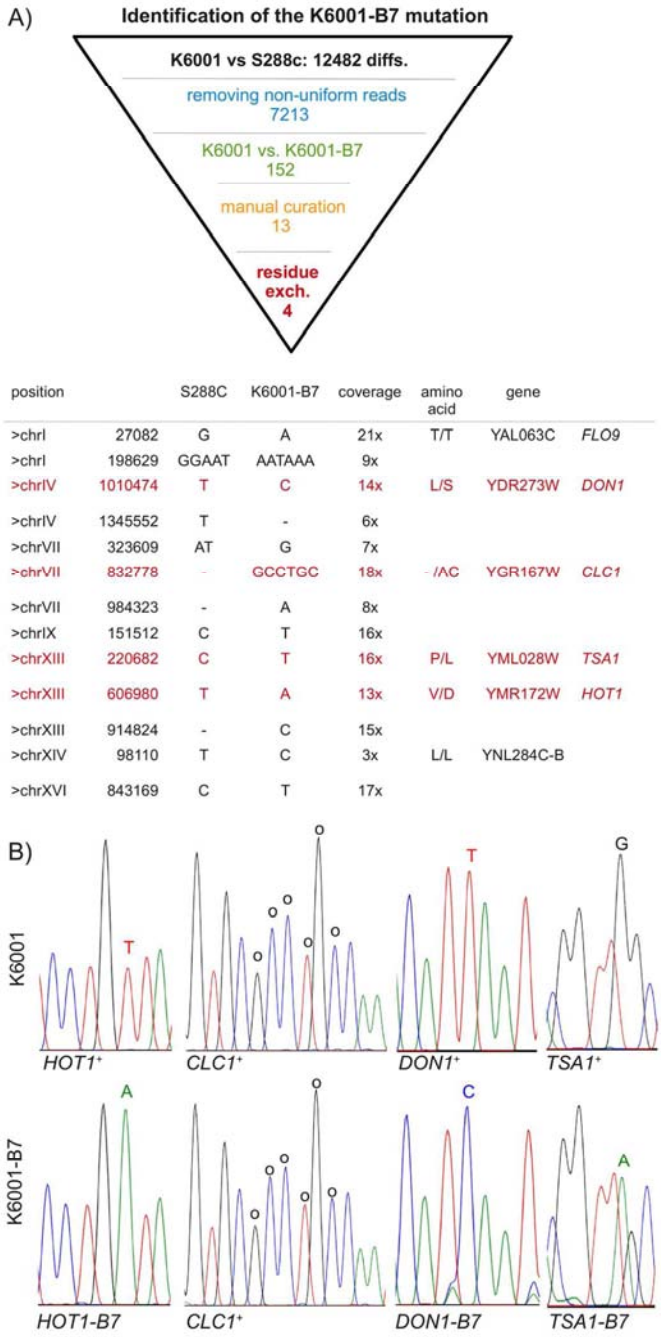


Figure 1. K6001-B7 carries a monogenic and dominant allele that confers resistance to cumene hydroperoxide (CHP). (**A**) *K6001-B7 is resistant to CHP.* Overnight cultures of the parent (K6001, *MAT*a), K6001-B7 and an isogenic *MAT*α (K6001-α32) strain were spotted as serial dilution on SC agar with and without CHP and grown at 28°C (upper panel). K6001-B7 grows on 0.1 mM CHP, a concentration lethal for K6001 and K6001-α32. (lower panel) Similar test were performed with oxidants tert-butyl hydroperoxide (t-*BTH)* (left) and H_2O_2 (right). (**B**) *K6001-B7 transmits with a monogenic Mendelian trait.* K6001-B7 was mated with K6001-α32, tetrads gained by sporulation and tested for CHP and t-BTH resistance. Shown is the 2:2 segregation of a representative tetrade. (**C**) *B7 is a dominant mutation.* (upper panel) K6001 and K6001-B7 *were* mated with K6001-α32, resultant diploids assayed for CHP resistance. +/B7 diploids retained the stress resistance of the B7 haploid. (lower panel) *B7 is dominant across backgrounds.* K6001 and K6001-B7 were mated with the distant *S. cerevisiae* strain BY4742. Similar to the pure K6001 background situation, BY/K6001-B7 diploids were more resistant to CHP (left) and sensitive to H_2O_2 (right) compared to the BY/K6001 diploids.

Identification of the K6001-B7 gene via whole-genome re-sequencing

Dominant mutations yielding stress-resistance are difficult to identify using classic yeast genetics; e.g. resistance phenotypes are commonly observed when anti-oxidant factors are over-expressed or over-active, leading to dominance of the mutant allele and limiting the possibilities and the specificity of complementation strategies with clone libraries that often over-express the inserted genes, even if they are centromeric. Furthermore, for yeast strains evolutionary distant from the reference genome of S288c, such as W303, the use of whole-genome-tiling arrays is limited due to potential unspecific hybridisation with primers designed for this distant genome.

Here, we decided on a whole-genome resequencing strategy using the Roche/454 platform to identify the B7 mutation. For this, we isolated genomic DNA from both K6001 and B7 and generated 454 sequencing libraries. Libraries were quality- controlled and sequenced by a Titanium sequencing kit (Roche). The run produced 662 MB of high-quality sequence at a median read-length of 527 bp (average read-length 499 bp). First, we aligned the sequence information to the S288c reference genome. 96.89 % of the coding and 96.83% of the noncoding regions were called at high quality in K6001 and 96,97%/96,98% in K6001-B7 (merged 97,36%/97,54%). Most of the non-called regions were neither sequenced in K6001 nor in K6001-B7, indicating that they were physically absent. (see Supplementary Figure 1 for an illustration of coverage uniformity). Based on this data, we calculated an average 19.2 fold coverage for K6001 and 19.5 fold coverage for K6001-B7. To compare the K6001/K6001-B7 genome with the S288c reference, we merged both sequence runs for an alignment, yielding an 38.8 fold total coverage. At this depth, we detected 12,482 SNVs and small (<50bp) insertions and deletions (that were sequenced at a minimum of 3x and on both strands) which distinguished K6001 from the reference genome. Surprisingly, although sequencing a haploid genome, we detected several mutations that had a calling rate between 5% and 95%. In a diploid genome, one would assume that these variants are heterozygotes. However, since we do not expect the existence of true heterozygotes in a haploid genome, we refer to these variants as presently unexplained non-uniform sequences. These variants were not entirely randomly distributed in the genome, many of these (62%) clustered close to the telomeric regions. This result could point to a natural variability of the genome in these regions. However, the result that the B7 phenotype was segregating in 2:2 manner allowed the



position		S288C	K6001-B7	coverage	amino acid	gene	
>chrI	27082	G	A	21x	T/T	YAL063C	*FLO9*
>chrI	198629	GGAAT	AATAAA	9x			
>chrIV	1010474	T	C	14x	L/S	YDR273W	*DON1*
>chrIV	1345552	T	-	6x			
>chrVII	323609	AT	G	7x			
>chrVII	832778	-	GCCTGC	18x	/AC	YGR167W	*CLC1*
>chrVII	984323	-	A	8x			
>chrIX	151512	C	T	16x			
>chrXIII	220682	C	T	16x	P/L	YML028W	*TSA1*
>chrXIII	606980	T	A	13x	V/D	YMR172W	*HOT1*
>chrXIII	914824	-	C	15x			
>chrXIV	98110	T	C	3x	L/L	YNL284C-B	
>chrXVI	843169	C	T	17x			

Figure 2. Identification of K6001-B7 by subtractive whole-genome resequencing. (**A**) 454 sequencing found 12,484 genetic differences between K6001 and the S288c reference genome. Four K6001-B7 candidates were identified by systematic narrowing of this list via the exclusion of non-uniform reads, substracting K6001 from K6001-B7, and the manual exclusion of alignment artefacts. Four of the remaining 13 variants were predicted to result in amino acid exchange. (**B**) Sanger resequencing of candidate regions. Mutant variants are highlighted. Please note that for *TSA1* the sequence trace of the reverse strand is shown.

exclusion of these non-uniform SNVs from the candidate list. By setting a threshold of a minimal calling rate >95%, we reduced the list to 7213 uniform SNVs and small insertion/deletions.

We continued our investigations by comparing K6001 with B7 by removing all variants which were found in both genomes. This subtractive analysis narrowed our candidate list to 152 K6001-B7 specific mutations. Finally, we were able to remove alignment artefacts of the mapping algorithm by manual curation; primarily, some two-nucleotide transversions were wrongly called. In addition, we excluded all nucleotide variations which distinguished K6001-B7 from the reference genome, but were simply not sequenced at sufficient quality in K6001 and therefore potential false-positives.

The final list contained 13 candidate mutations. As illustrated in Figure 2, four of these were predicted to cause amino acid exchanges. We amplified these regions by PCR and subjected them to Sanger sequencing. The three SNV resulting in residue exchanges could be verified by capillary re-sequencing (Figure 2B). The fourth difference, a six nucleotide insertion of the *CLC1* gene, was also real, but found in both the K6001 and K6001-B7 strain.

Oxidant resistance of B7 is mediated by a single residue exchange within the TSA1 gene

To identify the B7 gene among these candidates, we chose a classic strategy of cloning and phenotypic analysis. The four potential candidate genes were PCR-amplified from K6001-B7 genomic DNA, and sub-cloned into a yeast single-copy (centromeric) expression vector along with their endogenous promoter sequences. The plasmids subsequently verified by sequencing were transformed into K6001, and monoclonal descendants selected on SC Gal media lacking histidine. As illustrated in Figure 3A, the transformants ectopically expressing *HOT1-B7*, *CLC1-B7*, *DON1-B7* as well those harbouring the control plasmids were viable on SC-His/Gal media, but not more resistant to CHP than the wild type.

In contrast, the transformants of the *TSA1-B7* encoding plasmid grew perfectly well on media containing 0.05 mM CHP. To exclude in a second step, that this phenotype was the result of a gene dose effect caused by the extra copy of *TSA1*, we generated an additional, isogenic plasmid encoding for its wild type form as well. Along with the empty as well as the *TSA1-B7* vector, this plasmid was transformed into K6001 and the S288c derived BY4741 strain. Selected transformants were grown over night, and spotted as dilution series on agar plates with and without the oxidant. As illustrated in Figure 3B, only yeast containing the *TSA1-B7* plasmid, but not its wild-type form, were resistant to CHP. This phenotype was also observed in the S288c (BY4741) background, confirming the dominant and background-independent inheritance. Thus, a new dominant allele in *B7*, encoding for Tsa1p^{Pro182Leu}, was responsible for the increased oxidant tolerance of the K6001-B7 mutant.

Table 1. Oligonucleotide primers

CDS	fwd oligo	rev oligo	cloning
DON1	TAGAATTCAGGGTACAGGCGAAGAAATG	TAGTCGACCTACGTAAAACTTAATTCTT	*Eco*RI/*Sal*I
CLC1	GAAGAGCTCAACAATACAATAAACCCAATC	TGGTCGACTTAAGCACCGGGAGCCTTCG	*Sac*I/*Sal*I
TSA1	GAGAGCTCATACGCTACCCAAGTACAGAAG	TGTCTCGAGTTATTTGTTGGCAGCTTCGA	*Sac*I/*Xho*I
Hot1	GAGAGCTCATTATATCCATGTTAAGTTCG	TATCTCGAGCTATATTCCAGCAAGGCTCT	*Sac*I/*Xho*I
	Underlined sequences represent introduced restriction sites		

Table 2. MRM transitions

	Q1/Q3 transition	sequence	Charge/Fragment ion
TSA-pep_1	617.85 – 984.55	HITINDLPVGR	2+ / y9
TSA-pep_2	451.77 – 732.43	GLFIIDPK	2+ / y6
TPI-pep_1	762.37 -989.49	ASGAFTGENSVDQIK	2+ / y9
TPI-pep_2	758.93 – 864.46	KPQVTVGAQNAYLK	2+ / y8

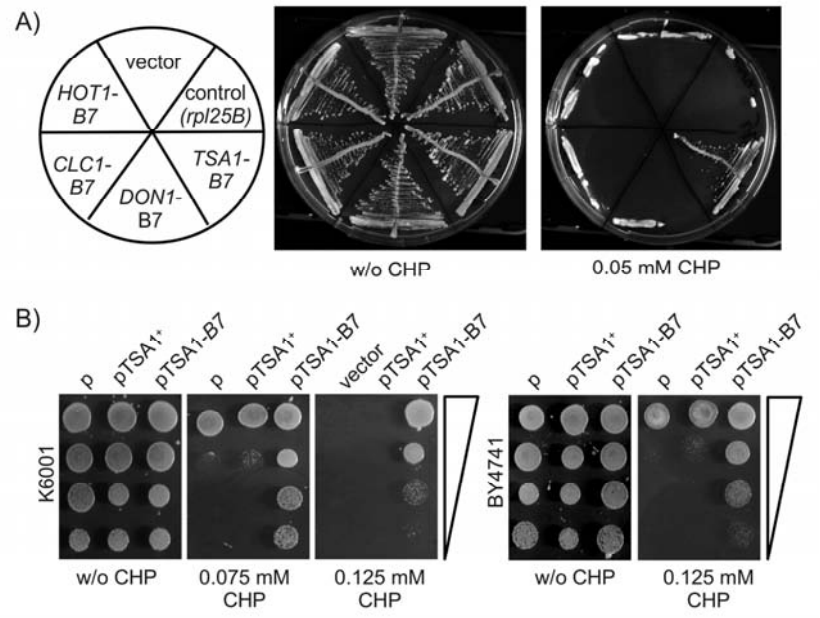


Figure 3. Ectopic expression of *TSA1-B7* mediates CHP resistance. (**A**) Candidate genes *HOT1, CLC1, DON1,* and *TSA1* were amplified from the K6001-B7 genome, subcloned, and ectopically expressed in the K6001 parent. Transformants expressing *TSA1-B7* were viable on CHP media. (**B**) *CHP resistance is specific to the expression of TSA1-B7*. Centromeric plasmids encoding wild type *TSA1*$^{+}$ and its B7 form were transformed into K6001. The additional copy of *TSA1*$^{+}$ had no effect on CHP resistance in K6001 (left) and S288c/BY4741 (right).

To gain insights, if *TSA-B7* represents a gain or a loss of function allele, we first assayed *TSA1-B7* mRNA levels in the K6001 parent and its B7 mutant. As shown in Figure 4A, via quantitative RT-PCR, we could not detect any difference in *TSA1* mRNA expression between the K6001 parent and K6001-B7. Next, using targeted mass spectrometry, we addressed Tsa1p protein levels. By quantifying two Tsa1p-specific tryptic peptides (GLFIIDPK and HITINDLPVGR) and normalization of their peak areas to peptides specific for Triosephosphate isomerase (ASGAFTGENSVDQIK and KPQVTVGAQNAYLK), we found that Tsa1p is expressed at identical levels in K6001 and K6001-B7 (Figure 4B). Thus, stress resistance caused by *TSA1*-B7 is not attributable to altered expression of Tsa1p.

Next, we questioned, if the stress resistance of B7 might be attributable to loss of function of Tsa1p. As shown in Figure 4C left, deletion of the *TSA1* gene in two haploid yeast strains decreased rather then increased yeast's resistance to CHP. Thus, depletion of *TSA1* does not result in the *TSA-B7* phenotype, indicating that the dominant *TSA1-B7* is not a loss of function allele. Next, we spotted diploid wild-type, *TSA1*/*Δtsa1* heterozygous and *Δtsa1*/*Δtsa1* homozygous strains on oxidant-containing agar. Whereas the homozygous *Δtsa1* / *Δtsa1* deletion strain was sensitive to CHP, this phenotype was not detected in the *TSA1*/*Δtsa1* heterozygotes (Figure 4C, right). Thus, partial loss of *TSA1* due to a haploinsufficiency does not resemble the B7 phenotype.

We continued by addressing the effects of ectopic Tsa1 and Tsa1-B7 expression in the *Δtsa1* background. BY4741 based *Δtsa1* yeast was transformed with *TSA1*- and *TSA1-B7*- as well as empty single-copy (CEN) plasmids, and assayed for CHP resistance. Expression of *TSA1* in the deficient background restored the wild-type phenotype, whereas B7 expression resulted in CHP resistant transformants (Figure 4D).

Finally, we questioned if the increased CHP resistance might be attributable to higher Tsa1 activity and thus, could be simulated by overexpression of wild-type Tsa1p. We subcloned *TSA1* and *TSA-B7* into a high-

copy 2µ plasmid (p423GPD) and transformed both K6001 and K6001-B7. As shown in Figure 4E, overexpression of wild-type *TSA1* did not simulate the effects of the B7 mutation. Unexpectedly, however, wild-type K6001 cells harbouring the multicopy overexpression plasmid for *TSA1-B7* resulted in less CHP resistant cells than single-copy counterparts (Figure 4E left). This effect was not seen in the mutant background (Figure 4E right). However, here similar effects were caused by multicopy expression of the wild-type allele. Thus, heterogeneous overexpression of *TSA1* with *TSA1-B7* reduces the stress resistance mediated by the *TSA1-B7* allele.

In sum, these results indicate that TSA1-B7 is a gain of function allele. *TSA1-B7* is dominant above *TSA1* across yeast backgrounds, its phenotype is not simulated by *TSA1* over- or underexpression nor its deletion, and finally, heterogeneous *TSA1/TSA1-B7* overepxression diminishes the stress resistant phenotype.

K6001-B7 has a premature aging phenotype

Since altered resistance to oxidants has often been observed in yeast strains with aging phenotypes, we screened for potential alterations in the replicative lifespan of the B7 mutant. K6001 offered the possibility to screen for alterations replicative aging via the generation of growth curves in galactose- and glucose, an alternative to time-consuming micromanipulation experiments [16]. As illustrated in Figure 5A, K6001 and K6001-B7 both showed a comparable doubling time in galactose. However, when shifted to glucose, linear growth stagnated at a lower biomass in K6001-B7. The final biomass reached by the wild type and mutant was significantly different. This indicated that – although oxidant-resistant – K6001-B7 has a premature aging phenotype. To confirm this result, we assayed replicative aging of K6001 via removal and counting of daughter cells by micromanipulation (Figure 5B). On galactose media, the median lifespan of K6001 was 17 generations, and the lifespan of K6001-B7 shortened to 13 generations.

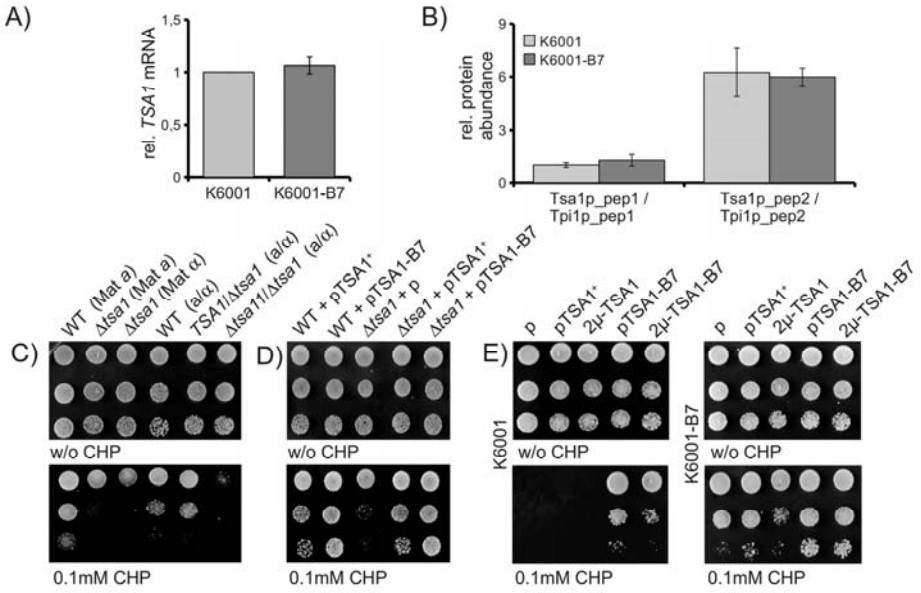
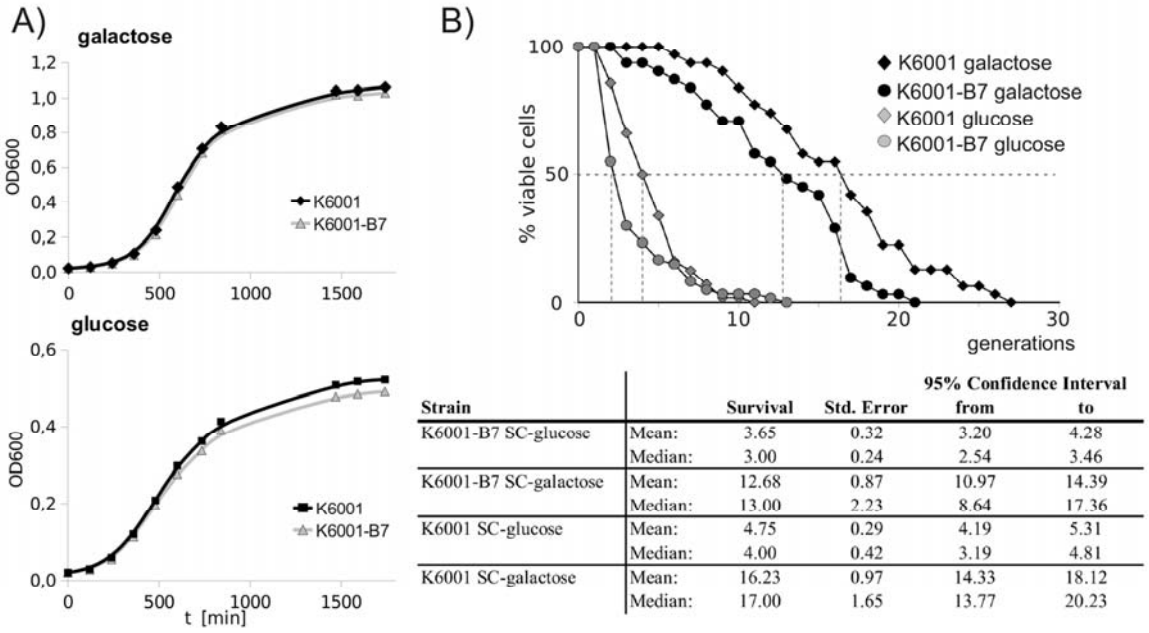


Figure 4. CHP resistance is mediated by a *TSA1* gain of function. (**A**) *TSA1 mRNA expression does not differ between K6001 and K6001-B7. TSA1* mRNA levels were assayed by classic qRT-PCR and normalized to the geomean of reference transcripts *ACN9, ATG27* and *TAF10*. (**B**) *Tsa1p protein levels do not differ between K6001 and K6001-B7.* Two Tsa1p specific tryptic peptides were quantified by nanoLC-MS/MS and set in relation to two peptides of triosephosphate isomerase (Tpi1p) in both K6001 and K6001-B7 extracts. (**C**) *Deletion of tsa1, but not TSA1/Δtsa1heterozygosity, decreases CHP tolerance. Tsa1* deletion strains of *MAT*a and *MAT*α mating types, as well as corresponding heterozygote and homozygote diploids of the S288c background, were spotted on CHP containing agar and grown at 28°C. (**D**) *CHP sensitivity of Δtsa1 is restored upon ectopic expression of TSA1$^+$. MAT*a *Δtsa1* cells were transformed with single-copy expression vectors encoding *TSA1+* and *TSA1-B7* and assayed for growth on CHP containing media. (**E**) *Mulicopy overexpression of TSA1$^+$ does not mimic TSA1-B7, heterogeneous overexpression of TSA1$^+$/B7diminishes the oxidant resistance phenotype.* K6001 (left) and K6001-B7 (right) were transformed with single-copy (p) or high-copy (2µ) *TSA1$^+$* and *TSA1-B7* expression plasmids, spotted on CHP containing yeast agar, and grown at 28°C for three days.



Strain		Survival	Std. Error	95% Confidence Interval from	to
K6001-B7 SC-glucose	Mean:	3.65	0.32	3.20	4.28
	Median:	3.00	0.24	2.54	3.46
K6001-B7 SC-galactose	Mean:	12.68	0.87	10.97	14.39
	Median:	13.00	2.23	8.64	17.36
K6001 SC-glucose	Mean:	4.75	0.29	4.19	5.31
	Median:	4.00	0.42	3.19	4.81
K6001 SC-galactose	Mean:	16.23	0.97	14.33	18.12
	Median:	17.00	1.65	13.77	20.23

Figure 5. K6001-B7 has a shortened replicative lifespan. (**A**) K6001 and K6001-B7 have similar growth in galactose (upper panel), but K6001 stagnates at a lower biomass in glucose media (lower panel). (**B**) Lifespan assay by micromanipulation. Daughter cells from K6001 and K6001-B7 were continuously removed by micormanipulation and counted (upper panel) and analyzed statistically (lower panel). The shortened lifespan of K6001-B7 in both glucose and galactose media was tested statistically significant using Mantel-Cox, Breslow as well as Tarone/Ware statistics.

.

The genetic manipulations of K6001 dramatically shorten its lifespan on glucose media [16]. Nonetheless, we also compared the lifespan of K6001 and K6001-B7 mothers under these conditions. Here, the median replicative lifespan was 3 for K6001-B7, and 4 generations for K6001. Thus, the premature aging phenotype of K6001-B7 is not galactose-specific. Differences in the aging phenotypes on both media were statistically significant as tested by Log Rank-, Breslow- and Tarone Ware statistics.

DISCUSSION

Here, we describe the generation, identification and initial functional analysis of a dominant peroxiredoxin allele which causes oxidative stress resistance and premature aging in yeast. The oxidant-resistant mutant was isolated after EMS treatment of the yeast aging model strain K6001, a descendant of the broadly used yeast strain W303. Subsequent experiments demonstrated that B7 is transmitted in a monogenic and dominant pattern.

The classical genetic and molecular genetic approach for the dissection of a dominant monogenic trait is constructing a genomic clone bank representative for the genome of the mutant, transforming this bank into wild type cells, selecting for clones that show the phenotype (in our case, CHP resistance), cloning and subcloning, and identifying the gene by capillary sequencing (see [18] for an overview on classic screenings methods). For the case presented here, this strategy would be burdened by several limitations, like occasionally a non-selectable phenotype, or if transformats from the clone bank confer a selectable phenotype even if they contain overexpressed wild type genes. A second classical strategy is cloning via mapping using determination of linkage with a large number of genetic markers on all chromosomes in meiotic tetrads. This strategy, however, is tremendously laborious. Here, we tested if with the advent of ever improving sequencing methods, resequencing of whole yeast genomes has become a serious alternative method to isolate and confirm the gene in which a given dominant mutation resides.

The genome of both, the parent strain and the mutant derived from it, were sequenced using the Roche/454 system. At an average read length of 499bp, this yielded a 38.8 fold average total coverage of the genome, and divided into 19.2X coverage for K6001 and 19.5X coverage for K6001-B7. Compared to the S288c reference, we detected 7660 uniform small nucleotide variations. 7499 of these, proven by the very stringent criteria of a >10X coverage are available in the Supplementary material. This rate indicates one nucleotide exchange per 1600 bp, which points to a high degree of evolutionary divergence between S288c and W303/K6001. Interestingly, the rate of divergence differed between the individual chromosomes (Supplementary Figure 2).

An interesting observation of this study was the identification of many non-uniform variants in the haploid genome. Many of these were called with high coverage and confidence, indicating that they might be biologically relevant and not the result of technical artefacts. 62.1 % of the variants detected at a calling rate between 25 and 75% were located in or close to the telomeric regions; however, they clearly were located in unique sequences. We will intensively discuss the nature of these variants in a future publication.

In the last three years, next generation sequencing has been used for mapping of epigenetic mutations [19], identification of spontaneous mutations [20], or for evolutionary considerations [21] in yeast. The accuracy of next generation technologies is key to this approach and in the course of this study we identified less than 1 difference (0.86) in 1,000,000 nucleotides between both genome sequences. Lynch et al. [20] used a pyrosequencing approach and reported an average depth of sequence coverage of 5X and restricted their analyses to sites within each genome with at least 3X coverage. The authors pointed out that especially for homopolymeric sites, a higher coverage depth is necessary. Recently, Araya et al. [22] used a short read sequencing-by-synthesis approach for their whole genome sequencing of a laboratory-evolved yeast strain. At an average depth of 28X, they covered 93% of the yeast genome. The coverage required for an in-depth analysis of individual genomes highly depends on the used sequencing technology. According to Wheeler et al. [23] sequencing of diploid organisms demonstrated a minimum of 15X coverage for pyrosequencing, and a 30X for sequencing-by-synthesis for accurate detection of heterozygous variants. For haploid genomes, the minimal coverage depth required is lower. In this study we reached a calling rate of 98% of the genome with an average coverage of 19.4X for both strains. For the detection of single nucleotide variations which distinguish K6001 from the reference genome, we merged the genomes of both K6001 and K6001-B7, which resulted in a 38.8x fold average genome coverage.

The number of detected variants which distinguished our strain from the Reference genome (12,482) was substantially large. To delimit candidate genes to identify the mutation causing the phenotype observed, we used a rigorous subtraction strategy. Surprisingly, the 12,482 SNV contained 5269 non-uniform variations which had a calling rate < 95 %. We excluded these non-uniform SNVs, because such a frequency would be incompatible with the observed 2:2 segregation of the B7 mutant. Next, we removed all variations that were shared between K6001 and K6001-B7, and all differences to the S288c genome that were not present in the K6001 genome sequence. At the end, we curated the final list manually to identify alignment artefacts of the mapping algorithm. We ended with a list of 13 differences, of which 6 were located in protein coding regions. Four of these mutations were predicted to cause residue exchanges.

Therefore, EMS mutagenesis had caused one mutation per ~950,000 nucleotides in the genome, and thus, indicated that the applied protocol was efficient to create a monogenic trait. However, we have to consider that mutant B7 was isolated after strong selection on CHP, and therefore the unbiased primary mutation rate immediately after EMS mutagenesis could have been substantially different.

We started the experimental verification strategy focusing on the four residue exchanging mutants. This choice was hypothesis driven; at this stage of the project, also non-exonic sequences were potential, although less probable, candidates for the B7 gene. We amplified *CLC1*, *DON1*, *HOT1* and *TSA1* genes by PCR from both K6001 and K6001-B7. Sanger sequencing verified the B7-specific mutations in *DON1*, *HOT1*, and *TSA1;* the six nucleotide insertion (relative to S288c) into *CLC1* was also real, but found in the parent strain as well. These four genes were then subcloned with their native promoters into a single-copy yeast expression vector, transformed into strain K6001, and transformants tested for their phenotype on CHP containing media. These experiments identified and confirmed *TSA1-B7* as the phenotype-causing gene.

Tsa1p (*thiol specific antioxidant*), a 2-Cys peroxiredoxin, is an important enzyme of the cellular antioxidative machinery and catalyzes H_2O_2 reduction in the presence of thioredoxin, thioredoxin reductase and NADPH [24]. Tsa1p furthermore acts as a

molecular chaperone, and is associated with ribosomes, it prevents oxidative damage of newly synthesized polypeptides [25]. It was previously shown that disruption of Tsa1p diminishes oxidative stress resistance of yeast; cells lacking this enzyme were highly sensitive to tert-butyl hydroperoxide, hydrogen peroxide, and CHP [26]. Peroxiredoxin 1 has been associated to the aging of mammals, since it interacts and stimulates the activity of the lifespan regulator protein p66Shc [27], and disruption of *Tsa1p* in yeast shortens its chronological lifespan [28]. The results presented here show that the involvement of Tsa1 in aging phenotype is complex. In contrast to the *TSA1* knock-out, yeast cells expressing the *TSA1-B7* allele gain resistance to oxidants, but are also compromised in a lifespan phenotype. Thus, changes in the natural antioxidative capacities of the peroxiredoxin system in either direction can accelerate yeast aging. These results add and substantiate other observations that highlight the importance of a natural redox balance, rather than the total antioxidative capacity, as an important determinant of cellular lifespan [29-31]. For instance, *C. elegans* requires a natural rate of free radical generation for lifespan extension by caloric restriction [32], and yeast cells, which are oxidant-resistant due to an excessive NADPH production caused by mutations in triose phosphate isomerase gene, are also replicatively short lived [8]. Thus, understanding the influence of free radicals and oxidative stress on the complex phenotype of aging requires examination of these processes in the context of the highly evolved regulation of the cellular redox environment.

Our study demonstrates that whole-genome re-sequencing is suitable to identify a functional single nucleotide exchange generated by random mutagenesis. Although all commercially available sequencing platforms (Genome Analyzer (Illlumina), SOLiD (Life Technologies) and FLX Genome Sequencer (454/Roche)) would provide appropriate workflows, we decided on 454 sequencing because of large read-lengths, sequence accuracy and the available software tools. All data was collected in a single Titanium run on the FLX sequencer. Thus, whole genome re-sequencing strategies have the potential to increase the efficiency and flexibility of random strategies, highly increasing their attractiveness for addressing current biological problems.

MATERIALS AND METHODS

Yeast cultivation and mutagenesis. Yeast was cultivated on yeastextract peptone dextrose (YPD) or galactose (YPGal), synthetic complete glucose (SC) or galactose (SCGal) media at 28°C. For EMS mutagenesis, logarithmically growing K6001 cells were washed twice with 50 mM potassium phosphate (pH 6.8), and resuspended in 10 ml of this buffer supplemented with 300µl EMS. After one hour incubation at 28°C, where 10% of the cells were still alive, mutagenesis was stopped by adding 10ml 10% w/v sodium thiosulfate [33].

qRT-PCR. Yeast total RNA was extracted using RiboPure-Yeast Kit (Ambion). After quality control, cDNA was synthezised using 12–18 oligo dT primers and Moloney Murine Leukemia virus (*M-MuLV*) reverse transcriptase (NEB) according to the manufacturer's instructions. Real-time PCRs were performed in triplicates in a final volume of 5 µl containing 1 µl cDNA, 1 µl 5x combinatorial enhancer solution (CES) [34], 0.5 µl primer mix and 2.5 µl 2× SybrGreen master mix (Fermentas). Reactions were run on a Prism 7900HT sequence detection system (ABI). The thermal cycling conditions comprised 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s/60°C for 1 min. The relative expression ratio of the target gene *TSA1* was normalized to the geometric mean of three endogenous reference transcripts (*ACN9, ATG27, TAF10*) by the method of Pfaffl [35].

Cloning and sanger sequencing. Candidate genes were amplified by oligonucleotides given in table 1 using Phusion polymerase (Finnzymes). The resulting products were gel-purified and used a) for Sanger resequencing and b) subcloned into a pRS413-derived yeast single-copy centromeric expression vector. For this, the products were treated with the endonucleases (New England Biolabs) given in Table 1, and ligated into compatible sites of the yeast vector. 2µ overexpression plasmids p423-*TSA1* and p423-*TSA1-B7* were generated by excision of the corresponding fragments with *Sac*I/*Xho*I from the cen-plasmid, and their ligation into the backbone of the *Sac*I/*Xho*I digested 2µ vector p423GPD [36]. Plasmids were verified by endonuclease digestion and sequencing.

Targeted protein quantification by mass spectrometry. Tsa1p levels were quantified by the means of determination of their relative abundance relative to the expression of triosephosphate isomerase (Tpi1p). Whole-proteome tryptic digestes were generated and analyzed by nanoflow liquid chromatography tandem mass spectrometry (nanoLC-MS/MS) on an QTRAP5500 hybrid triple quadrupole/ion trap mass spectrometer (AB/Sciex) as described earlier [37]. Analyzed peptides and the MRM transitions used for quantification are given in Table 2.

Lifespan assay by micromanipulation. Logarithmically growing yeast cultures were plated at low density, and

at least 50 daughter cells were isolated as buds with a MSM micromanipulator (Singer instruments). After their first division, mothers were removed and 2nd generation virgin cells kept for analysis. The lifespan of these cells was determined by counting and removing all subsequent daughters at 28°C. Cells were shifted to 8°C overnight to allow resting of the investigator; depending on the age of the cells, 1-2 generations were completed at this temperature per night. Statistical calculations for lifespans were conducted with SPSS 11.0 (SPSS) and Excel with Winstat. By applying Kaplan-Meier statistics the standard deviations of the median lifespan at a confidence level of 95% were calculated. To show if two given survival distributions are significantly different at a 95% confidence level, logrank (Mantel-Cox), the modified Wilcoxon test statistic (Breslow), and Tarone & Ware statistics were used as described earlier [16].

454 Sequencing and data analysis. DNA was sheared by sonication to a fragment size of 500 - 800bp, and adaptors ligated. The amplified template beads were recovered after emulsion breaking and selective enrichment. Sequencing primer was annealed to the template and the beads were incubated with *Bst* DNA polymerase, apyrase, and single-stranded binding protein. Template beads, enzyme beads (required for signal transduction) and packing beads (for *Bst* DNA polymerase retention) were loaded into the wells of a 70 x 75 mm two compartment picotiter plate. The picotiter plate was inserted in the flow cell and subjected to pyrosequencing on the Genome Sequencer FLX instrument (454/Roche).

The system flows 200 cycles of four solutions containing either dTTP, alphaSdATP, dCTP or dGTP reagents, in that order, over the cell. For each dNTP flow, a single image was captured by a charge-coupled device (CCD) camera on the sequencer. The images were processed to identify DNA bead-containing wells and to compute associated signal intensities. The images were further processed for chemical and optical cross-talk, phase errors, and read quality before base calling was performed for each template bead.

Raw data processing. After default raw data processing, we used a resequencing trimming filter to increase the data output. (parameters doValleyFilterTrimBack = false, vfBadFlowThreshold = 6, vfLastFlowToTest = 168, errorQscoreWindowTrim = 0.01).

Mapping 454 reads to a reference genome. We generated 1.3 million sequences which produced 662 million bases that were aligned to the *Saccharomyces cerevisiae* reference genome [38], using GS reference mapper version 2.3. The best match in the genome was used as the location for the reads with multiple matches. SNPs and small insertion-deletions were included in the analysis.

Filtering small nucleotide variations. To identify the phenotype-causing SNP of the B7, we first removed non-uniform SNVs and insertions/deletions by introducing a cutoff-filter of <95%. Next, we eliminate those variants which were equal in B7 and K6001. The remaining 152 differences were controlled by hand, taking into account the covered position of the yeast genome, the reference base and the consensus base for each position ('454AlignmentInfo' file). Of these, we manually compared all variants called with at least three non-duplicate reads in both directions with the consensus base position in K6001.

ACKNOWLEDGEMENTS

We thank Beata Lukaszewska-McGreal for help with sample preparation and our lab-members for critical discussion. We are grateful to the EC (Brussels, Europe) for project MIMAGE (contract no. 512020; to M.B.) and to the Austrian Science Fund FWF (Vienna, Austria) for grant S9302-B05 (to M.B.) and S9306 (to JG).

CONFLICT OF INTERESTS STATEMENT

The authors of this manuscript have no conflict of interests to declare.

REFERENCES

1. Finkel T, Holbrook NJ: Oxidants, oxidative stress and the biology of ageing. Nature. 2000; 408:239-247.

2. Muller FL, Lustgarten MS, Jang Y, Richardson A, Van Remmen H: Trends in oxidative aging theories. Free Radic Biol Med. 2007; 43:477-503.

3. Harman D: Aging: a theory based on free radical and radiation chemistry. J Gerontol. 1956; 11:298-300.

4. Aguilaniu H, Gustafsson L, Rigoulet M, Nystrom T: Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. Science. 2003; 299:1751-1753.

5. Klinger H, Rinnerthaler M, Lam YT, Laun P, Heeren G, Klocker A, Simon-Nobbe B, Dickinson JR, Dawes IW, Breitenbach M: Quantitation of (a)symmetric inheritance of functional and of oxidatively damaged mitochondrial aconitase in the cell division of old yeast mother cells. Exp Gerontol. 2010; 45:533-542.

6. Blagosklonny MV: Aging: ROS or TOR. Cell Cycle. 2008; 7:3344-3354.

7. Heeren G, Rinnerthaler M, Laun P, von Seyerl P, Kössler S, Klinger H, Hager M, Bogengruber E, Jarolim S, Simon-Nobbe B, Schüller C, Carmona-Gutierrez D, Breitenbach-Koller L, et al: The mitochondrial ribosomal protein of the large subunit, Afo1p, determines cellular longevity through mitochondrial back-signaling via TOR1. Aging. 2009; 1:622-636.

8. Ralser M, Wamelink MM, Kowald A, Gerisch B, Heeren G, Struys EA, Klipp E, Jakobs C, Breitenbach M, Lehrach H, Krobitsch S: Dynamic rerouting of the carbohydrate flux is key to counteracting oxidative stress. J Biol. 2007; 6:10.

9. Sanz A, Fernandez-Ayala DJ, Stefanatos RK, Jacobs HT: Mitochondrial ROS production correlates with, but does not directly regulate lifespan in Drosophila. Aging. 2010; 2:220-223.

10. Postma L, Lehrach H, Ralser M: Surviving in the cold: yeast mutants with extended hibernating lifespan are oxidant sensitive. Aging. 2009; 1:957-960.

11. Blagosklonny MV, Hall MN: Growth and aging: a common molecular mechanism. Aging. 2009; 1:357-362.

12. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, et al: Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science. 1999; 285:901-906.

13. Powers RW, 3rd, Kaeberlein M, Caldwell SD, Kennedy BK, Fields S: Extension of chronological life span in yeast by decreased TOR pathway signaling. Genes Dev. 2006; 20:174-184.

14. Kaeberlein M, Powers RW, 3rd, Steffen KK, Westman EA, Hu D, Dang N, Kerr EO, Kirkland KT, Fields S, Kennedy BK: Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. Science. 2005; 310:1193-1196.

15. Bobola N, Jansen RP, Shin TH, Nasmyth K: Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. Cell. 1996; 84:699-709.

16. Jarolim S, Millen J, Heeren G, Laun P, Goldfarb DS, Breitenbach M: A novel assay for replicative lifespan in Saccharomyces cerevisiae. FEMS Yeast Res. 2004; 5:169-177.

17. Laun P, Pichova A, Madeo F, Fuchs J, Ellinger A, Kohlwein S, Dawes I, Frohlich KU, Breitenbach M: Aged mother cells of Saccharomyces cerevisiae show markers of oxidative stress and apoptosis. Mol Microbiol. 2001; 39:1166-1173.

18. Breitenbach M, Dickinson R, Laun P: Smart Genetic Screens. In *Methods in Microbiology. Volume* Volume 36. Edited by Stansfield I, Stark M: Academic Press; 2007: 331-367

19. Irvine DV, Goto DB, Vaughn MW, Nakaseko Y, McCombie WR, Yanagida M, Martienssen R: Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing. Genome Res. 2009; 19:1077-1083.

20. Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK: A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci U S A. 2008; 105:9272-9277.

21. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ: The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 2008; 4:e1000303.

22. Araya CL, Payen C, Dunham MJ, Fields S: Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomics. 2010; 11:88.

23. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, et al: The complete genome of an individual by massively parallel DNA sequencing. Nature. 2008; 452:872-876.

24. Wood ZA, Schröder E, Robin Harris J, Poole LB: Structure, mechanism and regulation of peroxiredoxins. Trends Biochem Sci. 2003; 28:32-40.

25. Trotter EW, Rand JD, Vickerstaff J, Grant CM: The yeast Tsa1 peroxiredoxin is a ribosome-associated antioxidant. Biochem J. 2008; 412:73-80.

26. Chae HZ, Kim IH, Kim K, Rhee SG: Cloning, sequencing, and mutation of thiol-specific antioxidant gene of Saccharomyces cerevisiae. J Biol Chem. 1993; 268:16815-16821.

27. Gertz M, Fischer F, Leipelt M, Wolters D, Steegborn C: Identification of Peroxiredoxin 1 as a novel interaction partner for the lifespan regulator protein p66Shc. Aging. 2009; 1:254-265.

28. Lee JH, Park J-W: Role of Thioredoxin Peroxidase in Aging of Stationary Cultures of Saccharomyces cerevisiae. Free Radical Research. 2004; 38:225-231.

29. Zuin A, Castellano-Esteve D, Ayte J, Hidalgo E: Living on the edge: stress and activation of stress responses promote lifespan extension. Aging. 2010; 2:231-237.

30. Ralser M, Benjamin IJ: Reductive stress on life span extension in C. elegans. BMC Res Notes. 2008; 1:19.

31. Rattan SI: Hormesis in aging. Ageing Res Rev. 2008; 7:63-78.

32. Schulz TJ, Zarse K, Voigt A, Urban N, Birringer M, Ristow M: Glucose restriction extends Caenorhabditis elegans life span by inducing mitochondrial respiration and increasing oxidative stress. Cell Metab. 2007; 6:280-293.

33. Burke D, Dawson D, Stearns T, Cold Spring Harbor Laboratory.: *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual.* 2000 edn. Plainview, N.Y.: Cold Spring Harbor Laboratory Press; 2000.

34. Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S: An efficient and economic enhancer mix for PCR. Biochem Biophys Res Commun. 2006; 347:747-751.

35. Pfaffl MW: A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001; 29:e45.

36. Mumberg D, Muller R, Funk M: Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene. 1995; 156:119-122.

37. Wamelink MM, Gruning NM, Jansen EE, Bluemlein K, Lehrach H, Jakobs C, Ralser M: The difference between rare and exceptionally rare: molecular characterization of ribose 5-phosphate isomerase deficiency. J Mol Med. 2010;

38. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, et al: Life with 6000 genes. Science. 1996; 274:546, 563-547.

SUPPLEMENTAL FIGURES

The Supplemental Figures are found in full text version of this manuscript.