ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

OPTIMIZATION AND USE OF DNA SEQUENCING BY CAPILLARY ELECTROPHORESIS FOR COMPARATIVE YEAST GENOTYPE ANALYSIS

M.Sc. THESIS

Cihan Erdinç GÜLSEV

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

MAY 2014

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

OPTIMIZATION AND USE OF DNA SEQUENCING BY CAPILLARY ELECTROPHORESIS FOR COMPARATIVE YEAST GENOTYPE ANALYSIS

M.Sc. THESIS

Cihan Erdinç GÜLSEV
(521101103)

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

Thesis Advisor: Prof. Dr. Zeynep Petek ÇAKAR

MAY 2014

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

KAPİLER ELEKTROFOREZİNE DAYALI DNA DİZİLEME YÖNTEMİNİN OPTİMİZASYONU VE MAYADA KARŞILAŞTIRMALI GENOTİP ANALİZİ İÇİN KULLANILMASI

YÜKSEK LİSANS TEZİ

Cihan Erdinç GÜLSEV
(521101103)

Moleküler Biyoloji-Genetik ve Biyoteknoloji Ana Bilim Dalı

Moleküler Biyoloji-Genetik ve Biyoteknoloji Programı

Tez Danışmanı: Prof. Dr. Zeynep Petek ÇAKAR

MAYIS 2014

Cihan Erdinç GÜLSEV, a **M.Sc.** student of ITU **Graduate School of Science Engineering and Technology** student ID 521101103, successfully defended the **thesis/dissertation** entitled "**OPTIMIZATION AND USE OF DNA SEQUENCING BY CAPILLARY ELECTROPHORESIS FOR COMPARATIVE YEAST GENOTYPE ANALYSIS**", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Zeynep Petek ÇAKAR** Istanbul Technical University	.....
Jury Members :	**Prof. Dr. Nazlı ARDA** Istanbul University	.....
	Assoc. Prof. Dr. Nevin GÜL KARAGÜLER Istanbul Technical University	.....

Date of Submission : 05 May 2014
Date of Defense : 27 May 2014

To my dearest family

FOREWORD

I would like to extend my sincerest thanks and appreciation to my supervisor Prof. Dr. Zeynep Petek Çakar, who is an excellent guide with her background knowledge, eternal patience and endless desire for answering all questions I have. I would also like to thank to Tanju Basmacı for his recommendations and statements about sequencing technique.

I would like to thank to all employees who work in GENOMED Diagnostic AG for their valuable support and understanding. I especially thank to Ecem Temelkaya for sharing her spare time with me to complete this thesis at best.

I would like to extend my gratitude to my labmates; to Ceren Alkım for her everlasting understanding and challenging attitudes during my studies and to Nazlı Kocaefe for her both materially and spiritually support whenever I have needed, to Can Holyavkin, Berrak Gülçin Balaban, Gökhan Küçükgöze and all my other labmates for sharing their valuable times and knowledge with me.

Lastly, I wish to express my deepest gratitudes to my family for their encouragement and limitless support during my studies. Their presence in my life makes everything more bearable than it is. They are wonderful beings in moving me to realize and achieve the things I have dreamt.

May 2014

Cihan Erdinç GÜLSEV
(Bioengineer)

TABLE OF CONTENTS

Page

ABBREVIATIONS

^{32}P : Phosphorus-32
A : Adenine
ABI : Applied Biosystems
bp : Base Pair
C : Cytosine
CCD : Charge-Coupled Device
cDNA : Complementary Deoxyribonucleic Acid
CGE : Capillary Gel Electrophoresis
ddNTP: Dideoxynucleotide triphosphate
DNA : Deoxyribonucleic Acid
dNTP : Deoxyribonucleotide triphosphate
ds : Double Strand
EMS : Ethyl Methane Sulphonate
et al. : And Others
EtBr : Ethidium Bromide
etc. : *et cetera*
G : Guanine
h : Hour
H_2O_2 : Hydrogen Peroxide
HSPs : Heat-Shock Proteins
kb : Kilobase
LIF : Laser-induced fluorescence
mg : Milligram
min : Minute
mL : Milliliter
mM : Micromolar
NCBI : National Center for Biotechnology Information
ng : Nanogram
NGS : Next-generation Sequencing
OD : Optical Density
ORF : Open Reading Frame
PBS : Phosphate Buffered Saline
PCR : Polymerase Chain Reaction
pmol : Picomole
POP : Performance Optimized Polymer
qPCR : Quantitative Polymerase Chain Reaction
RNA : Ribonucleic Acid
ROS : Reactive Oxygen Species
rpm : Revolutions per minute
SGE : Slab Gel Electrophoresis
SGRP : Saccharomyces Genome Resequencing Project
SNP : Single Nucleotide Polymorphism

T **:** Thymine
Ta **:** Annealing Temperature
TBE **:** Tris amino methane, Boric acid, EDTA
UV **:** Ultraviolet
VIS **:** Visible Spectrum
wt **:** Wild Type
µg **:** Microgram
µL **:** Microliter
µm **:** Micrometer

LIST OF TABLES

Page

LIST OF FIGURES

Page

OPTIMIZATION AND USE OF DNA SEQUENCING BY CAPILLARY ELECTROPHORESIS FOR COMPARATIVE YEAST GENOTYPE ANALYSIS

SUMMARY

DNA sequencing has been a main focus of technological development of molecular biology and genetic research since Nobel laureates Frederick Sanger and Walter Gilbert introduced sequencing by chain termination or chemical fragmentation techniques, coupled with gel electrophoresis-based separation. Remarkable improvements in data resolution were encountered when the separations of fragmented products were performed in capillaries versus slab gels. Since its invention, DNA sequencing by capillary electrophoresis technique has illuminated the future on genomics and proteomics studies. Although new sequencing strategies offering high-throughput data analysis have been developed, DNA sequencing by capillary electrophoresis technique still retains its global importance in genetic research.

Saccharomyces cerevisiae is one of the most studied organisms and a common model organism in molecular cell biology to understand the eukaryotic cell. With the advent of DNA sequencing approaches, *S. cerevisiae* has been the first eukaryote the genome of which was completely sequenced. In addition to its use in scientific research, it has also been used for industrial applications such as brewing and baking for ages.

In nature and/or industrial applications, microorganisms are exposed to a variety of stress conditions such as oxidative, temperature, dehydration, starvation, and metal ions stress. *S. cerevisiae* is suitable for stress-resistance studies, because of being a eukaryotic model organism and its industrial importance.

In the present study, *S. cerevisiae* was used as a model organism. Oxidative stress-resistant strain 'H7', cobalt-resistant strain 'CI25E', nickel hyper-resistant strain 'M9' and iron-resistant strain '8C' had been obtained previously by evolutionary engineering approach, using the reference strain CEN.PK 113-7D.

Before performing the experimental studies for comparative gene analysis between reference and mutant strains, all sections of the DNA sequencing by capillary electrophoresis method were optimized to obtain the most accurate sequencing data.

Optimized protocols of the sequencing method were then applied to the experimental studies. DNA extraction from the reference and mutant strains was performed using a commercial extraction kit. Polymerase Chain Reaction (PCR) of PSE1, GRX4, MSN2, MSN4, MSN5, OCA1 and CTT1 genes was carried out in a thermal cycler using the specifically prepared PCR Master Mix for each region of interest and thermal conditions. Electrophoretic analysis of the PCR products was performed using an agarose gel. PCR products confirmed on agarose gels were then purified using the enzymatic purification method. Cycle sequencing reaction was performed

using enzymatically purified amplicons in a thermal cycler. Sequencing reaction products were purified by spin-column purification method. DNA sequencing by capillary electrophoresis of the samples was performed on ABI 3130 Genetic Analyzer using the optimized sequence module.

Data analysis of the samples was carried out in two steps. Sequencing Analysis v.6.0 software was used to investigate sequence quality and baseline clearness before comparative genotype analysis. For each gene studied, designated project including sequence data of the samples, reference sequence file of the *S. cerevisiae* and analysis parameters were created in SeqScape v.3.0 software. Comparative sequence analysis of each gene was performed between reference and mutant strains in detail.

KAPİLER ELEKTROFOREZİNE DAYALI DNA DİZİLEME YÖNTEMİNİN OPTİMİZASYONU VE MAYADA KARŞILAŞTIRMALI GENOTİP ANALİZİ İÇİN KULLANILMASI

ÖZET

DNA dizi analizi, Nobel ödüllü iki araştırmacı Frederick Sanger ve Walter Gilbert'in geliştirdikleri ilk dizileme yöntemlerinden beri moleküler biyoloji ve genetik alanındaki teknolojik gelişmelerin daima odağında olmuştur. Temelde her iki yöntemin de amacı DNA zinciri üzerindeki belirli bir noktanın nükleotit dizisini belirlemek olsa da, teknikler arasında önemli farklılıklar mevcuttur. Allan Maxam ve Walter Gilbert, DNA'nın modifikasyonuna ve belirli kimyasalların kullanılması ile belirli nükleotitlerde kesilmesi esasına dayanan bir DNA dizileme yöntemi geliştirirken, Sanger yönteminin ana prensibi zincir sonlandırıcı nükleotit olarak ddNTP'lerin kullanılmasıdır.

Her iki yöntemin de kendine ait avantaj ve dezavantajları bulunmasına rağmen zaman içerisinde Sanger DNA dizileme yönteminin otomasyona daha çok izin veren ve kullanıcı kolaylığı sunan bir teknik olduğu anlaşılmıştır. Halbuki başlarda Maxam-Gilbert yöntemi, saflaştırılmış DNA'nın doğrudan dizilenebilmesine imkan verdiği için çok daha popülerdi. Ancak zincir sonladırma yönteminin zaman içinde iyileştirilmesi, otomatik sistemlere adapte edilmesi ve Maxam-Gilbert tarafından geliştirilen tekniğin çok karmaşık olması ve insan sağlığı açısından tehlikeli kimyasalların kullanımını içermesi gibi farklı etkenler nedeniyle Maxam-Gilbert DNA dizileme yöntemi eski önemini yitirmiştir.

DNA dizi analizinin işlem hacmindeki asıl kayda değer gelişim, kapiler elektroforezine dayalı ayırım yönteminin jel tabanlı sistemlerin yerine kullanılmaya başlanması sonucunda yaşanmıştır. Kapiler elektroforezine dayalı DNA dizi analizi yöntemi, halen günümüzde hızla gelişmekte olan genomik ve proteomik alanlarına ışık tutmuştur. Daha teknolojik ve daha ileri dizileme sistemlerinin geliştirilmesine rağmen, DNA dizi analizinde halen 'altın standart' olarak adlandırılan bu yöntem, önemini hiçbir zaman kaybetmemekle birlikte, halen günümüzde en yaygın olarak kullanılan dizi analizi yöntemidir.

Saccharomyces cerevisiae mayası, ökaryotik canlılarda gerçekleşen fizyolojik ve biyokimyasal olayları anlamak için dünyada en fazla ve en yaygın olarak kullanılan model organizmalardan biridir. Geliştirilen dizileme yöntemleri sayesinde, 6000 gen bölgesi içeren tüm genomunun dizi analizinin tamamlanmış olması ile ökaryotlar arasında öncü olmuştur. *S. cerevisiae,* sadece bilimsel çalışmalarda değil, ayrıca endüstriyel alanlarda da yüzyıllardır kullanılan tek hücreli bir ökaryottur.

Doğal ortamlarında ve/veya endüstriyel uygulamalarda, mikroorganizmalar oksidatif, sıcaklık, dehidrasyon, besinsiz kalma ve metal iyonlarının varlığı gibi farklı stres koşullarına maruz kalmaktadırlar. Hem tek hücreli ökaryotik bir model organizma olması, hem de endüstriyel açıdan önemi sebebiyle *S. cerevisiae* mayası, hücresel stres direnci ve tepkisine yönelik araştırmalar için uygundur.

Bu çalışmada *S. cerevisiae* mayasının referans suşu CEN.PK 113-7D ve bu suştan daha önce evrimsel mühendislik yaklaşımı ile elde edilmiş olan, çeşitli streslere dirençli suşlar kullanılmıştır. Oksidatif strese dirençli suş 'H7', kobalt-dirençli suş 'CI25E', nikel-dirençli suş 'M9' ve demir-dirençli suş '8C' olarak adlandırılmıştır.

Karşılaştırmalı genotip analizinin deneysel çalışmalarından önce, kapiler elektroforezine dayalı DNA dizi analizi yöntemi en kapsamlı ve en kesin dizileme verilerinin alınabilmesi için bu çalışmaya yönelik özel olarak optimize edilmiştir. Çalışılan suşlardan DNA eldesi ve saflaştırılmasından sonuçların analizine kadar tüm ara basamaklarda kullanılan kimyasalların ve tekniklerin optimizasyonu sayesinde deneysel çalışmalarda hedeflenen genotipleme analizi kalitesine ulaşılmıştır.

Optimizasyonu her adımı için tamamlanan yöntem, çalışılan suşların belirlenen genleri için karşılaştırmalı dizi analizinde kullanılmıştır. Çalışılan referans ve mutant suşların DNA izolasyonu ticari bir kit kullanılarak üretici firmanın belirlediği çalışma protokolüne göre gerçekleştirilmiştir. PSE1, GRX4, MSN2, MSN4, MSN5, OCA1 ve CTT1 genlerinin Polimeraz Zincir Reaksiyon'ları (PZR) için özel olarak hazırlanmış reaktif karışımları ve çalışmaya özgü sıcaklık döngüleri her bir gen için ayrı ayrı optimize edildikten sonra kullanılmıştır. PZR ürünlerinin doğrulanması agaroz jel elektroforez sisteminde gerçekleştirilmiştir. Amplifikasyonu gerçekleştirilen gen bölgelerinin artık primer ve kullanılmayan nükleotitlerden arındırılması için enzimatik parçalama yöntemi kullanılmıştır. Saflaştırılan amplifikasyon ürünleri dizileme yöntemi için kalıp dizi olarak kullanılmıştır. Dizileme reaksiyonu sonrası artık olarak kalan işaretli-zincir sonlandırıcı nükleotitler kolon saflaştırma yöntemi sayesinde ortamdan uzaklaştırılmıştır. Dizileme yürütmeleri ABI 3130 Genetik Analizör cihazında gerçekleştirilmiştir.

Dizi bilgilerinin analizi iki adımda gerçekleştirilmiştir. Dizi bilgisinin kalitesi ve dizileme yönteminin başarısının analizi için Sequencing Analysis v.6.0 programı kullanılmıştır. Çalışmanın güvenilirliği ve temizliği doğrulandıktan sonra, çalışılan her bir gen bölgesi için SeqScape v.3.0 programında bir proje açılmıştır. Her bir projeye, dizilemesi yapılan suşların ilgili gen bölgesi için dizi dosyaları, bilgi bankasından indirilmiş referans dizi ve analizler için özel olarak tanımlanmış analiz parametreleri yüklenmiştir. Açılan projelere yüklenen bilgiler kullanılarak, referans dizi dosyasının içeriği temel alınarak, referans suş CEN.PK 113-7D ve onun mutant suşları arasındaki nükleotit değişimleri analiz edilmiştir.

Yapılan karşılaştırmalı genotip analizi sonucunda PSE1, GRX4 ve MSN5 genleri için CEN.PK 113-7D referans suş ile kobalt-dirençli suş CI25E ve demir-dirençli suş 8C arasında herhangi bir nükleotit farklılığı tespit edilmemiştir.

CEN.PK 113-7D referans suş ile oksidatif strese dirençli suş H7 arasında yapılan ve MSN2, MSN4 ve OCA1 genlerini içeren karşılaştırmalı genotip analizi sonuçları ise her iki suş arasında nükleotit farklılığı bulunmamasına rağmen, analizler sırasında referans gen dizisi kullanılan diğer bir önemli referans suş olan S288C'nin çalışılan genler açısından CEN.PK 113-7D ve ondan elde edilmiş H7 mutant suşlarından farklılıklar gösterdiği tespit edilmiştir.

Daha çok oksidatif stres ortamlarında etkinliği bilinen fakat farklı birçok stres koşullarının varlığında da stres etkenine karşı direnç geliştirilmesinde aktif rol oynadığı yapılan çalışmalarda gösterilmiş olan CTT1 geninin, CEN.PK 113-7D referans suş ile kobalt-dirençli suş CI25E, oksidatif strese dirençli suş H7 ve nikel-dirençli suş M9 arasında yapılan karşılaştırmalı genotip analizinde, H7 ve CI25E mutant suşların evrimsel mühendislik yöntemi ile elde edildikleri referans suş ile

aynı genotipe sahip oldukları, nikel-dirençli suş M9'un ise DNA zinciri üzerinde transkripsiyonun başlatıldığı bölge olarak bilenen düzenleyici dizi üzerinde referans suş CEN.PK 113-7D'den bir nükleotitlik farklılık gösterdiği tespit edilmiştir.

1. INTRODUCTION

1.1 DNA Sequencing

Sequencing technology has advanced far beyond the analysis of DNA sequences, and is now routinely used to research other biological components such as RNA and protein, as well as how they interact in complicated networks. In addition, increasing throughput and decreasing costs are making medical and life science applications of sequencing a reality [1].

DNA sequencing has been a main focus of technological development of molecular biology and biochemistry since Nobel laureates Frederick Sanger and Walter Gilbert introduced sequencing by chain termination or chemical fragmentation techniques, coupled with gel electrophoresis-based separation [2, 3].

In the early 1970s, the first DNA sequences were obtained via challenging laborious techniques. An example is the sequencing of less than thirty base pairs of the *lac* operator [2]. The first revolution in the DNA sequencing came true in the second half of the 1970s through unique methods developed by Allan Maxam and Walter Gilbert and Frederick Sanger and colleagues [4, 5].

Both techniques significantly increased the throughput of DNA sequencing. The method introduced by Gilbert and Maxam, however, was more complex and involved the use of hazardous chemicals such as hydrazine and piperidine used for modification and degradation of specific nucleotide groups in DNA chain before electrophoresis. Figure 1.1 shows the basic stage of Gilbert and Maxam method (left) and electrophoretic image of slab gel after chemical treatment of DNA (right). The technique underlying sequencing strategy is to separate firstly Purine (Adenine and Guanine) and Pyrimidine (Cytosine and Thymine) groups by using a series of chemical modifications of DNA. Single base difference is subsequently achieved by the use of nucleotide-specific chemical agents. Radioactive isotope (^{32}P) labeled fragments are then run on slab gel electrophoresis system [6].

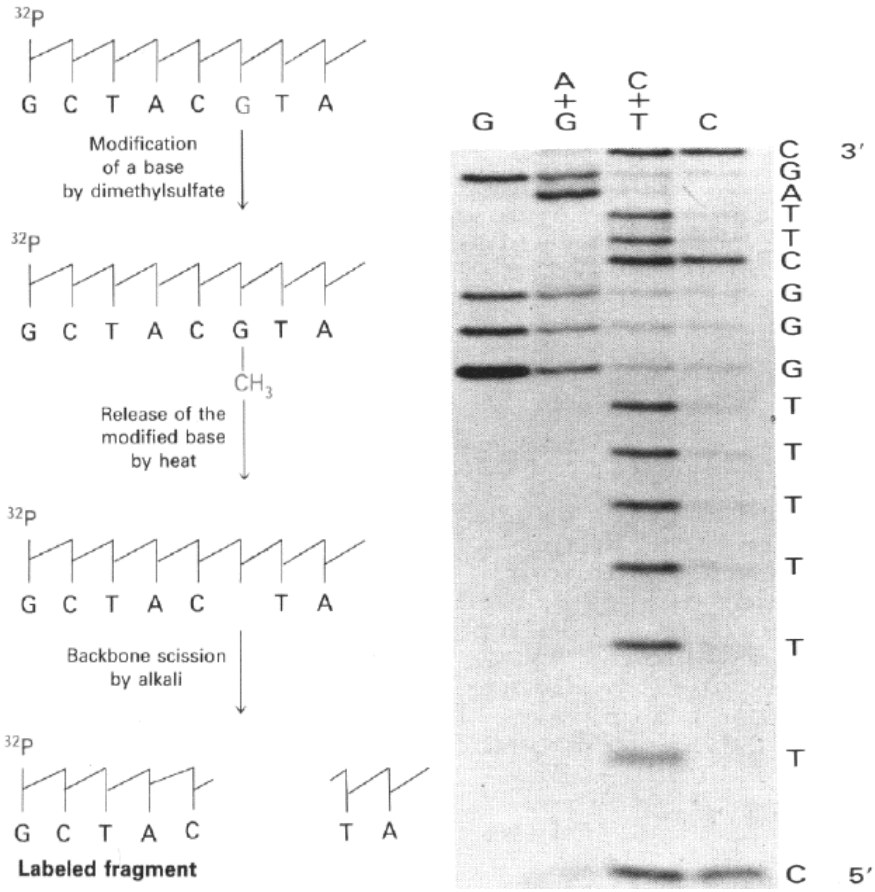

Figure 1.1: Basis of Gilbert and Maxam DNA sequencing technique. Chemical treatment of DNA for sequencing (left) and gel electrophoresis of labeled fragments (right) [4].

The Sanger sequencing method, on the other hand, offered improved efficiency after a series of optimizations, in particular switching from radioactive isotope to dye labeling of nucleotides and using capillary electrophoresis instead of slab gels. The key principle of the Sanger method, the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators, presented to be more efficient, requiring fewer toxic chemicals and lower amounts of radioactivity than that of the Gilbert and Maxam chemical fragmentation method [7]. Instead of amplifying a region of interest of DNA, the ddNTPS would cause amplification to terminate in a random number of amplified products. Figure 1.2 shows that chain extension of PCR product is achieved by dNTPs whereas ddNTPs that lack the 3′-hydroxyl group essential (Oxygen) in phosphodiester bond formation terminate elongation [7]. When a deoxynucleotide (A, C, G, or T) is added to the 3′ end, chain extension can continue. However, when a dideoxynucleotide (ddA, ddC, ddG, or ddT) is added to the 3´ end, chain extension terminates.

First applications of Sanger method also took place using slab gel electrophoresis, however, even then it was thought that it would be a more compatible method for high-throughput of automation. In the early studies, dye-labeled PCR primers were used [5]. The template was split into four different aliquots having all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and DNA polymerase. One of the four chain-terminating dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP) was included into each reaction to terminate the sequence during the chain elongation reaction. The method resulted in various length DNA fragments which were heat-denatured and size separated by polyacrylamide Slab Gel Electrophoresis (SGE). Radioactively or fluorescently labeled fragments were detected by CCD camera (Figure 1.3) [8].

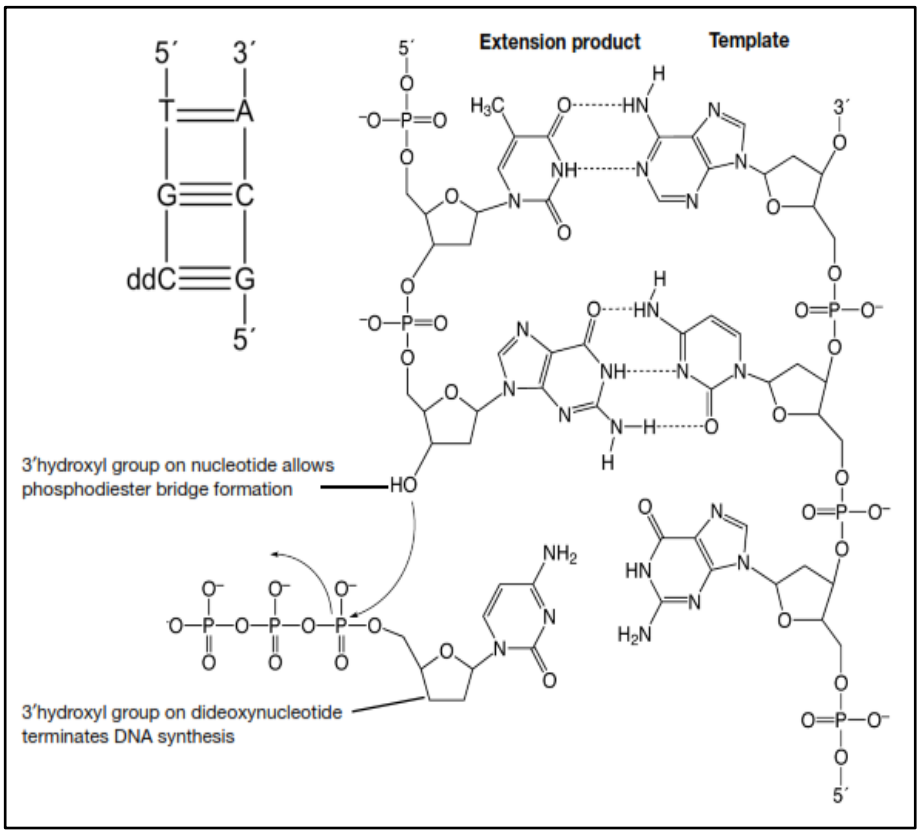

Figure 1.2 : DNA strand synthesis and termination [7].

A milestone advance was made by Smith and his colleagues in the form of fluorescently labeled ddNTPs that became the basis of the approach used to sequence the human genome [9]. Dye-terminator based DNA sequencing offered the use of four ddNTP chain terminators which was tagged with dyes of different fluorescent emission wavelengths, in a single sequencing reaction. For example, the ABI 377®

DNA sequencer released by Applied Biosystems in 1995, which employs a slab gel format and LIF (laser-induced fluorescence) detection using a CCD camera with base calling accomplished via a four-color approach, could produce sequencing data in the order of 100 kb/day [10]. Today, the dye terminating fluorescent tag is still the method of choice in automated long-read DNA sequencing by capillary electrophoresis [11].

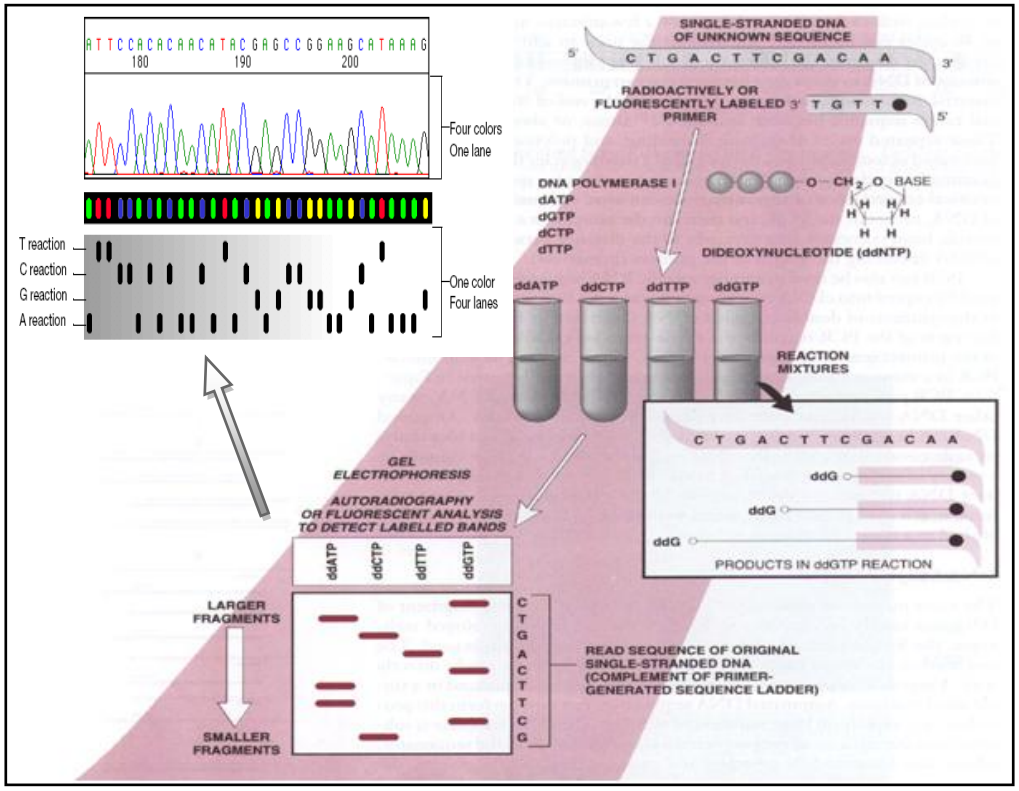

Figure 1.3 : The DNA sequencing method using dye labeled PCR primer [12].

Remarkable improvements in data resolution were also encountered when the separations of fragmented products were performed in capillaries versus slab gels. Capillary Gel Electrophoresis (CGE) was published almost 35 years ago by Jorgensen and coworkers as a new and automated alternative to slab gel electrophoresis [13]. One step ahead of 1989, Bob Brownlee and coworkers announced the first commercial instrument with on column UV/VIS detection, automatic injection and computerized data analysis for rapid, high-resolution capillary electrophoresis separation [14]. Studies have proved an about 3-fold advance in resolution and a 25-fold increase in the speed of the separation in Capillary Gel Electrophoresis (CEG) compared to slab gel electrophoresis due primarily to the ability of CGE to more effectively dissipate Joule heat, allowing

operation at high electric field strengths [15]. Several groups have also made improvements in data throughput by increasing the speed of DNA separation using CGE [15-16].

Since 1989, competition among several pioneer companies developing capillary-based DNA sequencers gradually increased to become a leader in manufacturing commercial capillary electrophoresis devices. In 1995, Applied Biosystems, being a definite leader company in capillary electrophoresis systems introduced ABI 310 Genetic Analyzer represented a breakthrough in DNA sequencing technology. The system offers continuous 24-hour hands-free automation sequencing through single capillary array filled with a novel polymer chemistry which allowed automated DNA sequencing.

Figure 1.4 demonstrates the internal constitution of ABI 310 Genetic Analyzer with basic parts. Florescent dye labeled sequencing products are collected from autosampler tray, electrophoresis of the fragments towards CCD camera (Laser detector) are achieved by the help of negative electricity field applied by cathode electrode. Before reaching the detection point, fragments are separated from each other according to their size by the help of performance optimized polymer (POP) filled inside single capillary array. At the end of each run, the used polymer inside the capillary is purged by forcing with a fresh volume [17].

One step beyond of ABI 310 Genetic Analyzer, several types of genetic analyzers have been introduced by Applied Biosystems, however, basic operation has been remained the same. Today, many of the researchers involved in DNA sequencing still use the same strategy but in a more automated way improved in high-throughput data acquisition by using multicapillary sequencers such as ABI 3130xl (16-capillary array) or ABI 3730xl (96-capillary array) Genetic Analyzers. Figure 1.5 shows multicapillary structure of ABI 3130xl instrument. In contrast to ABI 310, heat transfer of the oven has been carried out with convection created by an internal fan. Sensitivity of CCD camera has been also improved for high quality data acquisition [18].

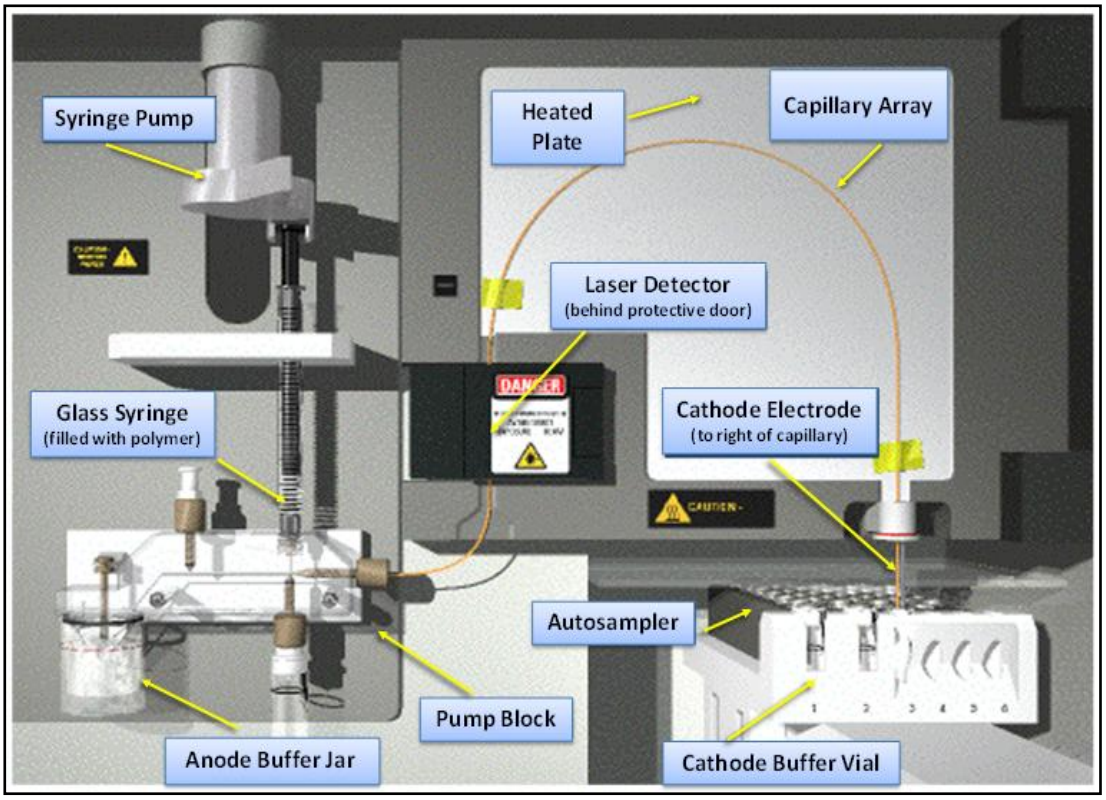

Figure 1.4 : Internal constitution of ABI 310 Genetic Analyzer.

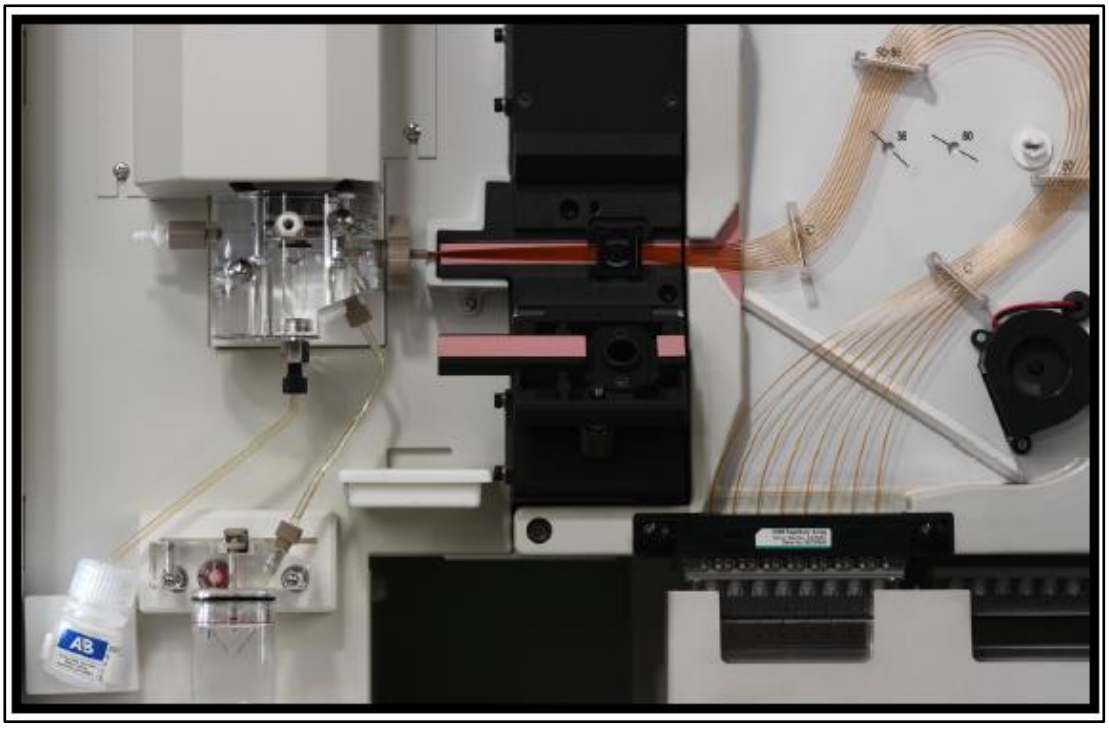

Figure 1.5 : Internal constitution of ABI 3130xl Genetic Analyzer.

1.2 Next Generation DNA Sequencing

After nearly three decades of gradual improvement, the Sanger biochemistry can be applied to obtain read-lengths of up to ~1,000 bp, and per-base 'raw' accuracies as high as 99.999%. In the context of high-throughput shotgun genomic sequencing, Sanger sequencing method costs in the order of $0.50 per kilobase (kb) [19].

Although there were numerous application fields for DNA sequencing and Fragment Analysis technologies such as de-novo sequencing, DNA profiling in human identification, chromosome copy number detection, sub-typing of infectious virus and bacteria groups, chimerism analysis, DNA microdeletion detection, SNP detection etc.; some disadvantages, nevertheless, remained with the capillary electrophoresis method. Overall, the process was still rather labour-, reagent- and time-consuming and thus involved major expenses [6]. The total costs of sequencing the genome of Craig Venter [20], for instance, using an automated Sanger instrumentation was estimated at $70 million [21]. Overall technique limitations and demands on high-throughput sampling led to investigation and engineering of alternative, more efficient methods in the 2000's.

Next-generation sequencing (also 'Next-gen sequencing' or NGS) refers to DNA sequencing methods which are highly parallelized processes that enable the sequencing of thousands to millions of molecules at once. Next-generation high-throughput DNA sequencing techniques, which are opening challenging new opportunities in biomedicine and life science, were selected by Nature Methods as the method of the year in 2007 [22]. Their innovative design, particularly with respect to the hands-on-time and reagents needed, and the competition between commercial vendors has resulted in a remarkable drop in sequencing costs since the introduction of those technologies. In 2011, the overall cost of sequencing a single human genome was about $15,000, through enthusiastic trend, this amount was dropped up to $6,000 by the early 2014 [23].

Alternative strategies for next-generation DNA sequencing can be grouped into few categories. These include Microelectrophoretic methods [24], Sequencing by Hybridization [25], Real-Time Observation of Single Molecules [26-27] and Cyclic-array Sequencing [28]. These techniques are quite various in sequencing biochemistry as well as in how the array is generated, however, their work flows are conceptually similar.

With the publication of more than hundreds of research articles in less than eight years, next generation sequencing has demonstrated its enormous potential for any one working in life sciences, at a time many believed the age of post-genomics has arrived [29].

1.3 Eukaryotic Model Organism: *Saccharomyces cerevisiae*

The use of model organisms for research in medicine and life sciences is a major requirement. Such organisms are used because a) they can help overcome ethical and experimental obstacles which hold for the target life form, b) they provide a framework to advance and optimize analytical methods that simplify and standardize analysis, and c) they are thought to be representative of a larger class of living organisms for whatever biological phenomenon or process the community is dealt with [30].

Saccharomyces cerevisiae is one of the most studied organisms and a common model organism in molecular and cell biology to understand the eukaryote cell. With the advent of DNA sequencing approaches, *S. cerevisiae* has been the first eukaryote whose genome, which is composed of 6000 genes, was completely sequenced [31]. Its simplistic structure enables easy understanding of major metabolic pathways in complex organisms. It provides essential molecular genetics tools to study signal transduction, control cell cycle progression, figure out the basis of switch from mitosis to meiosis, genetic recombination, intracellular distribution of proteins, response to stress (heat shock proteins, ethanol resistance, oxidative stress and heavy metal resistance etc.) and protein degradation [32]. For example, up to 30% of genes involved in several diseases in human beings may have orthologs in the yeast genome [33].

Genome sequencing of industrially relevant organisms such as *S. cerevisiae* S288C also provided a framework for gene annotation through functional genomics [34]. So far, only a few yeasts have been comprehensively studied in the laboratory, however, recent advances in molecular biology, genomics and post-genomic approaches have introduced several nonconventional yeasts in focus [35]. There are numerous *S. cerevisiae* strains all around the world and it has to be noted that all these strains are derived from the interbred stocks of Lindegren, Winge, and other valuable backroom scientists [36]. For instance, the ancestor strains for *S. cerevisiae* S288C was shared

initially with the other laboratories by C.C. Lindegren [31, 37]. Several strains of *S.cerevisiae* such as CEN.PK, A364A, S1278b, SK1, W303, FL100 and BY4716 were also mostly issued by different yeast researchers in the world [38].

One of the first comprehensive studies on genetic variation of yeast strains, achieved by using high-density oligonucleotide arrays containing about 200,000 oligonucleotide probes from the yeast genomic sequence, unveiled differences at the level of single nucleotide polymorphisms and gene copy number alterations [39]. Another study showed unexpected differences in 288 genes between the S288C and CEN.PK113-7D laboratory strains by involving differential gene amplification, gene absence or sequence polymorphisms [40].

In 2008, the Saccharomyces Genome Resequencing Project (SGRP) between the Sanger Institute and Institute of Genetics, University of Nottingham, completed the ABI sequencing of haploids of 37 *S. cerevisiae* strains and 27 paradoxus strains to a depth of between 1x and 3x, yielding a total of 1.42 million reads (1,292 megabases). In addition, Illumina-Solexa genome sequencing of four of the 37 *S. cerevisiae* strains, one of which included S288c, was accomplished [41]. This sequencing effort was based on exploration of genomic variation in the context of evolution, thereby using multiple strains from different *Saccharomyces* species [42].

1.4 *S. cerevisiae* CEN.PK113-7D as A Reference Strain

Although a high quality reference genome, thereby well-nown metabolic characteristics of the laboratory strain *S. cerevisiae* S288C has been available since 1996 [31], there are four main reasons to re-sequence the genomes and characterize the metabolic processes of other *S. cerevisiae* strains [43]. First, the considerable sequence divergence among *S. cerevisiae* species may lead to practical complications, for instance, the design of oligonucleotide arrays and cassettes for gene disruption in non-S288C strains [44]. In addition, although the genomes of *S. cerevisiae* strains seem to be much more strongly conserved than those of other organisms, such as *E. coli*, *S. cerevisiae* strains show physiologically relevant diversity in their gene complement [45]. For instance, the lack of a functional *MALx3* gene in *S. cerevisiae* S288C causes a maltose-negative phenotype, whereas an atypical *ENA* gene complement renders the laboratory strain CEN.PK 113-7D more sensitive to lithium ions [46]. Third, in addition to the absence or

presence of coding regions, differences may take place in regulative regions of the genes, such as promoter regions. Information of such differences is crucial for the analysis and modeling of regulatory networks in systems biology [47]. Lastly, laboratory evolution is rapidly gaining popularity as a tool to analyze genome function and to select for yeast strains with industrially relevant properties.

Over the past 15 years, the CEN.PK family of *S. cerevisiae* strains have been considered as a popular alternative platform for genomic and proteomic studies in many laboratories [48]. *S. cerevisiae* CEN.PK 113-7D is commonly used for metabolic engineering, evolutionary engineering and systems biology research for industrial and academic purposes.

1.5 Metabolic Engineering

Metabolic engineering is basically the enabling technology to improve cellular properties by targeted genetic modifications such as gene deletion, over-expression, or modulation. Recombinant DNA technology implemented in a host cell ideally will lead to redirection of fluxes for enhancement of cellular activities by manipulation of enzymatic, transport, and regulatory functions or robustness of organism, respectively [48, 49]. Genetic rearrangements are necessary for the intended phenotype, however, this requires extensive information about metabolic pathways of interest and/or regulatory systems. This can be a major limitation for the study [50]. Besides, understanding the entire metabolic network is highly challenging in the current era. Complexity of cellular physiological responses, such as alteration of the existing metabolic pathway by the cells as a response to the disturbance created by metabolic engineering can cause unexpected difficulties. Thus, to realistically estimate the responses to a perturbation, a high number of ‘stimulus–response’ experiments must be systematically performed [51].

1.6 Inverse Metabolic Engineering Approach

The previously mentioned limitations of standard metabolic engineering strategy have led to the emergence of alternative approaches [52]. An alternative strategy, a ‘bottom-up’ approach called ‘inverse metabolic engineering’ has been defined by Bailey and coworkers to overcome limitations of conventional metabolic engineering [49, 51]. The most significant advantage of inverse metabolic engineering is that,

unlike conventional metabolic engineering, extensive biochemical, genetic, or regulatory knowledge on the studied organism are not essential to obtain the desirable phenotype [51]. Three main steps should be sequentially followed to complete this approach. Firstly, identification and isolation of desired phenotype must be achieved. Second step which is also the most challenging step is the determination of genetic basis for the corresponding phenotype. Last but not least, the identified genetic characteristics must be transferred to a competent organism to obtain the desired phenotype in that organism [49, 50, 51].

1.7 Evolutionary Engineering

Continuous environmental treatments and selections which are commonly called 'evolutionary engineering' can be applied to obtain the target phenotypes [53]. A conventional method for organism development is physical or chemical mutagenesis followed by cultivation on solid or liquid media which consist of essential chemical components for the strain to grow, along with the selective condition such as high salt, oxidative agent, high metal concentrations and high/low temperatures [32]. Unlike repeated cycles of mutation and selection of colonies on solid culture plates, evolutionary engineering employs a more systematic approach: repeated batch cultivations can be achieved in the presence of a selective stress, or alternatively, extended chemostat cultivations can be carried out under selective conditions. Cultivations in question can be performed with a reference or, to improve genetic diversity, a mutant strain which is obtained by physical or chemical mutagenesis [51].

1.8 Understanding The Stress Response Genes

An important issue of evolutionary engineering research is the improvement the resistance of *S. cerevisiae* against a variety of industrial stress conditions [52]. Especially *S. cerevisiae* is commonly used as a model organism to understand the role of stress response genes in living organisms. So far, *S. cerevisiae* has been adapted to new environmental conditions by gaining cross resistance against different stress environments through evolutionary engineering strategy [51, 53, 54].
It was demonstrated that Msn2p and Msn4p induce a variety of genes in response to different types of stresses. Hence, these factors in question were considered as the

“general stress” transcription factors that were activated by stress conditions [55, 56]. It is also known that OCA1 gene expression level increases under oxidative stress conditions since OCA1 gene product provides cell cycle arrest at G1 phase. Cell cycle arrest may be essential to circumvent the damage during cell cycle phases [57]. Transcription of the CTT1 gene encoding the cytosolic catalase T in *S. cerevisiae* is activated under different stress conditions: it is downregulated by nitrogen starvation and upregulated by heat shock. Moreover, it is activated by osmotic and oxidative stress as well [58]. Another study has showed that both Grx4 has undertaken new role in actin cytoskeleton remodeling and in cellular defense against oxidative stress caused by reactive oxygen species (ROS) accumulation. The Grx4 protein has played a unique role in the maintenance of actin cable integrity [59]. It is also shown that nuclear import of Pho4 is mediated by Pse1/Kap121 which is a member of the importin β-receptor family when phosphate levels are limiting [60] whereas nuclear export of Pho4 is mediated by Msn5 when phosphate levels are abundant [61].

1.9 Aim of The Study

In light of the information provided in Section 1, the aim of the study was first to optimize the DNA sequencing by capillary electrophoresis method for the present study to obtain the most informative sequence data of the regions of interest, and second, to use the optimized method for comparative yeast genotype analysis. For this purpose, each individual step of the DNA sequencing by capillary electrophoresis method was examined in detail. In addition, different strains of *Saccharomyces cerevisiae* yeast which were previously obtained by evolutionary engineering method using continuous increasing stress selection strategy were selected for comparative genotyping. The mutant strains derived from the *S. cerevisiae* CEN.PK 113-7D reference strain were an oxidative stress-resistant strain 'H7' [62], a cobalt-resistant strain 'CI25E' [50], a nickel-resistant strain 'M9' [63] and an iron-resistant strain '8C' [64]. To perform the comparative analysis, PSE1, GRX4, MSN2, MSN4, MSN5, OCA1 and CTT1 genes which were considered to be related to the stress-resistant phenotypes of interest were chosen.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Strains

Saccharomyces cerevisiae CEN.PK 113-7D (Mat*a*, MAL2-gc, SUC2) was used as the haploid reference strain which was kindly provided by L. Benbadis from INSA-Toulouse. Oxidative stress-resistant strain 'H7' [62], cobalt-resistant strain 'CI25E' [50], nickel hyper-resistant strain 'M9' [63] and iron-resistant strain '8C' [64] which were obtained via evolutionary engineering from *S. cerevisiae* CEN.PK 113-7D (Mat*a*, MAL2-gc, SUC2) background were also used as mutant strains in this study.

2.1.2 Laboratory equipment, chemicals, kits and enzymes, solutions and buffers, softwares/databases and primer sequences for PCR

Materials which were used for the experiments are listed according to their trademarks, and country of origin. Laboratory equipments, kits and enzymes, buffers, solutions and chemicals, softwares/databases and primer sequences are indicated in Tables 2.1-2.11, respectively.

Table 2.1 Equipments used in the study.

UV-Visible Spectrophotometer	NanoDrop Lite, Thermo Scientific (USA)
Laminar Flow Cabinet	NÜVE LN 090/LN 120 Series (Turkey)
Centrifuge (1.5ml-2ml tubes)	Hettich MIKRO 120 (Germany)
Plate Centrifuge	Biosan LMC-3000 Laboratory Centrifuge (Turkey)
Plate Vortex	IKA Vortex 3 (Germany)
Pipette set	Eppendorf (Germany)
Thermal Cycler	ABI VERITI Thermal Cycler (Applied Biosystems, USA)
Transilluminator	UVP (Canada)
-20°C Freezer	Arçelik 2031D (Turkey)
+4°C Refrigerator	Profilo BD 2107 TE (Turkey)
Gel Electrophoresis System	Owl™ EasyCast B2 Mini (Thermo Scientific, USA)
Microwave Oven	Arçelik (Turkey)
Genetic Analyzer	ABI 3130 Genetic Analyzer (Applied Biosystems, USA)

Table 2.2 Kits and enzymes.

DNA Isolation Kit	Roche High Pure PCR Template Preparation Kit (Switzerland)
ExoSAP	GML ExoSAP PCR Purification Kit (Switzerland)
PCR Mix	GML PCR Mix (Switzerland)
Taq. DNA Polymerase	GML Taq. Pol. (Switzerland)
Cycle Sequencing Kit	BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA)

Table 2.3 Buffers, solutions and chemicals.

10X TBE (gel electrophoresis buffer)	Tris (Hydroxymethyl) amino methane, boric acid, EDTA (GML, Switzerland)
PBS (Phosphate Buffer Saline)	pH 7.4 50mM (ThermoFisher Scientific, USA)
Agarose gel (%2.0)	2,0% agarose in 1xTBE
Ethanol (Surface Cleaning)	%70 (Sigma-Aldrich, Germany)
Distilled Water	Eczacıbaşı (Turkey)
5X Sequencing Buffer	BigDye® Terminator v1.1/v3.1 Sequencing Buffer (Applied Biosystems, USA)
DNAZap™ RNA-DNA Degradation Solutions	(Lifetechnologies, USA)
TE Buffer	Tris-EDTA buffer solution (100X, pH 8.0) (Sigma-Aldrich, Germany)
GeneAmp® dNTP Blend (10 mM)	2.5 mM each of dATP, dCTP, dGTP, and dTTP. (Applied Biosystems, USA)
$MgCl_2$ (25 mM)	(Applied Biosystems, USA)
6X Loading Dye	(Thermo Scientific, USA)
DNA Ladder-1	Roche DNA Molecular Weight Marker XIV (Switzerland)
DNA Ladder-2	GeneRuler 100 bp DNA Ladder (Thermo Scientific, USA)
PCR Enhancer Buffer	GML Enhancing Buffer (5M)
Nuclease Free Water	GML Distilled Water (Switzerland)
Ethidium bromide	(Sigma-Aldrich, Germany)

Table 2.4 Softwares/databases and their functions.

Fast PCR	Primer control for Ta and dimerization
Microsoft Office 2007	Primer order, Mutation analysis
Sequencing Analysis v.6.0	Sequencing Analysis of raw data
SeqScape Software v.3.0	Comparative analysis with reference sequence
www.ensembl.org/	Gene and cDNA sequencing of *S. cerevisiae*
www.ncbi.nlm.nih.gov	Primer Blast
www.yeastgenome.org	Genome and gene database of *S. cerevisiae*

Table 2.5 Primer sequence used in PCR for PSE1.

Region	Primer Name	Primer Sequence (5'-----3')	Amplicon Size	Temperature of Annealing
1	PSE1_1_for	TCAAAGTGGCAGATAGCCCA	703bp	56^0C
	PSE1_1_rev	AACCCACGAAAGCCGTAACT		
2	PSE1_2_for	ACTGATGCAAGTGATAATGTCA	721bp	53^0C
	PSE1_2_rev	CATACTGTACTCTTGGATGAGG		
3	PSE1_3_for	CGTCAAGCTCTTGATCGTGT	701bp	54^0C
	PSE1_3_rev	AGAAATCAGCTCTTGAGAGTGC		
4	PSE1_4_for	TGTATGGAATGTGCAACTCTG	685bp	52^0C
	PSE1_4_rev	GCAATTGTCACCCATGA		
5	PSE1_5_for	CGGAGGCTTAATGTCAGAA	692bp	51^0C
	PSE1_5_rev	CTACGTTAGGAATGTTGGA		
6	PSE1_6_for	TCTTCAACAGAGAATGCC	633bp	51^0C
	PSE1_6_rev	ACTCGATAAGGTATGGTATC		

Table 2.6 Primer sequence used in PCR for GRX4.

Region	Primer Name	Primer Sequence (5'-----3')	Amplicon Size	Temperature of Annealing
1	GRX4_1_for	CGCGTTTTAATTGGTGAAGCCT	720bp	55^0C
	GRX4_1_rev	CATCACAGGTGCAGCTTGTAC		
2	GRX4_2_for	TAGATGCAGACGAACATCCAG	692bp	55^0C
	GRX4_2_rev	TTTGACGAGAAAACCCCAGGA		

Table 2.7 Primer sequence used in PCR for MSN2.

Region	Primer Name	Primer Sequence (5'-----3')	Amplicon Size	Temperature of Annealing
1	MSN2_1_for	ATATGCCAAAGCCGGTTTCCA	666bp	54^0C
	MSN2_1_rev	TATCTCCAAGGTTTCCGTCTG		
2	MSN2_2_for	TAAGAAAGCAGCACGAGCTCA	675bp	59^0C
	MSN2_2_rev	TTGATTTGCCCCAGAGCCACT		
3	MSN2_3_for	GCATTTGAAAGCTGATAGCAAC	746bp	54^0C
	MSN2_3_rev	GACTTGCATTGGAACTTGAGG		
4	MSN2_4_for	CAAGTTCCAATGCAAGTCCGA	742bp	54^0C
	MSN2_4_rev	TGATTTGTTGCAAGAGACGTG		
5	MSN2_5_for	TGGCTCCATTATCGCCTGCATC	632bp	54^0C
	MSN2_5_rev	TTCAAAATGTGAAGGTACCGGA		

Table 2.8 Primer sequence used in PCR for MSN4.

Region	Primer Name	Primer Sequence (5'-----3')	Amplicon Size	Temperature of Annealing
1	MSN4_1_for	TCTGAAGGCAAAGAAGTTGTGA	709bp	54^0C
	MSN4_1_rev	TGTGTAGCATCAAGAGGGTAGA		
2	MSN4_2_for	AGAGCGTTTAACAAAGGTCTCT	694bp	54^0C
	MSN4_2_rev	TTTCCACTTCCAGTTGGTAGC		
3	MSN4_3_for	GCGACAACTAATACGACTGCT	714bp	55^0C
	MSN4_3_rev	TGAAATGGTTGATCCACGCCT		
4	MSN4_4_for	ATCCCTGAAGGTACTGCCACT	671bp	56^0C
	MSN4_4_rev	TCTTGTGTTTGGGCTTACCGT		
5	MSN4_5_for	AGTCAAGCTCACCATGCAGCT	505bp	54^0C
	MSN4_5_rev	CCGAATGAAATGACCAACCTAC		

Table 2.9 Primer sequence used in PCR for MSN5.

Region	Primer Name	Primer Sequence (5'-----3')	Amplicon Size	Temperature of Annealing
1	MSN5_1_for	AGAGCAAGATTCCATTGAACAG	729bp	53^0C
	MSN5_1_rev	GTTCTTTTTGCCACCTCTACC		
2	MSN5_2_for	AGGATTACTAGACCATGCAGTG	698bp	54^0C
	MSN5_2_rev	TACTATAGGGACGGGTCAACA		
3	MSN5_3_for	ACTGGCCATTAACAGAGGTCA	721bp	55^0C
	MSN5_3_rev	GTGGAACTTAAGACCTGCTGAC		
4	MSN5_4_for	TAGGTTGATTTCTTGCGTCGA	731bp	53^0C
	MSN5_4_rev	GCAAATCAGAAAGACCAACAAC		
5	MSN5_5_for	TACCCAGACCTTGAAAGTGTC	684bp	54^0C
	MSN5_5_rev	CCTAATGATGTGTCCTACGGAG		
6	MSN5_6_for	ACATATGAAAGCGATGCAAACC	703bp	54^0C
	MSN5_6_rev	TCATCCACCGTGTTGTTTTG		
7	MSN5_7_for	TGAAATGCTGCTTAACCTCAGT	689bp	54^0C
	MSN5_7_rev	AGAAATACCTTGTTCAGACCCA		

Table 2.10 Primer sequence used in PCR for OCA1.

Region	Primer Name	Primer Sequence (5'-----3')	Amplicon Size	Temperature of Annealing
1	OCA1_1_for	AACGGCGAAGGATTCCTATT	675bp	53^0C
	OCA1_1_rev	TACCGCCGGGTGCAGCTTGCT		
2	OCA1_2_for	TAGTAAGGCCGGTCAAGTTTC	675bp	54^0C
	OCA1_2_rev	AGGTATCTTTCAACGGGACAG		
3	OCA1_3_for	CCCACAAGAACGTATCGTTCC	655bp	56^0C
	OCA1_3_rev	GCGTCGACTTTTGAATTTCATT		

Table 2.11 Primer sequence used in PCR for CTT1.

Region	Primer Name	Primer Sequence (5'-----3')	Amplicon Size	Temperature of Annealing
1	CTT1_1_for	GAGATGCTTGGAAGATGTACG	701bp	54^0C
	CTT1_1_rev	ACGAATAGTACTGGCAATGAC		
2	CTT1_2_for	TGCGTATTGTTTCCTCCTATCA	713bp	54^0C
	CTT1_2_rev	AAAGAAACACCTCTTGGGTCT		
3	CTT1_3_for	ATTACATACGCCGCTCCATAC	699bp	54^0C
	CTT1_3_rev	GTCAACATTCTCCGTTAGGGT		
4	CTT1_4_for	AAACAATGACACCCGAACAAG	703bp	54^0C
	CTT1_4_rev	TGGCAAACAACGTTATGAACG		
5	CTT1_5_for	AGTCCACTAGACTTCGAACAG	382bp	53^0C
	CTT1_5_rev	TTAGTGCTTGTACATACGACCA		

2.2 Methods

2.2.1 Design of primers

Primers were designed using Fast-PCR software. Annealing temperatures of the primers were set at values higher than 50^0C for amplification, since the subsequent sequencing reaction has an annealing temperature of 50^0C, as a default setting stated by the manufacturer of the cycle sequencing chemistry. Primer pairs were then confirmed in NCBI (National Center for Biotechnology Information) database to determine their specificity to the regions of interest in the genome.

2.2.2 DNA extraction from reference and mutant strains

DNA extraction was carried out using High Pure PCR Template Preparation Kit (Roche, Switzerland). A cell suspension with an OD_{600} of about 0.5-0.6 (4-6 x 10^6 cells/ml) was transferred into a 1.5 ml microfuge tube for DNA extraction and centrifuged at 2000xg for 5 min using a benchtop centrifuge (Hettich Mikro 120, Germany). The pellet was resuspended in 200 µl PBS 1X (ThermoFisher Scientific, USA). Subsequent procedure of extraction was completed according to the manufacturer's instructions (Roche, Switzerland).

Eluted DNAs in collection tubes were measured using UV-spectrophotometer (Nanodrop Lite, Thermo Scientific, USA) and stored at -20^oC until use.

2.2.3 Polymerase Chain Reaction (PCR) of regions of interest

DNA dilution was achieved to set initial DNA concentration for each strain as 10 ng/µl. Stock primers (100 µM) were diluted to 2.5 µM final concentration by using 1X TE (Sigma-Aldrich, Germany). PCR Master Mixes were prepared for PSE1, GRX4, MSN2, MSN4, MSN5, OCA1 and CTT1 as described in Tables 2.12-2.18, respectively.

Table 2.12 Preparation of PCR Master Mix for PSE1

Component	**Volume µL per Specimen**
GML PCR Mix	12.5
GML Taq. Pol.	0.2
Forward Primer	1.5
Reverse Primer	1.5
GML Enhancing Buffer	3.0
PCR grade water	3.8
Total	22.5

Table 2.13 Preparation of PCR Master Mix for GRX4

Component	**Volume µL per Specimen**
GML PCR Mix	12.5
GML Taq. Pol.	0.2
Forward Primer	1.5
Reverse Primer	1.5
GML Enhancing Buffer	2.0
PCR grade water	4.8
Total	22.5

Table 2.14 Preparation of PCR Master Mix for MSN2

Component	**Volume µL per Specimen**
GML PCR Mix	12.5
GML Taq. Pol.	0.2
Forward Primer	1.5
Reverse Primer	1.5
GML Enhancing Buffer	2.0
PCR grade water	4.8
Total	22.5

Table 2.15 Preparation of PCR Master Mix for MSN4

Component	Volume µL per Specimen
GML PCR Mix	12.5
GML Taq. Pol.	0.2
Forward Primer	1.5
Reverse Primer	1.5
GML Enhancing Buffer	2.0
PCR grade water	4.8
Total	22.5

Table 2.16 Preparation of PCR Master Mix for MSN5

Component	Volume µL per Specimen
GML PCR Mix	12.5
GML Taq. Pol.	0.2
Forward Primer	1.5
Reverse Primer	1.5
GML Enhancing Buffer	1.5
PCR grade water	3.8
$MgCl_2$	1.0
dNTP Blend	0.5
Total	22.5

Table 2.17 Preparation of PCR Master Mix for OCA1

Component	Volume µL per Specimen
GML PCR Mix	12.5
GML Taq. Pol.	0.2
Forward Primer	1.5
Reverse Primer	1.5
GML Enhancing Buffer	3.0
PCR grade water	3.8
Total	22.5

Table 2.18 Preparation of PCR Master Mix for CTT1

Component	Volume µL per Specimen
GML PCR Mix	12.5
GML Taq. Pol.	0.2
Forward Primer	1.5
Reverse Primer	1.5
GML Enhancing Buffer	2.0
PCR grade water	4.8
Total	22.5

2.5 µl extracted DNA (10 ng/µl) was added onto PCR Master Mix dispensed in 0.2 ml PCR tubes (Applied Biosystems, USA). PCR was performed in ABI 9700 thermal cycler (Applied Biosystems, USA) with standard mode using the following programs;

PSE1: 10 min for activation of hot start Taq. DNA Polymerase (GML, Switzerland) at 95^0C followed by 35 cycles of 30 s at 95^0C, 1 min at 53^0C, and 1 min at 72^0C, followed by 7 min at 72^0C. Reaction volume was set as 25 µl.

GRX4: 10 min for activation of hot start Taq. DNA Polymerase (GML, Switzerland) at 95^0C followed by 40 cycles of 40 s at 95^0C, 45 s at 55^0C, and 1 min at 72^0C, followed by 7 min at 72^0C. Reaction volume for was set as 25 µl.

MSN2: 10 min for activation of hot start Taq. DNA Polymerase (GML, Switzerland) at 95^0C followed by 35 cycles of 1 min at 95^0C, 1 min at 56^0C, and 1 min at 72^0C, followed by 7 min at 72^0C. Reaction volume was set as 25 µl.

MSN4: 5 min for activation of hot start Taq. DNA Polymerase (GML, Switzerland) at 96^0C followed by 40 cycles of 45 s at 95^0C, 1 min at 55^0C, and 1 min at 72^0C, followed by 7 min at 72^0C. Reaction volume was set as 25 µl.

MSN5: 10 min for activation of hot start Taq. DNA Polymerase (GML, Switzerland) at 95^0C followed by 35 cycles of 1 min at 94^0C, 1 min at 56^0C, and 90 sec at 72^0C, followed by 7 min at 72^0C. Reaction volume was set as 25 µl.

OCA1: 10 min for activation of hot start Taq. DNA Polymerase (GML, Switzerland) at 95^0C followed by 40 cycles of 45 s at 95^0C, 1 min at 54^0C, and 1 min at 72^0C, followed by 7 min at 72^0C. Reaction volume was set as 25 µl.

CTT1: 10 min for activation of hot start Taq. DNA Polymerase (GML, Switzerland) at 95 ^{0}C followed by 35 cycles of 40 s at 95 ^{0}C, 1 min at 54 ^{0}C, and 1 min at 72 ^{0}C, followed by 10 min at 72 ^{0}C. Reaction volume for PCR was set as 25 µl.

2.2.4 Confirmation of amplified PCR products

2% agarose gel was prepared by solving 2 g agarose powder (Sigma Aldrich, Germany) in 100 ml 1X TBE (GML, Switzerland). 2% agarose gel solution was boiled for 5 min in a microwave oven (Arçelik, Turkey) then cooled down up to 65^{0}C. Five µl EtBr (Ethidium bromide, Sigma Aldrich, Germany) was mixed into 2% agarose gel solution as an intercalating agent and the whole volume was immediately poured on electrophoresis tray which was placed inside the fume hood.

At the end of the cooling stage of the gel at ambient temperature, 5 µl of each PCR product was combined with 0.5 µl 6X Loading Dye (Thermo Scientific, USA) and the mixture was immediately loaded into the appropriate well. Agarose gel electrophoresis was realized on Owl™ EasyCast™ B2 Mini Gel Electrophoresis System (Thermo Scientific, USA) for 20 min at 100 mA. Agarose gel was then transferred onto transilluminator tray (UVP, Canada) to visualize amplicons.

2.2.5 Purification of PCR products

PCR products were purified using the GML ExoSAP PCR cleanup kit (GML, Switzerland). Five µl of each PCR product was mixed with 2.5 µl of GML ExoSAP in a 0.2 ml PCR tube. Reaction was carried out using a ABI 9700 thermal cycler (Applied Biosystems, USA), with the following settings: 30 min for activation of Exo I Nuclease and Alkaline phosphatase at 37^{0}C followed by 15 min for inactivation at 80^{0}C. Preparation of the mixture was completed using racks placed on ice, since enzymes included in the GML ExoSAP were heat sensitive.

2.2.6 Cycle sequencing of purified PCR products

Purified PCR product was used for sequencing reaction. Two PCR tubes were prepared for forward and reverse sequencing reaction of each specimen, separately. Sequencing master mixes were prepared as defined in Table 2.19 and Table 2.20.

Table 2.19 Preparation of sequencing master mix for forward sequence

Component for forward sequence	**Volume µL per Specimen**
BigDye Terminator v.3.1 Cycle Sequencing Kit	2
5X Sequencing Buffer (v.1.1,v.3.1)	2
Forward Primer (2µM)	2
PCR grade water	2
Total	8

Table 2.20 Preparation of sequencing master mix for reverse sequence

Component for reverse sequence	**Volume µL per Specimen**
BigDye Terminator v.3.1 Cycle Sequencing Kit	2
5X Sequencing Buffer (v.1.1,v.3.1)	2
Reverse Primer (2µM)	2
PCR grade water	2
Total	8

2.0 µl purified PCR product was added onto Forward and Reverse sequencing master mixes which were already dispensed in 0.2 ml flat cap PCR Tubes (Applied Biosystems, USA) for each specimen. Manufacturer's instructions were used as default settings to perform sequencing reaction on ABI 9700 Thermal cycler (Applied Biosystems, USA) with 9600 Emulation mode (Ramp speed: 1^0C/s). Briefly, 1 min for activation of hot start Taq. DNA Polymerase at 96^0C followed by 25 cycles of 10 s at 96^0C, 5 s at 50^0C, and 4 min at 60^0C. Reaction volume for sequencing reaction was set as 10 µl.

2.2.7 Purification of sequencing products

DNA purification columns (filter-tipped spin column within receiver tube) (GML, Switzerland) were used for purification of sequencing products by filtrating through rigid Sephadex (GML, Switzerland) which was filled into spin column. One g Sephadex powder (GML, Switzerland) was added into a centrifuge tube of 50 ml volume including 14 ml molecular grade water (Eczacıbaşı, Turkey). This suspension was vortexed vigorously and kept in the refrigerator at least for 4 hours before use. For each sequencing product, 700 µl suspension was dispensed with the help of a wide orifice pipette tip into the spin column placed within 2 ml receiver tube. Tubes were centrifuged for 3 minutes at 4600 rpm after bottom cover of spin

columns were removed. Receiver columns filled with filtered water were replaced with the new ones. Whole volume of 10 µl sequencing reaction was absorbed from upside of the rigid Sephadex column by using a micropipette, without disturbing the column. Tubes were centrifuged again for 3 minutes at 4600 rpm to capture residual ddNTPs inside Sephadex column, while the purified sequencing product was recovered at the bottom of collection tubes.

2.2.8 Running of the samples on genetic analyzer

ABI 3130 Genetic Analyzer (Applied Biosystems, USA) with POP7 (Performance Optimized Polymer 7) and 36 cm multicapillary was used to image the sequence profile of the samples. Samples purified by Sephadex column were loaded into MicroAmp 96-Well Reaction Plate (Applied Biosystems, USA). Plate was covered with a septa mat and placed in the instrument. Run module with appropriate settings was used as shown in Figure 2.1.

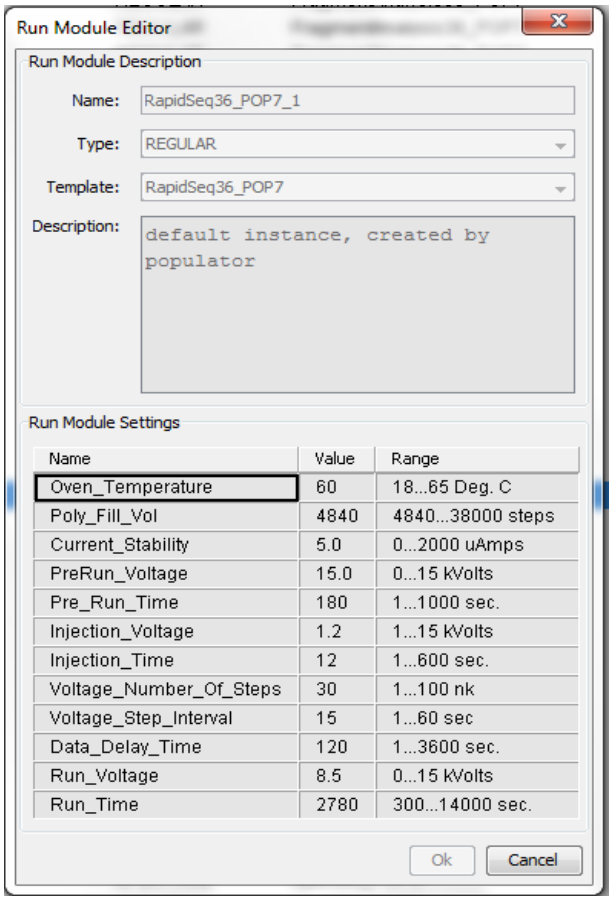

Figure 2.1 Run module for sequencing on ABI 3130 GA.

2.2.9 Interpretation of the results

Sequencing Analysis v.6.0 software (Applied Biosystems, USA) was used for the assessment of raw data to determine sequencing quality. SeqScape v.3.0 software (Applied Biosystems, USA) was used for surveying the possible mutations by comparing sequence data with the reference sequence of *S.cerevisiae* S288C which was obtained from NCBI (National Center for Biotechnology Information) database. Project for each gene studied was created in SeqScape v.3.0 software for comparative genotyping of strains used in the study. Sequence data of samples, reference sequence of *S.cerevisiae* S288C and analysis parameters were imported to the related project.

3. RESULTS AND DISCUSSION

3.1 Optimization of The Sequencing Method

Sanger DNA sequencing by capillary electrophoresis involves DNA or clone isolation, amplification of the region of interest via Polymerase Chain Reaction (PCR), PCR product purification, chain termination sequencing method of Sanger which uses fluorescently labeled dideoxynucleotide triphosphates (ddNTPs), purification of sequencing reaction product and electrophoretic analysis in a capillary tube with 50μm internal diameter. Each step to be followed sequentially requires its unique optimization procedure to obtain unambiguous sequencing data.

Conventional PCR which was performed using a thermal cycler was optimized by using a PCR enhancer reagent (GML Enhancing Buffer, Switzerland) at an overall primer annealing temperature for each gene tested. PCR Enhancer not only facilitates amplification of G/C reach regions, but also generally stimulates PCR efficiency. DNA concentration was measured to 10 ng/μl as starting material of PCR to achieve optimal amplification intensity for all genes studied. Determination of reasonable concentrations for template-specific primers which bind different chains of DNA double helix as forward and reverse is crucial to define the region of interest. Excessive amount of primers cannot be used by Taq Polymerase for amplification, hence unused primers remain intact in the reaction volume at the end of the PCR. These residual primers reduce sequencing quality unless they are handled. Therefore, primers used in the studies were diluted to 2.5 μM (2.5 pmol/μl) and added only 1.5 μl into each reaction volume.

Excess primers have to be either removed from the reaction mixture or degraded before sequencing reaction. Insufficient removal or degradation of unused primers can affect the optimization of sequencing reaction which should be performed in one direction by using forward or reverse primer for each region of interest. Peak quality of sequence data depends on baseline clearness and compression-free separation. Residual primers coming from PCR interfere with the sequence data by creating background peaks (Figure 3.1).

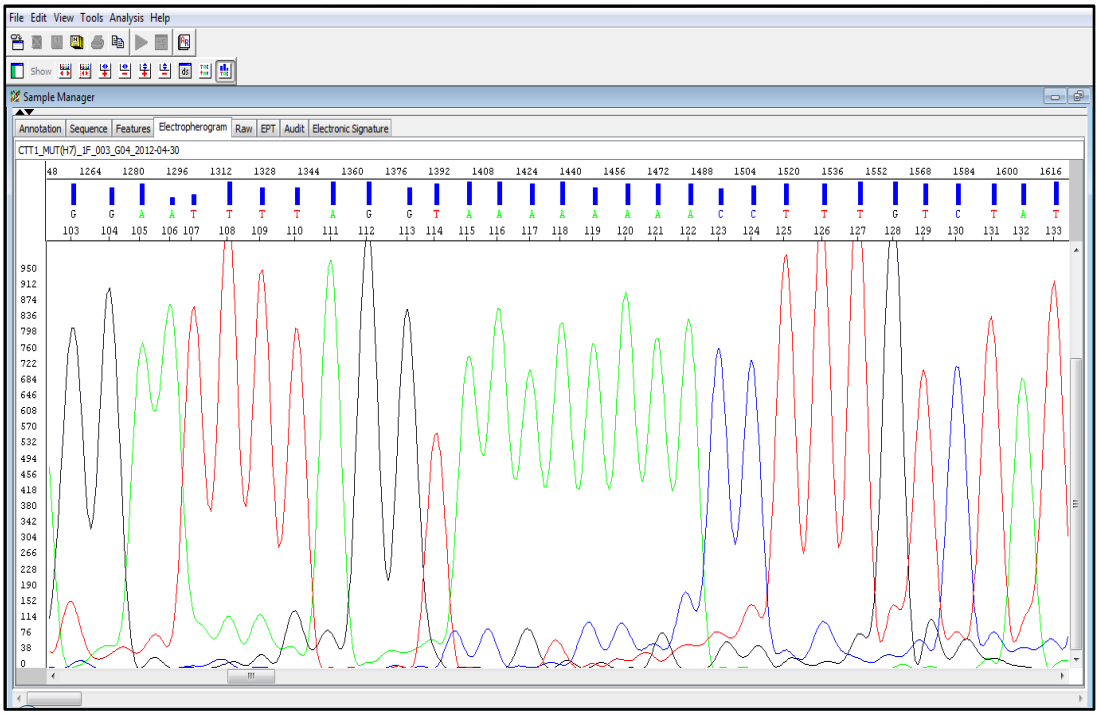

Figure 3.1 : Background peaks due to inadequate cleavage or removal of PCR primer.

Removal of the excess primers from PCR mixture is achieved by using commercial gel extraction kits which isolate amplification products (double strand amplicons) from agarose gel with high purity at the end of the electrophoresis. However, gel purification strategy was not used in this study, since their average recovery yield is not more than 80% [65]. On the other hand, enzymatic degradation of unused primers is a well-known and the most commonly used method in molecular studies, because it is straightforward, easy-to-use and offers high efficiency. Some commercial enzyme blends such as GML ExoSAP (GML, Switzerland) which offer optimal proportion of Alkaline Phosphatase and Exo I Nuclease handle the cleavage of all residuals from the reaction mixture. Alkaline Phosphatase is used for degradation of unused dNTPs, whereas Exo I nuclease directly cleaves single strand PCR primers in 3' → 5' direction, releasing deoxyribonucleoside 5'-monophosphates one after another.

Investigation of residual primers on agarose gel electrophoresis (mostly seen at size 20-30 bp) performed at the end of the PCR to confirm amplification is crucial to determine the necessary amount of the enzyme mixture that can degrade all residual primers. GML ExoSAP volume should be increased according to the intensity of residual primers. Although it is stated in the kit instruction that 2 µl ExoSAP is added

into 5 µl PCR product, as a result of the optimization, 2.5 µl GML ExoSAP mixture was used for each 5 µl PCR product to obtain clear sequencing data as shown in Figure 3.2.

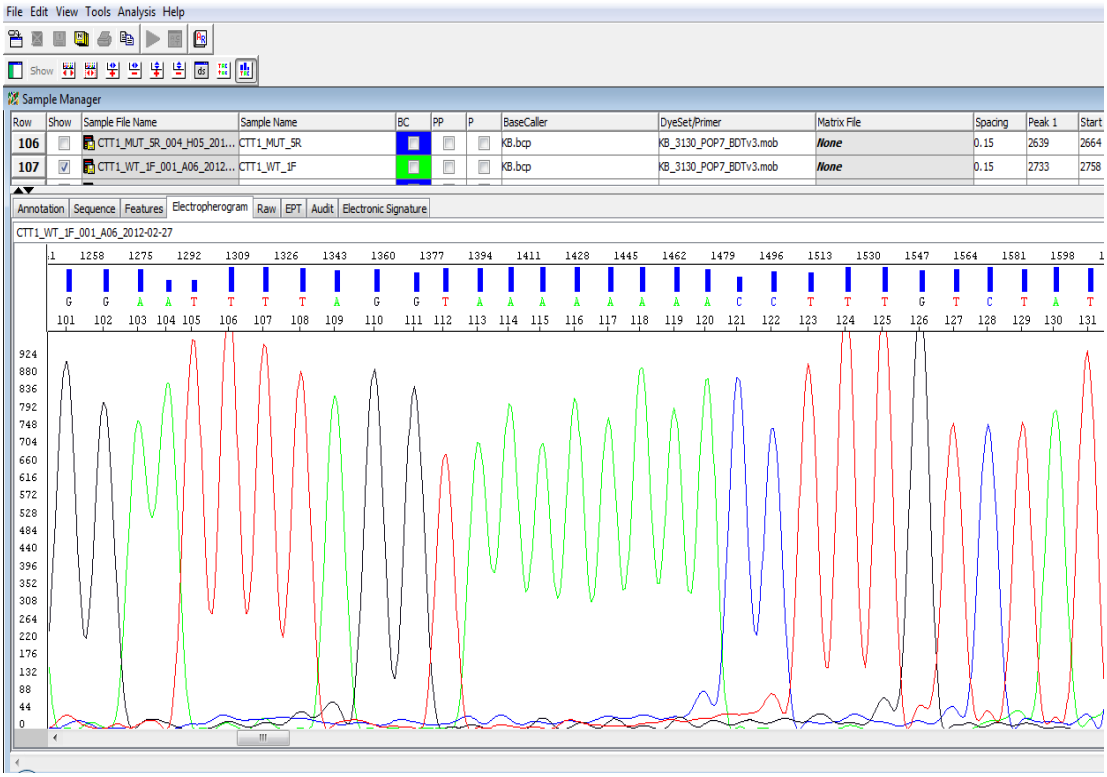

Figure 3.2 : Clear sequencing data via complete degradation of PCR primers.

Background peaks may appear due to several reasons including inappropriate cleavage of PCR primers. Thus, it may be difficult to identify the real cause of the background peaks. Another reason for obtaining noisy data, which highly resembles to phenomena owing to existence of PCR primers in the sequencing reaction, is the contamination of sequencing primer with n+1 or n-1 sequencing primer, as well as the presence of a run of identical nucleotides in the sequencing primer. Both reasons can cause same background pattern and that is quite confusing to understand the exact source of the problem. However, minor and valuable details present in the sequence data can help illuminate the real cause of the background patterns. If the problem is caused by sequencing primer, each background peak should belong to next main peak sequentially (Figure 3.3), thereby tail peaks in question cannot be observed under mononucleotide repeats, as indicated by black arrows in Figure 3.3.

This kind of minor sequencing problem would be easily fixed, if re-synthesized sequencing primer was used. However, all analyses were performed in the presence of the background peaks since it was not an obstacle for unambiguous analysis of sequence. Region of interest was confirmed with both forward and reverse sequencing data. In addition, it would create extra cost and would also lead to the time loss.

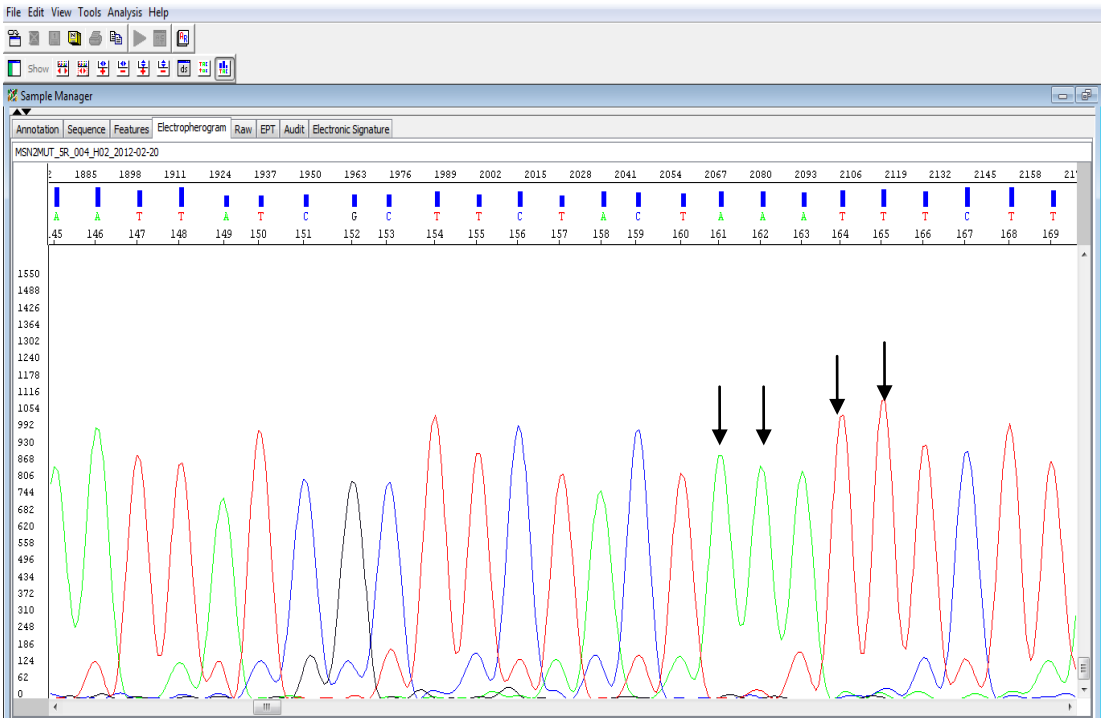

Figure 3.3 : Tail peaks due to contamination of sequence primer with n+1 or n-1.

In addition to the importance of the PCR product purification, optimization of the sequencing reaction is also essential to obtain clear sequence data. Use of excess amount of BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) which incorporates dNTPs and fluorescent dye labeled ddNTPs, or existence of residual ddNTPs during capillary electrophoresis due to lack of Sephadex purification causes peak pull-ups at the specific point.

Although dideoxynucleotide triphosphates have single nucleotide structure, molecular weight and molecular structure of dye labeled ddNTPs lead to unique mobility which is equivalent to 75-90 bp inside capillary filled with POP7 (Performance Optimize Polymer 7, Applied Biosystems), hence dye pull-ups were observed about 75-90 bp in the electropherogram (Figure 3.4). Avoidance of this background noise depends on the use of optimal concentration of cycle sequencing

mix, appropriate expansion of Sephadex powder via water and proper centrifugation force (rpm speed should be adjusted according to the rotor size of centrifuge)

The standard volume of cycle sequencing chemistry in the reaction mix which has been defined by manufacturer is 8 µl, however, sequencing reaction was optimized through the use of only 2 µl from cycle sequencing chemistry instead. This reduction provided an advantage to perform much more reactions using the same kit while decreasing the probability of residual ddNTPs at the end of the sequencing reaction.

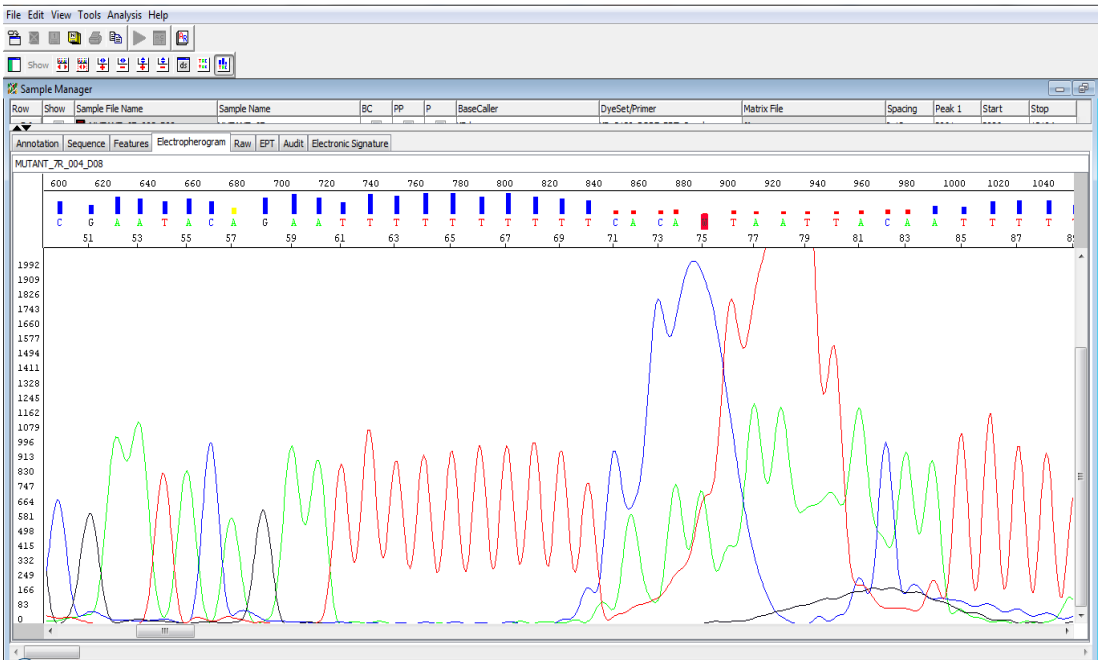

Figure 3.4 : Pulled-up peaks due to residual ddNTPs.

Ideal expansion of Sephadex molecules which absorb the water depends on the proportion of the mixture and storage time. Storage times shorter than 2 hours or longer than 4 days caused the formation of inappropriate pore size. Viscosity of the mixture was examined at certain time intervals to determine the exact proportion of powder in water and storage time. Optimal mixture was obtained by adding 1 g Sephadex powder into 14 ml molecular grade water and the suspension was stored at $+4^0$C for 4 hours after vigorously shaking to form an acceptable pore size.

It was also noticed that longer centrifugation times and slower rotation speeds could form more durable and rigid Sephadex column by gently removing the whole volume of water from the system. Therefore, it was decided to centrifuge the tubes at 4600 rpm for 3 min, instead of 5000 rpm for 2 min. Thus, the problem caused by residual

dye labeled ddNTPs was completely solved via optimization of all related steps (Figure 3.5).

Another factor which also impairs sequencing quality and integrity is the occurrence of DNA slippage due to long mono- or di-nucleotide repetitions in the region of interest. The exact mechanism of slippage is not known but it is accepted as a natural phenomenon of Sanger sequencing. Presumably, the two strands do not stay paired correctly during polymerization through the repetition region. This generates fragments with repetition regions of different length that have the same sequence after that region. The occurrence of slippage is length-dependent and short repetitions are rarely problematic in sequencing reactions. Slippage is more of a problem with T regions (A in the template strand) in BigDye® Terminator reactions due to the use of dUTP instead of dTTP in the deoxynucleotide mixture (Figure 3.6).

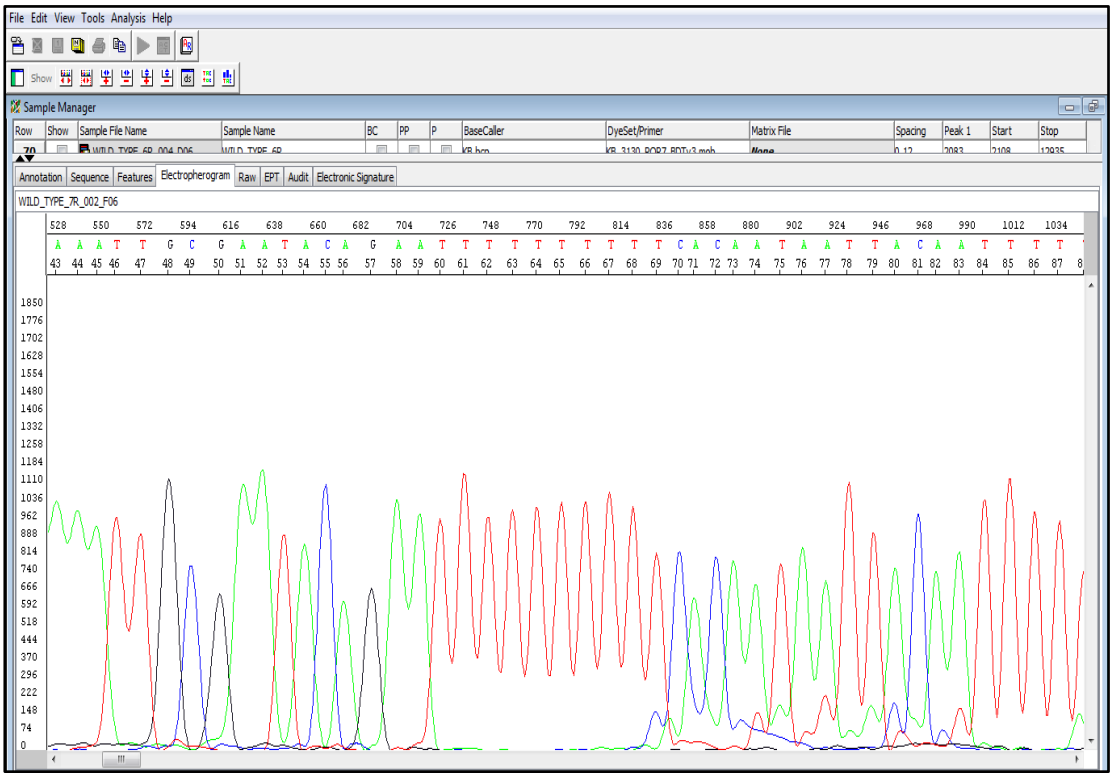

Figure 3.5 : Clear sequencing data via removal of residual ddNTPs.

Although the sequence data can be clean through the repetition region, the data after this region is noisy due to the presence of multiple sequences as indicated with a black arrow in Figure 3.6. Overlapping sequences in data are generally interpreted as heterozygous insertion or deletion. A more detailed analysis of the sequence data can

easily reveal whether these overlapping sequences are caused by mutation or repetition.

In the presence of heterozygous insertion or deletion, double peak morphology starts along with the mutation point whereas repetition-dependent sequence disruption occurs after the repetition, moreover, this pattern can be proven by performing the reverse sequencing reaction (Figure 3.7).

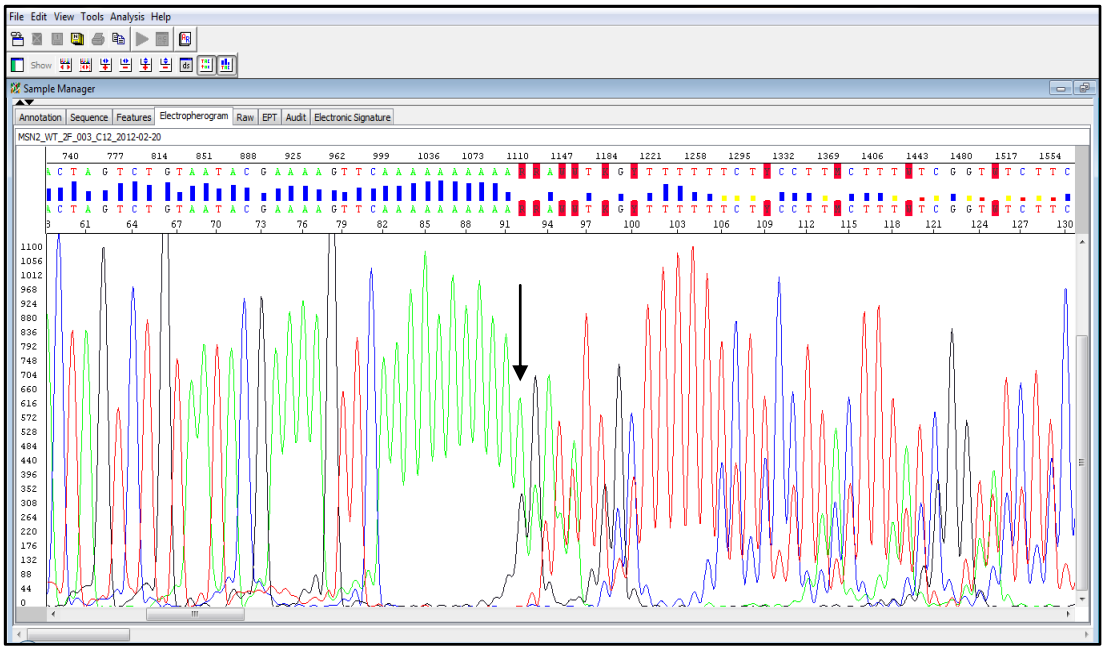

Figure 3.6 : Sequence disruption by mononucleotide repetition of Adenine (forward sequence)

Although there are alternative sequencing chemistries available, such as dRhodamine Cycle Sequencing Kit (Applied Biosystems, USA) which have been optimized to circumvent sequence disruption due to repetition terminators, this kit was not used in this study because there is no heterozygosity in these studies: all yeast strains used were haploid ones. Therefore, it was known that disruption of the sequence data was caused by nucleotide repetition and it was directly focused on obtaining the sequence integrity by performing both forward and reverse sequencing reaction for the region of interest which consists of mono- or di- nucleotide repetitions.

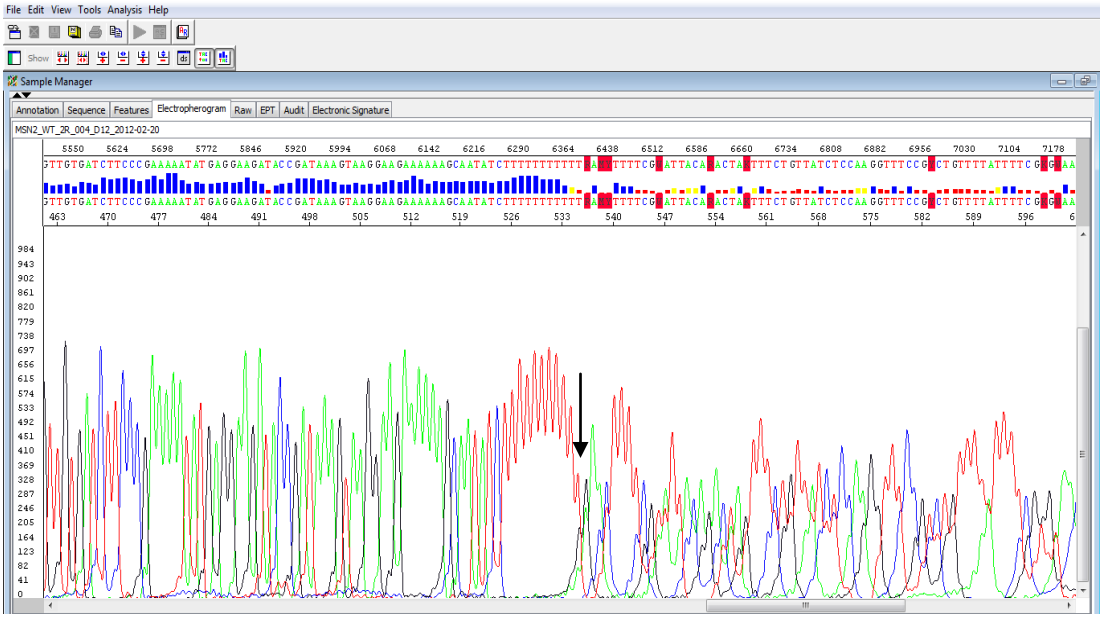

Figure 3.7 : Sequence disruption by mononucleotide repetition of Thymine (reverse sequence)

3.2 Genomic Comparison of S288c and CEN.PK 113-7D Reference Strains with CI25E and 8C Mutants, regarding PSE1, GRX4 and MSN5 Genes.

PCR was carried out as described in Section 2.2.3 with DNA extracted from *Saccharomyces cerevisiae* CEN.PK 113-7D reference strain (kindly provided by L. Benbadis, INSA-Toulouse), the cobalt-resistant mutant strain CI25E and the iron-resistant mutant strain 8C [50, 64]. Complete gene sequence of PSE1 (5' start point: -20, 3' end point: +246, Total length: 3539bp) was amplified in a fragmented way as 6 parts, GRX4 (5' start point: -237, 3' end point: +128, Total length: 1100bp) was amplified in a fragmented way as 2 parts and MSN5 (5' start point: -370, 3' end point: +165, Total length: 4210bp) was amplified in a fragmented way as 7 parts. The reason for the fragments is that the analysis capacity of the sequencing instrument permits to clearly read sequences up to 700bp length. PCR products of PSE1, GRX4 and MSN5 genes were confirmed using agarose gel electrophoresis as described in Section 2.2.4 to verify whether amplification intensity and product specificity were suitable for subsequent steps (Figure 3.8, 3.9, 3.10, respectively). PCR products purified with ExoSAP (GML, Switzerland) as described in Section 2.2.5 were sequenced with BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) as described in Section 2.2.6 and run onto 3130 Genetic Analyzer (Applied Biosystems, USA) after the Sephadex (GML, Switzerland) clean-up procedure as described in Sections 2.2.7 and 2.2.8, respectively. Sequence raw data in '.ab1' format were

obtained from the computer which was linked to Genetic Analyzer. All data were analyzed in Sequencing Analysis v.6.0 software (Applied Biosystems, USA) in order to investigate sequence quality and baseline clearness before sequence comparison.

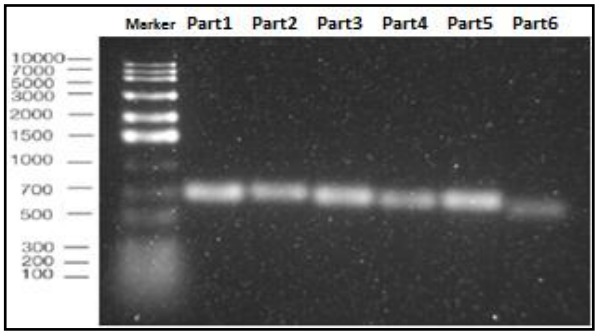

Figure 3.8 : DNA gel electrophoresis of PSE1 gene amplification products (CEN.PK 113-D).

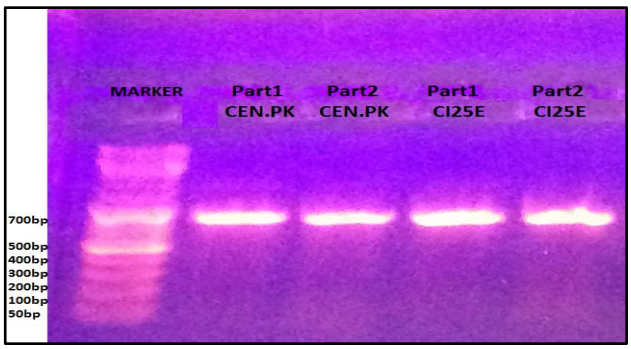

Figure 3.9 : DNA gel electrophoresis of GRX4 gene amplification products.

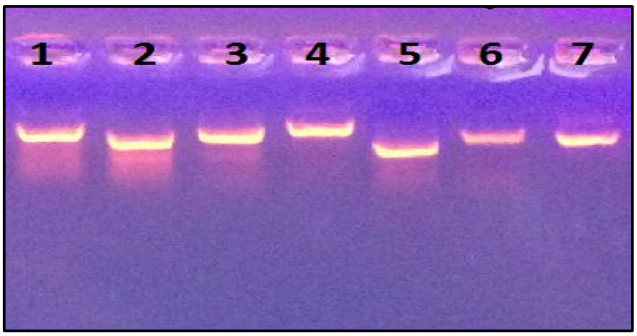

Figure 3.10 : DNA gel electrophoresis of MSN5 gene amplification products (CEN.PK 113-D). Product sizes (from lane 1 to 7): 729bp, 698bp, 721bp, 731bp, 684bp, 703bp, 689bp, respectively.

Sequence files of the reference strain (CEN.PK 113-7D) and the cobalt-resistant mutant (CI25E) were uploaded to projects that were created in the SeqScape v.3.0 software (Applied Biosystems, USA) to compare the sequences of PSE1 and GRX4 genes including their up and downstream regions as described in Section 2.2.9.

Another project was created for MSN5 gene to investigate possible mutations of CEN.PK 113-7D, mutant CI25E and mutant 8C. Reference nucleotide sequences which belong to another *S. cerevisiae* strain (S288C) used in nucleotide sequence databases were also downloaded from NCBI (National Center for Biotechnology Information) database with 'NC_001145.3', 'NC_001137.2' and 'NC_001136.3' access numbers for PSE1, GRX4 and MSN5 gene, respectively. Reference sequence files in 'GeneBank (.gb)' format were imported to designated projects. Report manager of the software did not indicate any kind of mutation for regions of interest at the end of the analysis using all necessary parameters. Manual inspection which was performed nucleotide by nucleotide also proved that *Saccharomyces cerevisiae* S288C, CEN.PK 113-7D and CI25E strains have the same genotype for PSE1 and GRX4 genes (3539bp/3539bp, 1100bp/1100bp, respectively), likewise, S288C, CEN.PK 113-7D, CI25E and 8C strains have also the same genotype for MSN5 gene (4210bp/4210bp).

3.3 Genomic Comparison of CEN.PK 113-7D Reference Strain with H7 Mutant for MSN2, MSN4 and OCA1 Genes.

PCR was carried out as described in Section 2.2.3 with DNA extracted from CEN.PK 113-7D reference strain (kindly provided by L. Benbadis, INSA-Toulouse) and the oxidative stress-resistant strain H7 [62]. Complete gene sequence of MSN2 (5' start point: -885, 3' end point: +123, Total length: 3118bp) was amplified in a fragmented way as 5 parts, OCA1 (5' start point: -1069, 3' end point: +123, Total length: 1909bp) was amplified in a fragmented way as 3 parts, whereas MSN4 (5' start point: -980, 3' end point: +106, Total length: 2980bp) was amplified in a fragmented way as 5 parts. PCR products of MSN2, MSN4 and OCA1 genes were confirmed by agarose gel electrophoresis (data not shown) as described in Section 2.2.4 and the same experimental strategy aforementioned in Section 3.2 was applied to obtain clear sequencing data. Sequencing Analysis v.6.0 software (Applied Biosystems, USA) was used for investigation of sequence quality and background clearness before comparison. Different projects were created on SeqScape v.3.0 (Applied Biosystems, USA) software for analysis of MSN2, MSN4 and OCA1, following this, relevant sequence files were imported to projects. Reference nucleotide sequences which belong to *S. cerevisiae* S288C strain used in nucleotide sequence databases were downloaded from NCBI (National Center for Biotechnology Information) database

with 'NC_001145', 'NC_001143.1' and 'NC_001146.8' access numbers for MSN2, MSN4 and OCA1, respectively. Downloaded reference files were imported to the appropriate projects. As a result of analysis, there were no nucleotide changes/differences between the H7 mutant and its reference strain CEN.PK 113-7D, in the region of interest for MSN2, MSN4 and OCA1 genes.

However, when the sequences of CEN.PK 113-7D and mutant strain H7 for these MSN2, MSN4 and OCA1 genes were compared to those of the other reference strain S288C, differences/changes in the nucleotide sequences were observed. It was found that there were two single nucleotide polymorphisms (SNP) in the MSN2 gene of the CEN.PK 113-7D and H7; first located at the position of -864 (T>C) in the 5' flanking region (Figure 3.11), while the second located at the position of c.356 (G>A, Ser119Asn), (Figure 3.12) the alteration of which resulted in a missense variant.

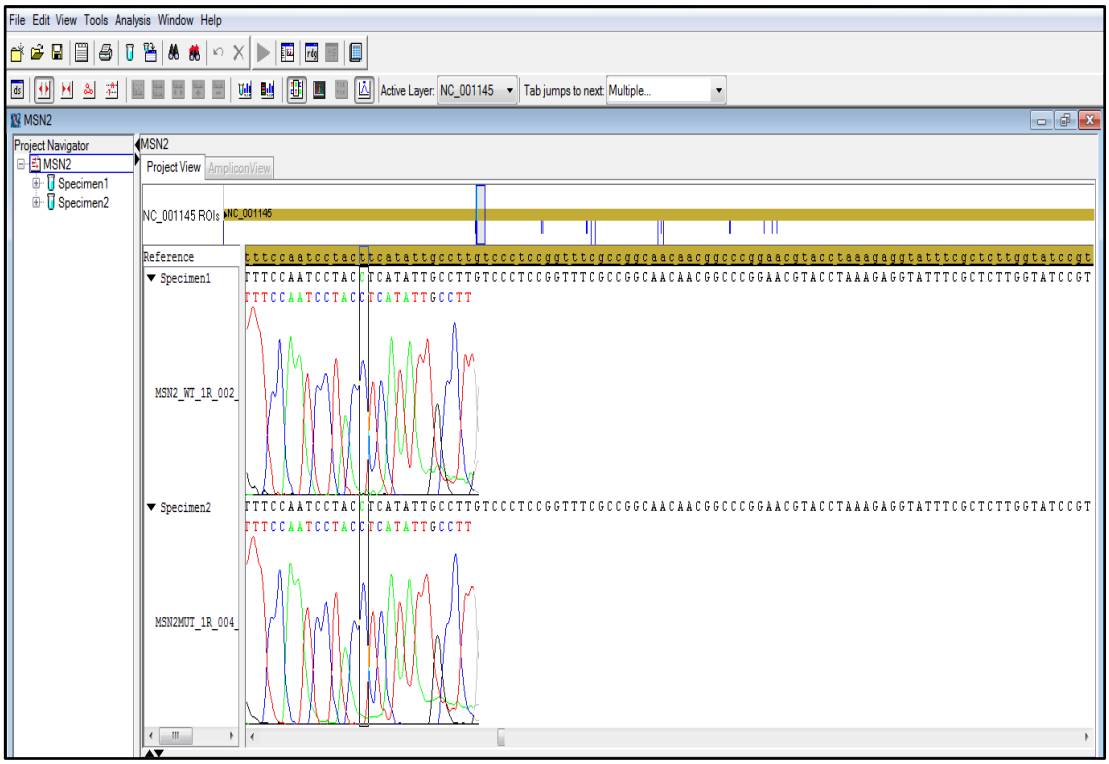

Figure 3.11 : SNP detection at 5' Flanking region (-864) of MSN2 gene using SeqScape v.3.0 software. CEN.PK 113-7D (upside, reverse sequence) and H7 (bottom, reverse sequence) have Cytosine (C), whereas S288C reference sequence (highlighted with ocher) has Thymine (T) base.

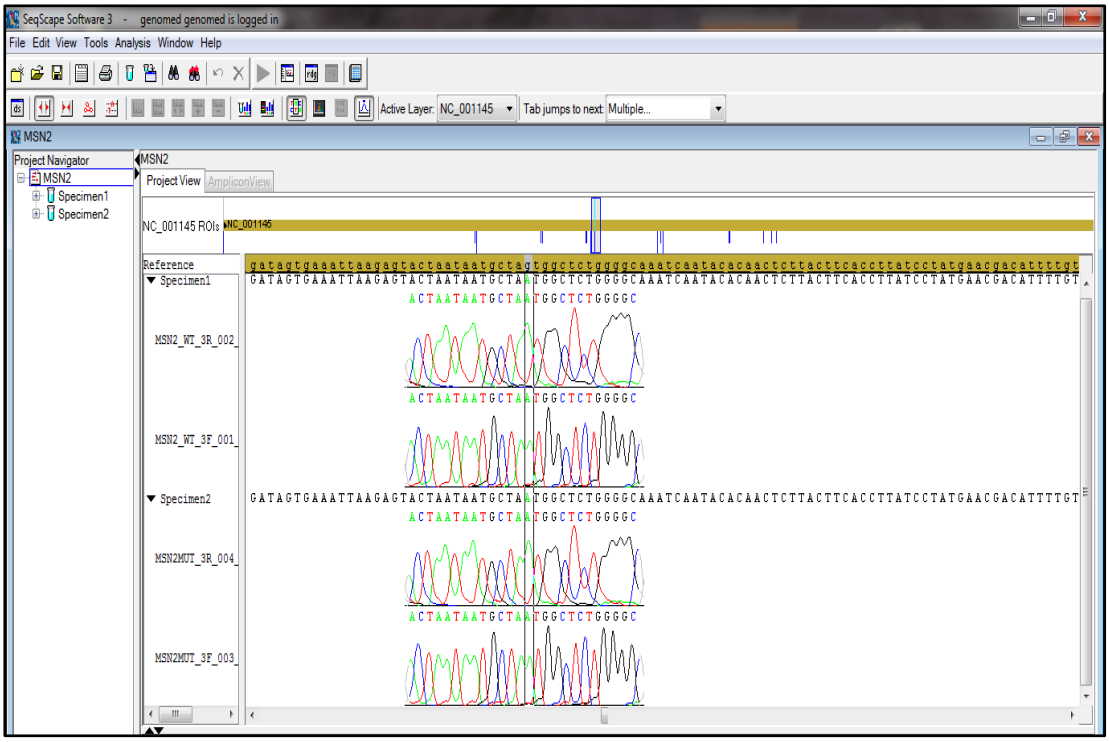

Figure 3.12 : SNP detection at c.356 of MSN2 gene using SeqScape v.3.0 software. CEN.PK 113-7D (upside, first row is reverse, second is forward sequence) and H7 (bottom, first row is reverse, second is forward sequence) have Adenine (A) whereas S288C reference sequence (highlighted with ocher) has Guanine (G) base.

MSN4 gene has also showed variations for CEN.PK 113-7D and H7 in comparison with S288C at a single nucleotide whose location is -209 (A>T) in the 5' flanking region (Figure 3.13).

In contrast to MSN2 and MSN4, the mutation which occurred in OCA1 gene has been stood out as an insertion of Adenine (A) nucleotide at the position of -115 (insA) in the 5' flanking region. The DNA slippage phenomenon which occurs in mono- or di-nucleotide repetitions and overlapping sequence pattern starting with the occurrence of heterozygous nucleotide deletion or insertion as described in 'Section 3.1' can be excluded in this case, as the yeast strains tested were haploids. Additionally, the presence of this Adenine nucleotide insertion was completely verified by both forward and reverse sequencing (Figure 3.14).

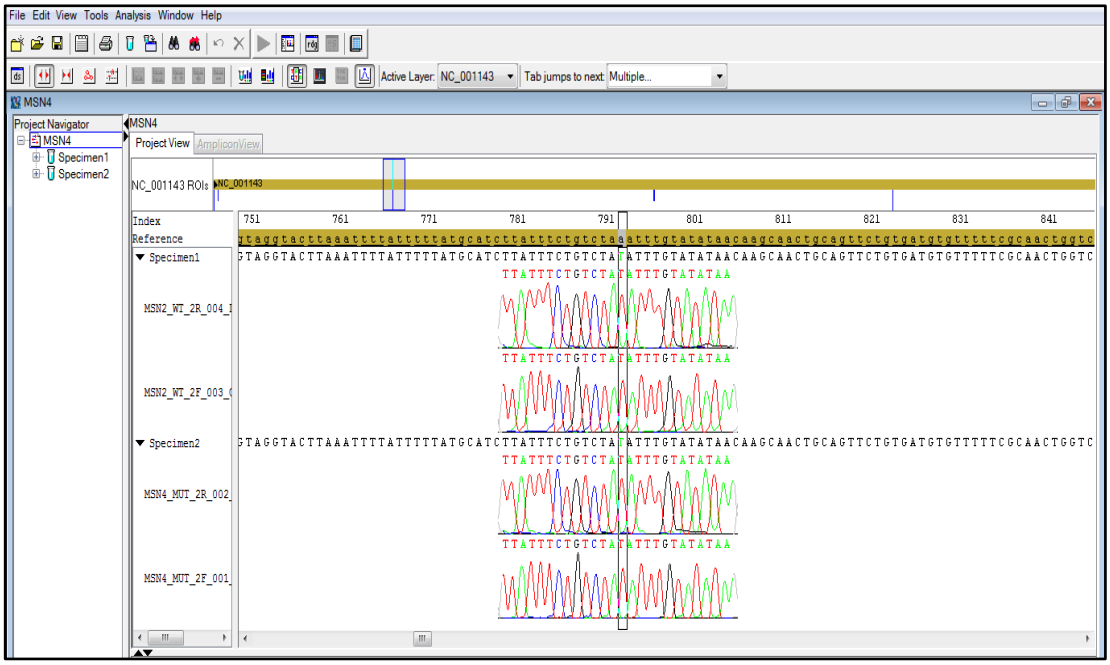

Figure 3.13 : SNP detection at 5' Flanking region (-209) of MSN4 gene using SeqScape v.3.0 software. CEN.PK 113-7D (upside, first row is reverse, second is forward sequence) and H7 (bottom, first row is reverse, second is forward sequence) have Thymine (T) base whereas S288C reference sequence (highlighted with ocher) has Adenine (A) base.

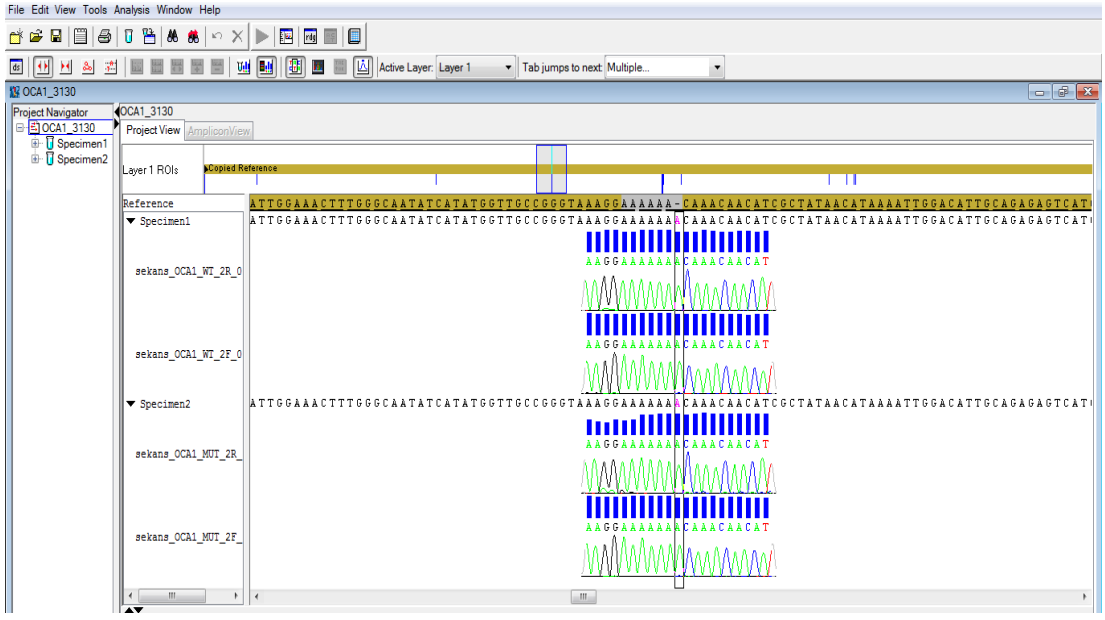

Figure 3.14 : Insertion of Adenine nucleotide in the 5' Flanking region (-115) of OCA1 gene. CEN.PK 113-7D (upside, first row is reverse, second is forward sequence) and H7 (bottom, first row is reverse, second is forward sequence), S288C reference sequence was highlighted with ocher.

Transcriptomic analysis of MSN2, MSN4 and OCA1 genes which was previously carried out using microarray and qPCR methods gave supporting information about expression level of the H7 mutant strain [32, 55]. It has been found that expression level of MSN2 and MSN4 genes was almost the same for reference and mutant strains under stress conditions performed in the presence of 0.5 mM and 1mM H_2O_2. Unlike MSN2 and MSN4, it has been reported that OCA1 expression of H7 as determined by qPCR was 1.7 fold of the reference strain upon continuous H_2O_2 stress application, however, this change was not regarded as a significant difference for microarray analysis [32]. As an overall result of comparative genotyping, qPCR and microarray studies, it can be concluded that MSN2, MSN4 and OCA1 genes do not seem to be directly related with the oxidative stress-resistant phenotype (H7) derived from CEN.PK 113-7D background, despite the 'general stress response' feature of these genes [55, 66].

It has been previously reported that genome characteristics of CEN.PK 113-7D are very similar to those of the S288C for genome parameters including total size, chromosome length, GC content, and the number of predicted genes [41, 67]. Another study involving restriction fragment length polymorphism (RFLP) and microarray techniques has shown that 96% of the genes represented on the arrays yielded a comparable signal, indicating a high degree of genetic relatedness according to a genomic comparison of these two strains [40]. On the other hand, many physiologically important nucleotide changes/differences in non-coding regions and subtle polymorphisms may be unnoticed in a rapid gene expression analysis with DNA microarrays [38, 40].

Unlike their genomic similarity, physiological characteristics of the two reference strains in both batch glucose- and galactose-supplemented fermentations were completely different. It was observed that CEN.PK 113-7D exhibited a 32% higher specific growth rate than S288C, while extracellular metabolic specific productivity rates were 32.6%, 392%, and 17.9% higher for ethanol, acetate, and glycerol production compared to S288C [34].

High-throughput sequencing studies also revealed that there are numerous nucleotide differences between the two strains, regarding the genes involved in key metabolic pathways which include carboxylic acid, organic acid, and carbohydrate metabolism, followed by nitrogen, amino acid, lipid, aromatic compound, and glycoprotein

metabolism [38, 41, 68]. Last but not least, it was shown that nucleotide changes that occurred in coding or non-coding regions of general stress response genes resulted in formation of specific characteristic for resistance against stress conditions in comparison of CEN.PK 113-7D with S288C strain [40, 69].

These may imply that CEN.PK 113-7D and its mutant strain H7 may be naturally more resistant than S288C strain to oxidative stress conditions, without requiring extra genetic alteration for MSN2, MSN4 and OCA1 genes, since their existing nucleotide configuration may provide higher tolerance against H_2O_2 than that of S288C. Besides, post-transcriptional regulation can also play an important role in protection of the CEN.PK 113-7D and its derived mutants against stress conditions better than that of S288C. Moreover, although it has not been the major focus of this thesis study, comparative genomic, transcriptomic and physiological analysis of S288C, CEN.PK 113-7D and H7 strains can give more detailed information about the influence of genetic alteration on the resistant phenotype.

3.4 Genomic Comparison of CEN.PK 113-7D Reference Strain with CI25E, H7 and M9 Mutant Strains for CTT1 Gene.

PCR was carried out as described in Section 2.2.3 with DNA extracted from CEN.PK 113-7D reference strain (kindly provided by L. Benbadis, INSA-Toulouse), the cobalt-resistant CI25E, the oxidative stress-resistant H7 and the nickel-resistant M9 mutant strains [50, 62, 63]. Complete gene sequence of CTT1 (5' start point: -945, 3' end point: +124, Total length: 2758bp) was amplified in a fragmented way as 5 parts. PCR products of CTT1 gene were confirmed by agarose gel electrophoresis (Figure 3.15) as described in Section 2.2.4 and the same experimental strategy aforementioned in Section 3.2 was applied to obtain clear sequencing data. Sequencing Analysis v.6.0 software (Applied Biosystems, USA) was used for investigation of sequence quality and background clearness before comparative genotyping of strains studied for CTT1 gene. Project was created in SeqScape v.3.0 (Applied Biosystems, USA) software for comparative analysis of CTT1 gene and sequence files were then imported to the CTT1 project. Reference nucleotide sequence which belongs to *S. cerevisiae* S288C strain used in nucleotide sequence databases was also downloaded from NCBI (National Center for Biotechnology Information) database with 'NC_001139.9', access number for CTT1 gene. However, reference sequence data was customized before downloading due to the fact that the

default section of the reference data did not cover the whole length of the region of interest that was defined to be compared. Reference sequence data was extended from start and end points as 300bp and was then downloaded in GeneBank format (.gb). Downloaded reference file was imported to the CTT1 project.

At the end of the analysis, a nucleotide change (-326 T>A in the 5' flanking region) was detected when comparing CEN.PK 113-7D with M9, whereas there was no nucleotide change/difference between the CTT1 sequences of H7, CI25E, CEN.PK 113-7D and the reference sequence of S288C (Figure 3.16).

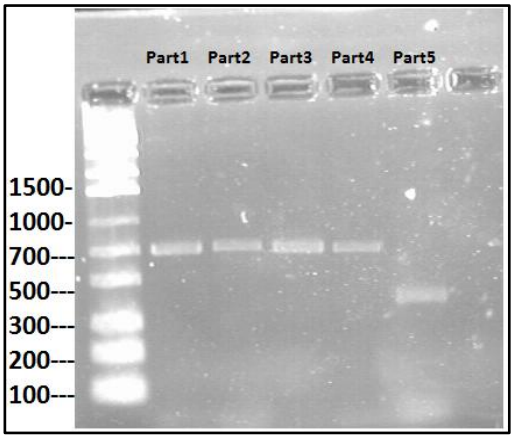

Figure 3.15 : DNA gel electrophoresis of CTT1 gene amplification products of CEN.PK 113-7D.

According to basic principles of evolutionary engineering, it was expected to find a genetic difference between the reference strain and the mutants. However, this nucleotide change indicated a difference between the reference strain CEN.PK 113-7D and the M9 mutant derived from it. On the other hand, it also revealed that M9 mutant strain had the same CTT1 sequence as the other reference strain S288C, as a result of this nucleotide change. M9 mutant and S288C contain Adenine (A) nucleotide at the position of -326 in the 5' flanking region, whereas CEN.PK 113-7D, H7 and CI25E strains contain T (Thymine) nucleotide.

Nevertheless, in addition to the nucleotide change at -326 (T>A) in the 5' flanking region, other differences were observed in 3 distinct nucleotide positions: -625 (A>G), c.150 (T>C, Ile50Ile) and c.1615 (T>G, Cys539Gly) on CTT1 gene sequence showed that all strains used in this study carried the same genotype, different from S288C reference sequence (Figures 3.17, 3.18 and 3.19, respectively).

These data provided strong evidence that M9 strain was indeed derived from the CEN.PK 113-7D reference strain, by the help of evolutionary engineering.

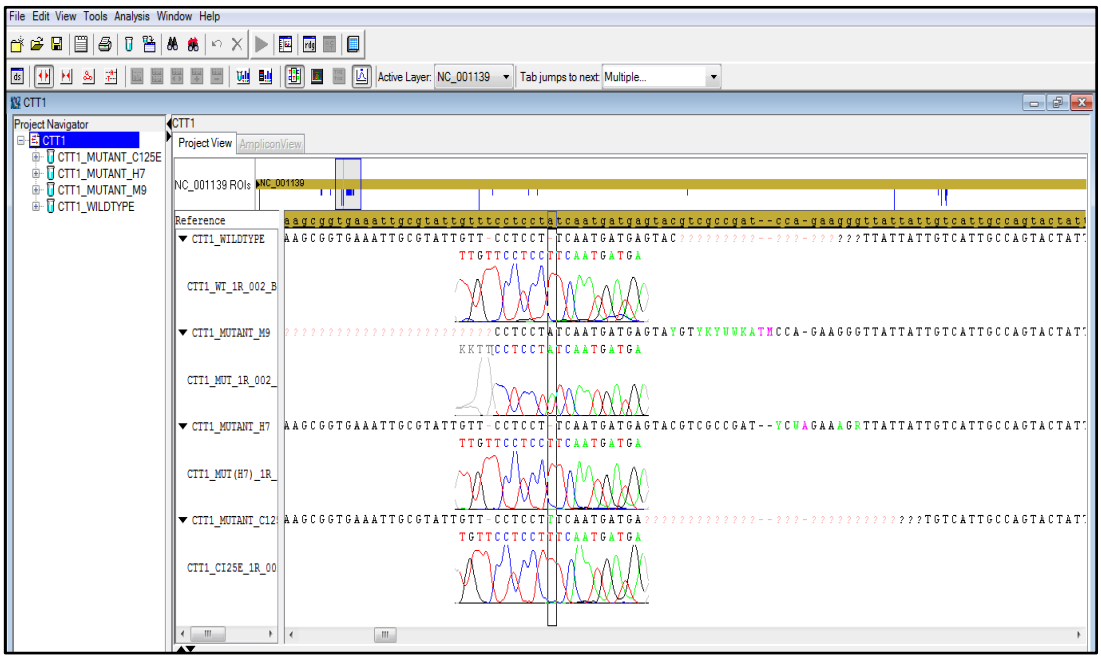

Figure 3.16 : Nucleotide change in the 5' Flanking region (-326) of CTT1 gene. CEN.PK 113-7D (first row), M9 (second row), H7 (third row) and CI25E (fourth row), S288C reference sequence was highlighted with ocher.

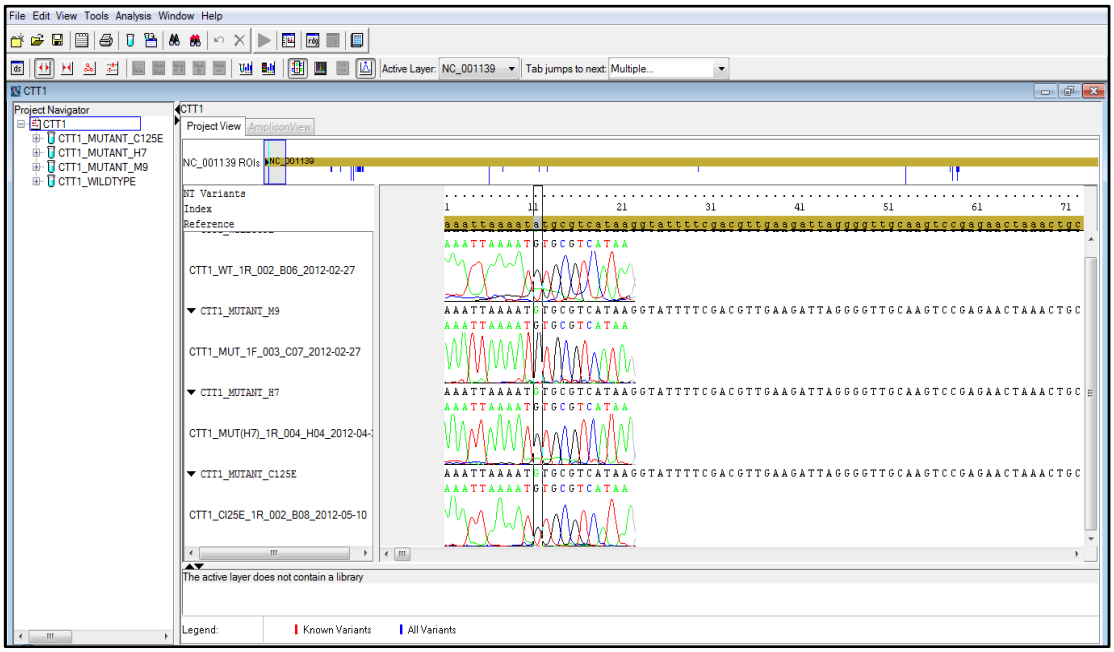

Figure 3.17: Nucleotide change in the 5' Flanking region (-625, A>G) of CTT1 gene. CEN.PK 113-7D (first row), M9 (second row), H7 (third row) and CI25E (fourth row), S288C reference sequence was highlighted with ocher.

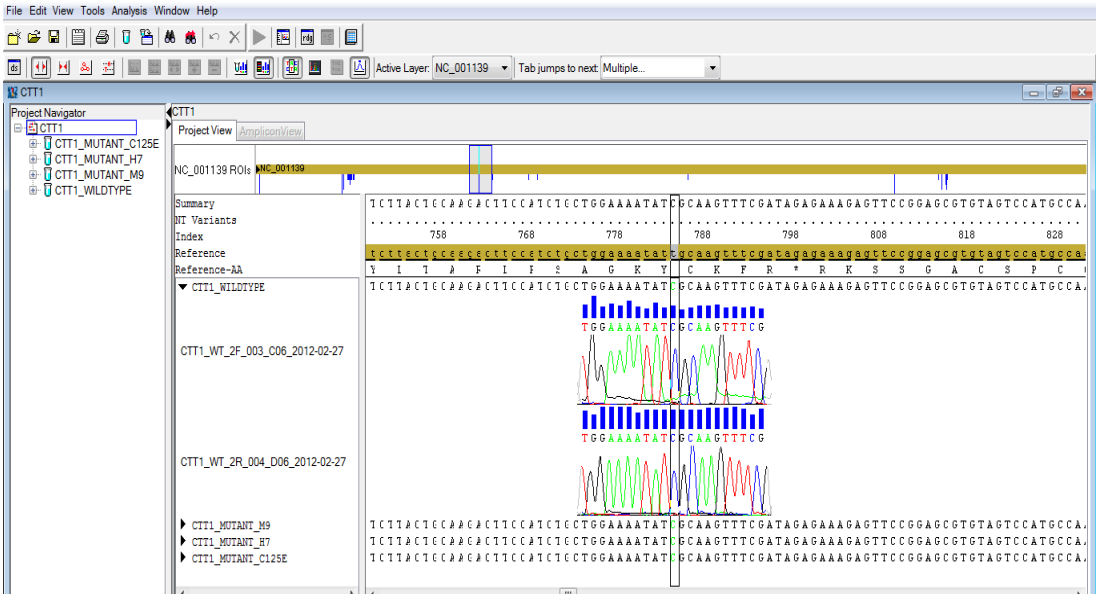

Figure 3.18 : Nucleotide change (c.150, T>C, Ile50Ile) in the CTT1 gene. Forward sequence (first row) and reverse sequence (second row) of CEN.PK 113-7D are shown by electropherogram, whereas sequences of M9, H7 and CI25E in FASTA format. S288C reference sequence was highlighted with ocher.

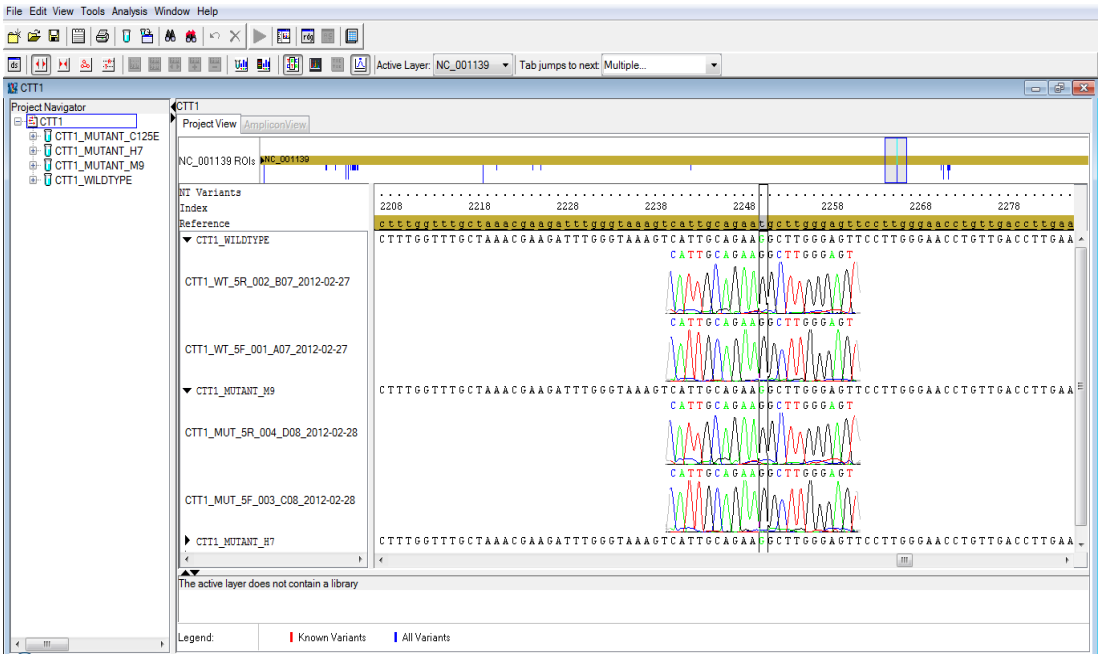

Figure 3.19 : Nucleotide change (c.1615 (T>G, Cys539Gly) in the CTT1 gene. Reverse and forward sequences of CEN.PK 113-7D and M9 were shown with electropherogram (1th, 2th, 3th, 4th rows, respectively) whereas sequence of H7 in FASTA format. S288C reference sequence was highlighted with ocher.

Transcription of the *Saccharomyces cerevisiae* CTT1 gene encoding the cytoplasmic catalase T is activated by a variety of stress conditions: it is derepressed by nitrogen starvation and induced by heat shock, and activated by osmotic and oxidative stress [63, 70].

A minimally addressed area in *S. cerevisiae* research is the mapping of transcription start sites (TSS). Mapping of TSS in *S.cerevisiae* is of importance to contribute to our understanding of gene regulation, transcription, mRNA stability and aspects of RNA biology [71].

It is known that nucleotide changes can be mainly effective on the change of protein structure if they cause the alteration of the amino acid chain during mRNA synthesis, whereas mutations occurring in regulatory regions such as Promoter region, upstream activator sequences (UAS), stress-response elements (STRE) may directly change mRNA expression levels thereby the production level of the related protein [72].

STRE sequences have been identified in many stress-induced genes and a computer search of the entire yeast genome predicts as many as 186 potential STRE-regulated genes [73]. However, the existence of a STRE-like sequence in the promoter region does not imply the functionality of this element [74]. A single copy of this sequence is mostly sufficient to induce a reporter gene by different types of stress, although two or more copies of this element provide a greater than additive effect on stress-induced gene expression [75].

It has been shown that CTT1 upstream region found to be involved in nitrogen, cAMP and heat control (upstream sequence at -382 to -325) contains different UAS elements (STREs) which are sufficient for the activation of a reporter gene by all types of stress acting on CTT1 [58]. Gel retardation experiments also demonstrated the existence of a factor specifically binding to STRE, but to a lesser extent to mutated elements having partly or entirely lost the ability to mediate stress control [58].

Expression analysis of CTT1 gene using microarray and qPCR methods supplied valuable information about the M9 mutant strain. qPCR experiment previously showed that M9 expressed CTT1 about 15-fold of the reference strain in the absence of nickel stress, and about 8-fold of the reference strain in nickel-containing growth medium [63]. Expression data for CTT1 in microarray experiments in the absence of

nickel stress also showed that CTT1 was overexpressed [63]. Based on CTT1 upregulation and the nature of the cytoplasmic catalase T, it can be expected that M9 might also have gained resistance to oxidative stress conditions. However, experimental work showed that M9 is surprisingly hypersensitive to oxidative stress [63].

The hypersensitive phenotype of M9 against oxidative stress can imply these facts: Firstly, the physiological characterization of the evolved yeast strains after evolutionary engineering should be carried out carefully, because improvement of resistance to a particular stress type by evolutionary engineering may result in sensitivity to another stress type, the so-called 'trade-off situation' in evolutionary engineering [51]. Secondly, there can be other factors than the CTT1 gene product which may confer resistance to oxidative stress, as mentioned previously [32, 63]. Thus, performing cross-resistance tests for other stress types could be useful to obtain more detailed stress response information on the evolved strains [54].

In light of DNA sequencing data results in this study for total gene structure including up and downstream regions, expression analysis of CTT1 gene, as well as information on molecular and functional characterization of CTT1 gene and its product obtained from different studies, it is likely that nucleotide change at -326 (T>A) in the UAS region of CTT1 gene could have changed nickel-stress tolerance of M9 strain. However, further complementary genomic, proteomic and functional characterization studies should be performed to test this hypothesis.

4. CONCLUSION

In this study, *Saccharomyces cerevisiae* CEN.PK 113-7D reference strain and the oxidative stress-resistant mutant H7, cobalt-resistant mutant CI25E, nickel-resistant mutant M9, and iron-resistant mutant 8C, that have all been obtained from CEN.PK 113-7D by *in vivo* evolutionary engineering strategy, were investigated for their genetic characteristics, regarding PSE1, GRX4, MSN2, MSN4, MSN5, OCA1 and CTT1 genes. With this aim, DNA sequencing by capillary electrophoresis method optimized and used for comparative genotyping of the reference and mutant strains for the regions of interest. Inspections of the nucleotide changes/differences were carried out in SeqScape v.3.0 software by creating designated project for each gene of interest.

Even though DNA sequencing by capillary electrophoresis method has been used for more than three decades, it should still be optimized in accordance to the sequence chemistries used and new sequence approaches. Any step which is skipped or any reagent which is used excessively or insufficiently may prevent obtaining the desired sequence data. This study showed that DNA sequencing by capillary electrophoresis method offers a gold standard by giving valuable information about nucleotide sequence of specific genes chosen. However, a more comprehensive study including numerous genes and strains to be compared will require new sequencing approaches like high-throughput genome sequencing. It will lead not only to very expensive operating costs, but also it will take months or even years to be completed if capillary electrophoresis strategy were to be used for large scale DNA sequencing studies.

Although the genes sequenced in this study were initially thought to be potentially important for conferring the mutant phenotypes obtained by evolutionary engineering, it was generally not possible to identify them as the main genes responsible for these specific phenotypes, based on gene expression and gene sequence data. For genetically complex phenotypes like stress resistance, multiple genes may simultaneously play an important role.

For both industrial and scientific purposes, many researchers use different reference strains of the yeast *Saccharomyces cerevisiae* in their studies. However, the present study revealed that selection of the most suitable reference strain is of great importance for the success of the study. Even if reference strains are phylogenetically close relatives, they can show remarkable differences at genomic level in terms of a specific characteristic. Otherwise, inappropriate selection of the reference strain may cause the problems in advanced stages of the study. In fact, as observed in this study, even these differences cannot be detected in exonic (coding region) region, nucleotide changes/insertions/deletions can occur in regulatory regions such as promoter/activator binding regions or splicing regions.

S288C has been used for systematic sequencing of the *S. cerevisiae* genome in scientific studies, whereas CEN.PK 113-7D is mainly used for physiological studies and functional genomics for both industrial and scientific purposes. Like other microorganisms, significant genotypic and phenotypic differences exist between different strains of *S. cerevisiae.* Even if *S. cerevisiae* genomic sequence information which is accessible via databases such as NCBI is obtained from the S288C background, especially for functional industrial studies, this strain background has some drawbacks [31].

Despite the fact that there are notably informative studies comparing genomic and/or physiological characteristics of S288C and CEN.PK 113-7D, there are no such studies comparing these reference strains along with their stress-resistant mutants for their genetic information and functional behavior in stress conditions [34, 38, 41]. Since both reference strains are extensively used in a variety of *S. cerevisiae* research studies, performing a comprehensive and comparative study which includes genetic, physiological and biochemical analyses of these strains has great importance.

The present study showed some genetic differences between both reference strains. These differences may significantly affect the results of physiological and genetic experiments to be performed using these reference strains. All genome-level differences between the reference strains used in yeast studies can be revealed by more comprehensive work using more powerful sequencing techniques which offer whole genomic information.

In this study, genetic comparative analysis was carried out between reference strain CEN.PK 113-7D and its derived mutant strains obtained by evolutionary engineering, and another reference strain S288C which was previously used in whole genome sequencing study of *S. cerevisiae* yeast [31]. On the other hand, previously obtained physiological information and results of experimental studies include data analysis which was obtained by comparing mutant strains with their CEN.PK 113-7D reference strain only.

Stress resistance and numerous physiological characteristics are too complex to be explained with only one or a few genes. Therefore, complementary experimental studies should be performed by using genomics, proteomics and metabolomics tools with the approach of systems biology to understand the molecular basis of these characteristics in more detail.

REFERENCES

[1] **Soon, W. W., et al.** (2012). High-throughput sequencing for biology and medicine. *Molecular Systems Biology*; 9:e640.

[2] **Gilbert, W. and Maxam, A.** (1973). The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences, USA*; 70:3581-3584.

[3] **Sanger, F.** (1975). Nucleotide sequences in DNA. *Proceedings of the Royal Society of London, Series B: Biological Sciences*; 191:317-333.

[4] **Gilbert, W. and Maxam, A.** (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences, USA*; 74:560-564.

[5] **Sanger, F., et al.** (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA*; 74:5463-5467.

[6] **Fritz, M. H.** (2011). Exploiting high throughput DNA sequencing data for genomic analysis, *PhD Thesis*, University of Cambridge, European Bioinformatics Institute, United Kingdom.

[7] **Applied Biosystems** (2009). DNA Sequencing by Capillary Electrophoresis. *Chemistry Guide, Second Edition*, Foster City, USA.

[8] **Karger, B. L. and Guttman, A.** (2009). DNA Sequencing by Capillary Electrophoresis. *Electrophoresis*; 30:196-202.

[9] **Smith, L. M., et al.** (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*; 321:674-679.

[10] **Soper, S.A., et al.** (1998). Sanger DNA-sequencing reactions performed in a solid-phase nanoreactor directly coupled to capillary gel electrophoresis. *Analytical Chemistry*; 70:4036-4043.

[11] **Glazer, A. N. and Mathies, R. A.** (1997). Energy-transfer fluorescent reagents for DNA analyses. *Current Opinion in Biotechnology*; 8:94-102.

[12] **Gelehrter, T. D., Collins, F. S., Ginsburg, D.** (1998). *The Principles of Medical Genetics. Second Edition*, p78.

[13] **Jorgenson, J. W. and Lukacs, K. D.** (1983). Capillary zone electrophoresis. *Science*; 222:266-272.

[14] **Kasper, T.J., et al.** (1988). Separation and detection of DNA by capillary electrophoresis. *Journal of Chromatography*; 458:303-312.

[15] **Drossman, H., et al.** (1990). High-speed separations of DNA sequencing reactions by capillary electrophoresis. *Analytical Chemistry*; 62:900-903.

[16] **St. Clair, R. L.** (1996). Capillary Electrophoresis. *Analytical Chemistry*; 68:569-586.

[17] **Applied Biosytems** (2010). *ABI 310 Genetic Analyzer User's Manual*. Foster City, USA.

[18] **Applied Biosytems** (2010). *ABI 3130 Series Systems Specifications*. Foster City, USA.

[19] **Shendure, J. and Ji, H.** (2008). Next-generation DNA sequencing. *Nature Biotechnology*; 26:1135-1145.

[20] **Levy, S., et al.** (2007). The diploid genome sequence of an individual human. *PLOS Biology*; 5:e254.

[21] **Metzker, M. L.** (2010). Sequencing technologies-the next generation. *Nature Reviews Genetics*; 11:31-46.

[22] **Schuster, S. C., et al.** (2008). Method of the year, next–generation DNA sequencing. Functional genomics and medical applications. *Nature Methods*; 5:11-21.

[23] **National Human Genome Research Institute**, [http://www.genome.gov/sequencingcosts], date retrieved: 19.05.2014.

[24] **Blazej, R. G., et al.** (2006). Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. *Proceedings of the National Academy of Sciences, USA*; 103:7240-7245.

[25] **Gresham, D., et al.** (2008). Comparing whole genomes using DNA microarrays. *Nature Review Genetics*; 9:291-302.

[26] **Soni, G. V. and Meller, A.** (2007). Progress toward ultrafast DNA sequencing using solid-state nanopores. *Clinical Chemistry*; 53:1996-2001.

[27] **Healy, K.** (2007). Nanopore-based single-molecule DNA analysis. *Nanomed*; 2:459-481.

[28] **Margulies, M., et al.** (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*; 437:376-380.

[29] **Schuster, S. C.** (2008). Next-generation sequencing transforms Today's biology. *Nature Methods*; 5:16-18.

[30] **Karathia, H., et al.** (2011). *Saccharomyces cerevisiae* as a model organism: a comparative study. *PLOS ONE*; 6:e16015.

[31] **Goffeau, et al.** (1996). Life with 6000 Genes. *Science*; 274:546-567.

[32] **Kocaefe, N.** (2012). Molecular characterization of oxidative stress resistant yeast, *M.Sc Thesis*, Istanbul Technical University, Institute of Science&Technology, Turkey.

[33] **Foury, F.** (1997). Human genetic diseases: a cross-talk between man and yeast. *Gene*; 195:1-10.

[34] **Otero, J. M., et al.** (2010). Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications, *BMC Genomics*; 11:e723.

[35] **Beck, H., et al.** (2008). *Saccharomyces kluyveri* as a model organism to study pyrimidine degradation, *FEMS Yeast Research*; 8:1209-1213.

[36] **Sherman, F.** (2002). Getting started with yeast. *Methods in Enzymology*; 350:3-41.

[37] **Mortimer, R. K. and Johnston, J. R.** (1986). Genealogy of principal strains of the yeast genetic stock center. *Genetics*; 113:35-43.

[38] **Schacherer, J., et al.** (2007). Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLOS ONE*; 2:e322.

[39] **Carreto, L., et al.** (2008). Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC Genomics*; 9:e524.

[40] **Daran-Lapujade, P., et al.** (2003). Comparative genotyping of the *Saccharomyces cerevisiae* laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays. *FEMS Yeast Research*; 4:259-269.

[41] **Carter, D.** (2008). Saccharomyces Genome Resequencing Project: User Manual. [http://www.sanger.ac.uk/Teams/Team118//sgrp/], date retrieved: 24.05.2014.

[42] **Nijkamp, J. F., et al.** (2012). De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microbial Cell Factories*; 11:e36.

[43] **Liti, G., et al.** (2009). Population genomics of domestic and wild yeasts. *Nature*; 458:337-341.

[44] **Lukjancenko, O., et al.** (2010). Comparison of 61 sequenced *Escherichia coli* genomes. *Microbial Ecology*; 60:708-720.

[45] **Daran-Lapujade, P., et al.** (2009). An atypical PMR2 locus is responsible for hypersensitivity to sodium and lithium cations in the laboratory strain *Saccharomyces cerevisiae* CEN.PK 113-7D. *FEMS Yeast Research*; 9:789-792.

[46] **Fendt, S. M., et al.** (2010). Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. *Molecular Systems Biology*; 6:e432.

[47] **Van Dijken, J. P., et al.** (2000). An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme and Microbial Technology*; 26:706-714.

[48] **Bailey, J.E.** (1991). Toward a science of metabolic engineering. *Science*; 252:1668-1675.

[49] **Bailey, J. E., et al.** (1996). Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnology and Bioengineering*; 52:109-121.

[50] **Alkim, C.** (2012). Evolutionary engineering and molecular characterization of stress-resistant *Saccharomyces cerevisiae* mutants, *PhD Thesis*, Istanbul Technical University, Institute of Science&Technology, Turkey.

[51] **Cakar, Z. P., et al.** (2012). Evolutionary engineering of *Saccharomyces cerevisiae* for improved industrially important properties. *FEMS Yeast Research*; 12:171-182.

[52] **Cakar, Z. P.** (2009). Metabolic and evolutionary engineering research in Turkey and beyond. *Biotechnology Journal*; 4:992-1002.

[53] **Butler, P. R., et al.** (1996). Improvement of antibiotic titers from *Streptomyces* bacteria by interactive continuous selection. *Biotechnology and Bioengineering*; 49:185-196.

[54] **Cakar, Z. P., et al.** (2005). Evolutionary engineering of multiple-stress resistant *Saccharomyces cerevisiae*. *FEMS Yeast Research*; 2:569-578.

[55] **Martinez-Pastor, M. T., et al.** (1996). The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO Journal*; 15:2227-2235.

[56] **Hohmann, S. and Mager, W. H.** (2003). Yeast Stress Responses. *Springer*; Vol.1.

[57] **Alic, N., et al.** (2003). Lipid hydroperoxides activate the mitogen-activated protein kinase Mpk1p in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*; 278:41849-41855.

[58] **Marchler, G., et al.** (1993). A *Saccharomyces cerevisiae* UAS element controlled by protein kinase A activates transcription in response to a variety of stress conditions. *The EMBO Journal*; 12:1997-2003.

[59] **Pujol-Carrion, N. and Angeles, M.** (2010). Grx4 and Grx3 of *Saccharomyces cerevisiae* play a role in actin dynamics through their Trx domains, which contributes to oxidative stress resistance. *Applied and Environmental Microbiology*; 76:7826-7835.

[60] **Kaffman, A., et al.** (1998a). Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121. *Genes and Development*; 12:2673-2683.

[61] **Kaffman, A., et al.** (1998b). The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus. *Nature*; 396:482-486.

[62] **Yilmaz, B.** (2009). Investigation of oxidative stress response in yeast and rat, *M.Sc Thesis*, Istanbul Technical University, Institute of Science&Technology, Turkey.

[63] **Kucukgoze, G., et al.** (2013). Evolutionary engineering and transcriptomic analysis of nickel-resistant *Saccharomyces cerevisiae*. *FEMS Yeast Research*; 13:731-746.

[64] **Balaban, B. G.** (2010). Evolutionary engineering of iron-resistant *Saccharomyces cerevisiae* for biomimetics, *M.Sc Thesis*,

Istanbul Technical University, Institute of Science&Technology, Turkey.

[65] **Dorman, N. and Patrick, C. H.** (2010). Citations. *BioTechniques*; 49:615-617.

[66] **Helen C., et al.** (2001). Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell*; 12:323-337.

[67] **Rougemonet J., et al.** (2008). Probabilistic case calling of Solexa sequencing data. *BMC Bioinformatics*; 9:e431.

[68] **Hernandez D., et al.** (2008). *De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Research*; 18:802-809.

[69] **Gasch, A. P., et al.** (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*; 11:4241-4257.

[70] **Bissinger, P. H., et al.** (1989). Control of *Saccharomyces cerevisiae* catalase T gene (CTT1) expression by nutrient supply via the RAS-cyclic AMP pathway. *Molecular Biology of the Cell*; 9:1309-1315.

[71] **Zhang, Z. and Dietrich F. S.** (2005). Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Research*; 33:2838-2851.

[72] **Chambers, A., et al.** (1988). The UAS of the yeast PGK gene is composed of multiple functional elements. *Nucleic Acids Research*; 16:8245-8260.

[73] **Treger, J. M., et al.** (1998). Transcriptional factor mutations reveal regulatory complexities of heat shock and newly identified stress genes in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*; 273:26875-26879.

[74] **Alepuz, P. M., et al.** (1997). Glucose repression affects ion homeostasis in yeast through the regulation of the stress-activated ENA1 gene. *Molecular Microbiology*; 26:91-98.

[75] **Kobayashi, N. and McEntee, K.** (1993). Identification of cis and trans components of a novel heat shock stress regulatory pathway in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell*; 13:248-256.

CURRICULUM VITAE

Name Surname: Cihan Erdinç GÜLSEV

Place and date of birth: Adana, 04/05/1986

Permanent Address: Hasanpaşa Mah., Sina Sok. 10/3, 34710 Kadıköy/Istanbul, Turkey

E-mail: erdincgulsev@yahoo.com

EDUCATION:

2004-2009 *Bachelor of Science* (Bioengineering), Ege University, Bornova, Izmir, Turkey.

PUBLICATIONS:

Göral, C., Korkmaz, A. K., Kanıt, N., **Gülsev, E.**, Alkım, C., Çakar Z. P. (2012). Genetic and Physiological Characterization of a SO_2-Resistant *Saccharomyces cerevisiae* Mutant. Poster Presentation. *1th International Congress of the Molecular Biology Association of Turkey*. 23-24 November, Istanbul, Turkey. Abstract book, page: 70, poster no: 27.