TRANSCRIPTIONAL PROFILE OF AN ANTARCTIC YEAST, *Rhodotorula* sp. USM-PSY62, IN RESPONSE TO TEMPERATURE CHANGES

CLEO NICOLE CHAI WEI SUN

UNIVERSITI SAINS MALAYSIA

2015

TRANSCRIPTIONAL PROFILE OF AN ANTARCTIC YEAST, *Rhodotorula* sp. USM-PSY62, IN RESPONSE TO TEMPERATURE CHANGES

By

CLEO NICOLE CHAI WEI SUN

Thesis submitted in fulfillment of the requirements

for the degree of Master of Science

JUNE 2015

ACKNOWLEDGEMENTS

My sincere and heartfelt gratitude as well as appreciation goes to Professor Dr. Nazalan Najimudin. Thank you for your inspirations that brought me into a wonderful journey of exploration and discovery in Science. None could be achieved without your perseverance and your dedication in training and teaching towards developing fine young researchers. A special note of appreciation also goes to Professor Razip Samian for his advice and support throughout this study.

The time has passed so quickly that it has been so many years that I have stayed with Lab 414. I am greatly indebted to all the members of Lab 414, especially my seniors: Abg Chai, Kak Aini, Kak Kem, Luan, Fey and Syafiqi. Not forgetting the other lab members as well. Thank you all for the support, wonderful experience, cheerfulness and warmth brought to the family of Lab 414.

Special thanks to my best friend, Flory, for being there for me throughout the years during the ups and downs in my life. Thank you also to my special one, Abe, thank you for everything!

Lastly to my beloved family, I really appreciate the endless patience, moral support and understanding that you have put on me in finishing my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

%	percentage
˚C	degree Celcius
bp	base pair
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
et al	and others
NGS	next generation sequencing
Csp	cold shock protein
Cap	cold acclimatization protein
AFP	antifreeze protein
PCR	polymerase chain reaction
RIN	RNA integrity number
OD	optical density
mRNA	messenger RNA
µg	microgram
µl	microliter
DEPC	diethylpyrocarbonate
ng	nanogram

PROFIL TRANSKRIPSI YIS ANTARTIK, *Rhodotorula* sp. USM-PSY62 SEBAGAI RESPONS TERHADAP PERUBAHAN SUHU

Abstrak

Rhodotorula sp. USM-PSY62 ialah yis bersifat psikrofilik yang mampu hidup di persekitaran sejuk. Yis ini boleh menjadi organisma contoh yang baik untuk kajian mekanisme molekul intrinsik yang terlibat dalam penyesuaian dan adaptasi dalam keadaan sejuk. Dalam kajian ini, teknik RNA-seq telah digunakan untuk mengkaji perubahan transkripsi global *Rhodotorula* sp. USM-PSY62 terhadap perubahan suhu. Sel-sel yis ini pada mulanya telah dikultur pada suhu optimum 15 °C dan kemudiannya dipindah ke suhu 21 °C, 5 °C dan 0 °C untuk merangsang respons adaptasi yis tersebut terhadap kejutan haba dan sejuk. Sel-sel yis daripada empat suhu pengkulturan kemudiannya dituai dan analisis RNA-seq menggunakan platform Illumina telah dijalankan. Data RNA ekoran daripada analisis tersebut telah disusun kembali secara "*de novo*" menggunakan program yang dikenali sebagai Trinity. Analisis perbezaan ekspresi juga telah dijalankan. Sejumlah 286 transkrip telah didapati menunjukkan perbezaan ekspresi yang nyata untuk sekurang-kurangnya empat kali ganda apabila perbandingan di kalangan sel daripada empat suhu tersebut dilaksanakan. Sebanyak 72 transkrip telah dianotasi dengan fungsi-fungsi masing-masing dan dibahagikan kepada tiga kumpulan berfungsi utama berdasarkan kumpulan-kumpulan orthologous eggNOG: 38 transkrip (Metabolisme); 17 transkrip (Proses-proses Selular dan Pengisyaratan); 17 transkrip (Penyimpanan Maklumat dan Pemprosesan). Sebagai kesimpulan daripada perhatian perubahan-perubahan transkripsi *Rhodotorula* sp. USM-PSY62 terhadap perubahan suhu, adaptasi terhadap kejutan sejuk yis ini termasuk kemotilan sel yang ditingkatkan, pengeluaran helicase secara berlebihan untuk mengurangkan kestabilan DNA dan RNA, peningkatan pengambilan ion serta pengaktifan mekanisme transduksi isyarat yang berkaitan dengan adaptasi persekitaran sejuk. Selain itu , transkrip yang berkait rapat dengan pengangkutan membran dan organisasi kromosom juga

telah diturunkan pengawalaturannya sebagai reaksi terhadap persekitaran sejuk. Respons kejutan haba telah diperhatikan dengan peninggian pengawalaturan heat shock factor protein 1 apabila suhu persekitaran meningkat. *Rhodotorula* sp. USM-PSY62 telah membentuk pelbagai mekanisme adaptasi yang tersendiri untuk melindungi sel-selnya daripada ancaman stress suhu.

TRANSCRIPTIONAL PROFILE OF AN ANTARCTIC YEAST, *Rhodotorula* sp. USM-PSY62, IN RESPONSE TO TEMPERATURE CHANGES

Abstract

Rhodotorula sp. USM-PSY62 is a psychrophilic yeast that is competent in withstanding and surviving in extreme cold environment. It is a good model organism to study the intrinsic molecular mechanisms that are involved in cold adaptation. In this study, a robust RNA-seq approach was employed to study the global transcriptomic response of *Rhodotorula* sp. USM-PSY62 upon temperature shifts. The cells were initially cultured at optimum growth temperature of 15 °C and subsequently shifted to 21 °C, 5 °C and 0 °C in order to induce the cold and heat adaptive responses of cells toward heat and cold shock conditions, respectively. The cells were then harvested from the four growth temperatures and subjected to RNA-seq analysis using Illumina platform. The resultant RNA reads were *de novo* assembled using Trinity and the differential expression analysis were executed. A total of 286 transcripts were found to be differentially expressed for at least four-fold when a pair wise comparison amongst four temperature conditions was performed. A number of 72 out of 286 transcripts have been functionally annotated and divided into three main functional groups based on the eggNOG orthologous groups: 38 transcripts (metabolism); 17 transcripts (cellular processes and signalling); 17 transcripts (information storage and processing). As deduced from the transcriptional alterations, cold shock responses of *Rhodotorula* sp. USM-PSY62 include enhanced cell motility, overproduction of helicase to reduce the increased stability of DNA and RNA, risen of ion intake as well as activation of a cold related signal transduction mechanism. Besides, transcripts associated with membrane trafficking and chromosome organization were also down-regulated as a response to the cold. Heat shock response was observed with the up-regulation of heat shock factor protein 1 when the growth temperature increased. *Rhodotorula* sp. USM-PSY62 has developed wide-

ranging and distinctive adaptive mechanisms to protect its cells from damage and to survive under thermal stresses.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Deep oceans (covering more than 70% of the Earth's surface), the Arctic and Antarctic polar regions, and high mountain ranges are considered as the main cold regions of the Earth. Antarctica is generally known as the coldest continent. It covers an estimated area of 14 million square kilometres, making it larger than Australia or the continent of Europe. Antarctica is all but about 1% buried by ice and snow. Nonetheless, there are several different environmental niches that can be found such as different types of soil, sediments, rocks, as well as snow and ice. These niches differ in their characteristics as well as the behaviour and pattern of its surrounding temperature, status of nutrient, water activity and salinity (Russell, 2006).

It has been reported that environments that are cold are normally colonized by microorganisms which include archae, yeasts, fungi, cyanobacteria, Gram-negative and Gram-positive bacteria, and protists. These microorganisms become representatives of the most abundant cold-adapted living things on earth with a high level species diversity and biomass (Feller & Gerday, 2003). These microorganisms are known as psychrophiles due to their cold-loving nature. In order to survive in extreme environments, especially with temperature playing as one of the most crucial survivability factor for psychrophiles, Antarctic microbes require a huge range of adaptive ability to retain their metabolic rates and maintain growth that is compatible with life (Morgan-Kiss *et al.*, 2006). In present times of ozone depletion and global warming, it is essential to understand microbial life in Antarctica and how the associated effects of increasing temperature may challenge this diversity of life. It is also of considerable significance how psychrophilic microbes survive and adapt to temperatures below and above their growth temperature range (Deegenaars & Watson, 1998).

Various studies were carried out to study heat and cold shock responses in Antartic organisms such as plants (Gidekel *et al.*, 2003), fish (Hofmann *et al.*, 2000; Chen *et al.*, 2008), insect (Rinehart *et al.*, 2006), fungi (Vishniac, 2006; Gocheva et *al.*, 2009), bacteria (Médigue *et al.*, 2005) and yeast (Deegenaars & Watson 1997; Deegenaars & Watson 1998; Boo *et al.*, 2013) in the effort of further understanding the response of these organisms towards a shift in temperature.

In this study, the organism of interest is a yeast isolated from Antarctic sea ice samples collected at Casey Station, Antarctica (by the late Mr. Omar Pozan during the austral summer of year 2001-2002). It has been identified as *Rhodotorula* sp. USM-PSY62 (Ong, 2006). Experimentally, it is psychrophilic in nature and it has an optimum growth temperature of 15 °C. In order to comprehend and expand the knowledge of cold and heat shock responses of Antarctic yeast, a transcriptomic approach using Next Generation Sequencing (NGS) technologies was employed to study the heat and cold shock responses of *Rhodotorula* sp. USM-PSY62 when the cells were exposed to cold shock (0 °C), mild cold shock (5 °C) and heat shock (21 °C) conditions. The heat shock condition was set to at least 5 °C above the optimum growth temperature because for organisms such as psychrophiles, the maximum response of heat shock could occur within this range (Lindquist, 1986). Previously, a similar study was conducted by Boo *et al.* (2013) where the cold and heat shock responses of Antarctic yeast, *Glaciozyma antarctica* PI12 was investigated under the same heat and cold shock conditions using qPCR. The recent advances and the advantages of Next Generation Sequencing (NGS) technologies compared to the traditional qPCR method have driven the choice of technique utilized in this study. NGS have unlocked the path to study any organism at the genetic and molecular level at an unparalleled scale and speed (Haas, 2013). Hence, transcriptome sequencing (RNA-seq) was utilized in this study to capture both coding and multiple forms of non-coding RNA which enable researchers to determine when and where genes are turned on or off in various types of cells and tissues. The number of transcripts can be quantified to get some idea of the amount of gene activity

or expression in a cell. Next-generation sequencing is most likely to radically speed up biological research, by allowing the complete analysis of genomes, transcriptomes and interactomes to become economical and common (Shendure & Ji, 2008).

1.2 Objectives

The general aim of this study is to uncover how this psychrophilic yeast responds towards changes in its environmental temperatures. By employing transcriptome sequencing (RNA-Seq), this study was set out to achieve the following goals:

1. To sequence and generate *de novo* assembled transcripts of *Rhodotorula* sp. USM-PSY62.
2. To identify transcripts that are significantly differentially expressed in *Rhodotorula* sp. USM-PSY62 when thermal stresses are introduced.
3. To perform functional annotation on the differentially expressed transcripts.

CHAPTER TWO

LITERATURE REVIEW

2.1 The Antarctica Continent

Antarctica is the southern-most continent located at the South Pole, surrounded entirely by the Southern ocean. With an estimated area of 14 million square kilometres, it is one of the largest continents, nearly twice the size of Australia (Russell, 2006). Reports have shown that Antarctica is almost completely covered by thick ice that is at least 1.9 kilometres in thickness (Fretwell *et al.*, 2013).

Antarctica is widely known as a terrestrial of extremes. It is notoriously one of the coldest, highest, windiest, driest, remotest and most untouched areas on earth despite being a region that is in actual fact a continent that is full of life. In addition to being the most commonly known habitat for inhabitants such as penguins, Antarctica is a sanctuary that offers a unique and wide range of biodiversity which includes microorganisms for discovery and exploration (Johnston & Lohan, 2005).

In terms of temperature, Antarctica has the lowest mean recorded temperature on earth as the Russian research station at Vostoc had documented an average yearly temperature of -55.4 °C, with the minimum temperature recorded as low as -85 °C during austral winter (Stonehouse, 2002). At Casey station which was the site where *Rhodotorula* sp. USM-PSY62 was collected, the annual mean highest recorded temperature was -5.9 °C with 9.2 °C being the highest recorded temperature. On the other hand, the annual mean lowest recorded temperature was -12.5 °C with -37.5 being the lowest recorded temperature (Australian Government Bureau of Meteorology, 2015).

2.2 Antarctic Sea Ice

Sea ice is defined as frozen seawater that floats on the surface of the ocean. It emerges during winter in both the Arctic and the Antarctic and recedes during summer. However, they do not fully disappear. In contrast to Arctic, which is described as an ocean container enclosed by land, the Antarctic is in some ways the exact opposite of Arctic as its land is basically surrounded by the ocean instead. Therefore, this geography has led to more expansion room for the Antarctic sea ice during winter. The formation of Antarctic sea ice reaches its highest point in September, which is the end of winter in the Southern Hemisphere and then retreats to a minimum in February (Scott, 2009).

Antarctic sea ice plays a vital role in polar ecosystems. During winter, when ice freezes, the salinity of the basal water increases. A mixing of the water column occurs as the basal water sinks and the nutrients are brought up to the surface (Scott, 2009). However, dissolved salts are excluded from the water matrix. Eventually these salts concentrate in the microscopic seawater pockets and this has strongly affected the diversity of microbial species (Garrison, 1991). On the other hand, when the sea ice melts each year, a solid vertical gradient in temperature, salinity occurs and nutrients are also released into the water causing an outburst of nutrients which indirectly have supported the development of phytoplankton and bacterial blooms (Wynn-Williams, 1989). This occurrence forms the energy base of the Southern Ocean. Furthermore, the Antarctic sea ice influences the global carbon fluxes used by microalgae during photosynthesis (Constable *et al.*, 2003). According to the ideal gas behavior principal, lower temperatures prolonged the time of interaction between the carbon molecules and the water molecules. Hence, a hike in the environmental temperature of these ecosystems due to global warming can lead to drastic decrease in the global carbon fluxes. In the long run, this may have resulted in the vanishing of the whole glacier ecosystems (Oerlemans *et al.*, 1998).

As salinity, space, light, gradients of temperature and carbon supply are all in relation to the formation of sea ice, all of these characteristics have resulted in a variety of microenvironment for the establishment of microbial communities. Microbial communities in the permafrost and seawater channels are limited to very low amount of unfrozen water that contain high concentration of salts, high osmotic and hydrostatic pressure, high oxidative stress and low nutrient availability (Hofmann, 2008). Hence, the physiological responses of these microbes where they combine temperature, salinity and pH stress are of substantial importance to the understanding of the sea ice microbial community. As a terrestrial of extremes, most of Antarctica as well as its surrounding water comprising a huge variety of microbial biodiversity of mainly bacteria, fungi and microalgae that possesses various metabolic activities such as nitrogen fixation and photosynthesis are yet to be explored (Bowman *et al.*, 1997).

2.3 Extremophiles

Microorganisms (prokaryotes, eukaryotes and archae) are able to live in a variety of environmental conditions. These conditions are controlled by several variables or in other word, stressors which include both physical and biological variables. Physical variables consist of temperature, pH level, oxygen level, osmolarity, ion balance, access to water and pressure whereas the biological variables refers to food availability, competition among organisms, predation, presence of toxic metal, and radiation level. In general, these microorganisms are prone to the exposure towards all of the stressors, occasionally with the combination of two or more of the stressors (Wharton, 2002).

Extemophiles are organisms that are able to survive in extreme environments such as environments that are freezing cold, hot, dry and acidic. Extremophiles can be divided into several groups with different terms which are distinct to the characteristics of conditions that they exist (Table 2.1). Extremophiles can be found in all three domains of life (archae, bacteria, eukarya) (Carrasco *et al.*, 2012).

In recent years, researches involving extremophiles have been rapidly increasing. The biotechnological potential of these organisms has unlocked a new age in biotechnology (Raspor & Zupan, 2006). One of the major reason being that extremophiles are effective sources of metabolites due to their unique features acquired during the adaptation to severe conditions. The exploitation of these organisms gives rise to biomolecules with great commercial value such as cold active alkaline phosphatase, amylases, DNA polymerases, cellulases, pullulanases and lipase from psychrophiles (Margesin & Feller, 2010; Carrasco *et al.*, 2012). Moreover, the exploitation of the genomes of extremophiles has also led to the discovery of novel products, especially antibiotics (McDevitt & Rosenberg, 2001).

Table 2.1 Different groups of extremophiles and their characteristics (Adapted from Rainey & Oren (2006).

Characteristics	**Extremophile**
Temperature	Thermophiles (High Temperature) Psychrophiles/ Cryophiles (Low temperature)
pH	Acidophiles (Acidic) Alkaliphiles (Alkaline)
Rocks	Cryptoendoliths/ Endoliths
Extremely dry	Xerophiles
High salinity	Halophiles
High osmotic pressure	Osmophiles
High atmospheric pressure	Piezophiles/ Barophiles

2.4 Yeasts in Antarctica

2.4.1 Psychrophilic and Psychrotrophic Yeasts in Antarctica

Psychrophiles are microorganisms that have the abilities to grow at temperature 0 °C or below, grow optimally at temperature 15 °C with an upper limit temperature of 20 °C (Moyer & Morita, 2001). On the other hand, another group of microorganisms that are almost similar to psychrophiles are known as psychrotolerants or psychrotrophs. These are microorganisms that have optimum growth temperature above 15 °C with an upper limit temperature of more than 30 °C (Mäännistö & Puhakka, 2002).

The occurrence of psychrophilic or psychrotolerant yeasts inhabiting cold environments has been very well described and studied (Raspor & Zupan, 2006; Vishniac, 2006; Fell *et al.*, 2006; Shivaji & Prasad, 2009). Besides, other locations such as in the coldness of the extremely vast deep seas, polar and near-polar regions have also been reported to harbor psycrophiles (Pikuta *et al.*, 2007). Many of these psychrophiles are eukaryotes (dominated by algae) and a diverse physiological group of bacteria and archaea are found in these environments (Hoover, 2009).

Most of the yeast species found in cold environments are basidiomycetous. However, some ascomycetous yeasts have also been described and reviewed by some researchers (Connell *et al.*, 2008; Branda *et al.*, 2010; de Garcia *et al.*, 2010). The most common basidiomycetous yeast genus isolated from glaciers and related cold habitats are *Cryptococcus*, *Bulleromyces*, *Cystofilobasidium*, *Dioszegia*, *Erythrobasidium*, *Filobasidium*, *Guehomyces*, *Leucosporidiellla*, *Leucosporidium*, *Malassezia*, *Mastigobasidium*, *Mrakia*, *Mrakiella*, *Rhodosporidium*, *Rhodotorula*, *Sporobolomyces*, *Sporidiobolus*, *Trichosporon* and *Udeniomyces* (Connell *et al.*, 2008; Branda *et al.*, 2010; de Garcia *et al.*, 2010). Additionally, there were also a number of new yeast species isolated from cold environment: namely, *Mrakia robertii*, *Mrakia blollopis*, *Mrakiella niccombsii* (Thomas-Hall *et al.*, 2010), *Wickerhamomyces patagonicus* (de Garcia *et al.*, 2010), *Rhodotorula arctica* (Vishniac &

Takashima, 2010), as well as *Dioszegia antarctica* and *Dioszegia cryoxerica* (Connell *et al.*, 2010). This further illustrates the rich, and not yet fully explored yeast biodiversity that occurs in such extreme and disappearing habitats. Moreover, Turchetti *et al.* (2011) have proposed a novel genus *Glaciozyma* where several species namely, *Glaciozyma martinii* and *Glaciozyma watsonii* and those phylogenetically related to *Leucosporidium antarcticum* were grouped under this novel genus.

2.4.2 Mesophilic Yeasts in Antarctica

Apart from psychrophilic and psychrotrophic yeasts, there were also reports of mesophilic yeasts in Antarctica. Some of the mesophilic yeasts are psychrotrophic with maximum growth temperatures of more than 25 °C. A list of the distribution and diversity of mesophilic yeasts in Antarctica is presented in Table 2.2.

Table 2.2 Mesophilic yeasts in Antarctica (Adapted from Vishniac (2006).

Species	Substrate and site	Maximum growth temperature	Reference
Candida albicans	Soil adjacent to camp liquid waste barrel, Taylor Valley	>25˚C	Baublis *et al.* (1991)
Candida parapsilosis	Ice tunnel, South Pole base	>37˚C	Jacobs *et al.* (1964)
Candida saitoana	Soil containing photosynthesizers	>30˚C	di Menna (1966)
Cryptococus albidus	Soil, Edmonson Point	>25˚C	Tosi *et al.* (2005)b
Cryptococcus diffluens	Soil, Wright Valley	>25˚C	di Menna (1960)
Cryptococcus humicola	Water, Lake Vanda, lakeshore soil	>25˚C	Goto *et al.* (1969)
	Soil, Schirmacher Oasis	>25˚C	Ray *et al.* (1989)
Cryptococus macerans	Soil	>25˚C	di Menna (1966)
Rhodosporidium sphaerocarpum	Antarctic Ocean near Palmer Peninsula and Archipelago	30˚C	Newell & Fell (1970)

Table 2.2 Continued.

Rhodosporidium toruloides	Water, Lake Vanda	>30˚C	Goto *et al.* (1969)
Rhodotorula aurantiaca	Moss	<20˚C	Babyeva & Golubev (1969)
Rhodotorula glutinis	Moss, penguin rookery, Haswell Island and lichen sites at Molodozhnaya Station	>25˚C	Babyeva & Golubev (1969)
Rhodotorula graminis	Soil, Wright Valley and soil	>20˚C	di Menna (1960, 1966)
Rhodotorula marina	Soils with algae between Campbell and Mawson glaciers	>30˚C	di Menna (1966)
Rhodotorula minuta	Water, Lake Miers and Lake Vanda	30˚C	Goto *et al.* (1969)
Rhodotorula pallida	Soil	>25˚C	di Menna (1966)
Sporidiobolus johnsonii	Soil, Ross Desert	>25˚C	Vishniac & Hempfling (1979b)
Stephanoascus ciferrii	Microbial mat, Lake Hoare	>35˚C	Baublis *et al.* (1991)
Torulaspora delbrueckii	Soil, Victoria Land	>30˚C	del Frate & Caretta (1990)
Trichosporon cutaneum	Soil, Showa Base	>25˚C	Soneda (1961)
Trichosporon moniliiforme	Lake water, Lake Vanda	>30˚C	Goto *et al.* (1969)

2.5 *Rhodotorula* sp. USM-PSY62

Rhodotorula sp. USM-PSY62 is a yeast isolated from Antarctic sea ice samples collected at Casey Station, Antarctica (by the late Mr. Omar Pozan during austral summer of year 2001-2002). It was identified as *Rhodotorula* sp. USM-PSY62 (Ong, 2006). This yeast is psychrophilic in nature and it has an optimum growth temperature of 15 ˚C. According to Vishniac (2006), there are members of *Rhodotorula* that were psychrophilic, psychrotolerant and mesophilic. The taxonomy of *Rhodotorula* sp. USM-PSY62 shows that it belongs to the basidiomycetous yeast (Table 2.3). The distribution of the members of this genus (Table 2.4) can be found at wide ranges of habitat such as soil, lake water, estuaries, Antarctic sea water samples, lake sediment and worldwide glaciers (Buzzini *et al.*, 2012; Kutty & Philip, 2008).

Members of the genus *Rhodotorula* have been reported to colonize extreme environments (Sampaio, 2004). Interestingly, they have been widely known for their ability to degrade phenolic compounds (Sampaio, 1999; Fell *et al.*, 2001). Several novel species of *Rhodotorula* have been isolated from Alpine glaciars in Austria, namely *Rhodotorula psychrophila* sp. nov., *Rhodotorula psychrophenolica* sp. nov. and *Rhodotorula glacialis* sp. nov. (Margesin *et al.*, 2007a).

Table 2.3 Taxonomy of *Rhodotorula* sp. USM-PSY62.

Taxonomic Hierarchy	**Nomenclature**
Domain	Eukaryota
Kingdom	Fungi
Phylum	Basidiomycota
Class	Microbotryomycetes
Order	Microbotryomycetes incertae sedis
Family	mitosporic Microbotryomycetidae incertae sedis
Genus	*Rhodotorula*
Species	USM-PSY62

(Source: http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=858472 [Last accessed: 1st March 2015)

Table 2.4 Distribution of *Rhodotorula* (Adapted from Buzzini *et al.* (2012).

Species	**Isolation source**	**Locality**	**Reference**
Rhodotorula auriculariae	Antarctic soil	Schirmacher Oasis, Queen Maud Land	Ray *et al.* (1992)
Rhodotorula diffluens	Lake sediment	Dry Valleys, South Victoria Land	Goto *et al.* (1969)
Rhodotorula glutinis	Antarctic soil; ice cores; lake water	Dry Valleys, South Victoria Land; Vostok Research Station	di Menna (1966b), Goto *et al.* (1969), Abyzov (1993)

Table 2.4 Continued.

Rhodotorula graminis	Antarctic soil	Dry Valleys, South Victoria Land	di Menna (1966b), Cameron (1971)
Rhodotorula ingeniosa	Antarctic soil	Schirmacher Oasis, Queen Maud Land	Ray *et al.* (1992)
Rhodotorula laryngis	Antarctic soil; decaying wood	Dry Valleys, South Victoria Land; Ross Island, Ross Sea Bay	Arenz *et al.* (2006), Connell *et al.* (2008)
	Glacier basal ice	Kongsfjorden fjord, Spitsbergen, Svalbard archipelago, Norway	Butinar *et al.* (2007)
	Sediments and ice from glaciers	Calderone glacier, Apennines, Italy; Forni and Sforsellina glaciers, Italian Alps	Turchetti *et al.* (2008), Branda *et al.* (2010)
Rhodotorula mucilaginosa	Antarctic soil; decaying wood; lake water	Dry Valleys, South Victoria Land; Ross Island, Ross Sea Bay; Schirmacher Oasis, Queen Maud Land; Vostok Research Station; West Ongul Island, Lutzow-Holm Bay	di Menna (1966b), Goto *et al.* (1969), Cameron (1971), Atlas *et al.* (1978), Baublis *et al.* (1991), Ray *et al.* (1992), Arenz *et al.* (2006), Connell *et al.* (2008), D'Elia *et al.* (2009), Pavlova *et al.* (2009)
Rhodotorula pallida	Antarctic soil	Victoria Land	di Menna (1966b)
Rhodotorula glacialis	Small puddles near to meltwater stream from glaciers	Kongsfjorden fjord, Spitsbergen, Svalbard archipelago, Norway	Pathan *et al.* (2010)
	Glacier cryoconites; mud at glacier foot; sediments and ice from glaciers	Forni and Sforsellina glaciers, Italian Alps; Stubaier glacier, Tyrol, Austria	Margesin *et al.* (2007a, b), Turchetti *et al.* (2008)
Rhodotorula minuta	Arctic soil; glacier basal ice	East Greenland; Kongsfjorden fjord, Spitsbergen, Svalbard archipelago, Norway	di Menna (1966b), Butinar *et al.* (2007)
Rhodotorula bacarum	Sediments from glaciers	Forni and Sforsellina glaciers, Italian Alps	Turchetti *et al.* (2008)
Rhodotorula colostri	Sediments from glaciers	Calderone glacier, Apennines, Italy	Branda *et al.* (2010)

Table 2.4 Continued.

	Meltwater stream from glaciers, Frias river; water from lakes and lagoons of glacial origin	Nahuel Huapi National Park	De Garcí a *et al.* (2007), Brandão *et al.* (2011)
Rhodotorula ingeniosa	Soil and sediments	Brenner pass, Austria	Bergauer *et al.* (2005)
Rhodotorula psychrophenolica	Glacier cryoconites; mud at glacier foot; sediments from glaciers	Calderone glacier, Apennines, Italy; Forni and Sforsellina glaciers, Italian Alps; Stubaier glacier, Tyrol, Austria	Margesin *et al.* (2007a, b), Turchetti *et al.* (2008),
Rhodotorula psychrophila	Soil and sediments	Brenner pass, Austria	Margesin *et al.* (2007a, b)
Rhodotorula pinicola	Water of glacial origin	Nahuel Huapi lake	Brandão *et al.* (2011)

2.6 Survival Strategies of Psychrophiles

In terms of environmental adaptation, habitat selection of psychrophiles is essential. Psychrophilic microorganisms in glacial ice predominantly exist in veins or liquid films that contain metabolic substrates (Mader *et al.*, 2006; Miteva *et al.*, 2009). In sea ice, psychrophilic microorganisms are confined in the hypersaline pockets and channels within the ice. These seawater pockets provide a microenvironment that is rich in dissolved organic matter (Mock & Thomas, 2005).

Psychrophiles have successfully thrived at low temperature in aqueous environments of high viscosity. Some of the survival adaptive strategies towards cold stressors include seasonal dormancy, pigment production, exopolysaccharide production and sheath production (protection against UV) (Hoover, 2009). In the cold, specifically in frozen temperature, psychrophiles have the ability to preserve biochemical reactions in the cytoplasmic membranes and maintain functional enzymes that have the tendency to rigidify when a descent in temperature occurs (Ratkowsky *et al.*, 2005). This will cause a certain

degree of adverse effects on enzyme activity, membrane fluidity, rate of transcription and translation, cell division and eventually intracellular ice formation.

Pikuta *et al.* (2007) had also described several other adaptative strategies which also involve the changing of protein functionality (optimal at low temperature) and increasing fluidity of membranes by altering the unsaturated fatty acid ratio, modifying ante-iso-/iso-branching pattern and short fatty acid length. Increasing the membrane fluidity will give the membrane more flexibility during the freezing environment.

The existence of cold adapted enzymes is also crucial in the adaptation of psychrophiles for survival. Cold adapted enzymes traded their thermostability for more flexibility at lower temperature (Hoyoux *et al.*, 2004). They also possessed more accessible active sites for binding (Aghajari *et al.*, 2003; Hoyoux *et al.*, 2004). Moreover, a sudden change or decrease in temperature can also induce cold shock responses of psychrophiles which ultimately lead to the production of cold shock protein (Csps) and cold acclimatization protein (Caps). The production level of Csps is closely related to the extent of extremity in conditions of reduced temperature whereas Caps are constantly synthesized in long extended period of growth in the cold (Pikuta *et al.*, 2007).

Ice-active proteins that function to alter the structure of ice crystal have also been reported as one of the survival strategies (Bayer-Giraldi *et al.*, 2010). These ice-active proteins are known as antifreeze proteins (AFPs). They are able to create thermal hysteresis as an approach to hinder damage introduced by intra- and extra-cellular ice crystal formation where ice crystals are lethal to the cells (Jia & Davies, 2002). In terms of osmotic imbalance, trehalose is produced to reduce its shock towards the cells. Psychrophiles also produced exopolysaccharides as a cryoprotection of proteins from cold denaturation and aggregation (Tanghe *et al.*, 2003; Nichols *et al.*, 2005).

2.7 Next Generation Sequencing: RNA-Seq

In recent years, methodologically comprehensive technologies for transcriptome analysis are available. They have been extensively used since the development of microarray technology and the completion of human genome sequencing project. Previously, before the existence of this powerful technology, mRNA expression was measured by microarray techniques or real-time PCR techniques. However, these methods have several limitations, which include: low sensitivity in microarray and comparatively expensive for PCR techniques, plus it is not suitable to be used for a genome-wide survey of gene expression (Mardis, 2008). In contrary, next-generation sequencing (NGS) methods are rapid and inexpensive. The technology offers high throughput gene expression profiling, genome annotation and discovery of non-coding RNA. NGS technologies are based on the sequencing-by-synthesis (SBS) technology called pyrosequencing. The transcriptomics variant of pyrosequencing technology is known as RNA-Seq (Denoeud, 2008). RNA-Seq is a recent approach to transcriptome profiling that applies deep-sequencing technologies. The utilization of this method is able to change our view of the degree and complexity of eukaryotic transcriptome. Apart from that, comparative to other methods, RNA-Seq also provides a greater precision in the measurement of transcripts and their isoforms level (Wang *et al.*, 2009).

2.8 RNA-Seq as A Tool for Transcriptomics Analysis

The transcriptome refers to the complete set of transcripts in a cell, and their quantity, for a specific developmental stage or physiological condition. Comprehending the transcriptome is crucial for the elucidation of the functional elements of the genome. Moreover, it is also important to reveal the molecular components of cells and tissues which will be particularly useful in understanding the development of disease (Wang *et al.*, 2009).

Next-generation RNA-sequencing (RNA-Seq) is swiftly outcompeting microarrays as the technology of choice for whole-transcriptome studies. It has become a powerful tool to study global gene expression (Van Verk *et al.*, 2013). Numerous technologies have been developed to understand and quantify the transcriptome. These technologies include hybridization-or sequence-based approaches. Hybridization-based methods (involving the incubation of fluorescently labelled cDNA with custom-made microarrays) are high throughput and reasonably economical. However, these methods have several drawbacks which include the need to depend upon the existing knowledge about genome sequence, high background levels due to cross-hybridization and a limited dynamic sensitivity due to both background and saturation of signals (Wang *et al.*, 2009). Moreover, the comparison of expression levels across different experiments is made difficult plus complicated normalization methods are necessary. Therefore, the choice of sequencing-based method which is now termed NGS is far more preferable being the method that is able to solve these issues despite several complex statistical issues in data normalization during analysis (t Hoen *et al.*, 2008; Jiang & Wong, 2009).

RNA-Seq utilizes recently developed deep sequencing technologies. One of the aims of using RNA-Seq is in the gene expression analysis. Gene expression analysis is widely used in the effort of uncovering regulatory mechanisms that hold controls of cellular processes in plants, animals, and microbes. RNA-Seq covers the possibilities of transcriptome studies to the analysis of previously unidentified genes and of splice variants (Wang *et al.*, 2009). Moreover, RNA-Seq offers an indefinite dynamic range of quantification at a condition where technical inconsistencies are minimized. Hence, with these advantages, RNA-Seq ia an attractive method for whole-genome expression studies including species without sequenced genomes.

A RNA-Seq experiment typically begins with RNA isolation. The quality of starting RNA can strongly affect the success of a RNA-Seq experiment. RNA quality is normally assessed using methods such as the Lab-on-Chip assay (Agilent Bioanalyzer). This method

produces a RNA integrity number (RIN) with a value in the range of 1 to 10, which represents RNA in the range of fully degraded RNA to good intact RNA respectively. Generally, RNA-Seq protocols require approximately 100 ng to 4 mg starting total RNA with a RIN value of at least 8 (Schroeder, 2006). Apart from that, it is also important to note that the key in obtaining good quality RNA lies in the prevention of RNase contamination. In the early years, total RNA isolation from eukaryotic cells was done using guanidinium thiocynate, a protein denaturant and caesium chloride centrifugation (Glisin *et al.*, 1974). Those methods have been replaced by phenol-chloroform extraction, which has become more preferable nowadays (Chomczynski & Sacchi, 1987; 2006). As RNA degradation during the RNA isolation being the most likely reason in most cases that a good library was not successful, RNase inhibitors such as guanidinium thiocyanate, phenol, chloroform and isoamyl alcohol were used to resolve this matter. Apart from RNase contamination, protein contamination is also to be avoided. In general, the NanoDrop (Thermo Scientific) spectrophotometric was used to measure the concentration and purity of the RNA. By calculating the absorbance ratio between A260/280 and A260/230, the purity of the nucleic acid after RNA isolation stage could be determined. A260/280 was used to detect protein contamination (Winfrey *et al.*, 1997) and A260/230 for the detection of polysaccharides and polyphenols contamination (Loulakakis *et al.*, 1996). Pure nucleic acid will have an absorbance ratio of A260/280 at approximately 1.8 for DNA and 2.0 for RNA; whereas the A260/230 absorbance ratio was also used as a secondary measurement of nucleic acid purity. A higher value of A260/230 ratio than the A260/280 ratio indicates high purity. A lower ratio value may suggest the presence of co-purified contaminants. Each nucleotide produces varying A260/280 ratio independently which are 1.15 for Guanine, 4.5 for Adenine, 1.51 for Cytosine, 4.00 for Uracil and 1.47 for Thymine. Hence, a higher A260/280 ratio will be observed for RNA due to the higher of Uracil compared to Thymine.

The isolated RNA is subsequently converted into cDNA to form a RNA-Seq library with the standard protocols based on the NGS platform used (MacLean *et al.*, 2009).

Subsequently, the millions of DNA fragments in the library will be sequenced. Following the sequencing, a precise quantity of the relative abundance of each transcript and splice variants will be acquired. In the last couple of years, numerous bioinformatic tools have been developed. These tools can be used to process the sequencing output data as an approach to translate these data into meaningful information on gene expression levels.

At present, the most commonly applied next generation sequencing technology for RNA-Seq is the Illumina HiSeq platform which currently is able to yield up to 3 billion reads per sequencing run. The Illumina HiSeq platform consists of two sequencing units (flow cells). Each of these flow cells provides eight separate sequencing reactions (lanes). Depending on the total read length, a sequencing run on Illumina HiSeq usually takes between 1.5 and 11 days to finish (Van Verk *et al.*, 2013).

The most suitable variation of the RNA-Seq protocol has to be determined specifically referring to the research questions and the organism under study even though this platform follows a standard sequencing protocol. For instance, the gene expression profiling of relatively non-repetitive genomes, such as *Arabidopsis thaliana*, it is recommended to opt for single end sequencing of 50 bp from one end of the cDNA fragments. A sequencing depth of 50 million reads can yield a near-saturated coverage of expressed genes in *Arabidopsis thaliana*. On the other hand, the sequencing per gene should be increased when studying alternative mRNA splicing. This can be done by performing a paired end sequencing of 50 or 100 bp from both ends of the cDNA fragments (Van Verk *et al.*, 2013).

2.9 Transcriptome *de novo* Assembly

Transcriptome sequencing (RNA-seq) has unlocked the path to study the genetic and functional information within any organism at an extraordinary scale and speed. Many genomic applications depend on the accessibility of high-quality genome sequences.

However, only a very small percentage of known organisms have these high-quality genome sequences. As sequencing and assembling an organism's genome is still considered a relatively expensive effort, conversely, RNA-seq data provides a more rapid and cheaper alternative as only a fraction of the genome is transcribed. This has enabled the definition of a reference transcriptome for downstream applications, such as alignment, phylogenetics or marker construction (Haas *et al.*, 2013).

Theoretically, RNA-seq enables the identification of all expressed transcripts. However, the reconstruction of all of the transcripts from short sequencing reads requires computational methods that can assemble these transcripts even without the availability of a genome sequence. Primarily, there are two alternative computational strategies for transcriptome reconstruction. First is through the guidance of a high quality genome sequences and secondly, it is through *de novo* assembly (Haas *et al.*, 2010; Martin & Wang, 2011). For most model-organisms, the genome-guided assembly method has rapidly becoming the standard method in RNA-seq analysis. Several software packages have also been developed along the course of time (Guttman, 2010; Trapnell, 2012). The genome-guided method, or in other words, mapping-first method principally promises maximum sensitivity, but it is still relying on the correct read-to-reference alignment which are complicated by splicing, sequencing errors and the incompleteness of many reference genomes (Haas *et al.*, 2013). However, this approach is unsuitable and cannot be applied to non-model organisms without a reference genome which are also equally of substantial ecological or evolutionary importance. Plus, even for organisms that have a well-assembled reference genome sequences, the results may still differ across genome assembly versions. Hence, in such cases, a *de novo* transcriptome assembler is required (Haas *et al.*, 2013).

Several *de novo* assembly tools which are mostly extensions of earlier developed genome assemblers are now available such as Trans-ABySS (Robertson, 2010), Velvet-Oases (Schulz, 2012) and SOAPdenovo-trans (http://soap.genomics.org.cn/SOAPdenovo-Trans.html). A novel alternative *de novo* assembly method was described by Grabherr *et al.*

(2011), which was known as Trinity. Trinity is an efficient and robust method in the reconstruction of transcriptomes and it consists of three software modules: Inchworm, Chrysalis and Butterfly. Trinity partitions the RNA-seq data into many independent *de Bruijn* graph and uses parallel computing to reconstruct transcripts based on the graphs (Haas *et al.*, 2013). In a *de Bruijn* graph, a node represents a sequence of a fixed length of *k* nucleotides, or simply known as *k*-mer, with *k* considerably shorter than the read length. The nodes on the graphs are connected by edges, if a perfect overlap by $k - 1$ nucleotides occurs and also supported by the sequence data information. Given the overlaps of $k - 1$, linear sequences can be reconstructed with the enumeration of all possible solutions (Grabherr *et al.*, 2011). Trinity provides an easy and user-friendly interface that involves very little parameter tuning. The efficiency of Trinity has been documented and reported in several independent studies. They have demonstrated that Trinity, as similar to other alternatives, is also highly efficient in recovering full length transcripts (Duan *et al.*, 2012; Xu *et al.*, 2012; Zhao *et al.*, 2011).

2.10 Transcriptomics Downstream Analysis

A *de novo* transcriptome assembly is the initial step towards transcriptome analysis. One of the common goals in studying transcriptomes in non-model organisms includes the identification of differentially expressed transcripts in different samples. As RNA-seq is quantitative, it is more preferable to be used in the determination of RNA expression levels as it is more accurate compared to microarrays. In principle, the determination of the absolute quantity of every molecule in a cell population and the direct comparison of results between experiments has been made possible (Wang *et al.*, 2009). RNA-seq has been used extensively in numerous studies such as in monitoring gene expression during yeast vegetative growth (Nagalakshmi *et al.*, 2008), yeast meiosis (Wilhelm *et al.*, 2008) and mouse embryonic stem-cell differentiation (Cloonan *et al.*, 2008). RNA-seq was used to

track changes in gene expression during development. Moreover, it can also provide a 'digital measurement' of gene expression difference between different tissues (Mortazavi *et al.*, 2008). Due to these various benefits, RNA-seq will absolutely be valuable for the understanding of transcriptomic dynamics during development and normal physiological changes, plus the analysis of biomedical samples, where strong comparison between diseased and normal tissues can be done.

As a prerequisite before differential expression analysis in different samples, transcript quantification has to be done. A normalization step, normalizing for depth of sequencing and the length of transcripts of RNA-seq data is necessary before the estimation of the transcripts abundance. Normalization of read counts enables the comparison of expression level between different transcripts as well as different experiments. Several metrics were proposed which include reads per kilobase of target transcript length per million reads mapped (RPKM) (Mortazavi *et al.*, 2008) for single-end sequences, and an analogous computation based on counting whole fragments (FPKM) (Trapnell *et al.*, 2010) for paired-end RNA-seq data. Bioinformatic tools have been developed to count the number of reads mapped to the transcripts (Park, 2009).

For differential gene expression analysis, numerous statistical methods have been developed for data generated from microarray data. However, these approaches are not suitable for use with RNA-seq data. This is due to the different nature of both data from microarray and RNA-seq. Microarray data are based on fluorescence intensity and hence they are continuous data. In contrast, RNA-seq data are discrete data as they are derived from read counts. As a result, these have led to the development of software tools to statistically test for differential gene expression between samples that are analyzed by RNA-Seq (Van Verk *et al.*, 2013). Transcripts that are differentially expressed under different condition can be detected by computational tools using the normalized expression scores and statistical tests. These tools can be classified as parametric or non-parametric. Parametric algorithms that are commonly used apply probability distributions such as Binomial or

Poisson, whereas non-parametric algorithms model the noise distribution based on the actual data (Li & Tibshirani, 2011). When the total number of reads per transcript has been summarized, tools such as DESeq, edgeR, and baySeq (Anders & Huber, 2010; Robinson *et al.*, 2010; Hardcastle & Kelly, 2010) model count data using a negative binomial distribution. These tools have been reported to perform well in several studies (Kvam *et al.*, 2012).

2.11 Transcriptomic Studies on Yeast Cold and Heat Shock Responses

The transcriptome of the model organism, *Saccharomyces cerevisiae* had been studied on its responses upon cold and heat shock (Becerra *et al.*, 2003). DNA microarrays were used to measure the changes in transcript levels as the cells of *S. cerevisiae* responded to temperature shocks from 30 ˚C to 37 ˚C or 45 ˚C. The heat and cold shock affected genes related to metabolism, cell growth and division, transcription, ribosomal proteins and protein synthesis (Becerra *et al.*, 2003). In a separate study, the heat shock responses of psychrophilic and psychrotrophic yeasts in Antarctica were studied focusing on the effect of heat shock towards the production of heat shock proteins (Deegenaars & Watson 1997; Deegenaars & Watson 1998). Apart from the production of heat shock proteins, the accumulation of trehalose was reported to be induced upon heat shock (Deegenaars & Watson 1997). Another transcriptomic study on the heat and cold shock of Antarctic yeast, *Glaciozyma antarctica* PI12 had been performed by Boo *et al.* (2013). The qPCR technique was utilized in the investigation of the expression patterns of 14 thermal stress-related genes, namely *afp4*, *hsp70*, *hsp90*, *hsp100*, *nth1*, *tps1*, *tps2*, *fad,prx*, *gst*, *grxA*, *grxB*, *cat* and *MnSOD* upon cold and heat shock. The findings of the study revealed that the expression profiles of the 14 genes varied under different thermal stress conditions which indicated the specific roles of these genes in the survival of the yeast.

CHAPTER THREE

MATERIALS AND METHODS

3.1 General Overview

The overview of the methodological steps is outlined in Figure 3.1.

3.2 Subject Organism

The subject organism that was used in this study was originally isolated from Antarctic sea ice samples that were collected at Casey Station, Antarctica by the late Mr. Omar Pozan during austral summer of year 2001-2002. It was then identified and characterized as *Rhodotorula* sp. USM-PSY62 (Ong, 2006).

3.3 Standard Method

3.3.1 Sterilization

All culturing media, glassware and thermostable solution that were required to be used under sterile condition were sterilized by autoclaving at 15 psi (121°C) for 15 minutes except otherwise mentioned.

3.4 Yeast Culturing Media

3.4.1 Yeast Extract-Peptone-Dextrose Agar (YEPD)

Yeast Extract-Peptone-Dextrose agar (YEPD) was prepared by dissolving 10 g/L of yeast extract, 20 g/L of peptone , 20 g/L of D-glucose and 15 g/L of agar in 1 L of distilled water. Glucose was autoclaved separately in another flask. This was to prevent it from caramelising and causing precipitation of the agar. After the medium was autoclaved, it was let to cool down. The D-glucose solution was then added into the flask. The solution was swirled gently before being poured into sterile petri dishes. The plates were let to solidify before being stored at 4 °C.

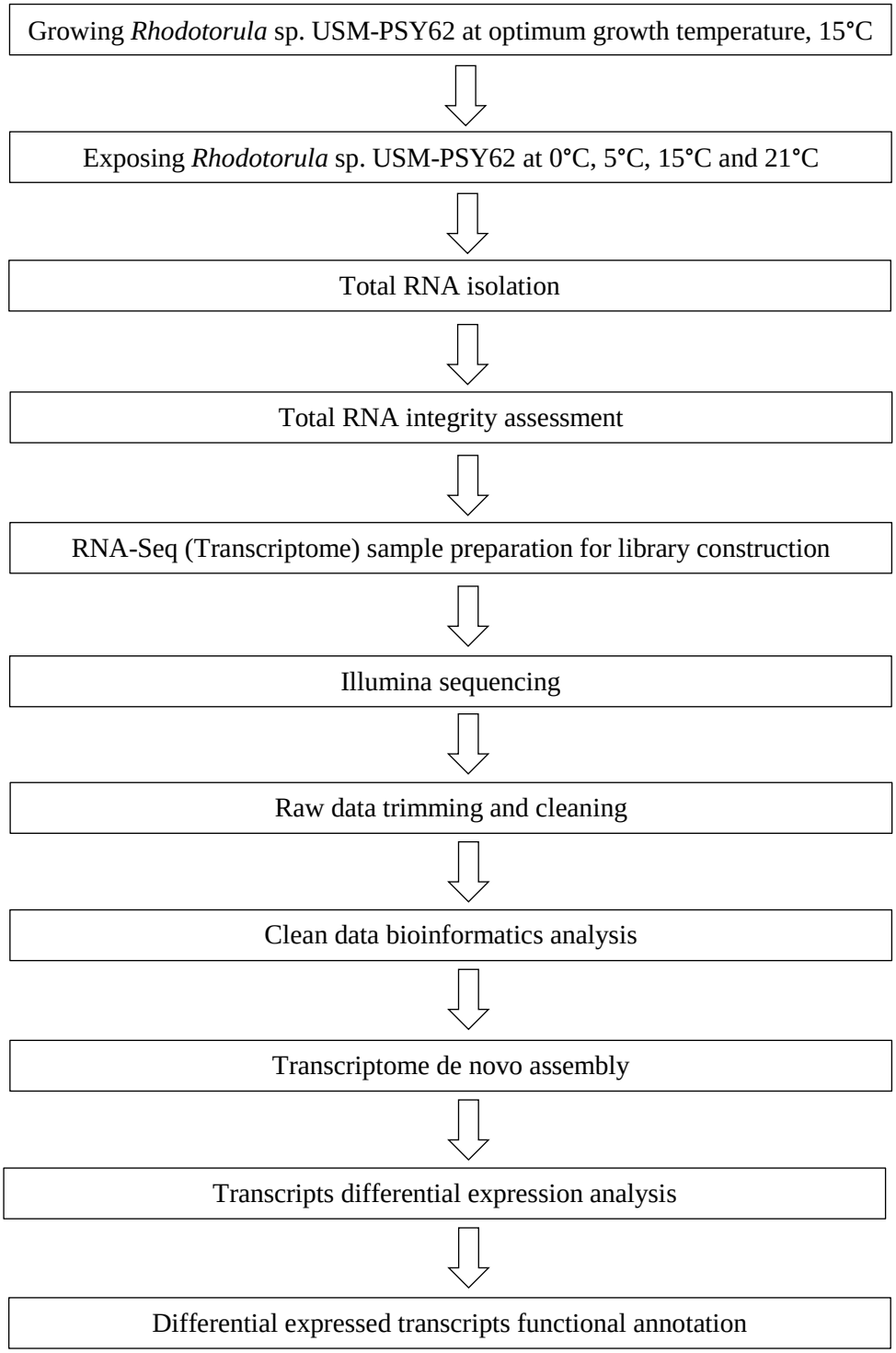

Figure 3.1 General flow chart of methodology.