Ice activities in New Zealand *Chionochloa* species

Haoyu Xiong

A thesis submitted for the degree of Doctor of Philosophy at the University of Otago, Dunedin, New Zealand

November 2017

Abstract

Overwintering plants produce ice active agents with low thermal hysteresis but a higher ability to inhibit the recrystallization of ice crystals already formed. Many plants also minimize the damage from freezing by triggering ice formation at high sub-zero temperatures. These processes in plants differ from those in fishes and insects, which prevent intracellular ice formation and maintain a supercooled liquid state by secreting low molecular weight cryoprotectants and thermal hysteresis proteins (THP) to depress the freezing point of body fluid.

Chionochloa tussocks are mainly endemic to New Zealand and dominate the alpine grasslands. Alpine areas in the South Island can experience snow and freezing at any time of the year due to the incursion of polar air masses. Alpine plants here have developed various mechanisms to overcome the cold weather. A previous study showed significantly high ice nucleation activity (INA) in two alpine *Chionochloa* species but provided no more information about other ice activities in this genus. Moreover, there are only a few studies on the phylogeny of the *Chionochloa* group based on ribosomal DNA and/or mitochondrial markers and the lack of genetic information has hampered further research in this genus.

Ice activities were investigated in *Chionochloa* species collected at the Dunedin Botanic Garden and significant seasonal variations were reported with winter collections showing the highest and summer collections showing the lowest ice activities. Non-bacterial intrinsic proteins were essential in maintaining all three types of ice activities, which showed different responses to physical and chemical treatments including heat, pH, high salt, proteinase K, lysosome, reducing and oxidizing agents.

Ice-shell enriched ice active proteins (IAP) were isolated from C. *macra* and *C. spiralis* respectively. This technique showed better purification efficiency than the traditional ice-finger method. Recrystallization inhibition proteins (RIP), class II endochitinase antifreeze proteins (EAF) and other potential ice-binding proteins were identified in the ice fraction (IF). Interestingly, ice nucleation activity (INA) was not present in the ice fraction (IF) suggesting weak ice-binding ability of ice nucleation protein (INP) or the requirement of other compounds such as metal ions to maintain the INA. INA was separated by a protocol of 100 kDa MWCO centrifugal device,

size exclusion chromatography and ion exchange chromatography. MALDI-TOF mass spectrometry failed to identify any potential INP as fewer peptide fragments were present.

The transcriptome of winter *C. macra* was annotated for the better understanding of essential proteins and metabolic pathways involved in the species-specific features including high polysaccharide contents, triterpenoid expression, water conservation and cold tolerance in this genus. In addition, alternative splicing events, SSRs markers, SNPs, miRNAs were reported and gene expression profiles were provided for further genetic research in *C. macra* and other species in this genus. The *C. macra* transcriptome also served as a personalized MASCOT database in assist with the search for ice-binding proteins in *Chionochloa* species assessed by mass spectrometry.

Finally, potential ice active protein (IAP) genes in the *C. macra* transcriptome were investigated by Blast search against current known IAPs. Two putative IAP genes, IRI2 and EAF2, which showed high similarity to the mentioned RIP and EAF respectively, were expressed in the prokaryotic *E. coli* system and corresponding proteins were purified with ice activities confirmed. Phylogenetic analysis of these IAPs and *Pooideae* specific cold stress-related genes (C-repeat binding factors and fructosyltransferase) in *C. macra* indicated these genes from *C. macra* showed different phylogenetic relationships with core *Pooideae* and *Brachypodium distachyon*.

Taken together, these findings confirmed that utilizing transcriptome combined with putative gene expression, and protein isolation combined with ice activity assays, were useful to study IAPs in New Zealand *Chionochloa* species. The application of a set of improved methods including DSC, ice-shell and transcriptome based MASCOT database not only saved time and labor but also gave accurate results in studying ice activities. Moreover, the first annotated transcriptome of winter *C. macra* provided a clear genetic map for the further genetic research in *Chionochloa* specific features and cold tolerance in this genus.

Acknowledgments

Firstly, I would like to acknowledge and thank my primary supervisor, Associate Professor Craig Marshall for his kind guidance and support throughout my project, and my co-supervisor, Professor David Wharton, and PhD advisor, Dr Janice Lord, for their insightful discussions and suggestions. I would also like to thank Kate and Amy from the Dunedin Botanic Garden for their help in my *Chionochloa* species collections. Without them, I cannot image how difficulty for me to recognize those tussocks by myself.

I would like to thank my wife, Lizhou, for standing by me and supporting me during my four-year time of PhD study. Without her encouragement, her patience and understanding, I may not have made it.

I thank Robyn for helping me through the tough RNA extraction in the tussocks, Elizabeth and Peter for their kind suggestions in my next generation sequencing data analysis; Monica, Max and Gene for their kind help in my protein expression and purification; Roman and Abhishek for the discussion of my experiments. I also want to thank all members of the Marshall Lab and Wharton Lab: Anna and Dom for teaching me with molecular biology techniques and ice activity assays. Moreover, many thanks go to the Department of Biochemistry, the Department of Zoology and the University of Otago for supporting my PhD study.

Finally, special thanks go to mum and dad, although you are thousand miles away, your encouragement and supports have been appreciated more than what I can put into words.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AFP	Antifreeze protein
AFGP	Antifreeze glycoproteins
AS	Alternative splicing
Blast	Basic local alignment search tool
bp	Base pair(s)
CBF	C-repeat binding factor
DSC	Differential scanning calorimetry
DTT	Dithiothreitol
EAF	Class II endochitinase antifreeze proteins
EDTA	Ethylenediaminetetraacetic acid
FPKM	Fragments per kilobase of transcript per million mapped reads
FPLC	Fast protein liquid chromatography
FST	Fructosyltransferase
IAP	Ice active protein
IF	Ice fraction
INA	Ice nucleation activity
INP	Ice nucleation protein
IPTG	Isopropyl-ß-D-thiogalactopyranoside
kb	kilobase(s)
LB	Lysogeny broth
LF	Liquid fraction
LRR	Leucine rich repeat
mQ	Type 1 (ultrapure) Milli-Q® water
mRNA	Messenger RNA
MSC	Multiple cloning site
MWCO	Molecular weight cut off
Mya	Million years ago
NCBI	National center for biotechnology information
PCR	Polymerase chain reaction
QLL	Quasi-liquid layer
RI (P)	Recrystallization inhibition (protein)

RLK	Receptor-like kinase
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SSR	Simple sequence repeats
STT	Sodium tetrathionate
T_c	Temperature of crystallization
TH (P)	Thermal hysteresis (protein)
T_m	Temperature of melting point
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol, THAM, Tris base

Chapter 1 General introduction

1.1. Background

Approximate 85 million years ago (Mya), the Zealandia subcontinent started to break with the supercontinent Gondwana, and became isolated around 65 Mya (Landis *et al.* 2008; Wegener 1966; Wallis and Trewick 2009). From 60 to 24 Mya, Zealandia was warm with most parts gradually submerged under the ocean except for a few low altitude islands and this was considered to greatly influence the descendant biota structure (Wallis and Trewick 2009). New Zealand originated from the Zealandia subcontinent, was formed by the collision of the Indo-Australian and Pacific tectonic plates with frequent volcanic activities, and became well separated from Australia, South America and Antarctica (Wallis and Trewick 2009). The foundation of modern New Zealand was accompanied by the uplift of mountains started from around 23 Mya (Graham 2008). New Zealand is a country dominated by mountains, with large alpine areas located mainly in the South Island (Mark and Adams 1996).

The ecology of New Zealand had changed with the geographical isolation, and the major climatic trend had shifted from warm to gradually cooling with repeated glacial periods (Bannister *et al.* 2005). The oldest glaciation in New Zealand was over 2.5 Mya and the most recent glacial period ended only around 10,000 years ago (Wellman 1955; Winkworth *et al.* 2005). There is subtropical weather in the North Island whereas the Southern Island is mainly cool temperate due to the incursion of polar air masses. Thus, mountains in the South Island can experience frost and snow at any time of the year and plants here have developed various mechanisms to overcome the cold weather (Bannister *et al.* 2005).

1.2. New Zealand alpine plants

Alpine areas in New Zealand are comparatively young and the alpine flora has evolved in a relatively isolated environment (Hugh 2012). There are more than 1,000 species of native mountain plants and over 500 species are found only in alpine areas (Hugh 2012; Mark and Adams 1996). Indigenous forests are located in the North Island and west coast of South Island whereas grass and shrub-land dominate the rest of the South Island (Forests 1976). Alpine plants are unique as most are endemic at

species level because of the isolation. Generally, those plants are not so colorful (mainly white and yellow) as those in other parts of the world, tolerate the cold environment and reproduce at low temperatures (Mark and Adams 1996; Raymond 1996). The New Zealand alpine flora are thought to have evolved from lowland plant adaption and originated either from the ancient New Zealand flora which dates back to the Gondwana supercontinent, or from precursors that arrived through long-distance dispersal across oceans, e.g. from Australia and South America (Heenan and McGlone 2013; Winkworth *et al.* 2005). There are different genetic models explaining the evolution and distribution of New Zealand alpine plants regarding the complexity of geographic changes, climate changes and the uncertainty of the flora origin (Heenan and McGlone 2013; Winkworth *et al.* 2005). Alpine plants have adapted in different ways to survive these changes which started with the formation of the Alpine Fault from about 5 Mya (Winkworth *et al.* 2002).

New Zealand alpine plants are mainly endemic and are included in a few genera. Many species of plants in New Zealand followed the genetic pattern of an origin by long-distance dispersal from Australia such as *Celmisia*, but the genus *Chionochloa* did not. *Chionochloa* belongs to the *Poaceae* family. Thirty-six of the currently recognized 38 species and subspecies of *Chionochloa* are endemic to New Zealand and dominate the New Zealand alpine grassland (Connor 1991; Pirie *et al.* 2010). Pirie *et al.* (2010) inferred a South Island origin of *Chionochloa* species based on ribosomal and mitochondrial DNA markers together with the rising of the southern alpine mountains. They suggested several sub-clades might exist during those repeat glacial periods. *Chionochloa* species could interbred and incomplete lineage sorting may make interpretation of the lineage difficult (Connor 1991). Snow tussocks, which refer to many alpine species of *Chionochloa*, dominated the low alpine area and expanded to high alpine areas where there is frequent snow. The snow tussocks showed various degrees of freeze resistance, with species such as *C. rigida* showing seasonal changes whilst species such as *C. rubra* do not (Bannister *et al.* 2005). The slim snow tussock, *C. macra*, and the narrow leaved snow tussock, *C. rigida*, are two of the most common high mountain tussocks in New Zealand (Connor 1991). *Chionochloa rigida* can grow at an altitude of above 1,000 m whereas its close relative, *C. macra*, can be found at even higher altitude (>1,250 m) and is more common in the east of the South Island (Mark 2012).

Generally, grass species were adapted to habitats that are warm and with high humidity whereas two subfamilies of grass, the *Danthonioideae* and *Pooideae*, include species adapted to cool weather (Kellogg 2014). The *Pooideae*, which include more than 4,000 species in 14 tribes, are the largest subfamily of *Poaceae* and are classified as C_3 cool season grass (Soreng *et al*. 2015). The C_3 *Danthonioideae*, which belong to the PACMAD clade of grass, contain around 300 species and are located in the southern hemisphere (Soreng *et al*. 2015). The cold tolerance of the *Pooideae* had been well studied and cold-related genes were characterized, but the cold tolerance mechanism in *Danthonioideae* was not well investigated. Kellogg (2014) concluded that higher cold tolerance in *Danthonioideae* was different from the rest of *Pooideae* as several main clades including *Chionochloa* had much higher cold tolerance but the detailed reason for this was unknown. Snow tussock grassland plays an important role in the ecosystem, as it has a high ability to retain water and it keeps a large part of its water through low evapotranspiration (Wharton *et al*. 2010). Water collection from the condensation of fog is also a significant way of snow tussocks to gain water and ice nucleators on the leaf surface may contribute to this ability (Ingraham *et al.* 2008).

1.3. Freezing tolerance and freezing avoidance

Three mechanisms are involved in the survival of creatures experiencing freezing environments: freezing tolerance, freezing avoidance and cryoprotective dehydration (Sørensen and Holmstrup 2011). Freezing avoidance involves the prevention of intracellular ice formation and maintenance of a supercooled liquid state at low temperatures. Polar fishes and insects facing cold environments are freeze avoiding creatures, which decrease the freezing point with either self-produced low molecular weight cryoprotectant analogues including polyhydroxy alcohols, polyols and ethylene glycol, or thermal hysteresis proteins (THP), or inactivating ice nucleation sites so as to decrease the ice nucleation points (Chen *et al.* 1997; Duman 2001; Fletcher *et al.* 2001; Ng and Hew 1992; Sørensen and Holmstrup 2011). Freezing tolerance involves reducing the freeze damage once ice formed and to balance the osmotic pressure created by the extracellular ice and the removal of water, in which high level of ice nucleation activity and recrystallization inhibition activity are essential (Hengherr *et al.* 2009; Wharton and To 1996).

Both intracellular freezing and extracellular freezing are harmful. Extracellular

freezing increases osmotic pressure outside membranes by concentrating the solute excluded from the ice crystals and damaging the membranes. However, a high sub-zero freezing temperature and a slow freezing rate inside cells reduce the injury to cells, compared with higher freezing rate, which results in the formation of the intracellular ice which may fracture the cells (Lorv *et al.* 2014). Most researchers considered intracellular ice crystals to damage cells but it does not lead to cell death directly but as the result of ice recrystallization (Lorv *et al.* 2014). Many plants and bacteria adapt to freezing tolerance by the expression of recrystallization inhibition proteins (RIP) (Antikainen and Griffith 1997; Beck 1994; Chun *et al.* 1998; Raymond *et al.* 2008; Raymond *et al.* 2007; Thomashow 1998; Wilson *et al.* 2006; Winfield *et al.* 2010). TH activity had also been found in plants, fungi and bacteria but maintained at comparatively low levels (Atici and Nalbantoglu 2003; Duman and Olsen 1993).

The cryoprotective dehydration strategy is to loss water steadily by evaporation when facing freezing and this method is found in small invertebrates including some nematodes and chironomid larvae (Sørensen and Holmstrup 2011). The mechanism related to cryoprotective dehydration has been well studied and the gene families involved and their regulation in the Arctic springtail (*Megaphorura arctica*) has been reported (Clark *et al.* 2009).

Generally, the adaption to cold requires a series of physiological changes to produce proteins and other chemicals such as sugars to stabilize the cell membrane when facing desiccation and shrinkage, and to maintain the transference of intercellular water to the extracellular compartment to avoid the formation of ice crystals accelerating (Hoch *et al.* 2003; Sakai and Larcher 1987). Studies of cold acclimated proteins and chemicals, and the phylogenies of related genes are of great help to understand the evolution of plants experiencing cold environment.

The phylogenetic characterization of a freeze tolerance-related protein, dehydrin-like proteins, in the dogwoods *Cornus* was reported and the origin of these dehydrin-like proteins was inferred prior to the development of antifreeze mechanisms (Karlson *et al.* 2004). A winter highly expressed dehydrin-like protein (*mddhn*) from apple trees (*Malus domestica*) was studied and its phylogeny revealed an early divergence from peach (Garcia-Bañuelos *et al.* 2009). The phylogeny of another cold environment related protein, the early light-induced proteins (ELIPs) families, from overwintering catawba rosebay *Rhododendron catawbiense* leaves was inferred and a phylogenetic split between angiosperms and gymnosperms was found (Peng 2007).

The freezing response was overlaid to phylogenetic trees and the origin of the antifreeze activity in the birch *Betula* and the dogwood *Cornus* species was deduced, which suggested that freezing avoidance was an "ancestral trait" in the woody genera (Byard *et al.* 2010). The phylogeny of C-repeat binding factor (CBF) transcription factors and cold responsive (COR) genes were analyzed and a complex gene family structure and the geographic origin of the whitlow-grasses *Draba* species (*Cruciferae*) were found (Meijenfeldt 2010). Preston and Sandve (2013) found multiple origins of cold tolerance across the phylogenetic tree of temperate seed plants (Preston and Sandve 2013; Smith *et al.* 2011). Qian *et al.* (2013) used a phylogenetic approach to study cold region woody plants and they found plants in colder regions had more complex and newer structures (Qian *et al.* 2013). The phylogenies of cold-related RIP, FST, and CBF gene families were compared and it was found that the CBF, RIP, and FST genes in *Pooideae* evolved after the separation from rice (Li *et al.* 2012a). Thus, the study of phylogenic relationships of cold-related proteins would throw new light on the evolutionary history of cold adapted plants and their tolerance forming.

1.4. Ice active proteins (IAP)

Wharton *et al.* (2005) suggested the term, ice active proteins (IAPs), to define proteins that affect the formation and stablity of ice crystals (Wharton *et al.* 2005). They pointed out that not all the proteins interacting with ice have high thermal hysteresis or inhibit ice recrystallization. They also indicated three different types of ice active proteins (IAPs) which are essential in different cold tolerance mechanisms are recognized in many organisms: 1) ice nucleation proteins (INP), triggering ice formation; 2) thermal hysteresis proteins (THP), inhibiting ice formation; and 3) recrystallization inhibition proteins (RIP), stabilizing ice crystals once formed (Hartmann *et al.* 2013; Wharton *et al.* 2005; Wharton *et al.* 2010).

IAPs may illustrate more than one type of ice activity under certain conditions. Both INPs and antifreeze proteins (AFP) (proteins with TH activity and/or RI activity) bind to the ice surface, and some researchers hypothesized that these two types of proteins utilized a similar ice-binding strategy with the only differences being in molecular weight and working concentrations (Lorv *et al.* 2014; Wilson *et al.* 2010; Xu *et al.* 1998). They also reported that some antifreeze proteins

enhanced ice nucleation when they reached a certain concentration, and some INPs showed RI activity.

1.4.1. Ice nucleation activity (INA)

Pure water stays in a liquid state when the temperature is below 0 °C but when cooled further (~-42 °C), the supercooling point, water will freeze quickly (Bigg 1953). Liquid contains a sufficient number of ice nucleation agents will freeze at a high sub-zero temperature, where the ice nucleation agents act as core sites of crystallization and catalyze the growth of ice crystals. Ice nucleation agents play an essential part in changing water from liquid to solid ice, where the H_2O molecules gather around the ice nucleation agents and form small ice crystals (Zachariassen and Kristiansen 2000).

Ice nucleation activity (INA) has been most studied in bacteria since the 1980s (Maki *et al*. 1974). Generally, ice nucleation active bacteria are gram-negative bacteria such as *Pseudomonas syringae* and *Erwinia herbicola*. (Gurian-Sherman and Lindow 1993; Wolber and Green 1990). Ice nucleation proteins (INP) are secreted on the outer membrane and lead to supercooling temperatures from -2 °C to -10 °C. This property are widely used in artificial snow by introducing freeze-dried *Pseudomonas syringae* cells to the water used in snow production (Tegos *et al.* 2000; Turner *et al.* 1990; Wolber 1993; Wolber *et al.* 1986; Hartmann *et al.* 2013; Gurian-Sherman and Lindow 1993). There are many variants of the bacterial INPs that mainly fall into three catalogs: type A (very rare, has an INA temperature no less than -4.4 °C), type B (major INA bacteria, induces an INA temperature between -4.8 °C and -5.7 °C) and type C (all INA bacteria, has an INA temperature lower than -7.6 °C) (Turner *et al.* 1990). These three types of INPs also show different level of resistance to water-soluble organic solvents or pH variations (Turner *et al.* 1990). Turner *et al*. (1990) also proposed that the all the three types of INPs are the products of the *inaZ* gene, but there is post-translational modification, which made type C structure change to type B structure, and type B structure changed to type A structure.

Studies by Lindow *et al*. (1982) showed bacteria residing on the plant leaf surface expressed INPs that promoted the freezing of water at high sub-zero temperatures resulted in leaf damage (Kieft 1988; Lindow *et al.* 1978; Lindow 1982). In general, the *ina* genes have open reading frames (ORFs) from 3,600 bp to 4,000 bp and show homology of ~77% to *Pseudomonas syringae* and the bacterial INP is a single

protein, which has more than 1,200 amino acid residues of about 120 kDa (Li *et al.* 2012b). The N-terminal covers about 15% of the amino acid residues and was determined to relate to the membrane transport and binding ability whereas the C-terminal covers about 5% of the amino acid residues and do not have such functions (Li *et al.* 2004; Li *et al.* 2012b). The N-terminal and the C-terminal can fuse to other proteins without changing its INA, which is used frequently in the bacterial exhibition system (Sarhan 2011). The rest of these proteins have a large number of repetitive tandem domains that cover most of the proteins and have the function of manipulating water molecule during water crystallization. Muryoi *et al.* (2004) identified that post-translational modifications such as glycosylation and lipidation were quite important in the production of INPs (Muryoi *et al.* 2004).

Normally, one INP molecule has the property of triggering ice nucleation at the temperature of -12 °C and the high sub-zero INA temperature is due to the oligomerization of more than 50 single INPs (Garnham *et al.* 2011b). Garnham *et al.* (2011b) listed two previous theories on how the INPs form oligomerizes. These considered the repeat ice-binding motifs generated as stair-like overlapping or anti-parallel ß-strand array respectively. Garnham *et al.* (2011b) pointed out that those theories were inconsistent with the INP peptide studies that tyrosine in the repeat motifs facing the interior sides. Instead, they proposed another theory that the extension of dimer structures of connected ice nucleation surfaces from two INPs could largely increase the ice nucleation function areas, which in return raises the INA temperature.

Ice nucleation was reported in many plant tissues such as wheat leaves and flowers from fruit trees, with the INA temperatures varying from -2 °C to -14 °C (Lindow 1983). Generally, INA in plants was low whereas many cold sensitive plants were reported to have supercooling temperatures around -5 °C but there were no INA bacteria detected (Brush *et al.* 1994; O'Brien and Lindow 1988). These endogenous INAs were observed at low concentration but with high activity, such as *Prunus* stems (Gross *et al.* 1988), *Citrus* fruit (Constantinidou and Menkissoglu 1992; Constantinidou *et al.* 1991), the Afroalpine plant *Lobelia telekii* (Beck 1994; Krog *et al.* 1979), and lichens (Kieft 1988). Hacker *et al.* (2011) investigated INA temperatures in four alpine cushion plants (*Saxifraga bryoides, S. caesia, S. moschata* and *Silene acaulis*) and hypothesized that intrinsic ice nucleation at high sub-zero temperature helped with freezing survival (Hacker *et al.* 2011). The composition of

intrinsic INA agents was characterized to be different from bacterial INA agents and fell into two groups: protein INA agents and non-protein INA agents (Griffith and Antikainen 1996). Intrinsic INA agents in *Lobelia telekii*, which induced nucleation at -4 °C was a heat stable (~100 °C) carbohydrate (Kishimoto *et al.* 2014). INA in *Prunus* was retained after protein denaturing treatments with organic solvents and proteases (Gross *et al.* 1988). Ice nucleation agents in the rock-posy lichen, *Rhizoplaca chrysoleuca*, which were separated by sonication and homogenization from cell walls, were proteinase sensitive but heat and pH stable (Kieft and Ruscetti 1990). INP was also discovered in winter rye, *Secale cereale*, which consisted of lipoprotein and carbohydrates (Brush *et al.* 1994). The composition of the ice nucleation agent in birch pollen was a large polysaccharide (formed by clusters of small polysaccharides), which was retained by a 100 kDa filter (Dreischmeier *et al.* 2017). Seasonal changes were found in plant INA and it was indicated that the plant INA was the highest in winter whereas reduced when the environmental temperature increased (Griffith and Antikainen 1996). It was also found that INA in plants including winter rye, *Rhododendron* species and *Prunus* species stayed constant without fluctuations (Griffith and Antikainen 1996).

Ice nucleation is important in freeze tolerant animals as ice nucleation is triggered at high sub-zero temperature and maintained in a controlled way, which slowed the ice formation rate and made it possible for vertebrates to adjust to the freezing environment (Wharton 2012). INA bacteria resident on the skin and in the gut of amphibians initiated ice nucleation at high sub-zero temperature. The frog *Litoria ewingi* secreted ice nucleators on their skin to ensure ice nucleation was initiated at a high sub-zero temperature (Wharton 2012). There were proteins which can trigger ice nucleation at -6 °C to -8 °C in the blood plasma and gland secretion of wood frogs in winter (Storey and Storey 2013). INPs were also found in the body fluid of several tardigrade, molluscs and insects with protein similarity (Costanzo and Lee 1996). Several invertebrate INPs were well characterized and showed great differences, such as the 74 kDa INP in overwintering queens of the hornet and the 800 kDa lipoprotein in the cranefly, *Tipula trivittata* (Storey and Storey 2013).

A study in 2010 reported that two of the New Zealand snow tussocks, *C. macra* and *C. rigida*, have ice nucleation activity but no antifreeze or recrystallization inhibition activity (Wharton *et al.* 2010). This work showed that there were endogenous INPs in these species and it was possible that their role was to encourage

the formation of condensation on the tussock at low temperatures and high humidity. However, nothing else was known of these proteins (Wharton *et al.* 2010).

1.4.2. Thermal hysteresis activity (TH)

Thermal hysteresis proteins (THPs), also known as antifreeze proteins (AFPs), decrease the freezing point through the Kelvin effect and lead to curve ice surfaces between those already bound, which raises the energy cost of attaching more water molecules to the ice surfaces, whereas the melting point is not affected (Wilson 1993). AFPs, which have been found in many bacteria, fungi, plants, arthropod and aquatic animals, had flat hydrophobic ice-binding surfaces (Jia and Davies 2002). AFPs irreversibly absorbed to the ice surface plane, integrated into the ice crystal matrix, and result in different non-round shaped ice crystals such as hexagons, columns or bi-pyramids, where the preferred axis face is different among fish, insert and plant AFPs. More than one type of AFPs is expressed so as to bind different faces of ice crystals (Jia and Davies 2002; Hew *et al.* 1988). AFPs can also act as an inhibitor of INPs with different protein working concentrations (Parody-Morreale *et al*. 1988). There is considerable sequence variability in the AFPs among species. AFPs have been shown to have either hydrophilic or hydrophobic ice-binding domains (Griffith and Yaish 2004). Moreover, some AFPs show structured changes when freezing at low temperatures. AFPs may inhibit ice crystal growth and prevent ice recrystallization so as to protect creatures from freezing injury (Atici and Nalbantoglu 2003). Most AFPs apparently depress the solution freezing point whereas the AFPs in plants have comparatively low thermal hysteresis. However, inhibition of the recrystallization of ice is essential in plants (Griffith and Yaish 2004).

AFPs were first discovered in Antarctic fish blood and are well-characterized (DeVries *et al*. 1970; Chao *et al.* 1995). The long α-helix structure and repeat ice-binding motifs in the AFPs from winter flounder were reported and it is deduced that the terminal caps were important to the structure (Sicheri and Yang 1995). Parallel ß-helix ice-binding structures were found in the 9 kDa AFPs from the eastern spruce budworm, *Choristoneura fumiferana* (Graether *et al*. 2000). Up till now, there are five types of AFPs identified in fishes: AFGP, Type I, II, III and IV AFPs. The antifreeze glycoproteins (AFGPs), which was first identified in Antarctic notothenioid fishes, has repeated tripeptide units (Ala-Ala-Thr/Arg) and fall into eight subtypes with sizes

ranging from 3 kDa to 34 kDa (DeVries and Wohlschlag 1969; Knight *et al.* 1984). Type I AFP is a single right-handed α-helix Ala-rich protein, in which the Ala-rich repeat site serves as the ice-binding surface with Thr and Asp on the opposite position of the ice-binding surface (Baardsnes *et al.* 1999). Type I AFPs are widely distributed in three groups of fishes and have many variant subtypes (Graham *et al.* 2013). Type II AFPs, which are Cys-rich globular proteins supported by multiple disulfide bonds, formed ß-sheet structures and showed similarity to the type-C lection families (Cheng 1998; Gronwald *et al.* 1998). Type II AFPs can be divided into Ca^{2+} dependent and Ca^{2+} independent groups. The Ca^{2+} is involved in the ice binding of the Ca^{2+} dependent type II AFPs and the change of ion results in the change of ice crystal shape, whereas the Ca^{2+} independent AFP II does not have the Ca^{2+} binding site (Barrett 2001; Nishimiya *et al.* 2008). Type III AFPs are small globular proteins containing ß-strand structures and show similarity to the C-terminal domain of the enzyme sialic acid synthase (SAS) but without amino acid preferences (Baardsnes and Davies 2001). Type IV AFPs are α-helix bundle structures, which might be analogues to the apoplipoprotein family (Deng and Laursen 1998).

AFPs are also found in many arthropods including insects since 40 years ago of the first insect AFP in darkling beetle *Tenebrio molitor* (Duman 2001; Ramsay 1964). Insects AFPs show the highest thermal hysteresis, of about 4 ℃ to 6 ℃ compared to fish AFPs (around 0.4 ℃ to 2.0 ℃), bacteria AFPs (0.1 ℃ to 0.3 ℃) and plant AFPs (0.15 ℃ to 0.7 ℃) (Lorv *et al.* 2014). The highly effective insect AFPs share similar ß-helix structures which serves as the ice binding surface and contains specific repeat motifs such as “threonine-x-threonine’ where one of the threonine may be replaced by other amino acids according to species difference (Friis *et al.* 2014; Graether and Sykes 2004). For instance, in *Tenebrio molitor*, the 8.4 kDa AFP was formed by disulfide bond linked loops, in which there are tandem repeat units consisted of threonine-cysteine-threonine motifs and formed the ß-sheet ice-binding surface at one side of the molecule (Liou *et al.* 2000). There are high diversities in insect AFPs and they can be divided into mainly two groups based on their tertiary structures: right-handed helix and left-handed helix (Jia and Davies 2002).

It is summarized that four species of bacteria produced AFPs including *Marinomonas primoryensis, Pseudomonas putida, Micrococcus cryophilus* and *Rhodococcus erythropolis* (Gilbert *et al.* 2004). More Arctic cryoconite bacteria were reported showing antifreeze activity and these AFPs have molecule weights around 22

kDa and belong to the cold-related IBP-1 family (Singh *et al.* 2014). Generally, bacterial AFPs are large molecules with low TH but Singh *et al*. (2014) found that several bacteria have TH activity no less than 2 °C. The AFPs of *Marinomonas primoryensis* has a molecular weight around 1.5 MDa and lost activity when EDTA was present, which removes the Ca^{2+} ions from its structure (Guo *et al*. 2017). This protein has two conserved repeat regions (II and IV) but only the region IV, which takes up only fiftieth of the entire protein, contributes to the antifreeze activity (Guo *et al.* 2012). In *Flavobacterium xanthum*, the TH of a 59 kDa AFP can be enhanced dramatically when malate is present, rising from 0.04 °C to 5.2 °C.

AFPs have been found in more than 50 species of overwintering plants (Ashworth and Kieft 1995; Franks 1985; Levitt 1980; Sinclair *et al.* 2003; Griffith and Yaish 2004). Antifreeze proteins (AFPs) found in winter rye had TH activity of around 0.3 °C and lead to a hexagonal shaped ice crystal in the nanoliter osmometry. (Griffith *et al.* 1992; Griffith and Yaish 2004; Hon *et al.* 1995). These proteins ranged from 9 kDa to 36 kDa. A 67 kDa AFP in the plant *Solanum dulcamara* was found containing carbohydrate and glycine components (Duman 1994). Besides these, cryoprotectants such as glucose and glycerol also have an influence on the cold tolerance of plants (and other organisms) (Kukal *et al.* 1989; Rexer-Huber *et al.* 2011). AFPs in plants are not continuously expressed but accumulated when plants facing the cold temperatures and short days overwinter (Griffith and Yaish 2004). Dehydrin, ethylene and Ca^{2+} are reported to be involved in the AFPs expression in winter rye (Yu *et al.* 2001). AFPs in plants are quite different from those of fishes and insects, but surprisingly show similarity to pathogenesis-related proteins, which usually have molecular weights from less than 4,000 to over 80,000 and are regulated by different gene families triggered by adverse conditions not only limited to cold temperature (Garnham *et al.* 2011a). Plant AFPs contained chitinases, thaumatin-like proteins, β-1,3-glucanases, and a polygalacturonase inhibitor proteins, and the encoding genes for several AFPs have been identified (Kumble *et al.* 2008).

1.4.3. Recrystallization inhibition activity (RI)

Small ice crystals already formed have the tendency to grow into large crystals, which is known as ice recrystallization (RI) (Knight *et al*. 1984). Two theories have been provided to explain this, which are grain boundary migration and Ostwald ripening

(Capicciotti *et al.* 2013). In the first theory, smaller ice crystals have more curved surfaces and higher energy than large ice crystals. The grain boundary migrates in order to reduce the overall curvature and energy status of the ice crystals by reducing the size of small ice crystals with convex surfaces and enlarging large ice crystals with less convex or concave surfaces. In Ostwald ripening, water molecules migrate from the surfaces of small ice crystals which are less stable to the bulk water whereas water in the bulk water move to large ice crystals which are more stable. In this theory, the total volume of ice crystals is constant and this process reduces the overall energy status (Capicciotti *et al.* 2013). The Ostwald ripening theory is more favored since the discovery of the quasi-liquid layer (QLL), which is an ordered water layer between the bulk water and the ice lattice. Ice recrystallization thus was interpreted as water molecules in the bulk water migrating to the QLL and then incorporated into the ice crystals.

RI activity was mainly observed and studied in cold tolerant creatures including plants, some insects and fishes (Doucet *et al.* 2000). The ability to inhibit ice recrystallization was found in many species although studies still continue to determine the exact mechanisms. Since the 1990s, RI activity was reported in AFPs that showed obvious thermal hysteresis but generally, the RI activity requires much lower concentration of AFPs compared with TH activity (Atici and Nalbantoglu 2003). Type III AFPs were proved to protect red blood cells due to RI activity during warming temperature and AFGPs were reported to show RI activity with no TH activity (Chao *et al.* 1996; Eniade *et al.* 2001). In the winter rye some AFPs with TH activity exhibited RI activity (Griffith *et al.* 1992). Later, these AFPs were found belonged to an anti-pathogen protein family, and anti-pathogen ability was correlated with the antifreeze ability in this species.

Unlike AFPs in fishes and insects that have high TH activity around 1 ℃ and 5 ℃ respectively, AFPs in plants have comparatively low TH activity, about 0.1 ℃ 0.5 ℃, and this indicates that the main antifreeze mechanism in plants is not preventing ice formation through TH activity (Griffith and Yaish 2004; Gupta and Deswal 2014). The majority of cold tolerant plants express proteins that interact with edges of ice crystals and stabilize the ice boundaries (John *et al.* 2009). Plant species from the UK and maritime Antarctic were screened, and it is reported that plants in the superorder *Commeliniflorae* expressed comparatively high RI activity (Doucet *et al.* 2000). However, it is indicated that RI activity was not found in the leaves of any

evergreen species including the *Pinopsida*, but in a large proportion of lichens (Doucet *et al.* 2000). In grass species including wheat, hair grass and barley, there are similar repeat domains in IAPs with RI activity (Lauersen *et al.* 2011). They found these IAPs were homologues and showed low TH activity and strong RI activity with no INA.

RIPs in plants are induced by cold and have a positive correlation with decreasing temperature. Until now, RIPs have been only well characterized in *Pooideae* grasses. They have leucine rich repeats (LRR) in the N-terminal and species-specific ice-binding repeats (NxVxxG and NxVxG) in the C-terminal was identified in various isoforms of *Pooideae* RIPs (John *et al.* 2009; Sandve *et al.* 2008; Tremblay *et al.* 2005). This structure is not like the ice-binding repeats in THPs but form two parallel hydrophobic β-sheets, which act as the ice-binding surface (John *et al.* 2009). The LRR repeats in the RIPs showed similarity with transmembrane receptor-like kinases (RLKs) proteins, which regulate plant growth and are involved in protein microbe interaction and the stress response.

Non-protein molecules were also found that contributed to the RI activity. Common non-toxic cryoprotectants such as glycerol and DMSO cross cell membranes and reduce the concentration of intracellular solutes (Briard *et al.* 2016). They also break the hydrogen bonds between water molecules and form stronger hydrogen bonds with water molecules, which disturb the ice lattice growth. The water-soluble polymer PVA interacts with ice crystals and inhibits ice recrystallization (Olijve *et al.* 2016). Following the discovery of a C-linked carbohydrate in an AFGP contributing to the RI activity, several carbohydrates including D-galactose and D-melibiose were reported as efficient recrystallization inhibitors with a high hydration index, which presents the hydration number per molar volume of carbohydrates, where the hydration number indicates the ability of the solute to firmly bind the surrounding water molecules (Tam *et al.* 2008). Tam *et al.* (2008) hypothesized RI activity as hydrated carbohydrates disturbing the ordering of bulk water and raising the energy of for water molecule to move to the QLL.

1.4.4. Applications

IAPs have already been widely commercially used. Bacterial INPs were well characterized and used widely in artificial snow production and food preparation such

as juice and beer to save energy during the concentrating process (Christner *et al.* 2008; Sun 2016). AFPs with TH activity and/or RI activity are mainly used in the cryopreservation industry. The total protein from winter wheat grass extract is added to the ice cream in order to reduce the ice recrystallization effect, which result in the loss of a creamy texture (Regand and Goff 2006). Meats were soaked in AFP solution from winter flounder before freezing to reduce the damage and the loss of food nutrition during the freezing and thawing process (Sun 2016).

1.5. General approaches to study ice activities

Various methods have been developed to screen these three types of ice activities. The splat cooling assay was initially described to measure recrystallization inhibition (RI) activity by directly observing the ice crystal size changes for 30 min in a thin ice layer covered by two round coverslips that placed on a temperature controlled cold stage under a microscope, where the ice layer was formed by dropping 10 μL test solution to a dry ice precooled aluminum cube from 2 m height (Knight *et al.* 1988). The sucrose sandwich assay was another approach with a similar principle but in the presence of 30% sucrose. Another instrument, an optical recrystallometer, was introduced in 2007 to measure recrystallization inhibition (RI) by measuring changes in the optical transmittance through frozen samples at programmed temperatures (Wharton *et al.* 2007). This technique indirectly analyzed the RI but saved much time compared with the splat cooling assay and sandwich freezing methods (Doucet *et al.* 2000). Other methods used in measuring RI activity include capillary assays and wide angle X-ray scattering (Abraham *et al.* 2015).

The measurement of INA can be either direct or indirect. An ice nucleation spectrometer was constructed to screen the ice nucleation activity (INA) indirectly by using a thermocouple to measure the temperature of crystallization (T_c), the supercooling point, of six glass capillary tubes of samples in one aluminum holder, four of which were gradient cooled by a programmed thermal circulator (Wharton *et al.* 2004). A drop freezing method was a direct way to measure the INA in which aliquots of test solution were placed into PCR tubes (2 μL per tube) that were sealed and placed in a thermal cycler maintained at 4 °C. Then the temperature was programmed to decrease at various constant rates and the number of frozen test solution tubes was recorded at each temperature gradients. Recent development of

INA assay platform consisted of thermal video camera, cooling unit and thermalyzer program enable batches of samples to be analyzed at the same time (Wisniewski *et al.* 2015).

Nanoliter osmometers were used to investigate thermal hysteresis by detecting the melting points, hysteresis freezing points and effects on ice crystal growth of test samples (Wharton *et al.* 2010). In this method, a droplet of sample solution was mounted in Cargille's A mineral oil (Cargille Labs), which filled the six wells of the aluminum sample holder. The sample solution was frozen quickly and melted slowly till there was only one single ice crystal left. The temperature of the sample was adjusted using the fine control and the point when the ice crystal started to grow or shrink were recorded as the hysteresis freezing point and melting point respectively. Milli Q deionized water and osmolality standards (Opti-mole, Wescor) were used to calibrate the readings.

1.6. Research aims and thesis overview

This study was to examine whether ice activites are typical for the *Chionochloa* species in New Zealand or not, what genes are involved and what these proteins are. The first part presents the investigation and characterization of ice activities in *Chionochloa* species and an attempt to identify related proteins by ice affinity purification from the raw extracts (chapter 3, chapter 4 and chapter 5). The second part demonstrates an annotated transcriptome of the *C. macra*, which was utilized to analyze ice active proteins and study the evolutionary history of cold response-related proteins (chapter 6 and chapter 7).

1.6.1. Aims and hypothesis

The first aim of this PhD thesis is to investigate the presence of ice activities in New Zealand *Chionochloa* species.

Hypothesis 1: Ice activities are commmon in most New Zealand *Chionochloa* species. (chapter 3)

The second aim is to examine patterns of ice activities in New Zealand *Chionochloa* species.

Hypothesis 2: There are seasonl pattern and geographic pattern of ice activities in New Zealand *Chionochloa* species. (chapter 3)

The third aim is to investigate the importance of plant intrinsic proteins in maintaining ice activities in *Chionochloa* species.

Hypothesis 3: Plant intrinsic proteins are mainly responsible for ice activities in *Chionochloa* species. (chapter 4)

The fourth aim is to investigate the characters of ice activities in *Chionochloa* species.

Hypothesis 4: The characters of INA in *Chionochloa* species is different from those of bacteria. Characters of TH and RI activity are similar to current known plant IAPs. (chapter 4)

The fifth aim is to examine various methods to purify and identify IAPs in *Chionochloa* species.

Hypothesis 5: All three types of IAPs can be harvested by ice affinity purification of *Chionochloa* species. Currently known plant IAPs are present in the ice fraction. (chapter 5)

The sixth aim is to investigate genes related to cold tolerance and water collection in *Chionochloa macra*.

Hypothesis 6: Ice active genes is present and high expressed in *C. macra* transcriptome. (chapter 6)

The seventh aim is to study the evolution of cold response-related genes in *Chionochloa* species.

Hypothesis 7: Cold response related genes in *Chionochloa* show quite different evolutionary history compared with the core *Pooideae* grass. (chapter 7)

1.6.2. Thesis overview

Chapter 2 describes the materials and general methods used in this study.

Chapter 3 provides an investigation on seasonal variations in ice activities from 17 *Chionochloa* species and subspecies.

Chapter 4 comprises the physical and chemical characterization of ice activities in *Chionochloa* species.

Chapter 5 presents the attempt to purify and identify the IAPs in *Chionochloa* species using various methods.

Chapter 6 presents the annotation and the analysis of the transcriptome of *C. macra* from individuals collected in winter that showed ice activities.

Chapter 7 describes the identification of putative IAP genes in the *C. macra* transcriptome and the use of the *E. coli* system to study the ice activities of expressed proteins. Moreover, phylogenies of cold response-related genes were inferred and compared to those from the *Pooideae*.

Chapter 8 presents the overall conclusions from this thesis and suggestions on further research of ice activities in this genus.

Chapter 2 Materials and general methods

2.1. Materials

2.1.1. Buffers

Type I water (mQ H_2O) (resistivity>18 MΩ at 25 °C) from Millipore Milli-Q Plus water system (Millipore) was used for all experiments and all buffers were filtered through 0.45 µm filters and degased.

2.1.1.1. Plant total protein extraction buffer

One tablet of cOmplete™, Mini Protease Inhibitor Cocktail tablet (Roche) was added to 10 mL buffer (25 mM Tris-HCl, 5 mM EDTA, 1 mM PMSF, pH 8.0)

2.1.1.2. Buffers for fast performance liquid chromatography (FPLC)

2.1.1.2.1. Ion exchange column (Resource Q, 5mL, GE Healthcare)

Binding buffer
25 mM Tris-HCl, pH 8.0
Elution buffer
25 mM Tris-HCl, 1 M NaCl, pH 8.0

2.1.1.2.2. Desalting column (HiTrap desalting, 5mL, GE Healthcare)

Elution buffer
25 mM Tris-HCl, pH 8.0

2.1.1.2.3. Hydrophobic interaction column (GE Healthcare)

Binding buffer
1.7 M $(NH_4)_3SO_4$, 25 mM Tris-HCl, pH 8.0
Elution Buffer
25 mM Tris-HCl, pH 8.0

2.1.1.2.4. Size exclusion column (Superdex 75, GE Healthcare)

Elution buffer

25 mM Tris-HCl, 150 mM NaCl, 0.02% Tween 20, pH 8.0

2.1.1.2.5. Immobilized metal affinity resin (TALON Superflow Resins, Clontech)

Binding buffer
20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.4
Elution Buffer
20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4

2.1.1.2.6. Immobilized metal affinity resin (Glutathione Sepharose 4B resin, GE Healthcare)

Binding buffer
140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3
Elution Buffer
50 mM Tris-HCl, 10 mM reduced glutathione, pH 7.3
Regeneration buffer I
100 mM Tris-HCl, 500 mM NaCl, pH 8.5
Regeneration buffer II
100 mM CH_3COONa, 500 mM NaCl, pH 4.5

2.1.1.3. Buffer for protein gel electrophoresis (SDS-PAGE)

Stacking buffer 4%
6.3 mL 0.5 M Tris-HCl pH 6.8; 250 µL 10% SDS; 15.9 mL mQ H_2O
Stack gel solution
1.12 mL 4% stacking buffer; 125 µL 40% Acrylamide; 4 µL TEMED; 6 µL 10% ammonium persulfate
Resolving buffer 12%
25 mL 1.5 M Tris-HCl pH 8.5; 1 mL 10% SDS; 43.5 mL mQ H_2O
Resolving gel solution
3.48 mL 12% Resolving buffer; 1.5 mL 40% Acrylamide; 10 µL TEMED; 15 µL 10% ammonium persulfate
Sample loading buffer 5X
10% SDS; 20% Glycerol; 0.2 M Tris-HCl; 0.05% Bromophenol blue; 5.0% β-mercapto-ethanol

Running buffer 5X

12 g Tris-base; 94 g Glycine; 5 g SDS; mQ H_2O to 1 Litre

Colloidal Coomassie Blue stain solution

50 mL O-phosphoric acid; 100 mL H_2O; 50 g $AlSO_4$ hexadecahydrate; 1.2 g Coomassie Brilliant Blue G250; 100 mL ethanol; mQ H_2O to 1 Litre

Coomassie R-250 Blue stain solution

250 mL ethanol; 80 mL acetic acid; 1.25 g Coomassie Blue R-250; 670 mL mQ H_2O;

Coomassie R-250 Blue destain solution

250 mL ethanol; 80 mL acetic acid; 670 mL mQ H_2O

2.1.1.4. Dialysis buffer

25 mM Tris-HCl, 50 mM NaCl, pH 8.0

2.1.1.5. Cell culture

LB broth

10 g Peptone; 5 g Yeast extract; 5 g NaCl; 1 Litre mQ H_2O;

Autoclave at 121 ℃ for 20 min

TB broth

12 g Peptone; 24 g Yeast extract; 4 mL glycerol; 12.54 g K_2HPO_4; 2.313 g KH_2PO_4; mQ H_2O to 1 Litre;

Autoclave at 121 ℃ for 20 min

Auto induced medium

6 g Na_2HPO_4; 3 g KH_2PO_4; 20 g Peptone; 5 g Yeast extract; 5 g NaCl; mQ H_2O to 1 Litre;

Autoclave at 121 ℃ for 20 min, cool to 55 ℃;

Then add 0.45 μm syringe filter filtered,

(10 mL 60% v/v glycerol; 5 mL 10% w/v Glucose; 25 mL 8% w/v Lactose)

LB plates (With appropriate antibiotics)

10 g Peptone; 5 g Yeast; 5 g NaCl; 15 g Agar; mQ H_2O to 1 Litre;

Autoclave at 121 ℃ for 20 min, cool to 55 ℃;

Then add,

Appropriate antibiotics; Plate for 15 mL each

Super optimal broth with catabolite repression medium (SOC medium)

2% (w/v) Tryptone; 0.5% (w/v) Yeast extract; 10 mM NaCl; 2.5 mM KCl;
Autoclave at 121 ℃ for 20 min, cool to 55 ℃, then add,
10 mM $MgCl_2$; 20 mM Glucose
Antibiotics
Antibiotics were added to cell culture media according to Table 2.1 below

Antibiotics	Stock concentration (mg/mL)	Working concentration (μg/mL)
Ampicillin	100	100
Kanamycin	50	50
Chloramphenicol	34	34

Table 2.1 Antibiotics and stock concentrations.

2.2. Methods

2.2.1. Sample collection

Seventeen species and subspecies of *Chionochloa* were sampled at the Dunedin Botanic Garden (lat. -45.858483°, long 170.524330°, alt. 80 m above sea level), Dunedin, New Zealand. These collections were initially derived from alpine regions within New Zealand and now form part of a working collection. Four seasonal collections were scheduled between 2014 and 2015 coinciding with spring, summer, autumn and winter, and *Chionochloa* species accession numbers were recorded to ensure the same species in each batch of collections (Accession numbers in chapter 3). *Chionochloa macra* was also sampled at Rock and Pillar Range (~1400 m) on 14th January, which allows to compare ice activities in *C. macra* with environment differences (Dunedin Botanic Garden vs Rock and Pillars).

Permits were granted from Dunedin Botanic Garden for all the sample collections in this work. Fresh plant leaves were cut off using autoclaved scissors, put in separate sealed sample bags then stored in dry ice immediately, and transferred to the -80 ℃ freezer when returned to the laboratory. Ethanol (70%) was used to sterile the scissors between each sample collection to avoid contamination.

2.2.1.1. Protein extraction

Leaves were cut into small pieces (~2 cm) and ground to fine powder in liquid nitrogen. Plant total protein extraction buffer (section 2.1.1.1) was used to resuspend the leaf powder. The raw extract solution was centrifuged at 16,000 × g, 4 ℃ for 20

min and the supernatant was kept. All solutions were stored at -20 ℃ for short term until required.

2.2.2. Ice activity assays

2.2.2.1. Quantification

All protein concentrations were estimated by A_{280} on a Nanodrop ND-1000 (Thermal Scientific) and Qubit protein assay (Invitrogen).

2.2.2.2. Ice nucleation activity (INA) assay

INA was screened by the INA spectroscopy, which measured the temperature of ice crystallization of test sample droplets (Wharton *et al.* 2004). Droplets of 10 µL protein extract solutions were filled into glass capillary tubes with mineral oil sealed at both ends and cristaseal sealed at the bottom. Six tubes were placed in an aluminum holder with a Nickel-Chromium/Nickel-Aluminium (NiCr/NiAl) thermalcouple inserted in the center of these tubes. Aluminum holders were placed in a metal cooling block, which was cooled by a Haake F8-C35 programmable refrigerated circulator. The block was cooled at a rate of 1 ℃/min from 1 °C to -25 ℃ and temperature changes were recorded continuously by a LabChart program (ADInstruments) on a computer which receiving the signals from the thermocouples interface through an A/D convertor. Temperatures in the block were recorded by an electronic thermometer, which was then used to calibrate the readings from the LabChart program. When the test droplets started to crystalize and release heat that was received by the thermocouple, which produced a peak in the LabChart program and indicated the temperature of crystallization (T_c). The temperature reading just before each peak was recorded as an individual T_c. T_cs were recorded except where peaks overlapped. Overlapped peaks were checked manually to assign individual T_cs. Twelve T_cs were recorded for each sample.

2.2.2.3. Thermal hysteresis (TH) assay

Thermal hysteresis activity was measured by nanoliter osmometry (Wharton *et al.* 2005). The sample stage was connected to a refrigerated circulator with a dry air pump to prevent condensation. The sample holder was first washed with xylene and

air dried, then attached to the cold stage of the osmometer with heat sink compound. Sample wells of the round sample holder were filled with Cargille's B mineral oil (Cargille Labs) then sample solution was injected to submerge in the mineral oil. Distilled water and 1000 mmol/kg standard were loaded into other wells for calibration. A Zeiss Axiophot Photomicroscope was used to observe the sample stage. The samples were rapidly frozen and then warmed to the point where there was only one single ice crystal in the well. The melting point was recorded as when the ice crystal started to shrink and the freezing point as when the crystal began to grow. The difference between the hysteresis freezing and the melting points is the thermal hysteresis. The shape of the ice crystal was recorded by a Cannon powershot A640 digital camera connected to the photomicroscope. Three TH measurements were taken for each sample.

2.2.2.4. Recrystallization inhibition (RI) assay

Recrystallization inhibition activity was examined by the splat cooling assay (Knight *et al.* 1988; Ramløv *et al.* 1996; Wharton *et al.* 2007). A 10 µL droplet of test solution was dropped from 2 m high onto an aluminum block precooled by dry ice. The resulting ice layer was transferred between two glass coverslips to a cold stage maintained at -20 °C. A Peltier module and a refrigerated circulator controlled the temperature of the cold stage, with condensation removed by a nitrogen gas flush. The cold stage was then raised to the annealing temperature -8 °C and the ice layer was observed under a 10X objective lens. Images of the ice layer was taken at the time when the cold stage raised to -8 °C and at 30 min later when stayed at -8 °C. Ten largest ice crystals were randomly selected and the diameters were measured with Zeiss AxioVision software. RI activity of each sample was assessed by calculating the mean diameters of ice crystals at 30 min after annealing at -8 °C. Ten largest ice crystals were measured for each sample.

2.2.3. Protein purification

2.2.3.1. Ice affinity purification

A metal finger connected to a programmable refrigerated circulator was used for the purification of IAPs. The filtered sample solution was precooled to ~4 °C and the metal finger was precooled to 0 °C. The cold finger was first dipped into filtered mQ

water (4 °C) and seeded with small ice crystals. Once a thin layer of ice formed around the metal finger, it was submerged in the precooled sample solution with a magnetic stirring bar stirring at the bottom. The temperature of the ice-finger was then programmed to decrease from 0 °C to -2 °C in 16 h. Ice hemisphere was formed and proteins interacting with ice crystals were included in the growing ice lattice. After 16 h, the ice hemisphere was washed with precooled mQ water to remove remain sample solution. The circulator was set to 1.5 °C and the ice hemisphere was released in a new autoclaved beak. When the ice hemispheres melt, a second round of ice-finger purification was performed using the ice fraction as the start material. The final ice fraction was centrifuged with a 3 kDa MWCO centrifugal device to concentrate.

Ice-shell purification was developed, which showed higher efficiency than the traditional ice-finger purification (Marshall *et al.* 2016). Briefly, 100 mL mQ was poured into a 500 mL round glass flask and rotated slowly in the 95% ethanol precooled by dry ice for 45 s. Around 15-20 mL water was frozen along the glass to form a hollow thin layer of ice sphere (ice-shell) in the round glass flask and the remain water was poured out. The flask was again rotated in the ethanol-dry ice bath to allow the ice-shell completely frozen. Precooled protein extracts (~0 °C) was then slowly poured into the mentioned ice-shell of the round flask which was then placed rotating in a -1 °C to -1.5 °C ethanol bath for 1-2 h to allow 50% of the solution incorporated into the ice-shell. With the growth of the inner ice surface of the ice-shell, proteins in the raw extract that interacted with the ice lattice were included whereas other proteins were not. The remaining solution was poured out and the inner of ice sphere was washed with 10 mL precooled mQ to get rid of remain sample solution. The ice-shell was melt in room temperature and harvested as ice purified fraction. A second round of the ice-shell purification was performed with the ice-purified fraction from the 1st round to increase the purity of ice-binding proteins.

2.2.3.2. Fast performance liquid chromatography

AKTA Prime^{TM} plus machine (GE Healthcare) was used for running the fast performance liquid chromatography (FPLC). Generally, the machine was first performed system wash by 20% ethanol and then filtered mQ. After that, the required column was attached to the system, washed with binding buffer and system was zeroed. Sample was injected in the correct sample loop and loaded on the column.

After another wash of binding buffer till a steady UV reading was observed, elution was performed by either step elution with designed concentration of elution buffer or gradient elution with set range of concentration of elution buffer.

2.2.3.2.1. Desalting

HiTrap Desalting 5 mL column (GE Healthcare) was used to remove high salt in the sample. Machine setting was followed the Application Template-HiTrap desalting option. Desalting was performed by balancing the column with 10-column volume of low salt elution buffer followed by sample injection. Peak fractions were collected at 1.5 mL fraction.

2.2.3.2.2. Ion exchange

Resource Q 5 mL column (GE Healthcare) was used as the anion-exchange chromatography in this study. Gradient elution was set as 0 to 100% elution buffer in 20 mL length, fraction size 1.5 mL.

2.2.3.2.3. Size exclusion

A 445 mm×15 mm Superdex 75 column was first equilibrated with elution buffer at a flow rate of 0.5 mL per min until the A_{280} absorbance reached a steady baseline. 1 mL sample was injected and the collected fraction size was set at 1.5 mL.

2.2.3.2.4. Hydrophobic interaction

The column was first balanced with start buffer (100 mM Na_3PO_4, pH 7.0), and then sample was injected into the loading loop. Gradient elution was performed with 0% to 100% elution buffer (1.7 M $(NH_4)_2SO_4$, 100 mM Na_3PO_4, pH 7.0) to wash out remained proteins. Three types of hydrophobic interaction column were tried: HiTrap Capto Butyl, HiTrap Capto Phenyl and HiTrap Octyl.

2.2.3.2.5. Immobilized metal affinity resin

TALON™ Superflow Resins and GST 4B resin in gravity flow columns were applied in the purification of His-tagged proteins and GST-tagged proteins respectively following the manufacturer's protocols.

2.2.3.2.6. Dialysis

Protein elution from high salt was dialyzed against 3,500 WMCO membranes with 20 mM Tris-HCl, 100 mM NaCl, pH 8.0 overnight. The desalted samples were freeze dried to concentrate.

2.2.3.3. Protein gel electrophoresis (SDS-PAGE)

Protein samples were analyzed on a Mini-PROTEIN®System (Bio Rad) with gel electrophoresis. Samples with a volume of 4-16 µL were mixed with 1-4 µL 5X sample loading buffer, which were denatured at 95 °C for 10 min. The processed protein samples were then loaded and separated on a 4% stacking/12% resolving SDS-PAGE at 80 V for the 4% stacking gel and 125 V for the resolving gel in 1X SDS running buffer. PAGEmark™ Protein Markers standard (G-Biosciences)/Precision Plus Protein™ All Blue Prestained Protein Standards (Bio Rad)/Blue Protein Standard, Broad Range (New England Biolabs) was used to estimate the molecular weight of sample proteins. The visualization was carried out using Colloidal Coomassie Blue stain. Gels were first fixed with 20% ethanol (microwaved for 10 s) and then washed in distilled water to reduce background. Gels were stained in Colloidal Coomassie Blue stain buffer for 4-5 h and destained in distilled water.

2.2.4. Mass spectrometry

Protein sample solution was lyophilized first and 40 µL of solubilization buffer (80% acetonitrile in 50 mM Tris-HCl, 10 mM $CaCl_2$, pH 7.8) was added with an incubation time of 20 min. Disulfides were reduced by adding 4 µL of 200 mM DTT with 20 min incubation. Alkylation solution (4 µL) was added and incubated in dark at room temperature for 15 min. Trypsin was added at a trypsin/protein ratio of 1:5 and incubated at 37 °C for 2 h to 4 h. The digestion was stopped by lyophilizing and was applied to C-18 ZipTip^{TM} (Millipore Corp). Peptides bound to the prepared ZipTip^{TM} were guanidinated with 0.5 M O-methylisourea hydrogen sulphate (MIS) in 0.25 M Na carbonate buffer (pH 11.7) and then sulfonated with 4-sulphophenyl isothiocyanate (SPITC) in 50 mM Tris-HCl buffer (pH 8.0). Peptides were eluted with 80% MeCN in 0.1% TFA with α-cyano-4hydroxy-trans-cinnamic acid (CHCA)

matrix and run on an Applied Biosystems MDS SCIEX 4800 MALDI TOF/TOF analyzer.

Target protein visualized on SDS-PAGE with Colloidal Coommassie G250 Blue Stain. In-gel tryptic digestion was performed on the gel slice based on published protocol (Shevchenko *et al.* 1996). Gel slice from the target band was cut into pieces and were de-stained and washed in 80% acetonitrile (ACN) and then 100 mM ammonium bicarbonate (ABC) buffer (pH 8.0) respectively. Destained gel pieces were dehydrated in 70% ACN, performed reduction of disulfides in 50 mM ABC buffer with 2 mM TCEP for one hour at 45 °C, and further performed alkylated in 50 mM ABC buffer with 20 mM iodoacetamide (IAM). Then the gel slices were washed repeatedly in 70% ACN and 50 mM ABC buffer, and dried with a vacuum concentrator. The dry gel pieces were digested with trypsin in 50 mM ABC buffer overnight at 37 ºC. Digested peptides were eluted by incubating in 70% ACN in 0.1% trifluoroacetic acid (TFA) for 30 min. Supernatant were mixed with CHCA matrix solution and run on an Applied Biosystems MDS SCIEX 4800 MALDI TOF/TOF analyzer. The MASCOT (Matrix Science) search engine was used to match the MS data with in silico digested amino acid sequences from SwissProt.

2.2.5. Molecular biology

2.2.5.1. PCR

PCR primers in this project were designed in CLC workbench 6.0, analyzed in IDT Oligo analyzer, and synthesized by IDT custom DNA oligos. Primers were re-suspended in mQ to10 µM after centrifuging for pelleting the powder before opening the tube. All primers were stored at -20 °C until required.

MyFi™ Mix (Bioline) was used in all PCR reactions on a MyCycler™ Thermal Cycler (Bio-Rad). PCR reaction (50 µL) was set up as follows, 10 ng plasmid, 1 µL forwards primer (20 µM), 1 µL reverse primer (20 µM), 25 µL 2X MyFi Mix and H_2O. The PCR temperature was set as: initial denaturation for 3 min at 95 °C; 30 cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 55-65 °C, extension for 1 min per kb at 72 °C; a final extension for 7 min at 72 °C.

2.2.5.2. Quantification

DNA and RNA were analyzed by NanoDrop ND-1000 spectrophotometer (Thermo Scientific), with the measurement of A_{260}/A_{280} and A_{260}/A_{230} indicating the quality. The Qubit 2.0 Fluorometer (Life Technologies) was applied to measure the concentration of RNA accurately.

2.2.5.3. Agarose gel electrophoresis

DNA products were run on a 1.0% agarose gel. DNA (10 µL) was mixed with 2 µL 5X loading dye before loaded on gels. A quick load purple 2-Log DNA ladder (New England BioLabs) or a 1 kb plus ladder (Invitogen) were used to estimate the molecular weight. Gel was run at 90 V for 30 min and then visualized under UV light with a Bio-Rad DOC system.

2.2.5.4. Purification

DNA products from PCR reactions or restrict enzyme digestions were purified by Isolate PCR and gel purification kit I (Bioline). PCR products or vectors after double digestion was run on 1.0% agarose gel and were cut with a sterile razor blade under UV lamp. The gel slice was processed to the same kit for gel purification.

2.2.5.5. Cloning

Target gene and vector were both double digested with restriction enzymes to expose stick ends at both 5' end and 3' end, which were later ligated together by T4 Ligase Kit (Roche) after gel purification.

The vector pET-28a(+) is an expression vector with T7 promoters and *lacZ* operon and the vector pGEX-6P-3 contains a *tac* promoter for high expression induced by IPTG. The inserted gene was fused with 6X His-tag with a thrombin cleavage site at N-terminal and 6X His-tag at the C-terminal in pET-28a(+), or GST-tag with a HRV 3C Protease cleavage site at the N-terminal in pGEX-6P-3. IPTG induction enabled the transcription of the insert gene in the MCS site.

2.2.5.5.1. Double digestion

Restriction enzymes, *NdeI* or *BamHI* and *XhoI*, were obtained from New England Biolabs. Double restriction enzyme digestions were performed according to the manufacturer's protocol in which 0.5-1 µg DNA, 1 µL *NdeI* (20,000 units/mL) or 1 µL *BamHI* (20,000 units/mL), 1 µL *XhoI* (20,000 units/mL), 3 µL 10X NEBuffer, were added in a 30 µL reaction and incubated at 37 °C for 1 h.

2.2.5.5.2. Ligation

Ligation was done in a 30 µL reaction with T4 DNA ligase (Roche). The the amounts of vector and insert DNA required was calculated with NEBiocalculator v1.5.0 (https://nebiocalculator.neb.com/#!/ligation) at a ratio of 4:1. Reaction was set as, 50 ng linearized vector, double digested DNA, 3 µL 10X ligation buffer, 1U T4 DNA ligase, add H_2O to 30 µL. Ligation was incubated at 4 °C for 16 h.

2.2.5.5.3. Protein expression

Ligation from section 2.2.5.5.2 were heat inactivated at 65 °C for 10 min. Then 2 µL ligation was mixed with 50 µL pre-prepared electro-competent cell, which were then processed to the electroporation by a MicroPulser™ electroporator (Bio Rad) with setting, bacterial EC2 (V=2.5 kV). SOC medium (500 µL) was added immediately and incubated at 37 °C, 200 rpm for 1 h to allow cell recovered from the electroporation. Transformation cell was then plated on LB agar plate at a volume of 150 µL (with appropriate antibiotics resistance) and incubated at 37 °C, 200 rpm, overnight. Single colony was picked up and cultured in 5 mL LB medium (with appropriate antibiotics resistance) overnight. Plasmid was extracted from 2 mL cells. Purified plasmid was digested with restriction enzymes and visualized on 1.0% agarose gel. Meanwhile, plasmid was sent to Sanger sequencing with primers to confirm the insertion.

After confirming the recombinant plasmid, cultured cells were resuspended in desired volume of LB medium and incubated at 37 °C until the OD_{600} reached 0.4 to 0.8. Protein expression was induced by adding IPTG to a final concentration of 0.4 mM and cultured at either 37 °C for 4 h or 18 °C overnight. Cells were harvested by centrifuging at 1,6000 × g, 4 °C for 20 min. Cell pellets were resuspended in immobilized metal affinity resin binding buffer at a ratio of 5 mL per gram wet

weight and lysozyme was added to a final concentration of 1 mg/mL. Cell suspension was incubated on ice for 1 h and then processed to 3 min sonication (30% amplitude, pulses 0.5 s on/0.5 s off for 1 min) (Sonics ultrasonic processor VCX500 with Tapered microtip 630-0418). The cell lysate was centrifuging at 1,6000 × g, 4 ℃ for 20 min. Supernatant was harvested as soluble protein fraction, which was processed to tagged protein purification.

2.2.6. Data analysis

One-way ANOVA analysis was performed with SPSS 22.0 (IBM) to determine differences in the ice activities of *Chionochloa* species. Mean values±standard deviations (SD) were reported by SPSS 22.0 (IBM). Meanwhile, biological significance was also considered in reporting difference.

Boxplot was drawn by SPSS 22.0 (IBM). Mean value was indicated by the horizontal black line in the box. The top and the bottom of the box indicated the 75^{th} and the 25^{th} percentile respectively. The vertical lines showed the distribution of the dataset. Ninety-five percent (95% confidence limits) of the data was included between the T-bar (whiskers).

Chapter 3 Seasonal ice activities in seventeen New Zealand *Chionochloa* species (Poaceae)

3.1. Introduction

Wharton *et al.* (2005) suggested the term, ice active proteins (IAPs), to define proteins that affect the formation and stability of ice crystals in biological systems. Three different categories of ice active proteins (IAPs) essential to different cold tolerance mechanisms were proposed: 1) ice nucleation proteins (INP) that trigger ice formation; 2) thermal hysteresis proteins (THP) that inhibit the growth of ice crystal seeds; and 3) recrystallization inhibition proteins (RIP) which inhibit annealing and growth of ice crystals once formed (Bar-Dolev *et al.* 2016; Hartmann *et al.* 2013; Wharton *et al.* 2005; Wharton *et al.* 2010).

Antifreeze proteins (AFP) typically have both thermal hysteresis and recrystallization inhibition (RI) activities, and have been characterized in fishes, insects and many plants (Bar-Dolev *et al.* 2016; Chen *et al.* 1997; Duman and DeVries 1976; Graham *et al.* 2008). Fish AFPs are classified into six groups including one comprising AFGP (glycopeptides). They have various structures that are thought to interact with different ice surfaces and demonstrate thermal hysteresis of 1.0 ℃ to 1.5 ℃ (DeVries and Cheng 2005; Storey and Storey 2005). Insect AFPs are typically rich in cysteine and form solenoids containing disulfide bonds and have even higher thermal hysteresis (> 5 ℃) than fish AFPs (Duman 2001). Plants, such as the widely-studied winter rye and wheat, produce AFPs with comparatively low TH activity (0.1 ℃ to 0.5 ℃) but high RI activity (Griffith *et al.* 1992; Yu and Griffith 1999).

Recrystallization of small ice crystals leading to larger crystals is likely to be harmful to freezing tolerant species (Bar *et al.* 2016). In plants, a major adaptation for freezing tolerance is the synthesis of RIPs to inhibit ice growth (Griffith and Yaish 2004; John *et al.* 2009).

Besides expressing RIPs, plants manipulate freezing by triggering extracellular ice formation in a controlled manner. This is a function of the third group of ice active proteins involved in interactions with ice crystals, the ice nucleation proteins (INP). These trigger ice crystal formation at high sub-zero temperatures (Wolber *et al.* 1986; Zachariassen and Kristiansen 2000). During this process, heat is released, as ice crystals form and the temperature remains constant at around 0 ℃ while ice formation

continues. Pure water can stay liquid at temperatures as low as -37 °C, but in the presence of ice nucleators which serve as core nucleation sites, water freezes at higher temperatures. A function of highly efficient ice nucleators is to ensure that ice forms relatively slowly at high sub-zero temperatures thus avoiding rapid freezing at lower temperatures with consequent cell damage (Franks 1985; Murray *et al.* 2010; Stan *et al.* 2009).

INPs in plants have two main sources: extrinsic proteins derived from microorganisms that colonize plants, and proteins produced by the plants themselves (Brush *et al.* 1994; Griffith and Antikainen 1996).

Extrinsic INP are commonly produced by ice nucleation active (INA) bacteria that reside on the surface of plant leaves. It is thought that ice formation caused by the INPs leads to freezing damage to plant epithelial cells and allows the bacteria to both invade plants and to utilize the cellular debris (Lindow 1982; Lindow 1983). There is considerable variation in the number of INA bacteria on different plant leaves and with the season (Mazarei and Kerr 1987; Nejad 2005). Crucial to this role is the susceptibility of the plant to freezing and consequent tissue damage: inducing freezing in freezing tolerant plants is unlikely to be of direct benefit to colonizing bacteria.

Common features of INA bacteria are that they are unstable outside of the range of pH 5-9, and at temperatures >30 °C (Kawahara *et al.* 1994; Kawahara *et al.* 1995).

Intrinsic INPs are more variable in character than those responsible for bacterial INA. Woody plants show seasonal changes of intrinsic INA that was highest in winter and decreased when the environmental temperature increased (Griffith and Antikainen 1996). However, intrinsic INA in winter rye and in *Rhododendron* and *Prunus* species did not follow a cold-induced expression pattern but was instead expressed constitutively (Kieft and Ruscetti 1990; Kishimoto *et al.* 2014). This may reflect differences in the likely seasonal exposure to freezing temperatures.

Intrinsic INPs seem to have different functions compared with those of bacteria INP and these presumably relate to maintaining cellular integrity rather than damaging it. Zachariassen and Kristiansen (2000) suggested that triggering ice nucleation at high sub-zero temperatures induced extracellular- rather than intracellular-freezing and thus controlled the extent of freezing and limited damage. Extracellular INPs have been found in many plants including winter rye, *Prunus* species, and lichens (Table 3.1), but these proteins have not been well characterized.

Other ice nucleation agents are based on polysaccharides, lipids or carbohydrates (Lorv *et al.* 2014; Pummer *et al.* 2014).

Both epiphytic INA bacteria and endogenous INA agents contribute to plant INA (Brush *et al.* 1994). Ice nucleation initiated by epiphytic INA bacteria on leaf surfaces leads to frost damage, however endogenous INA demonstrates different propertiesare (Brush *et al.* 1994; Griffith and Antikainen 1996). Endogenous INA has also been suggested to help water collection from fog in alpine regions (Wharton *et al.* 2010). Chemical composition varies among plant INA agents with both protein INA and non-protein INA agents such as carbohydrates and phospholipids identified (Brush *et al.* 1994; Griffith and Antikainen 1996). Different compositions have also been reported in INA agents from various plant species assessed by treatments including proteinase, heat and protein modification reagents (Brush *et al.* 1994).

This study examines the seasonality of ice activities in *Chionochloa* species in a common garden environment. *Chionochloa* (Poaceae, Danthonioideae) consists of 38 species and subspecies, 36 of which are endemic to New Zealand. Wharton *et al.* (2010) identified significant endogenous INA in two alpine species, *C. macra* and *C. rigida*, which are known for their fog interception abilities (Ingraham and Mark 2000). However, information is lacking for the rest of the genus. *Chionochloa* dominates New Zealand's alpine grasslands, but the ranges of several species extend, or are restricted to low elevations (Edgar and Connor 1999). Therefore this genus presents an opportunity to examine the extent and seasonal pattern of ice activities in relation to elevational range. We expect that strictly alpine species would show high ice activities generally, with higher year round levels compared with species characteristically found at lower elevations. This study is the first to examine INA more widely within the genus, as well as presenting the first full characterization of ice activities in *Chionochloa* species.

3.2. Materials and methods

3.2.1. Materials

Seventeen species and subspecies of *Chionochloa* were sampled at the Dunedin Botanic Garden (altitude 80 m), Dunedin, New Zealand. Seasonal collections were scheduled between 2014 and 2015, and *Chionochloa* species accession numbers were recorded to ensure same species in each batch of collection (Table 3.1 and Table 3.2).

Collection permits were granted from Dunedin Botanic Garden. *Chionochloa macra* was also sampled at the Rock and Pillar Range (chapter 2, section 2.2.1).

Garden_ID	Species	Altitudinal Range
20081507*A	*C. beddiei* Zotov	0-850 m
20080346*A	*C. cheesemanii* (Hack.)Zotov	900-1225 m
19956695E*A	*C. conspicua* (G.Forst.)Zotov	0-1500 m*
20100044*A	*C. defracta* Connor	-1000 m
20060590*A	*C. flavescens* Zotov	750-1550 m*
20021195*A	*C. flavescens ssp. brevis* Connor	-1500 m*
20040655*A	*C. flavescens ssp. lupeola* Connor	-1550 m*
20120603*C	*C. flavicans* Zotov	0-975 m
20100045*A	*C. macra* Zotov	500-1700 m*
20040665*B	*C. pallens ssp. cadens* Connor	-1700 m*
19946126E*A	*C. rigida* (Raoul)Zotov	-1700 m*
20040664*B	*C. rigida ssp. amara* Connor	500-1450 m*
20110867*A	*C. rubra* Zotov	550-1400 m
20040661*A	*C. rubra ssp. cuprea* Connor	0-1500 m
20040674*A	*C. rubra ssp. rubra* Zotov	550-1400 m
20120602*A	*C. spiralis* Zotov	-1000 m
20040663*A	*C. teretifolia* (Petrie)Zotov	300-1300 m*

Table 3.1 Dunedin Botanic Garden *Chionochloa* species accession numbers. Typical altitudinal range (meters above sea level) is from Edgar & Connor (1999). *Species reaching the alpine zone (Pirie *et al*. 2010).

Collection_season	Collection_date	Collection_time	Average_temperature
Spring	23/10/2014	Morning	10 ℃
Summer	22/01/2015	Morning	16 ℃
Autumn	6/05/2015	Morning	8 ℃
Winter	6/08/2015	Morning	6 ℃

Table 3.2 Garden *Chionochloa* collection schedules. All species were collected between 9 am and 10 am. Average temperature of the day was recorded.

3.2.2. Methods

3.2.2.1. Protein extraction

Total proteins of each sample were extracted following methods described in chapter 2, section 2.2.1.1.

3.2.2.2. Total protein quantification

Qubit protein assays (Invitrogen) were used to measure the concentration of each sample. According to the standard protocol, Qubit working solution was made by diluting the Qubit protein reagent 1:200 in the Qubit protein buffer. A Qubit 2.0 fluorometer (Invitrogen) was first calibrated with Qubit protein standards

(Invitrogen). Each 10 µL sample was mixed with 190 µL Qubit working solution, incubated at room temperature for 15 min and measured on the Qubit fluorometer. The final concentration of each sample was calculated by the formula, 20*QF, where QF stands for the actual reading of the Qubit fluorometer. All raw protein extracts were then standardized to a concentration of 5 mg/mL with plant total protein extraction buffer for the purpose of comparing ice activities with same total protein concentrations in the raw extracts.

3.2.2.3. Ice activity assays

INA, TH and RI activity of each sample were measured following methods described in chapter 2, section 2.2.2.

3.3. Results

3.3.1. Ice activities in the *Chionochloa* species at the Dunedin Botanic Garden

3.3.1.1. Ice activities in the spring collection (October 2014)

Eight species (*C. cheesemanii*, *C. conspicua*, *C. flavescens ssp. brevis*, *C. flavescens ssp. lupeola*, *C. rigida*, *C. macra*, *C. pallens ssp. cadens, C. rigida*, *C. rigida ssp. amara* and *C. teretifolia*) obtained mean supercooling temperature (T_c) above -6.5 °C in INA assays (Figure 3.1A). The remaining species showed mean T_cs below -10.0 °C but were still higher than the buffer control (T_c, -16.7±1.9 °C) (Figure 3.1A).

TH values in the *Chionochloa* species and ice crystal shapes during single crystal growth are shown in Figure 3.2 and Figure 3.3. Generally, mean TH value of the *Chionochloa* species were not significantly higher than the buffer control, 0.03 °C ($p>0.05$). Meanwhile, round ice crystals were observed in all *Chionochloa* species except the *C. macra* and the *C. spiralis*, which means proteins in the raw extracts did not interact with ice crystals already formed. However, the nanoliter osmometry indicated single typical hexagonal shaped ice crystals in the *C. spiralis* and single weak hexagonal shaped ice crystals in the *C. macra*, which indicate the interaction of proteins in the liquid fraction and the boundaries of ice crystals already formed.

Ten of the 12 *Chionochloa* species showing no obvious RI activity when annealed at -8 °C for 30 min in the splat cooling assays. All these ten species had mean ice crystals diameter of more than 40.0 µm (Figure 3.4). In the buffer control,

mean ice crystal diameter grew dramatically, from 74.4 µm at 0 min (t_{0min}) to 122.9 µm at 30 min (t_{30min}). However, there was strong RI activity in the *C. spiralis* with mean ice crystal diameter of 18.6 µm at t_{30min}. The RI activity in the *C. spiralis* was further confirmed by an additional 30 min annealing time with mean ice crystal diameter of 18.4 µm.

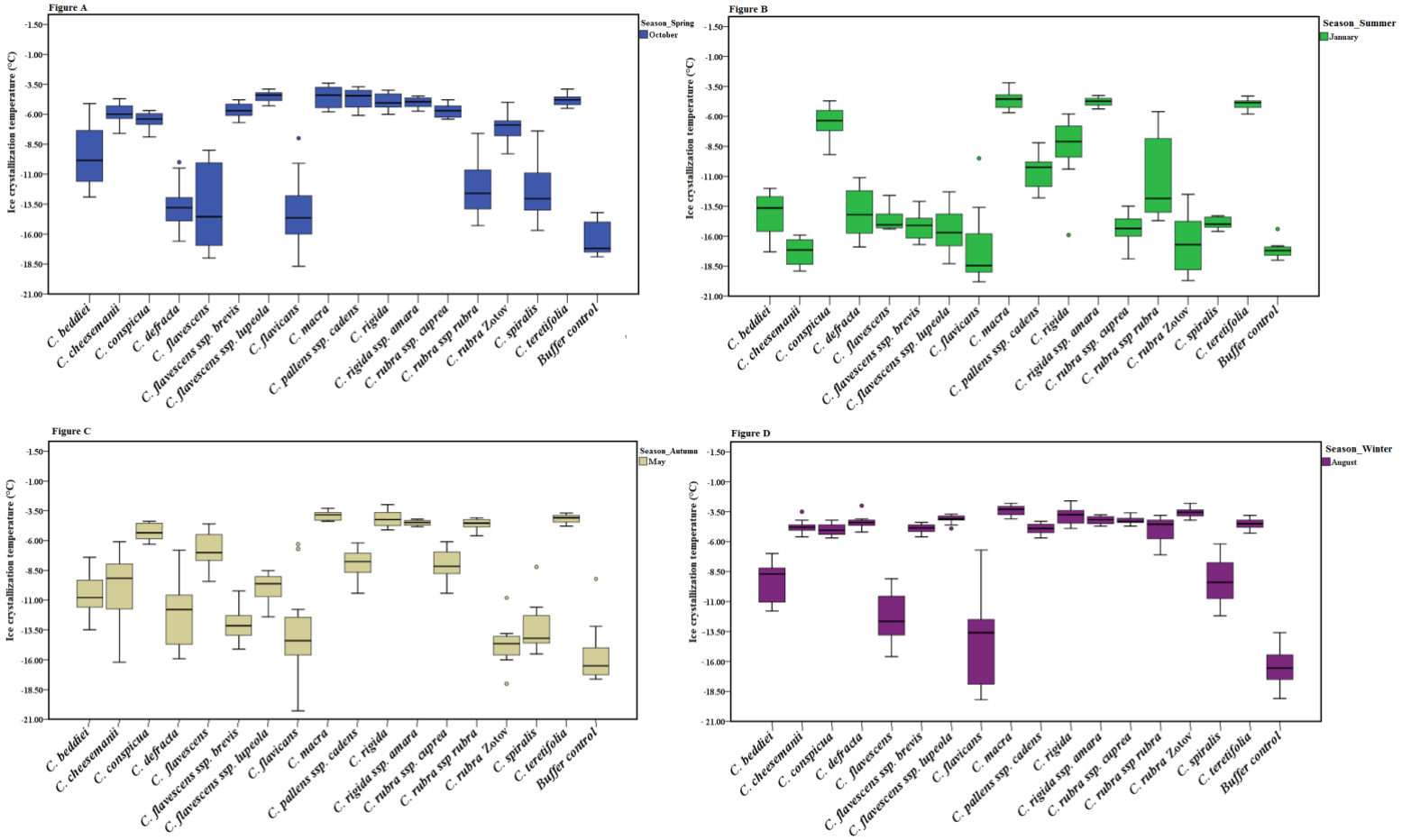


Figure 3.1 INA (ice crystallization temperature, T_c) in the *Chionochloa* species during four seasons. Figure A to D, season spring to winter. Middle dark line in boxplot, mean T_c from three individual mature leaves; Buffer control, n=3. Most species obtained highest INA in winter and lowest in summer.

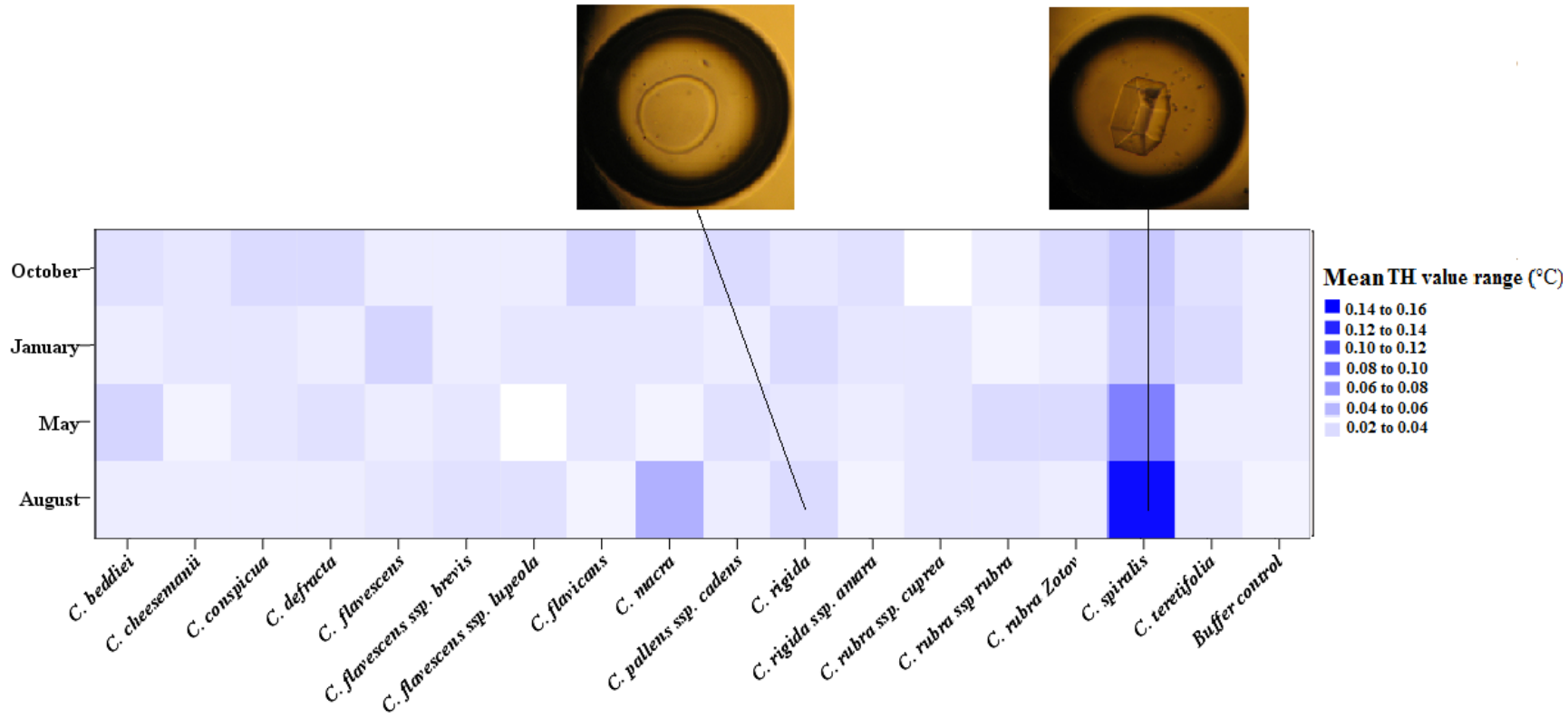


Figure 3.2 TH activity in the *Chionochloa* species during four seasons. Figure legends (right) indicate mean thermal hysteresis ranges. Heat map colors represent mean thermal hysteresis from three individual mature leaves; Buffer control, n=3. Typical hexagonal shaped ice crystal was present in the nanoliter osmometry of the winter *C. spiralis* extract. In the contrast, non-hexagonal shaped ice crystal was present in the winter *C. rigida*. Three observations of the ice crystal were performed. More than three images were taken per observation. Image of the ice crystal shown is representative for the species present.

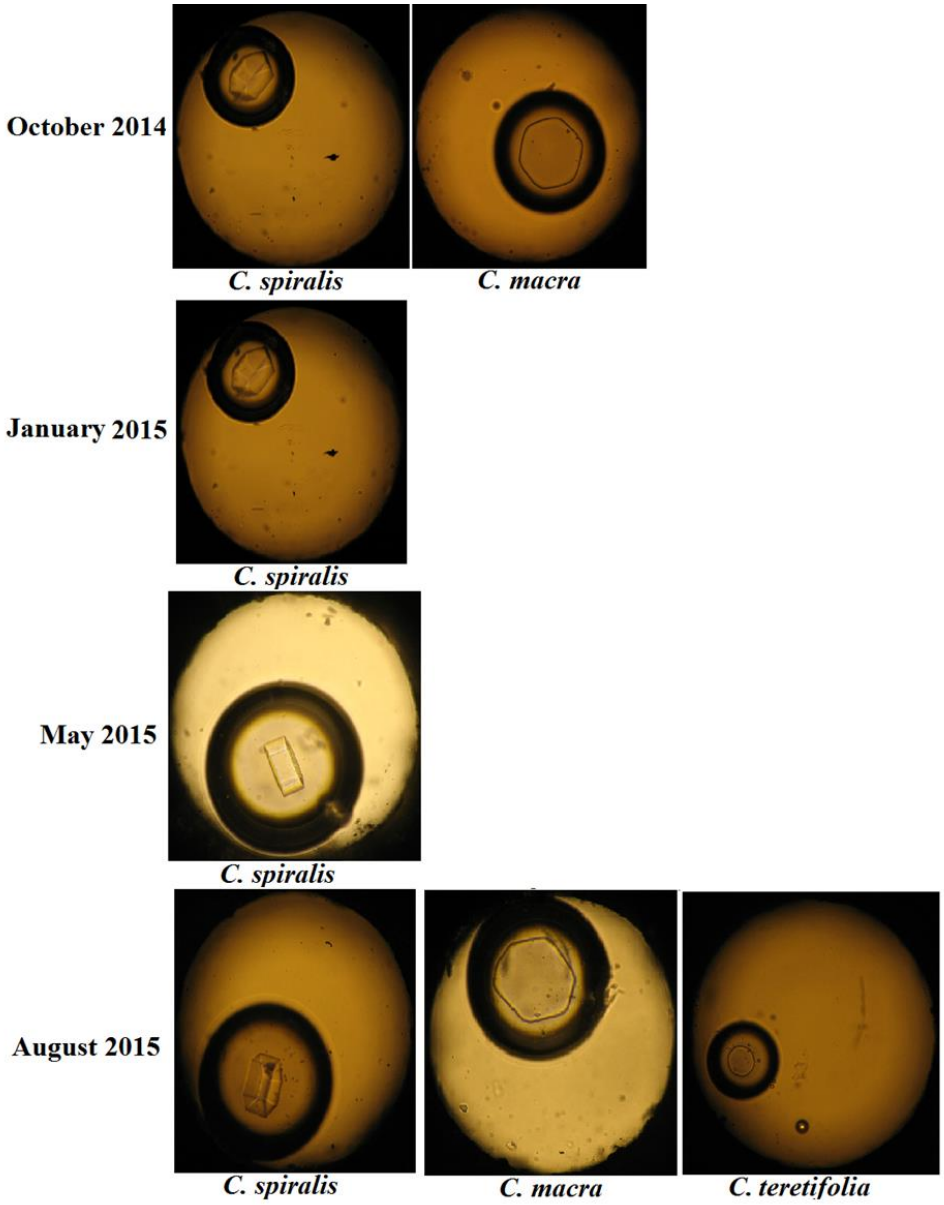


Figure 3.3 Morphology of ice crystals in the nanoliter osmometry analysis of the *Chionochloa* extracts. Only non-disc ice crystals were shown in the *Chionochloa* species sampled in each season. Three observations were performed per sample per season. More than three images were taken per observation. Image shown is representative for each species present. Typical hexagonal shaped ice crystal was observed in the *C. spiralis* throughout the year whereas weak hexagonal shaped ice crystal was only found in the winter and spring collection of the *C. macra* and in the winter collection of the *C. teretifolia*.

3.3.1.2. Ice activities in the summer collection (January 2015)

In the summer collection, there was high level of INA in the *C. macra* with a mean T_c of -4.7 °C, in the *C. rigida ssp. amara* (mean T_c of -4.8 °C) and in the *C. teretifolia* (a mean T_c of -5.0 °C) (Figure 3.1B). Comparatively high INA was also found in the *C. conspicua* (-6.6 °C) and the *C. rigida* (-8.6 °C). The other species had much lower mean T_cs, which were lower than -14.0 °C. Three species, *C. cheesemanii*, *C. flavicans* and *C. rubra* Zotov, had mean T_cs similar to the buffer control (-17.1 °C) (Figure 3.1B).

No obvious TH was observed in any of the species sampled whereas typical hexagonal shaped ice crystals were observed for the *C. spiralis* extracts (Figure 3.1B). High RI activity was only found in the *C. spiralis* with a mean ice crystal diameter of 28.8 µm in 30 min of the splat cooling assay. The other species obtained mean ice crystal diameters no less than 40.0 µm, which were still less than that of buffer control.

3.3.1.3. Ice activities in the autumn collection (May 2015)

In the autumn collection, *C. conspicua*, *C. macra*, *C. rigida*, *C. rigida ssp. amara*, *C. rubra ssp. rubra* and *C. teretifolia* showed mean T_cs above -6.0 °C (Figure 3.1C). *Chionochloa cheesemanii*, *C. flavescens*, *C. flavescens ssp. lupeola*, *C. pallens ssp. cadens* and *C. rubra ssp. cuprea* had mean T_cs between -6.0 °C and -10.0 °C. A typical hexagonal shaped ice crystal was observed only in the *C. spiralis* whereas there was a non-hexagonal shaped ice crystal in the other species. TH activity was comparatively high in the *C. spiralis* (0.09 °C) and there was no significant difference between the other 11 species and the buffer control. High RI activity was found only in the *C. spiralis* with a mean ice crystal diameter of 9.7 µm after 30 min of splat cooling assay.

3.3.1.4. Ice activities in the winter collection (August 2015)

In the August (winter) collection, nine species showed mean T_cs above -6.0 ℃ (*C. cheesemanii, C. conspicua, C. defracta, C. flavescens ssp. brevis, C. flavescens ssp. lupeola, C. macra, C. pallens ssp. cadens, C. rigida, C. rigida ssp. amara, C. rubra* Zotov, *C. rubra ssp. cuprea, C. rubra ssp. rubra* and *C. teretifolia*), two species (*C. beddiei and C. spiralis*) had mean T_cs around -9.0 ℃, and the remaining two species (*C. flavescens* and *C. flavicans*) had mean T_cs lower than -12.0 ℃ (Figure 3.1D). A typical hexagonal shaped ice crystal was observed in the *C. spiralis* and a weak hexagonal shaped ice crystal was found in the *C. macra* and the *C. teretifolia* (Figure 3.4). Comparatively high TH value was found only in the *C. spiralis* with 0.12 ℃. RI activity was also found only in the *C. spiralis* with a mean ice crystal diameter of 7.6 µm after 30 min of splat cooling assay (Figure 3.4).

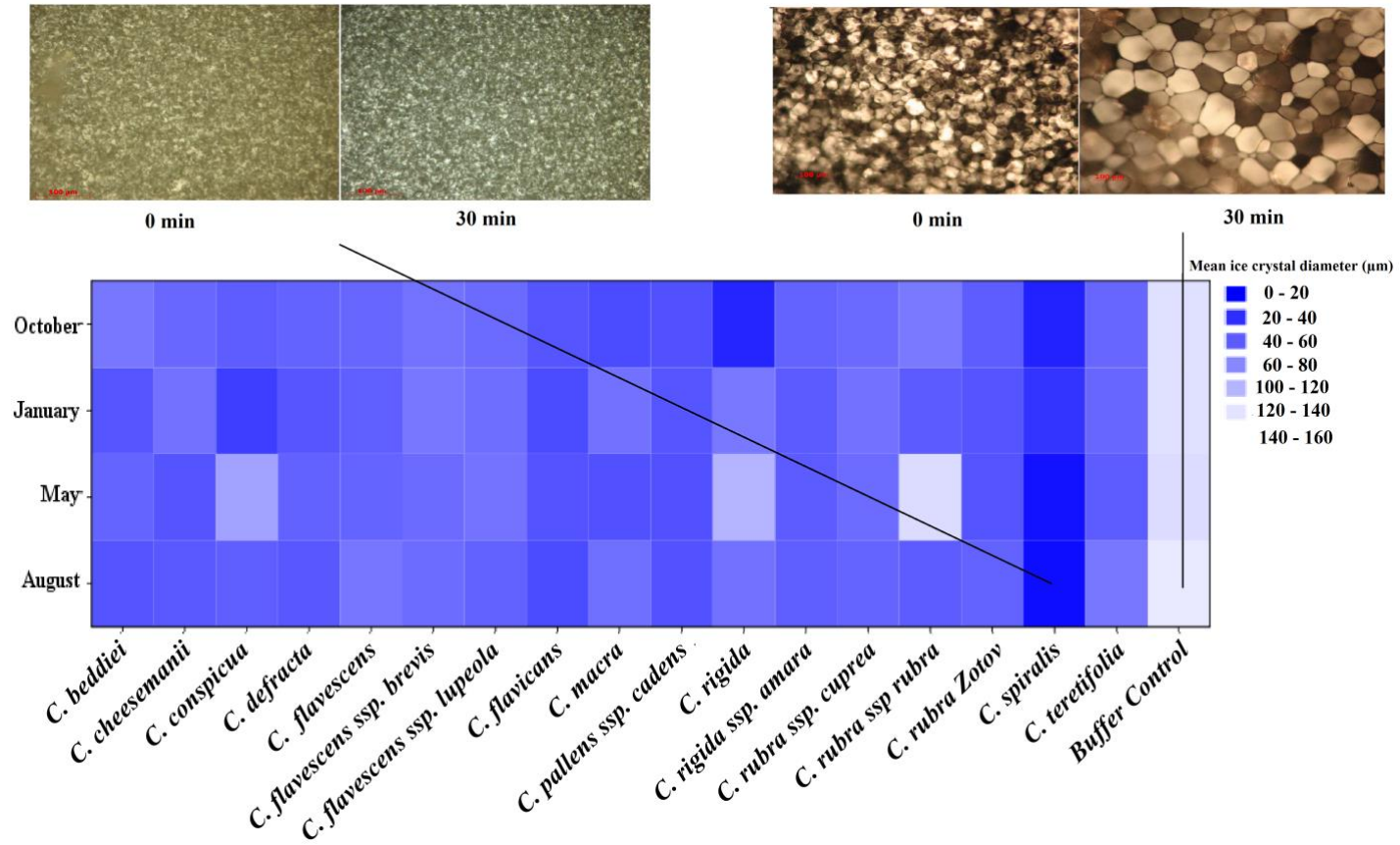


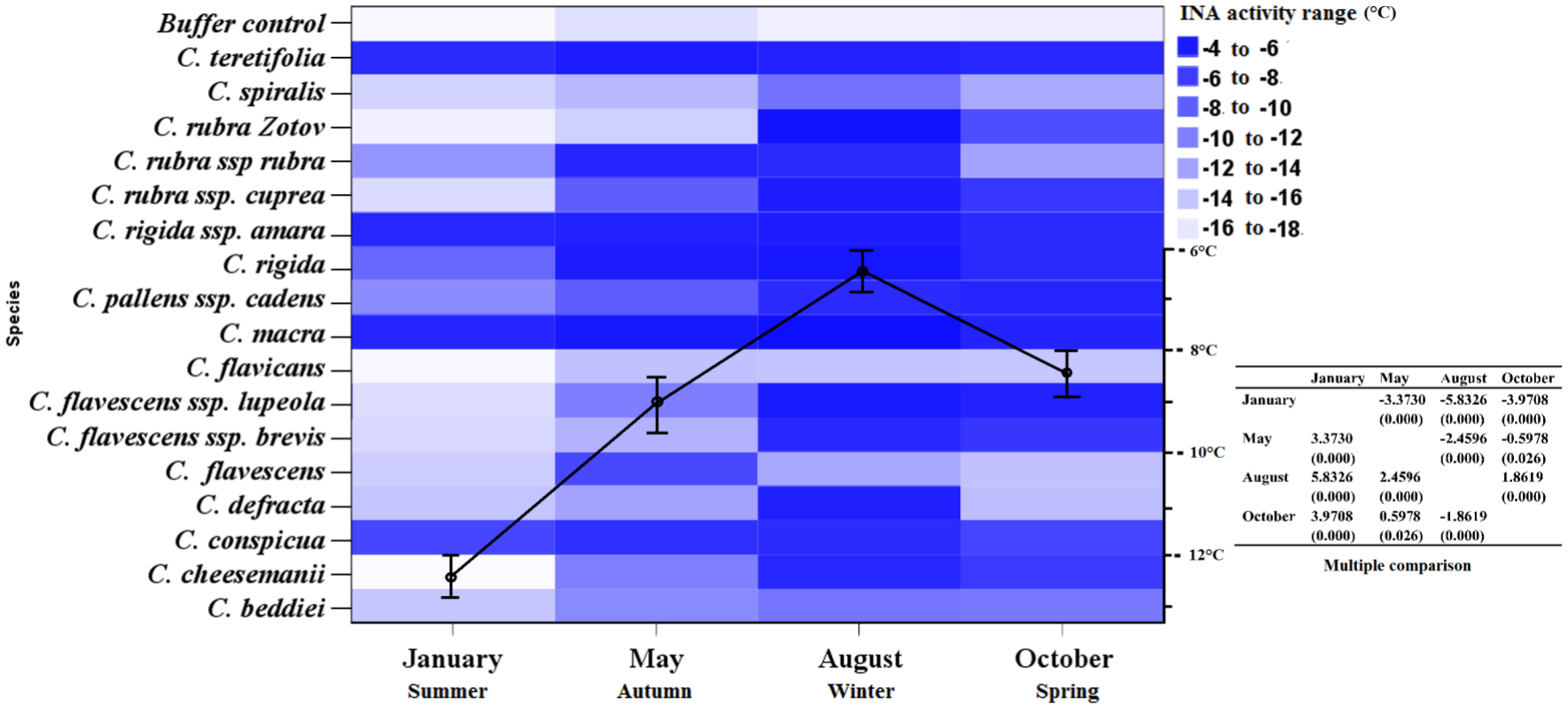
Figure 3.4 RI activity of the *Chionochloa* species during four seasons. Figure legends (right) indicate mean ice crystal diameter ranges. Heatmap colors represent mean ice crystal diameter for 10 randomly selected largest ice crystals from three individual mature leaves. *Chionochloa spiralis* obtained lowest mean ice crystal diameters in the splat cooling assays. Most species had mean ice crystal diameter smaller than that of buffer control. Images of ice crystals shown are representative.

3.3.1.5. Seasonal patterns of ice activities in each *Chionochloa* species

Chionochloa species investigated in this study showed significant overall seasonal differences in INA ($p<0.05$) (Figure 3.5; Figure S1), except for *C. flavicans*. Generally, *Chionochloa* species showed the highest INA in winter and gradually decreased activity in summer. The exception to this was *C. flavescens*, which showed highest mean INA (-6.7 °C) in autumn. For most species, the most active INA occurred within a narrow mean crystallization temperature range (~-4 °C to ~-5 °C). *Chionochloa beddiei*, *C. flavicans* and *C. spiralis* were exceptions and showed comparatively low INA throughout the year: the highest mean INA in these three species was -9.3 °C, -13.8 °C and -9.2 °C respectively. INA reduced slightly in most *Chionochloa* species in spring compared to that in winter, and decreased dramatically in *C. defracta* and *C. rubra ssp. rubra*. In autumn, INA started to increase in most species compared with those in summer (Figure 3.5).

Extracts of *C. flavicans* and the *C. spiralis* did not show high levels of mean INA (>-8 °C) in any of the four collections. In contrast, *C. macra* extracts showed the highest mean INA in all four collections, and the *C. spiralis* extract showed the highest mean TH and RI activities in all four seasons (Figure S2-S3). Most *Chionochloa* species showed no seasonal pattern in TH and RI activities, except for *C. macra* and *C. spiralis* (Figure S2-S3). TH values of *C. macra* in winter were different to those of the other three seasons whereas there was no difference among these three seasons. Mean TH values in the *C. spiralis* were low in October and January, increased from May and reached their highest values in August (Figure 3.2). It should be noted that the TH values measured here are very small and are of doubtful biological significance.

Only *C. spiralis* obtained recrystallization inhibition activity with biological significance (mean ice crystal diameter less than 30 μm in splat cooling assay). RI



	January	May	August	October
January		-3.3730 (0.000)	-5.8326 (0.000)	-3.9708 (0.000)
May	3.3730 (0.000)		-2.4596 (0.000)	-0.5978 (0.026)
August	5.8326 (0.000)	2.4596 (0.000)		1.8619 (0.000)
October	3.9708 (0.000)	0.5978 (0.026)	-1.8619 (0.000)	

Figure 3.5 Seasonal changes of INA in the *Chionochloa* species. Heat map color, level of INA (mean ice crystallization temperature from three individual mature leaves). Line chart in the figure indicates the overall seasonal pattern of INA in the *Chionochloa* species investigated. Open circles, estimated marginal means of INA in 17 species and subspecies per season. Error bars, 95% confidence intervals. Mean difference of seasonal INA was present in the multiple comparison table (Right). Significance (p-value) was in the bracket.

activity declined in *C. spiralis* after spring and was least active in summer, then increased reaching a maximum in winter (Figure 3.4, noting that small crystal sizes indicate higher RI activity).

3.3.2. Dilution resistance of ice activities in the *Chionochloa* species throughout a year

3.3.2.1. RI activities in the *C. spiralis*

RI activity in the *C. spiralis* showed increasing resistance to dilution with samples collected in winter obtaining the highest resistance in the *C. spiralis* and samples collected in summer the lowest resistance. RI activity of the *C. spiralis* collected in January 2015 was lost after dilution greater than 10-fold, whereas it resisted to at least 120-fold dilution for the *C. spiralis* collected in August 2015 (Figure 3.6).

TH activity in the *C. spiralis* collected in August 2015 reduced much quicker with 10-fold dilution (TH to control level) whereas hexagonal ice crystals disappeared only after 360-fold dilution.

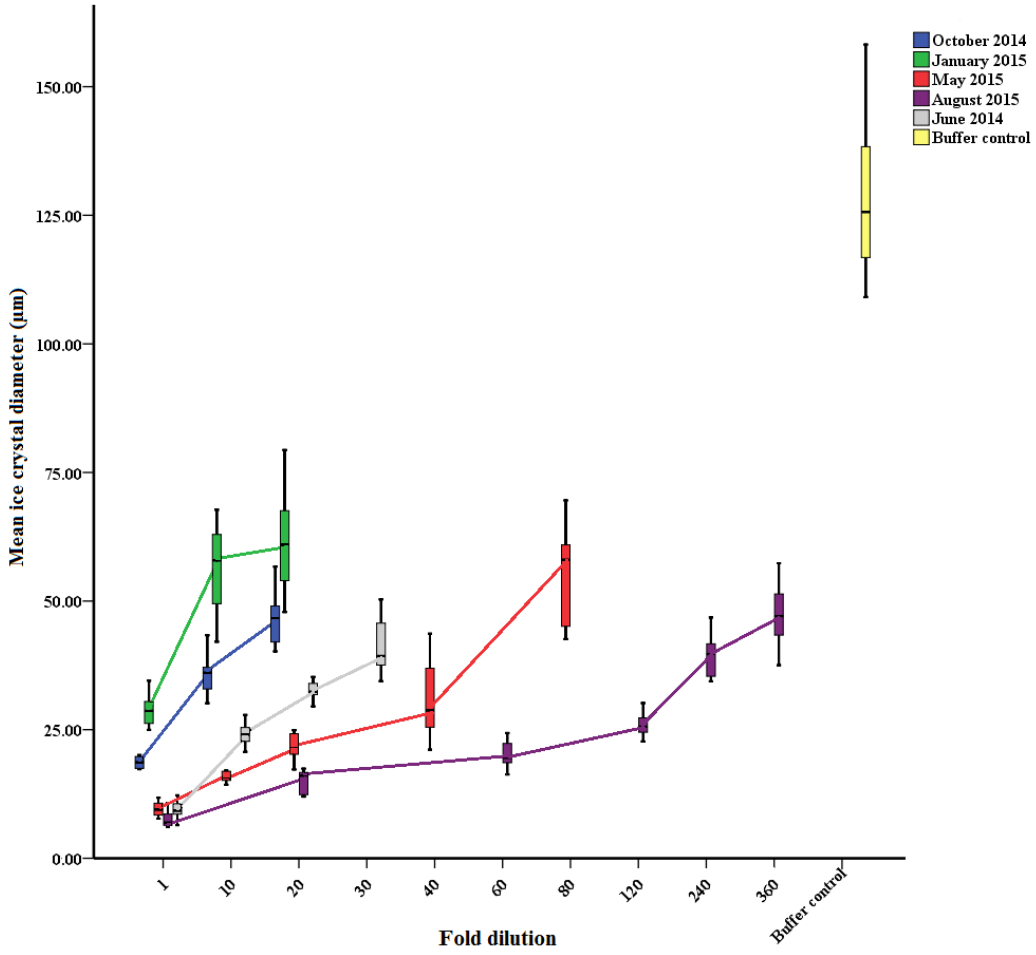
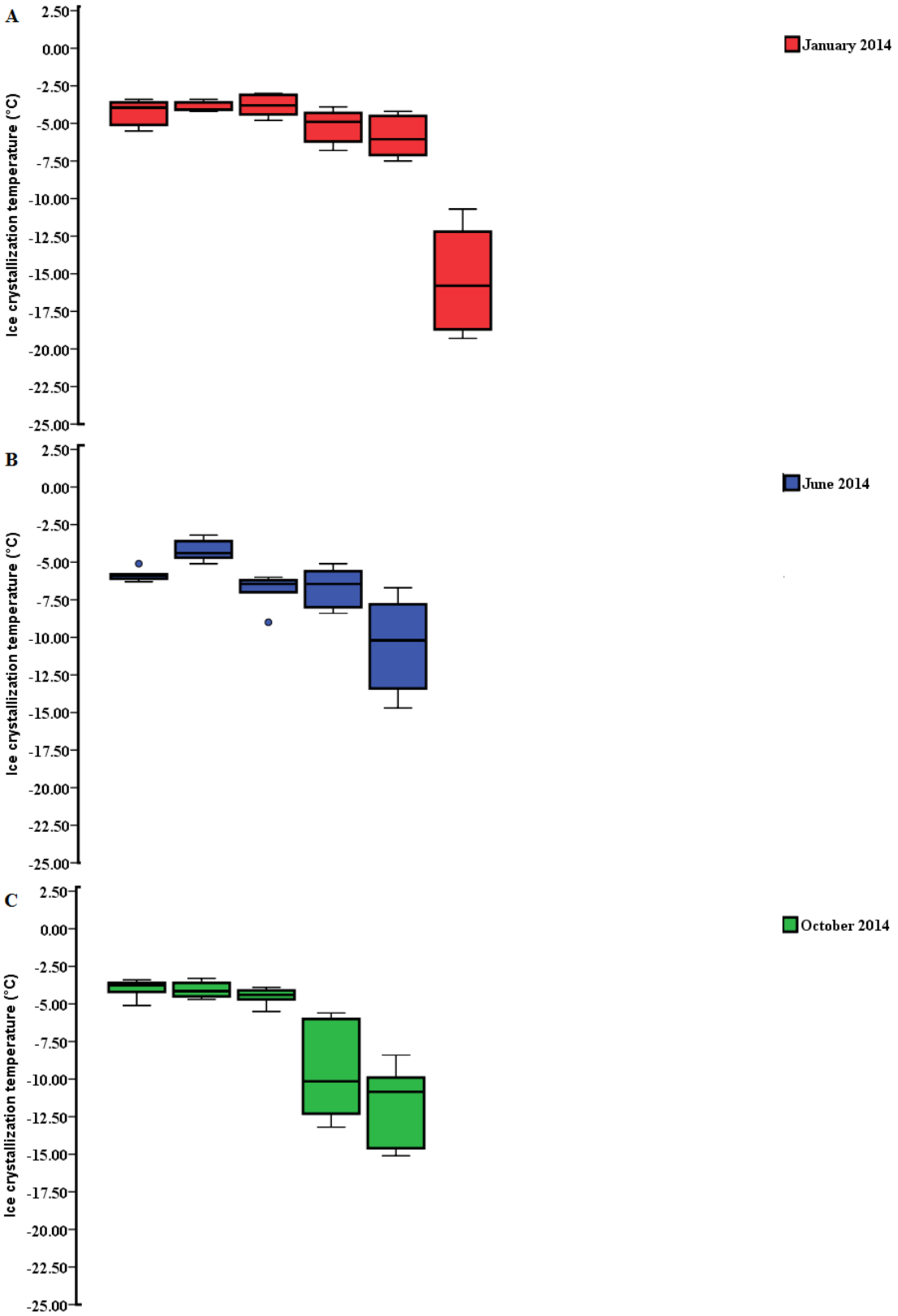


Figure 3.6 RI activities in the *C. spiralis* on dilution. The Y axis showed ice crystal diameter (µm) after annealing at -8 °C for 30 min. Middle dark line in boxplot, mean ice crystal diameter of 10 randomly selected largest ice crystals from three individual leaves.

3.3.2.2. INA in the *C. macra*

INA in the *C. macra* collected in January 2015 was lost when diluted more than 10-fold whereas the activity was not lost until diluted at least 6,000-fold in the species collected in August 2015. The *C. macra* collected at the Rock and Pillar Range in January was found to have higher dilution resistance than the *C. macra* collected at Dunedin Botanic Garden in January, with 2,000-fold dilution compared to the 10-fold dilution (Figure 3.7). This may reflect the different characters of these activities, or may be a consequence of different environments as the Rock and Pillars site experiences lower temperatures, particularly at night, even in summer, than the Botanic Gardens.


A
B
C
Ice crystallization temperature (°C)
2.50
0.00
-2.50
-5.00
-7.50
-10.00
-12.50
-15.00
-17.50
-20.00
-22.50
-25.00
January 2014
June 2014
October 2014

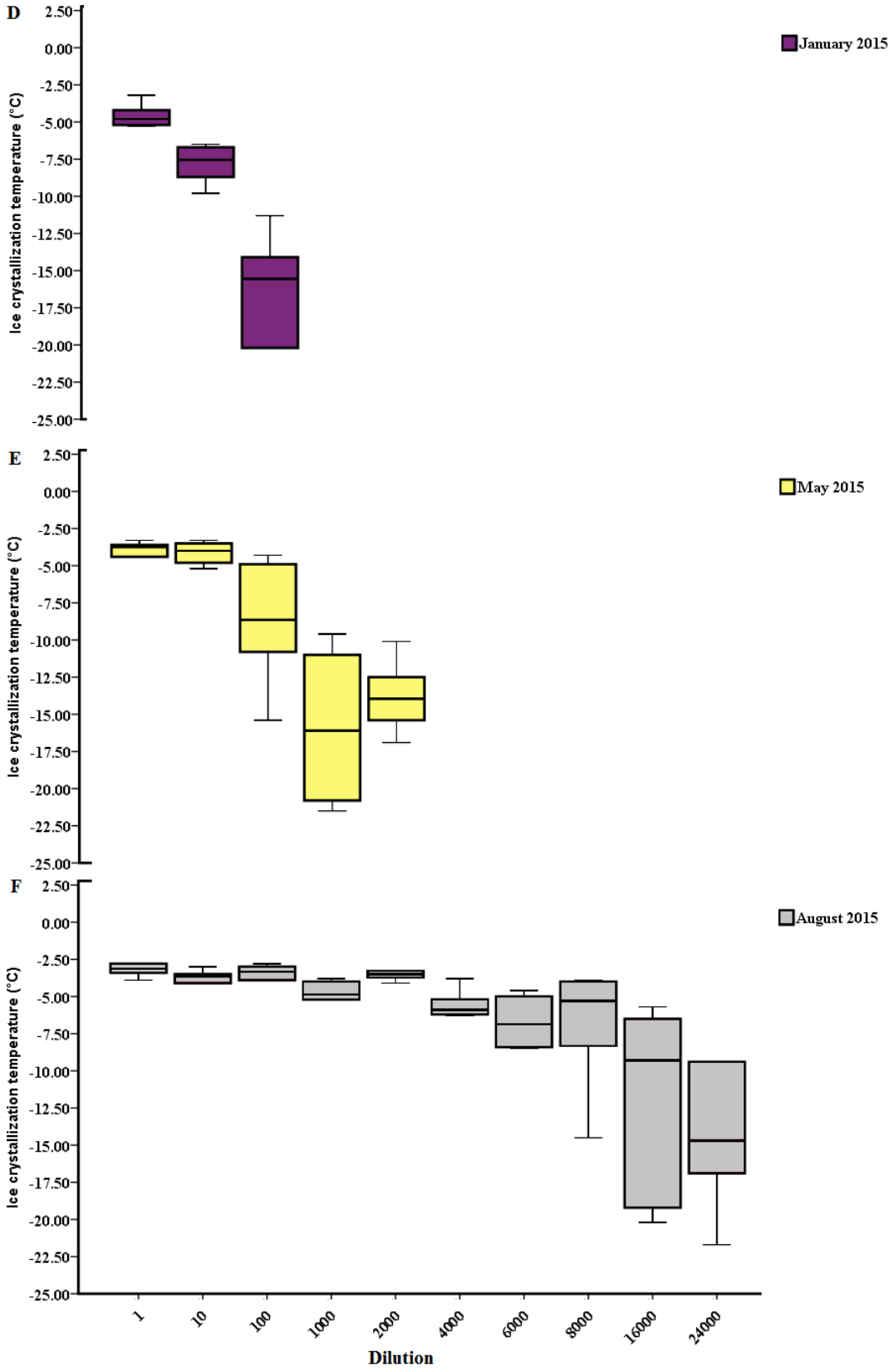


Figure 3.7 INA in the *C. macra* on dilutions. A, *Chionochloa macra* collected at Rock and Pillar; B to F, *Chionochloa macra* collected at Dunedin Botanic Garden. Middle dark line in each boxplot, mean T_c from three individual mature *C. macra* leaves.

3.4. Discussion

INA is reported in animals, plants and bacteria but only bacterial INA agents have been well characterized. INA in animals is not well understood but may come from bacteria in the gut and on the skin. In a few cases, INPs in the plasma or tissue have been identified which assist with freezing tolerance in such animals as the wood frog (Wolanczyk *et al.* 1990). Bacterial INA is caused by INPs, the structures of which are well characterized with repeat ice-binding domains playing key functions in the ice nucleation process (Lindow 1983; Wolber *et al.* 1986).

There have been few studies on plant intrinsic INA since its discovery in 1979 (Krog *et al.* 1979). Intrinsic INA have been reported in a few lichens, mosses and ferns, several grasses, and woody as well as nonwoody eudicots from a range of families (Table 3.3) (Gross *et al.* 1988; Kieft 1988; Embuscado *et al.* 1996; Ashworth and Kieft 1995; Wisniewski 1997; Wisniewski *et al.* 2004; Goto *et al.* 1993; Hacker and Neuner 2008; Ishikawa *et al.* 2015; Kishimoto *et al.* 2014). What is surprising is that intrinsic ice nucleators in *Rhododendron* bud resist high temperature to 121 ℃ for 15 min whereas intrinsic ice nucleators in the stem bark cannot tolerant such high temperature. Intrinsic ice nucleators in lichen and fungi which induced ice nucleation at ~-5 ℃ were found to be sensitive to proteinase and heat over 50 ℃ (Ashworth and Kieft 1992; Kieft and Ruscetti 1990).

Plant group	Ice activity (°C)	Characters of ice activity	Reference
Lichens: *Rhizoplaca chrysokuca*	INA	INA was sensitive to proteases, guanidine hydrochloride, and urea. Heat stable (40 °C), active from pH 1.5 to 12, and active without lipids. Proteinaceous ice nucleators, heat stable, released from the cell wall by sonication	(Kieft and Ruscetti 1990)
Non-angiosperm plants: Bryophyta, Lycophyta, Pteridophyta, Gingkophyta	TH 0.15 to 0.68	Proteinase sensitive.	(Duman and Olsen 1993)
Angiosperms: Monocots			
Poaceae: *Chionochloa rigida, C. macra*	INA -5	Nonbacterial INA.	(Wharton *et al.* 2010)
Deschampsia antarctica	RI	pH tolerant, proteinase sensitive, heat sensitive.	(Bravo and Griffith 2005)
Deschampsia antarctica	RI	Leucine-rich repeats (LRRs) and IRIP repeats. Heat stable.	(John *et al.* 2009)
Lolium perenne	TH 0.1	12 kDa. Six potential N-glycosylation sites containing the conserved N-X-S/T glycosylation motif.	(Sdebottom *et al.* 1997)
Lolium perenne	TH 0.1 RI	29 kDa. Boiling heat tolerant. Partial sequence identified. Sequence, DEQPNTISGXNNTVRXG.	(Pudney *et al.* 2003)
Lolium perenne	TH 0.1 to 0.5 RI	26 kDa. At least five isoforms. X-X-N-X-V-X-G repeat. Proteins were limited to the family *Pooideae*	(Middleton *et al.* 2009)
Secale cereale L.	TH 0.3	19, 26, 32, 34, and 36 kDa. Pathogenesis-related proteins (PRs), include endochitinases, endo-β-1,3-glucanases, and thaumatin-like proteins. Reduction in activity with DTT.	(Antikainen and Griffith 1997; Hon *et al.* 1995; Yeh *et al.* 2000)

Secale cereale L.	INA -7	Heat (90 °C) sensitive, urea sensitive, proteinase, DTT, boric acid and periodic acid treatments sensitive.	(Brush *et al.* 1994)
Triticum aestivum	RI	29 kDa and 43 kDa. Putative LRR motifs. N-terminal similar to the receptor kinase domain (RKD) of RLKs.	(Tremblay *et al.* 2005)
Angiosperms: Eudicots:			
Apiaceae: *Daucus carota*	TH 0.35; RI	36 kDa. Polygalacturonase inhibitor proteins and contains leucine-rich repeats.	(Meyer *et al.* 1999; Worrall *et al.* 1998)
Asteraceae: *Dendrosenecio* species	INA -3.5 to -5.7	INA in the polysaccharides fluid	(Embuscado *et al.* 1996)
Brassicaceae: Brassica oleracea	TH	66 kDa. Protein composition, Glycine (31.8 mol%), leucine (22.1 mol%), cysteine (20.6 mol%), serine (13.2 mol%).	(Huang 1995)
Cactaceae: *Opuntia humifusa, O. ficus-indica, O. streotacantha*	INA -2 to -3	Highly branched polysaccharides contributed to INA.	(Goldstein and Nobel 1994)
Ericaceae: Vaccinium corymbosum V. ashei	INA	Seasonal INA.	(Kishimoto *et al.* 2014)
Rhododendron flower	INA	INA in the bud tolerates 121 °C autoclave for 15 min.	(Ishikawa *et al.* 2015)
Fabaceae*: Ammopiptanthus mongolicus*	TH 0.35	50 kDa. No other information.	(Fei *et al.* 2001)
Ammopiptanthus mongolicus	TH 0.15	28 kDa. No other information.	(Wang and Wei 2003)
Arachis hypogaea	TH	33 kDa. N-terminal homology to Thaumatin-Like (TL) protein family. Hydrophilic protein and boiling stable.	(Dave and Mitra 1998)
Lobeliaceae: *Lobelia telekii*	INA	Large polysaccharides contributed to the INA	(Beck and Steer 1984; Krog *et al.* 1979)
Lobelia telekii, L. deckenii	INA -4	Carbohydrate contributed to the INA, heat stable to 100 °C.	(Krog *et al.* 1979)

African giant *Lobelia* species	INA -3.5 to -5.7	INA in the polysaccharides fluid	(Embuscado *et al.* 1996)
Rosaceae: *Prunus persica*	TH 0.1	60 kDa. Dehydrin-like proteins.	(Wisniewski *et al.* 1999)
Prunus wood	INA	Inactivated between 40 and 50 °C, but not affected by bacterial ice nucleation inhibitors, sulfhydryl reagents and Pronase.	(Gross *et al.* 1988)
Pyrus pyrifolia nakai	INA	Seasonal INA.	(Sekozawa *et al.* 2001)
Rutaceae: *Citrus sinensi*	INA	INA was proteinase sensitive but void of carbohydrate-like groups which is similar to bacterial INPs	(Constantinidou and Menkissoglu 1992)
Solanaceae: *Solanum dulcamara*	TH 0.2 to 0.5; RI	140, 62, 47, 40, 34 and 31 kDa. Proteinase and DTT sensitive. High similarity with insect THPs (by immunoelectroblots).	(Urrutia *et al.* 1992)
Solanum dulcamara	TH 0.2 to 0.5	67 kDa. Protein high in glycine. Carbohydrate including galactose is required for the activity.	(Duman 1994)
Solanum dulcamara	TH 0.3	64 kDa. Similar to WRKY proteins.	(Huang and Duman 2002)
Solanum dulcamara	TH	47 kDa. Chitinase-like protein; 29 kDa, Class I chitinase-like proteins.	(Huang and Duman 2002)
Theaceae: *Camellia sinensis*	INA	INA in young tea leaves.	(Goto *et al.* 1993)

Table 3.3 Summary of plant intrinsic ice activities (INA, TH and RI) identified. Numbers after INA or TH indicate INA or TH values (°C). Molecular weights (characters of ice activity) indicated molecular weight of related proteins identified.

The chemical compositions of plant intrinsic INPs were not well characterized. It is reported that polysaccharide and mucilage were functionally related to the INA of the African alpine plant whereas in winter rye, a protein complex with lipid and carbohydrate contributed (Table 3.3) (Griffith and Antikainen 1996). INA in woody stem was sensitive to organic solvent, heating and protease but the INA only reduced to -6 °C, which suggested a lower importance of protein or lipid in the INA of woody plants (Ashworth and Kieft 1995). However, heat and pH stable ice nucleators were reported in lichen with proteinase sensitivity (Kieft and Ruscetti 1990). Thus, it was concluded that the INA agent compositions were different in various plants, but mainly fell into two groups: non-protein ice nucleators and protein ice nucleators.

There are few studies analyzing the seasonal change of the three ice activities in plants, and those that do mainly analyze the seasonal change in INA including in the woody plant, blueberry, and in *Rhododendron* flowers (Brush *et al.* 1994; Griffith and Antikainen 1996; Ishikawa *et al.* 2015). INA in most *Chionochloa* species from this study showed different patterns of seasonal change. There are few studies analyzing the seasonal changes of the three ice activities in plants, and those that do mainly analyze the seasonal change in INA, for example in the woody shrub blueberry (*Vaccinium corymbosum*), and in *Rhododendron* flowers (Brush *et al.* 1994; Griffith and Antikainen 1996; Ishikawa *et al.* 2015). In this study INA was common in the *Chionochloa* species investigated but varied seasonally. Four species, *C. macra*, *C. conspicua*, *C. teretifolia* and *C. rigida ssp. amara*, had high levels of INA throughout the year investigated. Of these four species, only *C. macra* is strictly alpine; the ranges of the other three species extend into lowland-montane elevations (Pirie *et al*. 2010). This suggests that ice nucleators are present at quite high sub-zero temperatures and that they are constitutively expressed. The other *Chionochloa* species investigated showed different levels of INA at different seasons. There was no significant INA in *C. flavicans* in any season or in three of the four seasons (spring, summer, autumn) for *C. flavescens*, with mean INA lower than -12.0 °C. While *C. flavicans* is largely a lowland-montane species, *C. flavescens* extends into the alpine zone (Pirie *et al*. 2010). Frost is not just being an alpine phenomenon in New Zealand and lowland *Chionochloa* species often occur in frost hollows so are still exposed to freezing temperatures. Therefore, INA is a common feature in most species of this genus and lowland *Chionochloa* species may show similar patterns of INA, as species that can tolerate alpine environments (Table 3.1).

Interestingly, INA remained after 8,000-fold dilution in the August collections of *C. macra* and 4,000-fold dilution in the *C. rigida*, when there was a snowfall before the collections. In contrast, high levels of INA (~-6 °C) started to reduce when diluted 100-fold from the raw protein extract for the garden collections (January). It is assumed that a continuous cold stimulation will enhance the expression of INPs, with winter collection obtaining higher dilution resistance. INA remained after 1,000 fold dilution in the raw protein extract from the *C. macra* collected at Rock and Pillar in summer (14th January 2014) which is higher than Dunedin collections in summer, suggest that a high altitude collection has more INPs.

INA was more resistant to dilution in the *C. macra* than in the other *Chionochloa* species suggesting that this species has more INPs than other species. Higher INA was observed in the *C. macra* and the *C. rigida* than that found by Wharton *et al.* (2010), with the highest temperatures being -3.4 ℃ and -3.9 ℃ compared to -5.5 ℃ and -6.0 °C respectively (Wharton *et al.* 2010). Comparing the protein extraction methods, it is suggested a liquid nitrogen homogenization approach might be better than the chilled acetone method in retaining INA.

INA of the *Chionochloa* species investigated was found mainly in two branches of the phylogeny tree of figure 2 inferred by Pirie *et al.* (2010), which was built based on mitochondrial and rDNA markers (Pirie *et al.* 2010). Ice-shaping ability was only observed in *C. spiralis*, and *C. teretifolia*, which are lowland-low alpine species belong to the south of South Island in the figure 1 from Pirie *et al.* (2010), and *C. macra* which is a strictly alpine species belongs to the central Otago range in this figure 1. INA is absent in *C. beddiei*, *C. flavescens* and *C. flavicans* which belong to North Island and north of South Island. To achieve a whole view of the INA in this genus, other *Chionochloa* species in this genus need to be sampled. Meanwhile, the identification of related proteins and an ice activity-related maker might better solve the distribution and evolution of such activities.

The *Chionochloa* species investigated have no obvious TH activity throughout the year compared with buffer controls, except in the *C. spiralis* extract, in which a comparatively high TH activity was observed in the collections of June 2014, May 2015 and August 2015. The maintenance of TH activity requires high AFP concentration and dilution decreased protein concentrations, which reduced TH value quickly (Davies 2014). Typical hexagonal shaped ice crystals were observed in the nanoliter osmometry from the *C. spiralis* in all four collections. This suggested an

interaction between ice active proteins and the ice crystal whereas round shaped ice crystals in the buffer control suggest no interaction between proteins and ice crystals. Round shaped ice crystals were also observed in most of the other *Chionochloa* species in all four collections.

Generally, there was no obvious RI activity in most *Chionochloa* species investigated throughout the year except for in *C. spiralis*, with a mean ice crystal diameter of no more than 20 μm in 30 min and 30 μm in 60 min in the splat cooling assay. Dilution reduced the RI activity dramatically in the collection of October 2014 and January 2015, with the collection in January losing RI activity much more quickly on dilutions. *Chionchloa spiralis* collected in May 2015 and August 2015 had strong RI activity and it was still active after 40-fold and 240-fold dilution respectively. *Chionochloa spiralis* has needle-like leaves whereas the leaves of the many of the other *Chionochloa* species are comparatively broad and wide except for some species, such as *C. juncea*, *C. rubra* and *C. macra*, have narrow leaves. It is hypothesized that INPs on the wide leaves help to reduce freezing damage by crystallizing ice at higher sub-zero temperature but the *C. spiralis* has to utilize AFPs to both inhibit ice recrystallization and slightly depress the freezing point. Cold stress stimulates plants to enhance their tolerance to freezing and the expression of RI proteins is seasonal. INA in *Chionochloa* species were reduced when incubated at increasing temperatures. These characteristics were similar to intrinsic INPs in lichens and soil fungi but different to those of bacterial INPs (Fröhlich-Nowoisky *et al.* 2015; Kieft and Ruscetti 1990).

In the current study, I have found seasonal variation in ice activities among *Chionochloa* species in a common garden environment, indicating some amount of genetic control over these processes. However, comparison with wild collections from high elevations indicates that the environment also influences ice activities. Seasonal INA was present in most *Chionochloa* species. Thermal hysteresis and RI activity was also comparatively high in *C. spiralis*. The amounts and seasonal patterns of ice activities found here do not match patterns in the elevational ranges of the *Chionochloa* species examined, suggesting that ice activities have broad relevance in a range of environments.

Further studies were focused on characterizing (chapter 4) these three ice activities in this genus for the purposes of investigating whether extrinsic and/or intrinsic IAPs were involved and studying their characters and attempting to purify

and identifying (chapter 5) IAPs from *Chionochloa* leaves collected in winter utilizing various protein purification methods. The total RNAs derived from the winter *C. macra* was sequenced and the transcriptome was assembled (chapter 6) to create a pool of protein expression profile under cold environments to assist the identification of IAP genes. The function of candidate IAP genes from the transcriptome was investigated in *E. coli* system and the evolution history was deduced (chapter 7) to understand how IAPs from *Chionochloa* evolved in the grass family.

Chapter 4 Characterization of ice activities in *Chionochloa* species

4.1. Introduction

Ice nucleations can be either beneficial or harmful to the plants and these are quite different. Ice nucleation activity, which has been most studied in gram-negative bacteria, is mainly controlled by ice nucleation protein (INP) complexes expressed on the bacterial outer membrane with highly conserved ice-binding repeat units (Graether and Jia 2001; Lorv *et al.* 2014). *Pseudomonas syringae* and *Erwinia herbicola* are the most active INA bacteria, which have the ability of nucleating ice formation at as high a temperature as -1 ℃ (Kozloff *et al.* 1983).

INA bacteria have been reported in many plants. The number of INA bacteria on different plant leaves varied and seasonal variations were observed (Mazarei and Kerr 1987; Nejad 2005). Normally, *P. syringae* on leaf surfaces induced ice formation at temperatures from -1.5 ℃ to -6.0 ℃. Common features of bacterial INA that have been reported are that they are unstable in acid pH (pH<5), alkaline pH (pH>9) and warmer temperatures (>30 ℃). An exception was found in some *Pseudomonas* species such as KUIN-4, the INA of which was stable in acid environments and was comparatively resistant to denaturants and protein modifying reagents (Kawahara *et al.* 1994; Kawahara *et al.* 1995). It is found that *ina* gene families encoded INPs in *Pseudomonas* species whereas lipopolysaccharide proteins and other proteinic compounds associated with phospholipids or carbohydrates also seemed to be involved in INA from *Pseudomonas* species (Kawahara *et al.* 1995; Morris *et al.* 2004).

Both epiphytic bacteria and endogenous INA agents contributed to plant INA (Brush *et al.* 1994). In plants, ice nucleation initialized by epiphytic bacteria resident on leaves surfaces lead to frost damage whereas the properties of endogenous INAs were found to be different from those of INA bacteria (Brush *et al.* 1994; Griffith and Antikainen 1996). Instead of damaging plant leaves, endogenous INA was reported to help to collect water from the fog (Wharton *et al.* 2010). Chemical compositions varied in plant INA agents that both protein INA agents and non-protein INA agents such as carbohydrates and phospholipids were identified (Brush *et al.* 1994; Griffith and Antikainen 1996). Different compositions were also reported in INA agents from

various plant species assessed by treatments including proteinase, heat and protein modification reagents (Brush *et al.* 1994).

Thermal hysteresis (TH) in plants was induced by thermal hysteresis proteins (THP) or antifreeze proteins (AFP) secreted in the apoplast but the activity is usually low compared with that in fishes and insects (Griffith *et al.* 2005; Zhang *et al.* 2010). High diversity was observed and current known plant THPs do not show significant high sequence identity (Cheng 1998; Griffith and Yaish 2004). Plant THPs were found to be similar to plant pathogenesis-related proteins including glucanase-like (GLPs), chitinase-like (CLPs), thaumatin-like proteins (TLPs) and, posttranscriptional modifications were involved in inducing the activity (Antikainen *et al.* 1996; Yu and Griffith 1999).

RI activity in plants was studied in species including winter rye, Antarctic hair grass, wheat and carrot (John *et al.* 2009; Stressmann *et al.* 2004). It was reported that a comparatively low concentration of RIPs maintained high RI function compared with THPs (Brown *et al.* 2014; Carpenter and Hansen 1992). The structures of RIPs in the *Pooideae* subfamily have been well characterized where two highly conserved structures, the leucine rich repeats (LLRs), which show similarities to receptor regions of receptor-like protein kinases (RLKs), and the xxNxVxG ice-binding repeats (IRIP), were identified (Sandve *et al.* 2008; Tremblay *et al.* 2005). The LRR repeats were thought to have evolved from the LLR-RLKs in *Pooideae*, whereas there is no evolutionary information about the IRIP domains (John *et al.* 2009). Peptides such as melittin and glucagon also inhibit ice recrystallization (Knight *et al.* 1995; Mizrahy *et al.* 2013).

Nanoliter osmometry was developed to measure TH and observe the function of AFPs on ice crystals. Most of the work is based on the subjective observation and judgments of experienced technicians under microscopes (Hassas-Roudsari and Goff 2012). The essential measurement procedures that melt ice crystals to only one single ice crystal in the observing well are quite labor intensive and time-consuming. Recent development of the nanoliter osmometry used computer-based temperature control and video analysis of freezing points and melting points, and this assists the determination of freezing points but the process is still manual (Braslavsky and Drori 2013).

Differential scanning calorimetry (DSC) is a thermal analysis technique that measures the heat capacity versus the temperature change in the sample substance of a

known mass compared with a reference (Freire 1995). DSC is usually applied in studying the thermal characterization of compounds such as pharmaceuticals and polymers (Coleman and Craig 1996; Theeuwes *et al.* 1974). DSC has also been widely used in analysis of the process of freezing and melting involved by AFPs (Hansen and Baust 1988; Ramløv *et al.* 2005; Wharton and Block 1997). Lu *et al.* (2002) introduced a DSC based method to measure the TH in plant species and Hassa-Roudsari *et al.* (2012) reported an improved DSC measurement of plant AFPs as these are low in TH but high in RI (Hassas-Roudsari and Goff 2012; Lu *et al.* 2002).

Besides splat cooling assays, the optical recrystallometer measured the RI activity by detecting the optical transmittance change of a fast-frozen sample annealing at a defined temperature. Samples with RI activity showed less significant change whereas samples without RI activity had obvious optical transmittance changes (Wharton *et al.* 2007).

The agents involved in the ice activities of the *Chionochloa* species and their properties have not been well studied yet. In this chapter, those components were essential in the INA of the *Chionochloa* species, and in the TH and RI activity of the *C. spiralis* were investigated. Further characterizations of ice activities were also made in this chapter. In addition, the suitability of DSC based methods to screen ice activities were examined.

4.2. Materials and methods

4.2.1. Sample preparation

Chionochloa species collected on 6th August 2015 were selected as these had the highest ice activities of the four collections during 2014-2015 (chapter 3). Raw protein extracts for each sample were made following the method described in chapter 2 and 3. Briefly, leaves were ground to a fine powder in liquid nitrogen and resuspended in plant total protein extract buffer (pH 7.0). Protein solutions were adjusted by plant total protein extract buffer (pH 7.0) to a total protein concentration of 5 mg/mL.

4.2.2. DSC measurements of ice activities

A differential scanning calorimeter (DSC Q2000, TA Instruments) was used to analyze the RI activities of selected *Chionochloa* species (Hassas-Roudsari and Goff 2012). A volume of 2 µL (around 2 mg) of each sample was placed in an anodized aluminum T-Zero®DSC pan and sealed hermetically. The instrument's heat capacity response was calibrated with indium. The reference was set with an empty pan. The reference and the sample pan were cooled from 20 °C to -40 °C at 5 °C/min and then heated to -4 °C at the same rate with a holding time of 60 min. The slopes of isothermal thermograms (heat flow versus time) during one hour at -4 °C were measured. One-way ANOVA was applied to examine the influence of different treatments to the protein extracts. Samples with RI activity were expected to show higher mean slope values than those without RI activity based on the observations of Hassas-Roudsari and Goff (2012). Three measurements were taken per sample.

Thermal hysteresis was measured by DSC with a modified protocol (Lu *et al.* 2002). First, a 2 µL sample was cooled from 20 °C to -20 °C at 1 °C/min and held at -20 °C for 2 min. The sample was then heated to totally melt at the same rate. The melting point was recorded as T_m and the process was repeated as previous except for heating to a temperature T_p that is slightly lower than the melting temperature (Tm) where ice crystal underwent partial melting. The sample was cooled to -10 °C at 1 °C/min after holding at T_p for 15 min to let the whole system stabilize. Different T_p were repeated the melting point T_m. The temperature of onset recrystallization (T_o) was determined at the point where heat was released (continuously increased heat flow values). Thermal hysteresis was calculated as T_p-T_o. BSA (5 mg/mL) in the plant total protein extraction buffer was used as a control. Three measurements were taken per sample.

DSC was also used in analyzing ice nucleation processes which utilized similar methods compared with the INA spectrometer in that both instruments reported the temperatures of crystallization in test solutions during programmed temperature gradients by detecting the heat released either with thermocouples or by heat flows in the DSC (Parody-Morreale *et al.* 1986). In this analysis, INA was screened by cooling 2 µL sample under the temperature gradient (1 °C to -25 °C at 1 °C/min). Heat flow increased dramatically during the crystallization of ice and the temperature was recorded as T_c. Three measurements were taken per sample.

4.2.3. Optical recrystallometry

Briefly, 200 µL sample solution was added in a transparent glass tube (length of 100 mm, diameter 5 mm) and rapidly frozen in an ethanol/dry ice bath for 1 min. Tubes were then slowly warmed from -20 ℃ to -8 ℃ and held at -8 ℃. The optical recrystallometer was precooled to -8 ℃, calibrated with an empty glass tube (transmittance=100) and a glass tube containing a wooden skewer (transmittance=0). A first reading of the initial transmittance t_{0h} was recorded with the sample tube kept at -8 ℃. A second reading t_{12h} was recorded with the same sample tube annealing at -8 ℃ for 12 h. The RI activity was indirectly measured as $t_{12h} - t_{0h}$ (Wharton *et al.* 2007).

4.2.4. Effect of pH

The pH stabilities of ice activities were examined by treatments at room temperatures (~25 ℃) in buffers of various pHs. Ice activities were measured at pH 2, pH 7 and pH 12 with an incubation time of 20 min. INA was measured for INA positive *Chionochloa* species with a mean INA no less than -6.5 ℃. TH activity and RI activity were measured for the *C. spiralis* extract.

4.2.5. Effect of heat

To screen for the temperature stability, raw protein extracts were incubated at 50 ℃, 70 ℃ and 90 ℃ for 10 min respectively. INA was measured in INA positive *Chionochloa* species immediately. TH activity and RI activity were measured in the *C. spiralis*.

4.2.6. Effect of reducing agent and oxidizing agent treatment

Raw protein extracts were treated with dithiothreitol (DTT) to a final concentration of 0.1 M and incubated at room temperature for 10 min. Another set of raw protein extracts were treated with sodium tetrathionate (STT) to a final concentration of 0.1 M and incubated at room temperature for 10 min.

4.2.7. Effect of proteinase treatment

Raw protein extracts were treated with proteinase K to a final concentration of 1 mg/mL and incubated at room temperature for 4 h. Three types of ice activities (INA, TH and RI) were examined.

4.2.8. Ion strength treatment

Raw protein extracts in *C. macra* were treated with NaCl to various final concentrations (0.1 to 2.0 M) and incubated at room temperature for 10 min. INA were also examined of protein extract in 2.0 M NaCl after dialysis against 25 mM Tris-HCl pH 7.0 buffer for 12 h.

4.2.9. Attempt to isolate INA bacteria

A total of 0.2 g fresh whole leaves were submerged in 30 mL 0.22 µm filtered 25 mM ammonium carbonate buffer (pH 7.9), leaving the cut edge of the leaf blade about 1 cm above the buffer surface. Samples were sonicated for 2 min at 20% amplitude (pulse 2 s on/4 s off). Buffer after sonication was tested for INA. An aliquot of this buffer was exposed to 2 mg/mL lysozyme at 37 °C for 4 h and INA was also examined.

Another aliquot of the buffer after sonication was centrifuged at 16,000 × g 4 ℃ for 15 min and the pellet was resuspended with 0.22 µm filtered 25 mM ammonium carbonate buffer (pH 7.9). An aliquot of 150 µL resuspended buffer was plated on nutrient agar with 2.5% glycerol (NAG agar) and cultured at 10 ℃. Colonies were streaked and dissolved in 200 µL buffer and tested for INA.

4.3. Results

4.3.1. Characterization of INA in the *C. macra* extract

Significant heat flow release was detected during the ice nucleation process of 2 µL *C. macra* extract in the DSC pan (Figure 4.1A). DSC measurement of ice crystallization temperature (T_c) in the *C. macra* extract showed consistent results to the measurements with the INA spectrometry (~-4 °C in the fresh raw extract). No delay of increasing heat flows at various onset recrystallization points (T_os) was found in the TH scanning by the DSC in the *C. macra* extract indicated no thermal

hysteresis (Figure 4.1B). The slope value of heat flows in the DSC isothermal scanning for RI activity in the *C. macra* extract showed no difference with Tris buffer control in Figure 4.10 (Figure 4.1C).

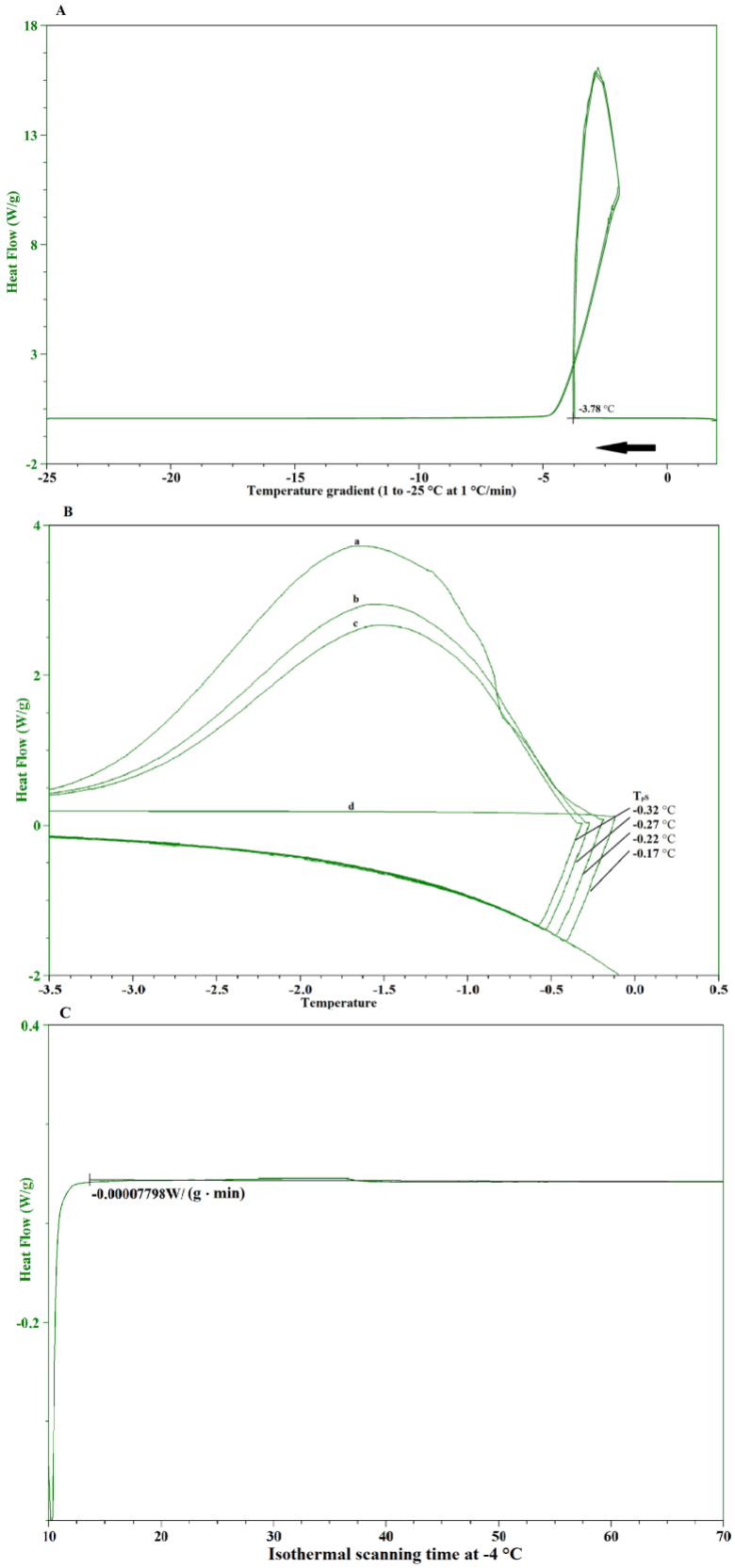


Figure 4.1 DSC measurements of INA (A), TH (B) and RI activity (C) in the *C. macra* extract. A, ice crystallization at -3.78 °C; B, refreezing of partially melted ice crystals at different T_{pS} (a, b and c) and no significant delay of freezing was observed. No refreezing in d indicates all ice crystals melt at -0.17 °C; C, slope value of heat flow in the sample vs isothermal scanning time.

Effects to INA of the *C. macra* extract under different treatments are shown in Figure 4.2. Briefly, high levels of INA in the *C. macra* slightly decreased in the 50 °C treatment for 10 min (-4.95±0.43 °C) and decreased in the 70 °C treatment for 10 min (-11.10±0.07 °C) and was inactivated in the 90 °C treatment for 10 min (-15.69±0.55 °C). This activity was stable in various pH conditions (pH 2, pH 7 and pH 12). Either 0.1 M reducing agent (DTT) or 0.1 M oxidizing agent (STT) could not inactivate INA in 10 min incubation but the activity was slightly decreased in either treatment compared with that in the raw extract. Proteinase K digestion reduced INA in the *C. macra* when incubated at 37 °C for 24 h but the activity was still higher than -9 °C. Mature leaves had higher INA than the stem and shoots (Figure 4.2). The buffer in which the *C. macra* leaves were subjected to low power sonication contained high levels of INA (~-5 °C).

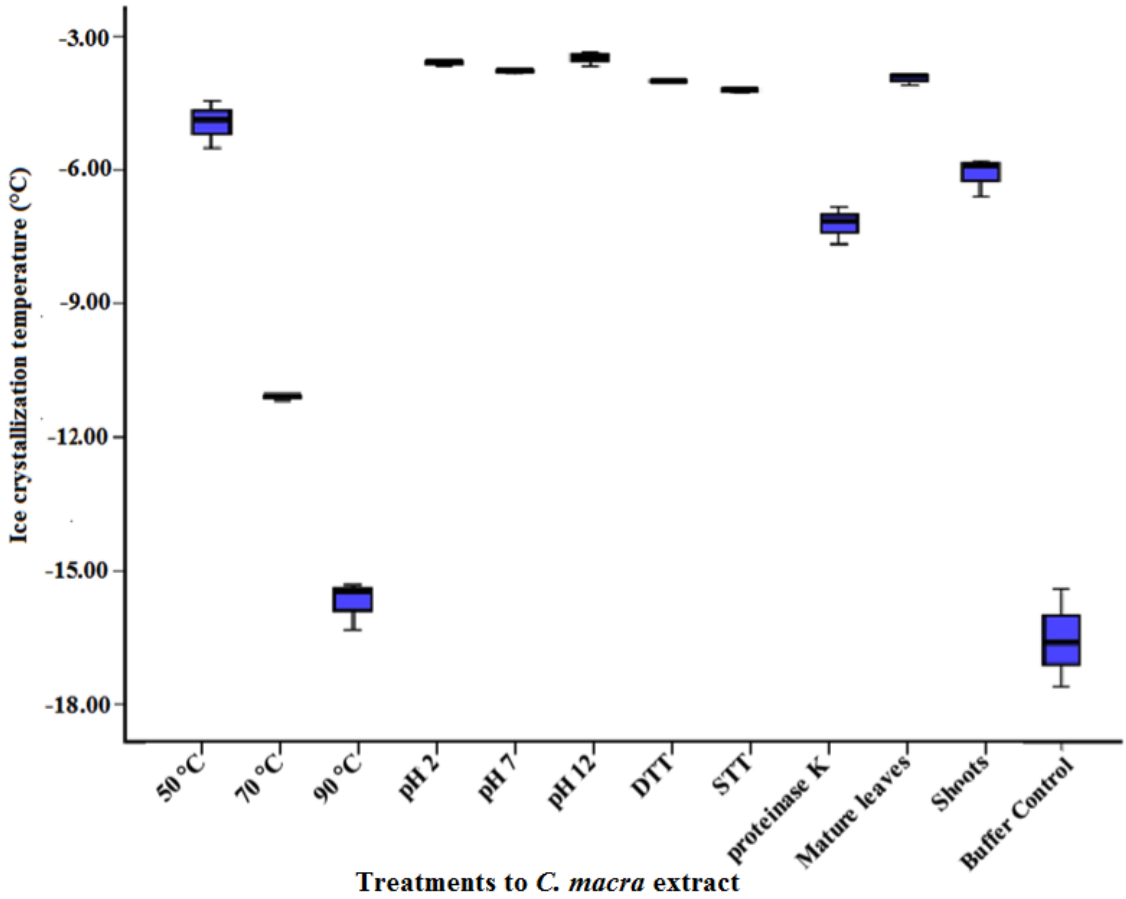


Figure 4.2 Characterization of the INA in the winter *C. macra* extract. Buffer control, 25 mM Tris buffer, pH 7.0. Middle dark line in boxplot, mean T_c from three individual leaves; Buffer control, n=3. T_c analyzed by DSC.

The effect of ionic strength on INA was examined in the *C. macra* extract at various concentrations (from 0 M to 2 M) of sodium chloride (NaCl) (Figure 4.2). INA started to reduce with increasing concentrations of salt in the solutions. It reduced slightly within 0.2 M NaCl and decreased moderately from 0.3 M to 0.9 M NaCl. When the concentration of NaCl reached 1 M, INA in the *C. macra* extract decreased to lower than -10 °C. INA decreased to -15.10±0.18 °C when at 2 M NaCl, whereas removal of salt by dialysis largely recovered the high INA in the *C. macra* extract (Figure 4.3).

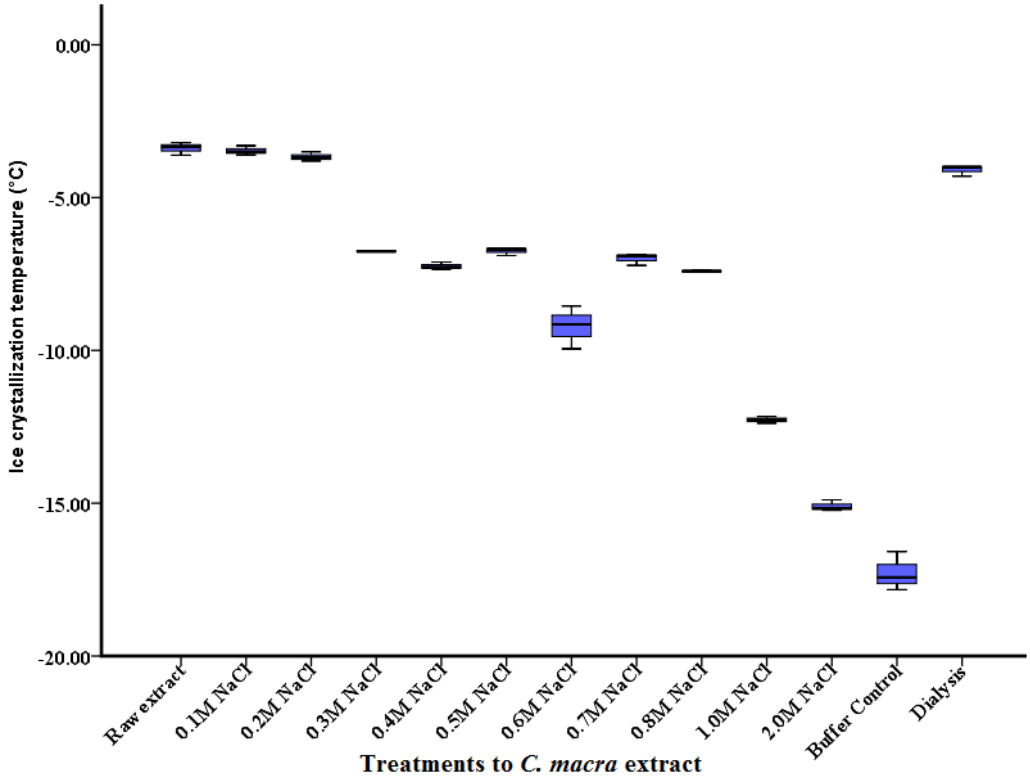


Figure 4.3 Effects of ionic strength on INA in the *C. macra* extract. Middle dark line in boxplot, mean T_c from three individual mature leaves; Buffer control, n=3. T_c analyzed by DSC.

4.3.2. INA in other *Chionochloa* species

Other *Chionochloa* species, which also had high INA from the 6th August 2015 collection, showed different patterns under the same treatments as the *C. macra* (Figure 4.4). INA in all these species gradually reduced with increasing temperatures but the resistance to heat varied among species. Generally, INA in all other INA positive *Chionochloa* species resisted the 30 °C treatment. Under the 50 °C treatment, high INA in the *C. conspiculia* remained and it reduced in the *C. rubra* Zotov but was

still higher than -7 ℃. INA in the *C. antarctica* and the *C. rigida* reduced to around -13 ℃ and in the other species it reduced to less than ~-15 ℃ after the 50 ℃ treatment.

Proteinase K digestion decreased INA by at least 5 ℃ in all test species except *C. flavescens ssp. brevis* and *C. flavescens ssp. lupeola* which still had INA over -12 ℃, with -8.18±0.61 ℃ and -11.41±0.47 ℃ respectively. In most *Chionochloa* species, proteinase K digestion decreased the INA to less than ~-15 ℃, which is different to the INA in *C. macra*.

STT and DTT slightly decreased INA in most species. Most *Chionochloa* species still had INA no less than -6.5 ℃ after DTT treatment except for *C. antarctica* (-7.47±0.07 ℃) and *C. rigida ssp. amara* (-8.36±0.42 ℃). Under the STT treatment, INA in *C. flavescens ssp. brevis*, *C. pallens ssp. cadens*, *C. rigida ssp. amara*, *C. rubra ssp. cuprea, C. teretifolia* and *C. antarctica* reduced to temperatures lower than -6.5 ℃.

Chionochloa species showed different tolerances to pH treatments in their INA. In the pH 2 treatment, INA reduced by around 1 ℃ in the *C. flavescens ssp. brevis*, the *C. rigid*a and the *C. rigida ssp. amara*, around 2 ℃ in the *C. rubra ssp. cuprea* and the *C. teretifolia*, around 3 ℃ in the *C. cheesemanii*, the *C. flavescens ssp. lupeola*, the *C. pallens ssp. cadens*. INA in the *C. antarctica* and the *C. conspicua* decreased by more than 3 ℃ but these INAs were still higher than the buffer control. In the *C. defracta* and the *C. rubra* Zotov, INA reduced to that of buffer control under pH 2 treatment. In the pH 12 treatment, INA reduced much quicker than in pH2 treatment except for the *C. conspicua*, the *C. defracta*, and the *C. rubra* Zotov, in which INA decreased less in the pH 12 treatment than in the pH 2 treatment. Only the *C. rigida* and the *C. pallens ssp. cadens* had INA similar to raw extract after pH 12 treatment, with -4.54±0.17 ℃ and -6.72±0.25 ℃ respectively.

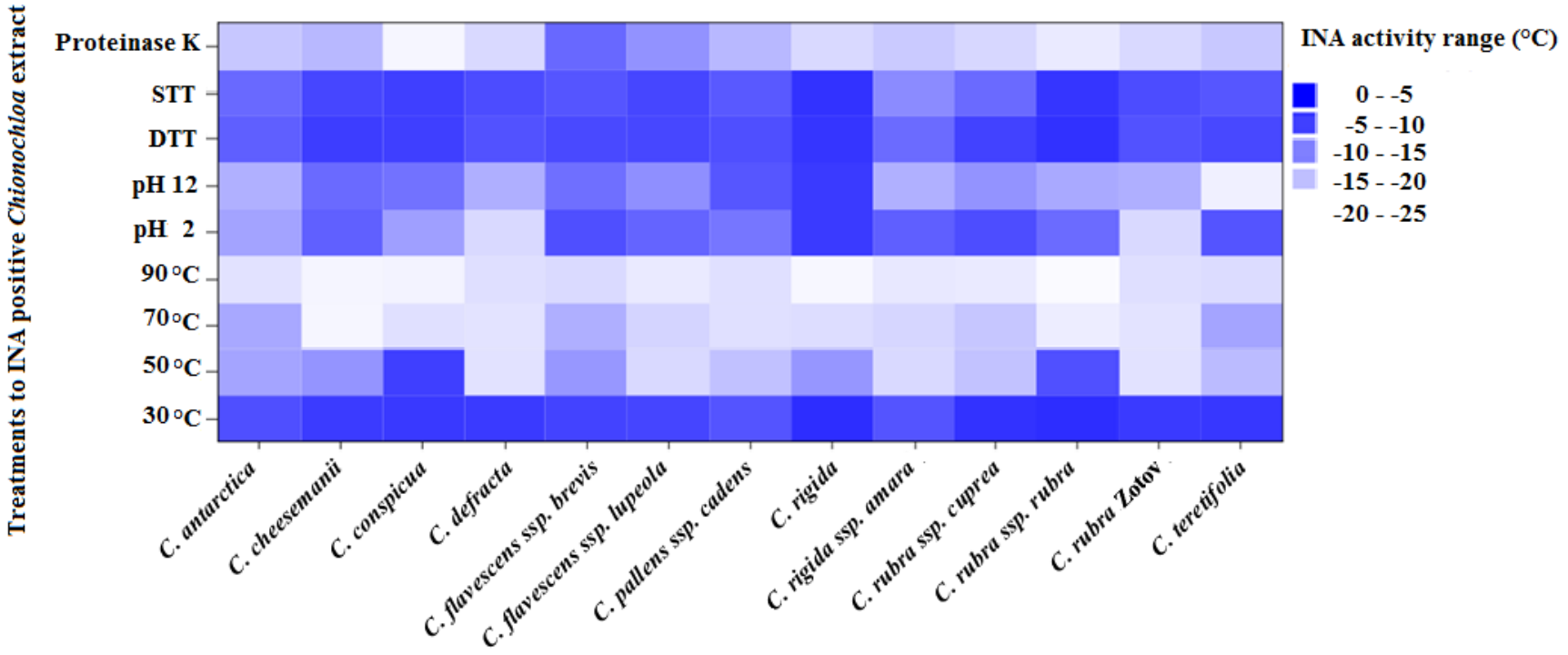


Figure 4.4 Characterization of INA in the *Chionochloa* species. Heat map figure legends (right) indicate mean INA (mean ice crystallization temperature) ranges. Heat map colors, mean INAs from three individual mature leaves; Buffer control, n=3. T_c analyzed by DSC.

4.3.3. TH and RI activities in the *C. spiralis*

TH activity and RI activity in the *C. spiralis* were screened for heat, pH, reducing, oxidizing agent and proteinase treatment stability (Table 4.1). The 50 °C treatment slightly changed TH value and the ice crystal shape of the *C. spiralis* in the nanoliter osmometry compared with raw extract. TH values were reduced significantly under the 70 °C treatment but hexagon ice crystals were still observed in the nanoliter osmometry. In the 95 °C and proteinase K treatments, TH values were reduced to the control level with round ice crystals observed. Strong RI activity was still remained after 50 °C, 70 °C, 95 °C, pH 2, pH 7, pH 12, DTT and STT treatments but reduced under proteinase K treatment.

Conditions	TH value	Ice shape	RI (0min) (μm)	RI (30min) (μm)
50 °C	0.09	Hexagon	9.7	10.2
70 °C	0.05	Hexagon	14.4	17.5
95 °C	0.05	Hexagon	13.7	17.8
pH 2	0.05	Hexagon	33.6	62.5
pH 7	0.12	Hexagon	11.7	12.8
pH 12	0.06	Hexagon	30.7	50.6
DTT	0.06	Hexagon	14.0	19.2
STT	0.07	Hexagon	15.1	19.7
Proteinase K	0.04	Round	13.7	43.0
Buffer	0.04	Round	75.6	109.0

Table 4.1 TH and RI activity in the *C. spiralis* extract after various treatments (measured by nanoliter osmometry and splat cooling assay). Mean TH was from three individual mature leaves. Mean ice crystal diameter was from 10 randomly selected largest ice crystals from three individual mature leaves.

TH activity in a protein extract of *C. spiralis* from the August 2015 collection was further confirmed by DSC. DSC measurements of freezing and melting processes of the *C. spiralis* and bovine serum albumin (BSA) control (5 mg/mL) were shown in Figure 4.5 and 4.6. The ice crystallization peak of the *C. spiralis* was at -11.35 °C and the peak covered a temperature range of around 3 °C. The melting peak was at -0.75 °C and it was smooth and covered a temperature range of around 7 °C (Figure 4.5). In the BSA control, the freezing peak showed a different pattern to that of the *C.*

spiralis. The melting peak of the BSA control was also different being narrower and covering less than 3 °C (Figure 4.6). In the *C. spiralis* extract, different T_ps from -2.50 °C to -0.75 °C were screened to detect delay freezing. Various degrees of delayed refreezing were observed from -1.50 °C to -0.83 °C, and the highest TH value reached 0.54 °C (T_m, -0.83 °C; T_o, -1.37 °C) in the DSC measurement, which was higher than what was observed in the nanoliter osmometry (0.1 °C) and TH of current known plant THPs (0.1 °C to 0.5 °C) (Figure 4.7). In the BSA control, TH was not detected and no delayed refreezing was observed. There was delayed freezing in the treatments including pH 2, pH 12, reducing agent, oxidizing agent. There was no delayed freezing in the protein extract after proteinase K treatment that suggested that the TH was destroyed.

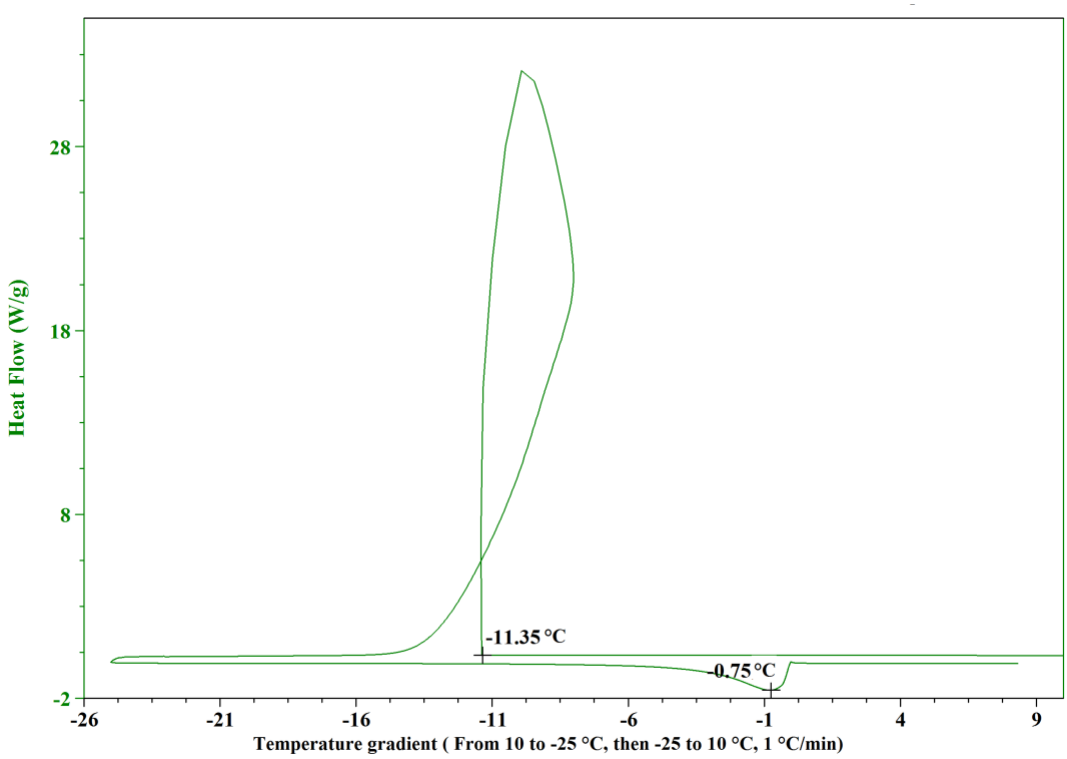


Figure 4.5 DSC scanning of freezing and melting in *C. spiralis* extract. Freezing point and melting peak were labeled. As there was slowly heat flow release during the temperature raised, it was difficult to detect the start point of melting. Green line presents the heat flow difference between sample pan and reference pan during the temperature gradients. Heat flow was identical with the reference pan till the ice crystallization point (-11.35 °C) where heat was released.

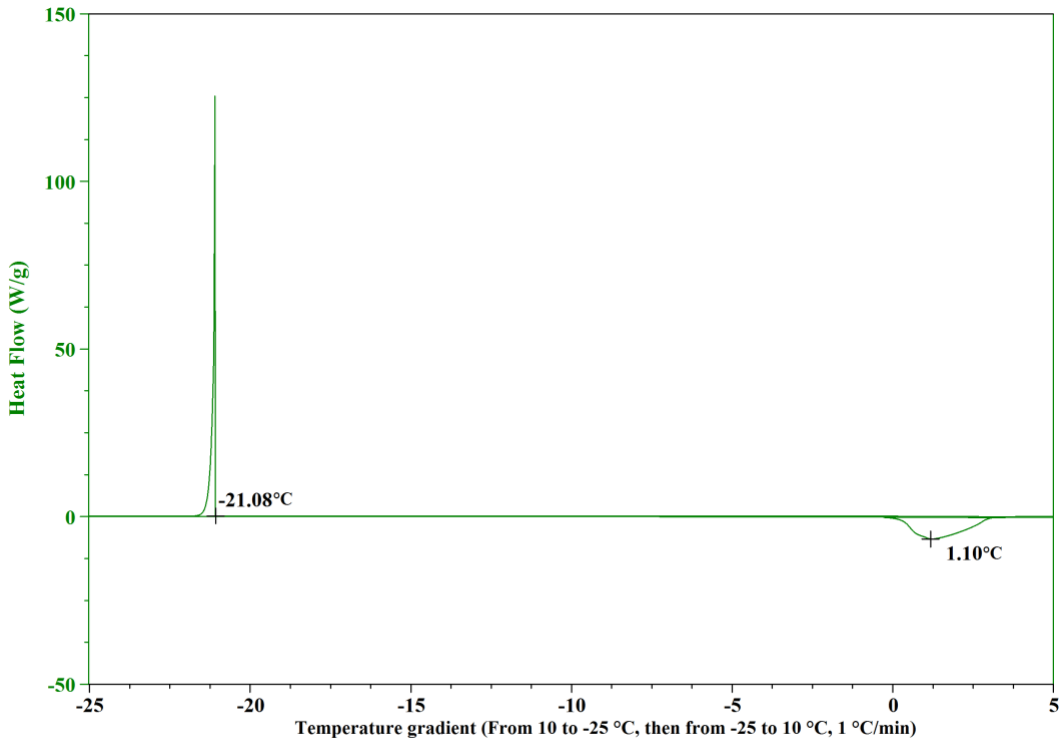


Figure 4.6 DSC scanning of freezing and melting in a BSA control. BSA froze at -21.08 °C and released heat (heat flows). BSA ice crystals started to melt at -0.58 °C (significantly heat flow release) and the heat flow release peak was at 1.10 °C. The melting process covered a temperature range of ~3 °C.

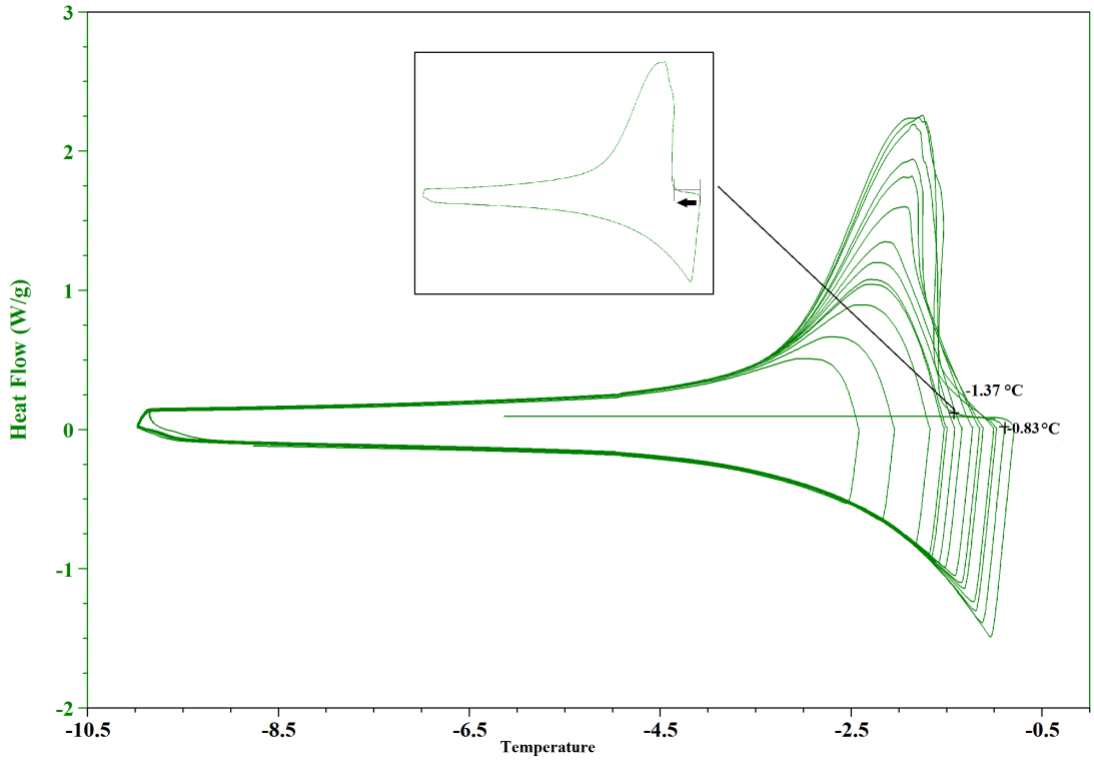


Figure 4.7 DSC scanning of TH activity in *C. spiralis* extract. The delay of heat flow increase indicates thermal hysteresis. Refreezing of partially melted ice crystals at

different T_ps (from -0.75 °C to -2.50 °C) and significant delay of freezing was observed at -0.83 °C. Between the labeled -0.83 °C and -1.37 °C points, no obvious heat flow was detected and a delay in crystallization was observed. After -1.37 °C, heat flow increased dramatically present as ice crystallization occurred.

RI activity in the *C. spiralis* extract was measured by the splat cooling assay, optical recrystallometry and DCS measurements following methods described by Hassa-Roudsari *et al*. (2012) (Figures 4.8-4.10). The slope of the *C. spiralis* extract obtained from the heat flow versus time isothermal diagrams between 25 min and 85 min at -4 °C ($1.3\times10^{-3}\pm1.7\times10^{-4}$ W/g.min) were significantly higher than those in the *C. macra* extract ($4.7\times10^{-4}\pm2.8\times10^{-5}$ W/g.min) and the buffer control ($9.0\times10^{-5}\pm3.1\times10^{-5}$ W/g.min) (Figure 4.10). Temperature treatment did not change the slope value significantly compared with those of the *C. spiralis* extract. Treatments including pH 2 ($5.0\times10^{-4}\pm9.0\times10^{-5}$ W/g.min), pH 12 ($4.8\times10^{-4}\pm6.3\times10^{-5}$ W/g.min), reducing agent ($4.4\times10^{-4}\pm3.7\times10^{-5}$ W/g.min), oxidizing agent ($4.9\times10^{-4}\pm3.9\times10^{-5}$ W/g.min) and proteinase K ($4.0\times10^{-4}\pm3.4\times10^{-4}$ W/g.min) decreased the slope values significantly compared with those of the *C. spiralis* extract without treatments (Figure 4.10).

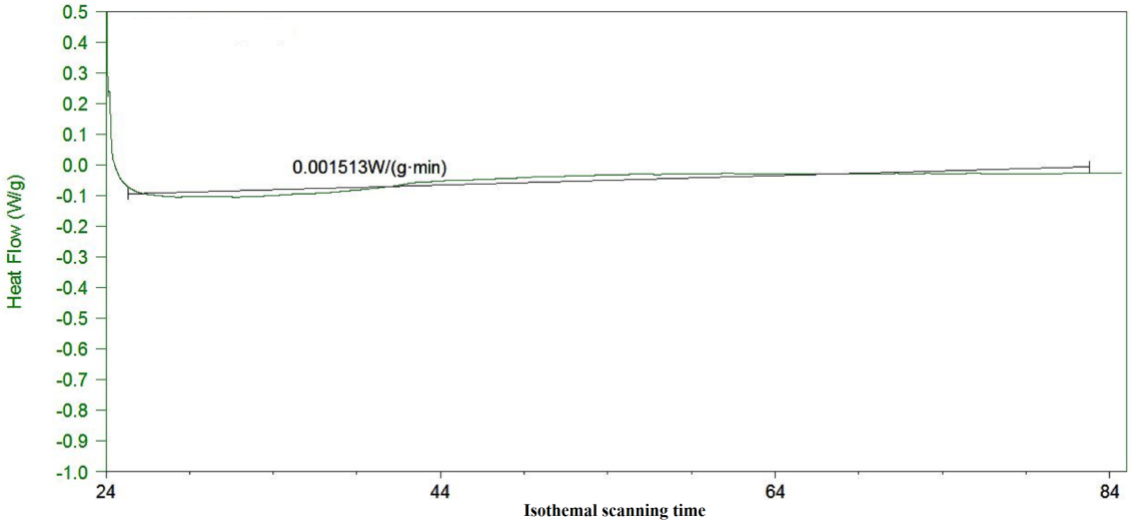


Figure 4.8 Measurement of RI activity in the *C. spiralis* extract by DSC. Heat flows of the *C. spiralis* extract under DSC isothermal scanning at -4 °C was recorded. The slope value of the thermogram between 25.0 min and 82.5 min was calculated by Universal Analysis 2000 version 4.5A, TA Instruments.

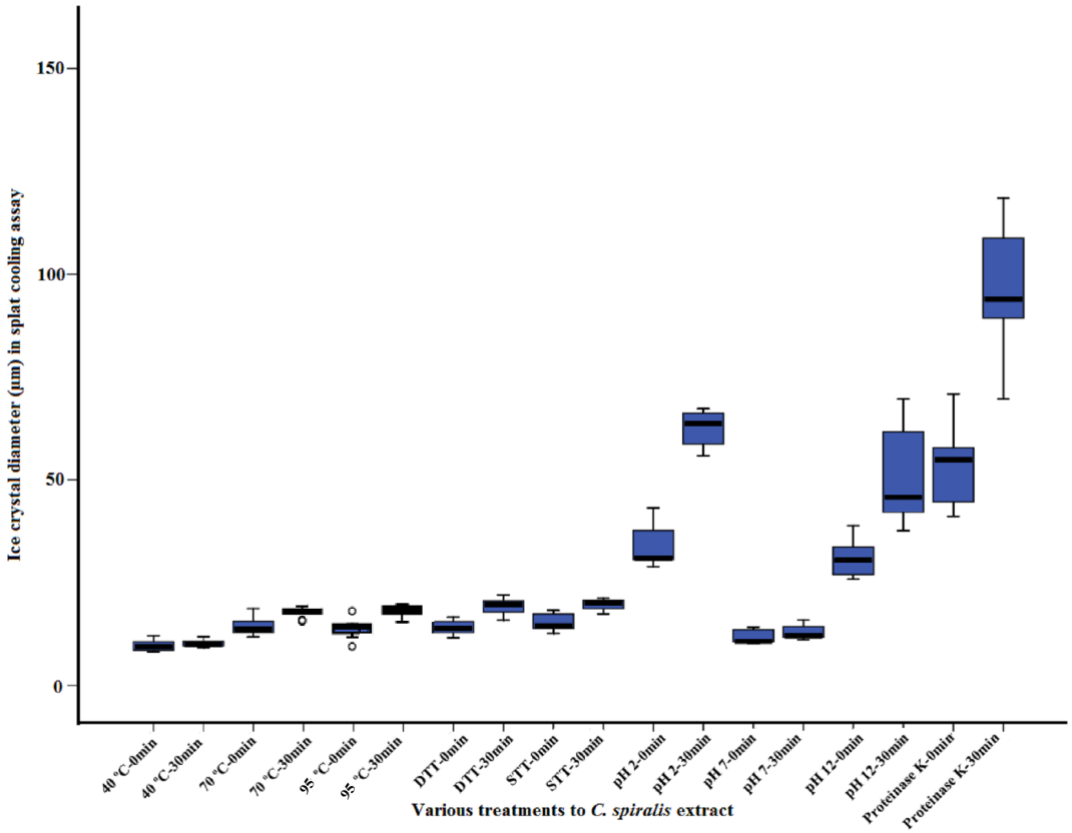


Figure 4.9 RI activity measured by splat cooling assay of the *C. spiralis* extract under different treatments. Y axis, ice crystal diameter (μm). Middle dark line in boxplot, mean ice crystal diameter for 10 randomly selected largest ice crystals from three individual mature leaves.

RI activity measured by optical recrystallometry showed similar trends to those of splat cooling assays (Figure 4.9; Figure 4.11).

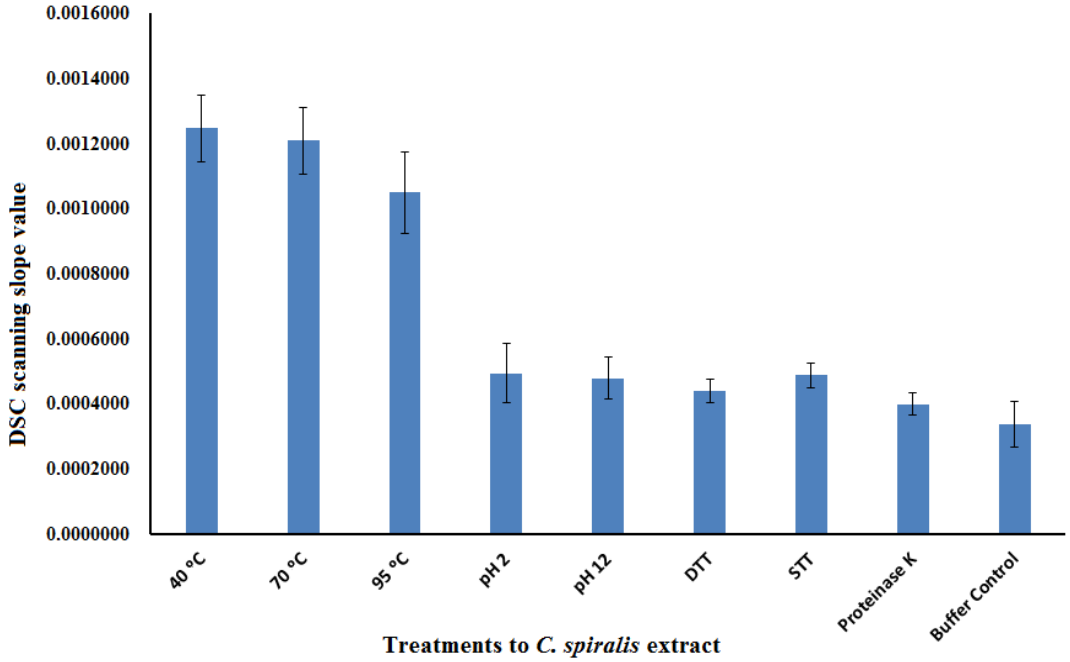


Figure 4.10 RI activity of the *C. spiralis* extract under different treatments measured by the DSC. Y axis, slope values of heat flow vs isotherm scanning time. Mean slope value±standard deviation, n=3.

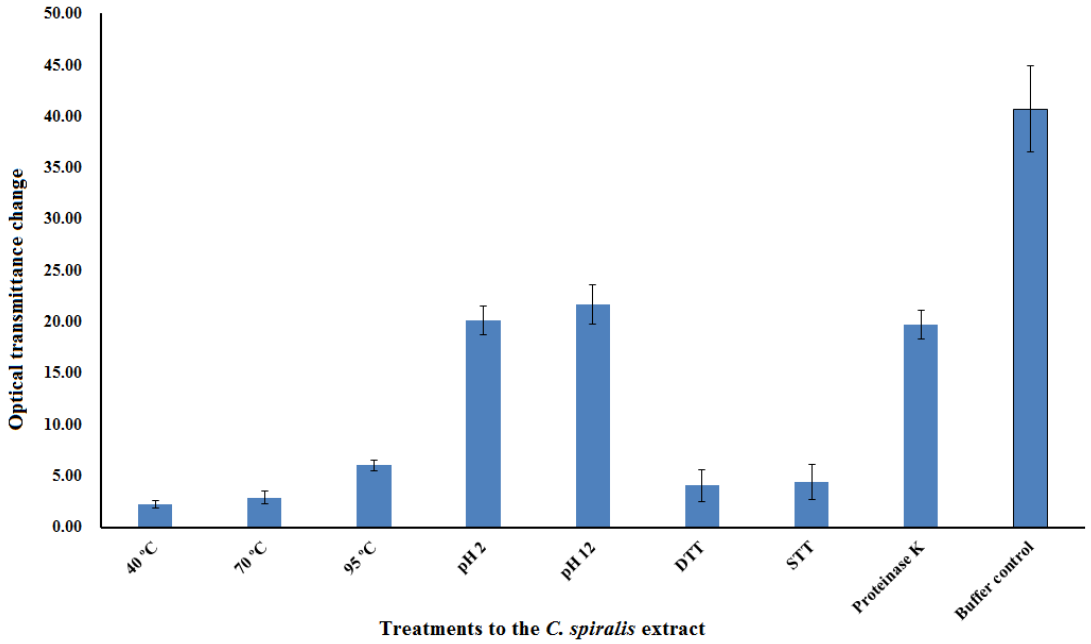


Figure 4.11 RI activity of the *C. spiralis* extract under different treatments measured by optical recrystallometry. Y axis, transmittance changes in 12 h. Mean transmittance changes ± standard deviations, n=3.

4.3.4. INA bacteria isolation

Many colonies were grown on the NAG plates cultured at 10 °C (for one week). Colonies were randomly selected to screen for the INA and bacteria with the highest mean INA (-8.2 °C) was chosen for further study (Figure 4.12). This highest bacteria

INA did not tolerate high temperatures, proteinase K and lysozyme treatments (Figure 4.12).

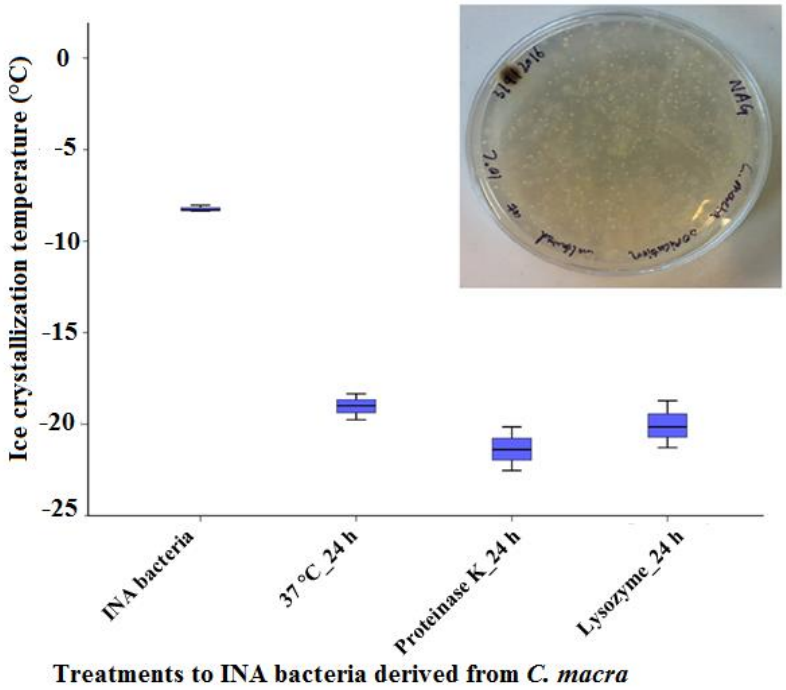


Figure 4.12 Bacteria derived from the *C. macra* leaf grown at 10 °C on a NAG agar plate and its INA under various treatments. Middle dark line in boxplot, mean T_c per treatment, n=3. The bacteria colony with the highest mean INA (-8.23 °C) was selected for downstream analysis. T_c analyzed by DSC.

INA in the *C. macra* extract was slightly decreased after lysozyme treatment (1 mg/mL) at 37 °C for 4 h and 24 h respectively but still higher than -5 °C, which is similar under proteinase K treatment (1 mg/mL). The buffer where the *C. macra* leaves were submerged and sonicated, was filtered by 0.45 µm and 0.22 µm syringe filter respectively. The INA in these fresh, 0.45 µm and 0.22 µm filtered buffers were slightly decreased and did not tolerant high heat, proteinase K and lysozyme treatments (Figure 4.13).

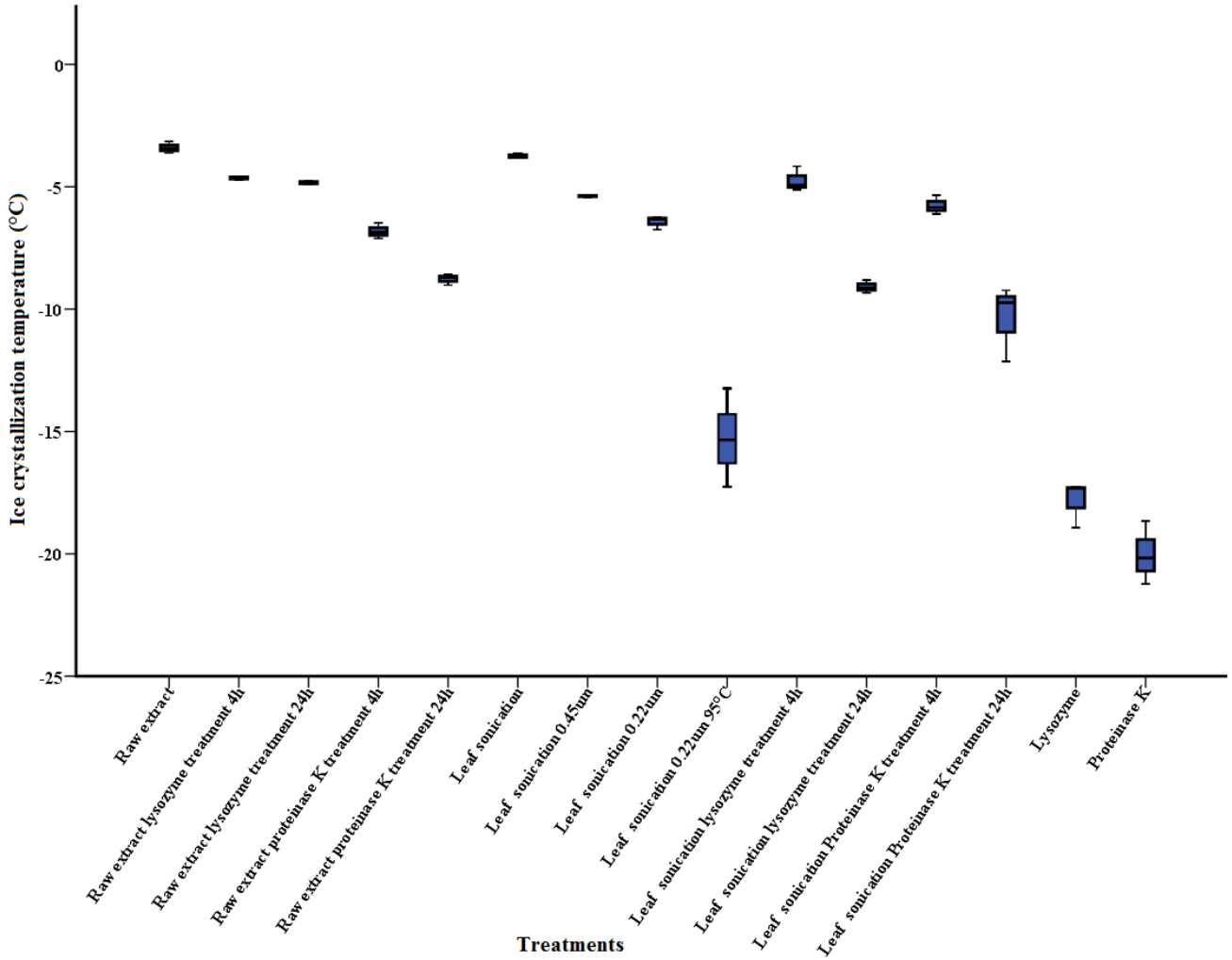


Figure 4.13 Bacterial INA screening in the *C. macra* raw extract. Middle dark line in each boxplot, mean T_c, n=3. T_c measured by DSC method.

4.4. Discussion

Ice activities, especially INA, are common features in *Chionochloa* species and protein components were confirmed to be contributing to all three types of ice activities in these species. Proteinase K digestion reduced ice activities including INA in all *Chionochloa* species and also TH and, RI activity in the *C. spiralis* extract. These activities after proteinase K treatments were still significantly higher than that of the buffer control suggesting that partial peptide fragments were still functional in maintaining these activities.

INA in the *Chionochloa* species were reduced when incubating at increasing temperatures. In the 50 °C treatment, INA remained around -5 °C in the *C. macra* and in the *C. conspicua* whereas it decreased in other *Chionochloa* species. In the 70 °C treatment, INA in the *C. macra* reduced to around -11 °C and in the 90 °C incubation INA was reduced to that of the buffer control level. Other *Chionochloa* species showed similar trends with the 70 °C treatment and the 90 °C treatment respectively. INPs were inactivated at higher temperatures and the numbers of available ice nucleators were gradually reduced. These characteristics were similar to intrinsic INPs in lichens and soil fungi but different to bacteria INPs (Fröhlich-Nowoisky *et al.* 2015; Kieft and Ruscetti 1990).

High salts lower the freezing points of solutions, which in turn inhibits the function of INPs (Banin and Anderson 1974). The removal of salts recovered the INA, which suggested dialysis was useful for the purification after using an ion exchange column (IEX) (Figure 4.3). DTT and STT slightly reduced INA at the concentration of 0.1 M, which was lower than INA under the treatment of 0.1 M NaCl. Increasing the concentration of DTT or STT may not be helpful in detecting the effect of DTT or STT on INA as the ion strength in the solution also contribute to the inhibition of INA (Figure 4.3). Ice activities are also pH stable in *Chionochloa* species including *C. macra* and *C. rigida* as it tolerant at pH 2, 7 and 12 whereas the INA decreased in other *Chionochloa* species investigated when at pH 2 or pH 12. Precipitation occurred when in pH 2 suggested that the raw extract is not stable in a strongly acidic environment.

Low power sonication was used to screen whether there were INA bacteria on the leaves contributing to the INA observed (Wharton *et al.* 2010). In this study, high

INA were detected in the buffer after leaf sonication but this may result from either the presence of INA bacteria or INPs on the leaf surfaces, which were separated by mechanical force, or both. Lysozyme digestion failed to eliminate INA suggest non-bacterial INPs are involved (Moffet *et al.* 2015; Warren 1987).

INA was found in fractions between 100 kDa and 300 kDa by MWCO centrifugal device and INA in the *C. macra* showed a similar molecular weight range with MWCO centrifugal device (Fröhlich-Nowoisky *et al.* 2015). Mature leaves were shown to have higher INA than the stem, which indicated more INPs on mature leaves. Together with previous studies, it is concluded that INPs in *Chionochloa* species is expressed on the leaf surface and can be released by sonication. INPs on the leaves may enhance the efficiency of water collection from fog condensation, which might be the function of these surface INPs in *Chionochloa* tussocks (Wharton *et al.* 2010).

DTT slightly reduced TH and RI activity, which suggests that disulfide bonds might play a role in maintaining these activities. Ice activities in *Chionochloa* species are comparatively heat stable in the short term, which may be because of abundant disulfide bonds and hydrogen bonds in the IAPs (Hassas-Roudsari and Goff 2012). RI activity in *Chionochloa* species showed a different pattern after heat treatments compared to that of INA. RI activity in the *C. spiralis* tolerated high temperature (95 °C for 20 min) with only a slight decrease. It is reported that RIPs in Antarctic hair grass also tolerate high temperature (John *et al.* 2009). Whether there are similarities between RIPs in *Chionochloa* grass and Antarctic hair grass need to be further investigated.

The presence of all three types of plant IAPs were screened by DSC. Ice nucleation by INPs release heat that is detected by the DSC during cooling and the temperature of ice crystallization (T_c) was measured. This method is similar to the INA spectrometer (Wharton *et al.* 2010). The DSC-based INA assay showed similar T_cs in the *Chionochloa* species compared with those measured by the INA spectrometer under same temperature gradient. DSC based INA assay requires less test solution than the INA spectrometer but still needs the presence of an ice nucleator to ensure ice crystallization.

The determination of the melting point and the hysteresis refreezing point of a single crystal is quite a challenge under the microscope of the nanoliter osmometer, which making accuracy and consistency difficult. Measurements are recorded by the

operator, which could induce experimental errors. Melting to one ice crystal in the observing well can also be a challenge. DSC provides an easier way to analyze the thermal hysteresis, which was detected by the occurrence of a delay in freezing during cooling after holding isothermally at the melting temperature to produce a situation where the ice crystals in the test solution are partly melted. Previous studies have proved that the DSC is suitable to study plant thermal hysteresis (Hansen and Baust 1988; Hassas-Roudsari 2011; Lu *et al.* 2002). Generally, TH values in plants are less than 0.5 ℃ measured by nanoliter osmometry and a TH value of 0.1 ℃ was reported in the *C. spiralis* extract. The readings of DSC based TH measurements were higher than those were recorded by nanoliter osmometry, with DSC measurement showing the highest TH value of 0.54 ℃ in the same *C. spiralis* extract. Higher TH values were also observed in DSC measurements of winter wheat leaves and *Ammopiptanthus mongolicus*, but these studies did not mention why DSC gave higher TH values than those measured by nanoliter osmometry (Hassas-Roudsar 2011; Lu *et al.* 2002). The TH value calculated by T_m-T_o in the DSC measurement indicates the melting and refreezing process of one or more ice crystals in the DSC sample pan but there is only a single ice crystal in nanoliter osmometry. The DSC measured the absorbance or release of heat, whereas nanoliter osmometry is based on the morphological changes of a single ice crystal. For these reasons, TH measured by DSC might be different from what was recorded by nanoliter osmometry.

The splat cooling assay analyzed changes in the size of ice crystals when held at a certain temperature. The optical recrystallometer reads the transmittance of light passing through ice crystals with samples with high RI activity having small crystals scattering the light more, with lower transmittance and showing less change of transmittance with time. During the DSC measurement, the reference pan is in a stable state with no heat being absorbed or released, and hence no heat flow change. In the sample pan freezing and melting of sample releases and absorbs heat respectively. The DSC measures the heat flow required to maintain the sample and reference pans at the same temperature (Hassas-Roudsari and Goff 2012). The presence of RIPs in the sample delay the melting of small ice crystals, making the time required to reach a stable state longer, whereas the stable state was reached quickly in the sample without RIPs. Hassas-Roudsari and Goff (2012) concluded that exotherm was observed during the isothermal scanning caused by the activity of IAPs, thus, a positive slope was present. Different slope values were found during

isothermal scanning of the *C. spiralis* extracts compared with those of the control buffer and in samples where the RI activity had been inactivated. This shows that the DSC is an accurate and fast way to examine the RI activity. However, in this study, slope values of the *C. spiralis* extracts after DTT and STT treatments were not found to be significantly different to controls in the DSC RI assays and for which I have no explanation. The results from splat cooling assay and the optical recrystallometry were consistent and showed a dramatically decrease in RI activity in pH and proteinase K treatments.

The common methods utilized in the ice activity measurements, which include INA spectrometry, nanoliter osmometry and the splat cooling assay, are time and labor consuming, whereas DSC provided a comparatively easy, accurate and quantitative method. The entire measurement process is automatically controlled by programmed commands on computers and samples used by DSC was less than in the other methods and can be reused for other measurements. However, no direct observation of protein-ice crystal interactions can be made in all measurements and comparatively high cost for the nitrogen gas and pans are drawbacks of the DSC technique.

Moffett *et al*. (2015) utilized lysozyme digestion and heat treatment (37 °C) to examine whether INA in lichen was from bacteria. Lysozyme destroyed bacterial INA by digesting cell walls and bacterial INA cannot tolerant heat over 30 °C (Moffet *et al.* 2015). Lysozyme digestion and the 37 °C treatment failed to significantly eliminate the INA in the *C. macra* extract and the hypothesis that plant endogenous INPs contributed to the INA in this species could not be rejected. INA from bacteria derived from the *C. macra* leaf was sensitive to heat (37 °C), lysozyme and Proteinase K treatment, which was similar to currently known bacterial INA.

To sum up, non-bacterial protein components were essential in maintaining all three types of ice activities in *Chionochloa* species. Low levels of INA were detected in the bacteria derived from the *C. macra* leaves. The *Chionochloa* species showed different levels of resistance in INA under various treatments. RI activities in the *C. spiralis* were comparatively resistant to heat, pH, reducing agents and oxidizing agents. There are obvious differences between the commonly used methods and DSC methods to measure TH and RI activity but DSC methods have some advantages.

Chapter 5 Identification of ice active proteins in the *Chionochloa* species

5.1. Introduction

Pure water can supercool to -37 °C without freezing but in the presence of efficient ice nucleators, water freezes quickly and at temperature much closer to the freezing point (Wilson *et al.* 2010). Ice nucleation proteins (INP), which help to trigger ice crystal formation, are found in many gram-negative bacteria including *Pseudomonas syringae* and *Pseudomonas fluorescens*, and their molecular genetics has been well investigated (Deininger *et al*. 1988; Hill *et al*. 2014; Kobashigawa *et al*. 2005). The mechanism of bacterial INP dimerization has also been well studied. Exposed tyrosine ladders in the central ice-binding repeats act as solvent hydrophobic interfaces, which allows INPs to closely contact with each other and results in high INA (Garnham *et al.* 2011). In contrast, INPs in plants, insects, intertidal invertebrates and amphibians (e.g. wood frog) are not well characterized and only partial physicochemical properties have been reported (Hew and Yang 1992).

Plant antifreeze proteins (AFP) have been reported in more than 60 plant species and over one fourth have been sequenced (Deswal *et al.* 2014). Generally, these proteins showed lower thermal hysteresis (TH) but higher recrystallization inhibition (RI) activity compared with the AFPs in fishes and insects (Huang and Duman 2002). Plant AFPs show high similarity with pathogenesis-related proteins including thaumatin-like proteins, class I and class II chitinases, β-1,3-glucanase protein and dehydrins (Gupta and Deswal 2014; Kuwabara and Imai 2009; Yu *et al.* 2001). Under cold stress, these antipathogenic proteins switched their pathogen resistance to antifreeze activity through posttranslational modifications (Hiilovaara-Teijo *et al.* 1999; Hiilovaara-Teijo and Palva 1999; Hon *et al.* 1995; Thomashow 1999). The β-sheet is the most common structure in plant AFPs, which provides a comparatively flat surface for ice binding, whereas there is no high homology in the plant AFPs among species, suggesting that these proteins evolved independent from each other (Gupta and Deswal 2014a; Kuiper *et al.* 2001; Zhang *et al.* 2004). Gupta and Deswal (2014) built a phylogenetic tree based on 47 AFPs and found two major groups with several sub-clades, in which the AFPs from the *Pooideae* subfamily belonged to a

group that contained the specific ice-binding domain, which might derive from a common ancestor in this subfamily (Gupta and Deswal 2014).

Endogenous plant INA other than bacterial INA were reported in two *Chionochloa* species (Wharton *et al.* 2010). Previous chapters have established the presence of seasonally expressed plant intrinsic INPs in most *Chionochloa* species and the occurrence of THPs and RIPs in the *C. spiralis*. The presence of INA in the moderate sonication of the *C. macra* leaf suggests leaf surface proteins may contribute to the activity.

Limited methods have been developed for the purification of IAPs and the main strategy is based on the affinity of proteins to the ice lattice (Davies 2014; Marshall *et al.* 2016). For a better understanding of the molecular genetics of IAPs in the *Chionochloa* species, I attempted to identify IAPs (INPs, THPs and RIPs) in *C. macra* and *C. spiralis* after ice affinity purification by homology with other ice-binding proteins or ice-shaping proteins.

5.2. Materials and Methods

5.2.1. Raw protein extract preparation

The preparation of *C. macra* and *C. spiralis* raw protein extracts was described in section 2.2.1.1 (chapter 2). In order to reduce the high concentration of salt after freeze-drying, the total protein extraction buffer (section 2.1.1.1, chapter 2) was replaced by 25 mM ammonium bicarbonate buffer (pH 7.5).

5.2.2. Ice affinity purification

Proteins in *C. macra* and *C. spiralis* that interact with ice lattices were harvested by both ice-finger and ice-shell, the detailed methods of which were described in section 2.2.3.1 (chapter 2). Ice fraction from the second round of the purification was concentrated by freeze-drying and processed by both SDS-PAGE and mass spectrometry.

5.2.3. Ice activities in ice affinity fractions

Three types of ice activities (INA, TH and RI) were measured for the ice fraction (IF) and liquid fraction (LF) with the methods described in section 2.2.2 (chapter 2).

5.2.4. Gel electrophoresis and identification of ice-binding proteins

Samples were concentrated by freeze-drying together with a TCA-DOC/acetone clean up protocol. Standard SDS-PAGE was used to resolve proteins included in the ice lattices. The detailed method of SDS-PAGE was described in section 2.2.3.3 (chapter 2).

5.2.5. Fast performance liquid chromatography (FPLC)

Size exclusion (SEC), ion exchange (IEX) and hydrophobic interaction (HIC) columns were used to separate IAPs. The detailed method was described in section 2.2.3.2 (chapter 2).

5.2.6. MWCO centrifugal device

The raw protein extract from *C. macra* was processed by 100 kDa MWCO centrifugal device (Pall Corporation). The retained fraction and the flow-through were visualized by gel electrophoresis

5.2.7. Leaf surface INP detection

Approximately 0.2 g of whole leaves were placed in 20 mL sonication buffer (20 mM ammonium carbonate pH 8.0) in a 50 mL Falcon tube with the leaf cut edges sealed with Vaseline and placed about 1 cm above the buffer. Leaves with the buffer were exposed to sonication at 20% amplitude for 2 min (0.5 s on/0.5 s off) and the liquid fraction was centrifuged at 16,000 × g, 4 ℃ for 20 min. Supernatant was concentrated to 2 mL by freeze-drying.

5.2.8. Mass spectrometry

Mass spectrometry was performed on the tryptic digest of the target fraction. The detailed method is described in section 2.2.4 (chapter 2).

5.3. Results

5.3.1. Ice binding

Two rounds of ice-finger purification were performed as the initial step to exclude proteins, and other substances that did not interact with the ice lattices during the

growth of the ice hemisphere. Different types of faceting were observed on the surface of the ice hemisphere of the *C. macra* and the *C. spiralis*, indicated ice-binding reactions, and whereas a smooth surface was observed in the control indicated no protein interaction with the ice lattices (Figure 5.1).

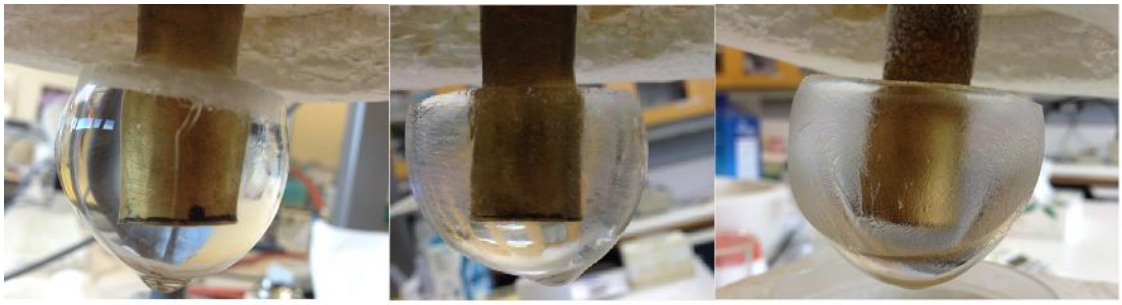


Figure 5.1 Ice-finger purification (2nd) of the control, *C. macra* and *C. spiralis* extract. Ice hemispheres were grown at the temperature gradient (-0.5 °C to -2 °C, over 16 h). A smooth ice surface was observed in the control. Irregular faceting was observed in the *C. macra* and the *C. spiralis*, with higher density in the *C. spiralis* extract.

Two rounds of ice-shell purification were also performed. Various shapes of ice crystals could be visualized in the inner surface of the hollow ice sphere from *C. macra* and *C. spiralis* indicating the interaction of ice-binding proteins and the ice lattices (Figure 5.2). As to the control, a comparatively non-transparent ice sphere was observed, which was different to those from *C. macra* and *C. spiralis* (Figure 5.2).

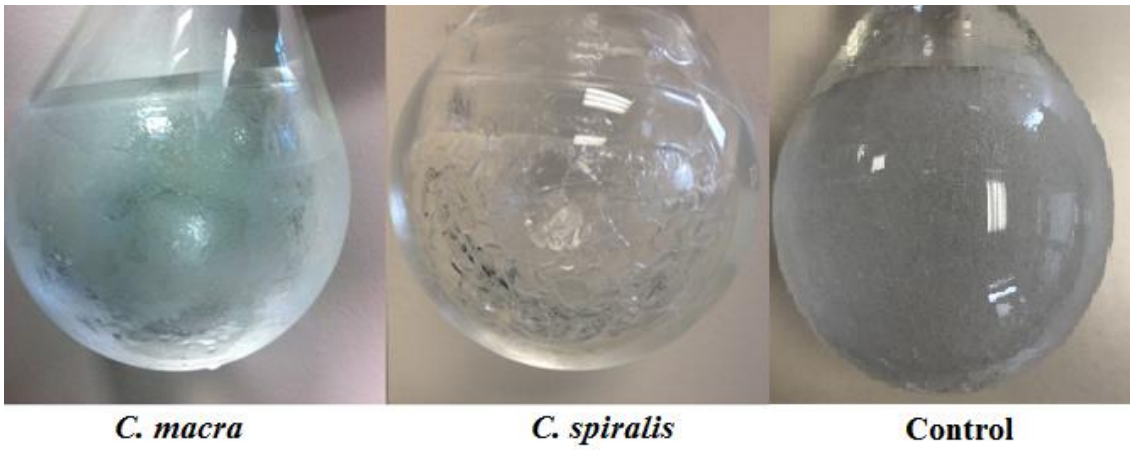


Figure 5.2 Ice-shell purification (2nd) from *C. macra* extract, *C. spiralis* extract and the control. Ice surfaces were grown at -0.5 °C for 2 h. Irregular shapes of ice were observed in both *C. macra* and *C. spiralis*.

5.3.1.1. Ice nucleation activity of the ice fraction from *C. macra*

The ice nucleation activity decreased to ~-6.5 °C in the 1st round ice affinity purified fraction whereas the INA stayed at the raw extract level (~-5.0 °C) in the 1st round

liquid fraction. There were no high levels of ice nucleation activity (less than -10 °C) in the 2nd round ice affinity purified fractions from the *C. macra* but the 2nd round liquid fraction still had INA of around -6.5 °C.

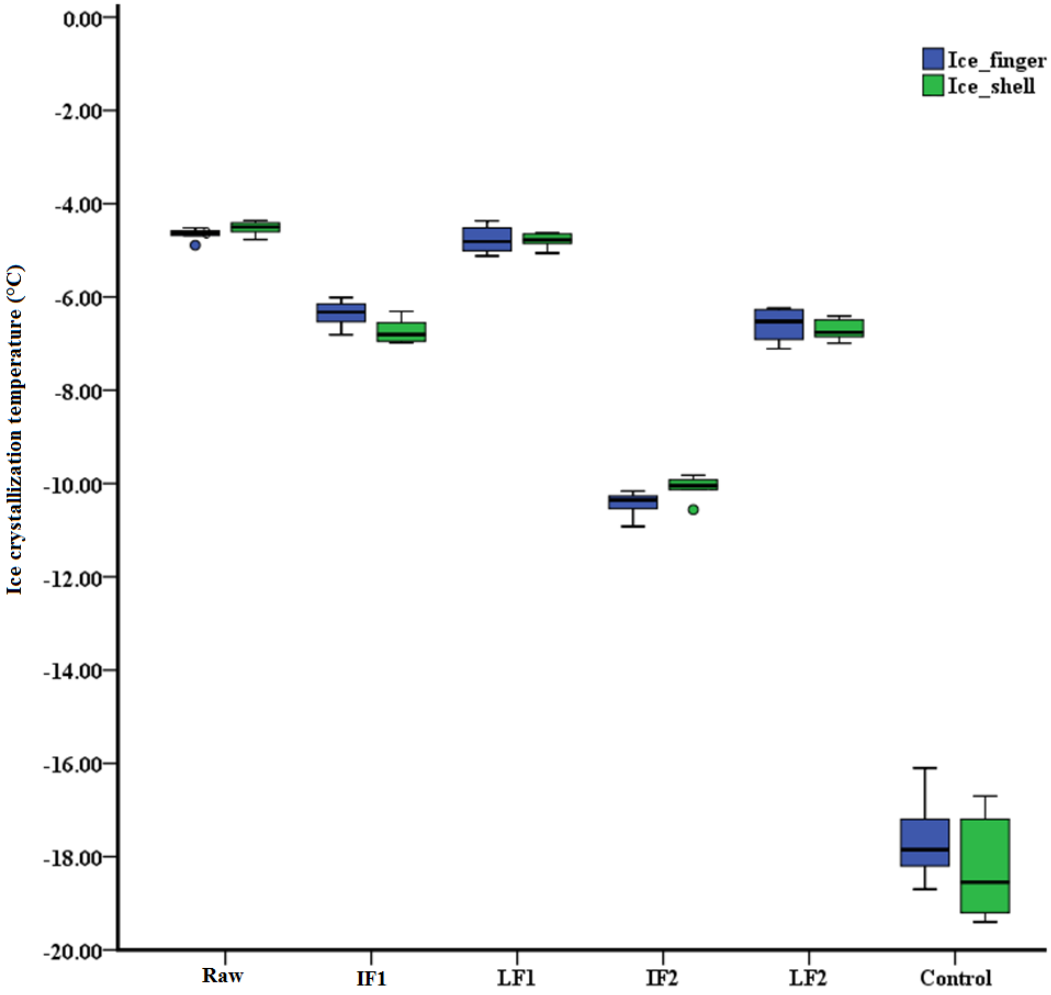
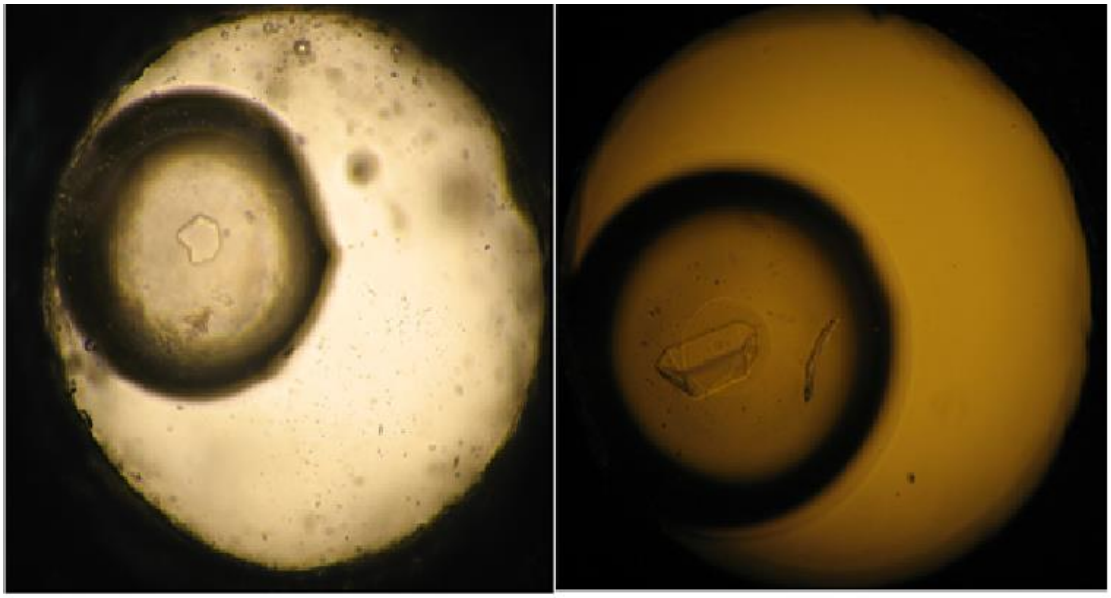


Figure 5.3 Ice nucleation activity of ice fractions (IF) and liquid fractions (LF) from the *C. macra* in two rounds of ice affinity purification by both ice-finger and ice-shell. Control, 25 mM ammonium bicarbonate buffer (pH 7.5). Middle dark line in each boxplot, mean T_c, n=3. T_c analyzed by DSC.

5.3.1.2. Ice-shaping ability of the ice fraction from *C. macra* and *C. spiralis*

In-regular shaped ice crystals were observed in the nanoliter osmometry of the raw extract, the 1st ice affinity purified fraction (IF1) and the 1st liquid fraction (LF1) from *C. macra* whereas only round-shaped ice crystals were found in the 2nd ice affinity purified fraction (IF2) and the 2nd liquid fraction (LF2) from *C. macra* in both ice-finger and ice-shell purification (Figure 5.4). Hexagonal shaped ice crystals were observed in the nanoliter osmometry of all fractions from *C. spiralis* in both ice-finger

purification and ice affinity purification (Figure 5.4). No obvious thermal hysteresis was found in all measurements.



In-regular shaped ice crystal in *C. macra* **Hexagonal-shaped ice crystal in *C. spiralis***

Figure 5.4 Ice shaping ability of fractions in the ice affinity purification from *C. macra* and *C. spiralis* in the nanoliter osmometry. Three observations were taken and the given images were representive. A non-typical shaped ice crystal was observed in the *C. macra* and a typical hexagonal shaped ice crystal was present in the *C. spiralis*.

5.3.1.3. Recrystallization inhibition activity of the ice fraction and the liquid fraction from *C. macra* and *C. spiralis*

RI activity was confirmed in the both the ice affinity fraction and liquid fraction from *C. spiralis* but the average ice crystal diameters were slightly larger than that of the raw extract, with ~20 µm in the raw extract, ~30-35 µm in IF, and ~35-40 µm in the LF (Figure 5.5). RI activity was low in all fractions from *C. macra* (Figure 5.6).

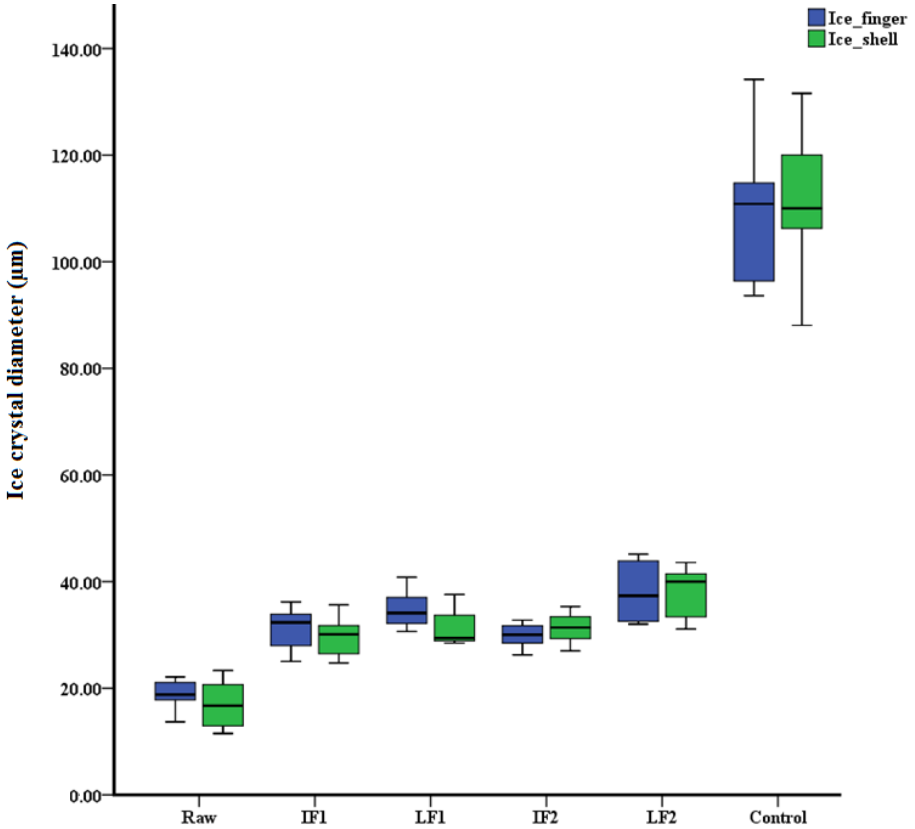


Figure 5.5 RI activity of ice fraction and liquid fraction in *C. spiralis*. Middle dark line in each boxplot, mean ice crystal diameter for 10 randomly selected largest ice crystals.

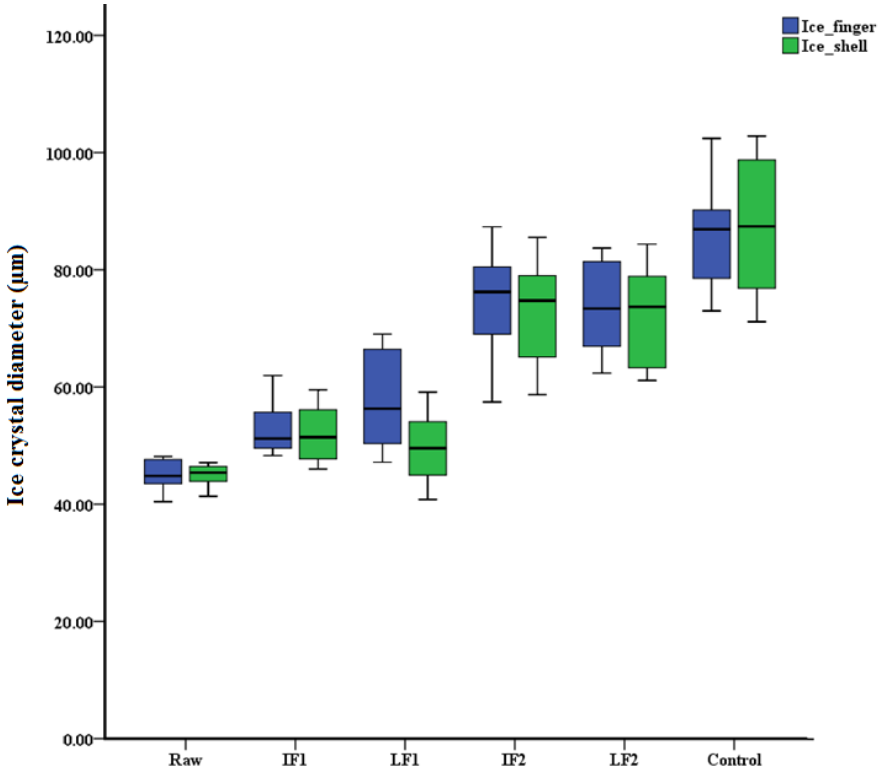


Figure 5.6 RI activity of ice fraction and liquid fraction in *C. macra*. Middle dark line in each boxplot, mean ice crystal diameter for 10 randomly selected largest ice crystals.

5.3.1.4. Resolving ice-binding proteins with gel electrophoresis

The ice fractions from *C. macra* (Figure 5.7) and *C. spiralis* (Figure 5.8) were visualized on 12% SDS-PAGE then stained with Colloidal Coomassie Blue stain. Comparatively clear protein bands ranged from ~18 kDa to ~150 kDa were present in the ice fractions from *C. macra* and *C. spiralis*.

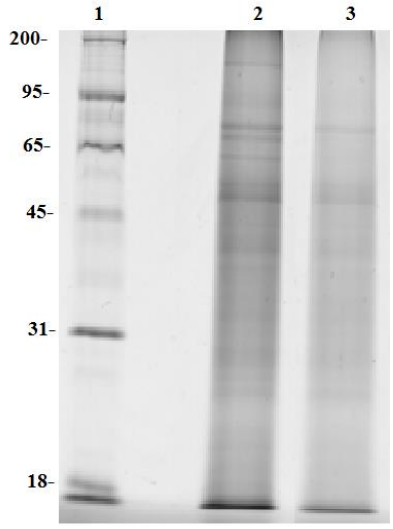


Figure 5.7 Gel electrophoresis of ice-shell purification of the *C. macra* raw extract. Lane 1, marker (kDa); Lane 2, 1st run ice fraction; Lane 3, 2nd run ice fraction;

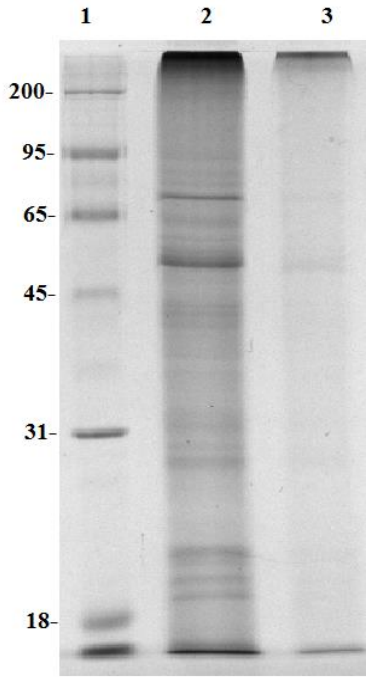


Figure 5.8 Gel electrophoresis of ice-shell purification of the *C. spiralis* raw extract. Lane 1, marker (kDa); Lane 2, 1st run ice fraction; Lane 3, 2nd run ice fraction.

5.3.1.5. Identification of proteins in the ice fraction from ice-shell purification

Proteins included in the ice lattice were digested with trypsin and assessed to MALDI-TOF mass spectrometry. Peptide fragments were searched against the currently known plant ice active proteins (IAP). Only the class II endochitinase antifreeze protein was identified in the ice-finger purification of *C. spiralis* with a peptide fragment of GPIQLTHNYNYGPAGR. Peptide fragments matched

recrystallization inhibition protein and class II endochitinase antifreeze protein in the ice-shell purification from both *C. macra* and *C. spiralis* (Table 5.1). The peptide fragment of AVHSIGVCNGVIGNNLPSPADVVK in the *C. macra* matched 1,3-ß glucanase. Meanwhile, Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) was found abundant in the ice fraction from both *C. macra* and *C. spiralis*.

Sample	Peptide fragment	Matched contig	Matched potential IAP
C. spiralis	ALAEQPNTIHGTNNAVRDGCNNAVSGNDNTVISGNNTTVSGSFNTIATGCYNTVSGSNQVVSGLNHIVTDNNN	c519574_g1_i1	ACN38299
	GLTGTIPGASLAGLAWLQSLNLANNRLVGTIPSFIGELDNLCYLDLSGNSLVGK	c90000_g1_i1	INC70254
	NQLPGSIPSSVGNLSELQYMTASLNSLSSNIPLSLWRLSK	c550777_g1_i1	ACN38300
	SLDSVDLSNNQLTGPVPSSFADLPNLR	c357883_g1_i1	ACG63783
	LQHNNLSGDIPAVLAALPSVTEVDLSWNALTGSVPPGFANCTTLETFDVSFNHLTPAGTSSGSPDPGEGSSPR	c359271_g5_i1	ACG63783
	LYGELPQSLSNLTYLADLDLSHNNLTGRIPSGGQLDTLYMQNPFMYDGNGGLCGPLLKKNCSNNSEPK	c370738_g1_i2	ANC70274
	YCQMLGVGLGDNVSCK	c319981_g1_i1	AAG53609
C. macra	VPPEVASLGALTRMELQHNGLTGSLPRMDAMRSLAHLDVSGNALSGSLLDAPER	c351587_g2_i1	ACG63783
	LSGEIPPELGEMTSMQSLDLSVNELGGDIPASL	c30486_g1_i2	ANC70259
	ELAAFFGQTSHETTGGTR	c352250_g2_i1	AAG53609
	APGYGVITNIINGGLECGHGADNR	c352250_g3_i1	AAG53609
	ATSPPYYGR	c352250_g2_i1	AAG53609
	AVHSIGVCNGVIGNNLPSPADVVK	c354194_g2_i1	AAP33176

Table 5.1 Peptide fragments of ice fraction (IF) from the ice-shell purification from *C. macra* and *C. spiralis* matched known IAPs. Generated peptide fragments were first searched against current known plant IAPs to identify potential IAPs in *Chionochloa*. Identified peptide fragments were then search against the transcriptome based datebase (chapter 6) to acheive corresponding contigs. Contigs and proteins with the highest sequence similarities to known IAPs were selected.

5.3.2. MWCO centrifugal device

The 100 kDa MWCO centrifugal device (Pall Corporation) was applied to the *C. macra* raw extract with high INA (no less than -5 °C). The retained fraction had high INA whereas there was no high INA (no more than -10 °C) in the flow-through. SDS-PAGE of retained and the flow through fractions showed different patterns. The unclear background of the retained fraction may due to the enrichment of interference from the plant extract by the 100 kDa centrifugal device. In the retained fraction, protein bands with molecular weights higher than ~50 kDa were clearly visualized and protein bands with molecular weights ranged from ~35 kDa to 150 kDa were enriched compared with that in the flow-through.

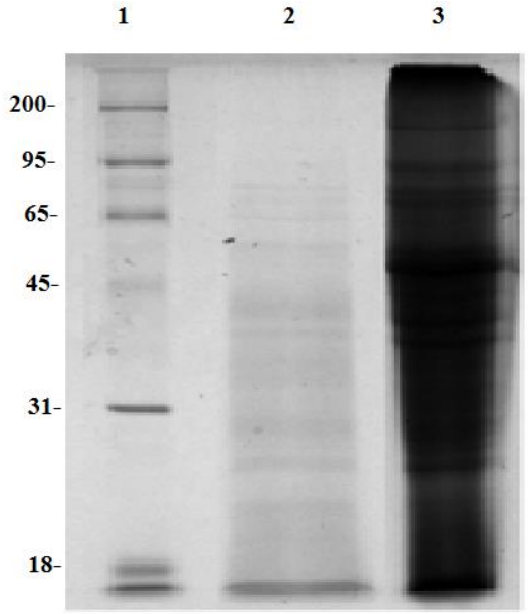


Figure 5.9 SDS-PAGE of the retained fraction and the flow-through by 100 kDa WMCO centrifugal device of *C. macra*. Lane 1, marker (kDa); Lane 2, flow through; Lane 3, retained fraction.

5.3.3. Fast performance liquid chromatography purification

The retained fraction of *C. macra* in section 5.3.2 was applied to a size exclusion column (Superdex 75, GE Healthcare) and INA was found in fraction 4-6 (Figure 5.10). The fractions 4-6 after the size exclusion column were concentrated with a 3

kDa MWCO spin-device and fractionated by ion exchange chromotography (Resource Q, GE Healthcare) for further purification (Figure 5.11). Ice nucleation activity was found in fractions 33-44 after dialysis. The hydrophobic interaction column (HiTrap Phenyl HP, GE Healthcare) was also tested with an aliquot of concentrated fractions 4-6 after size exclusion column. INA was present in the flow-through before gradient elution (Figure 5.12).

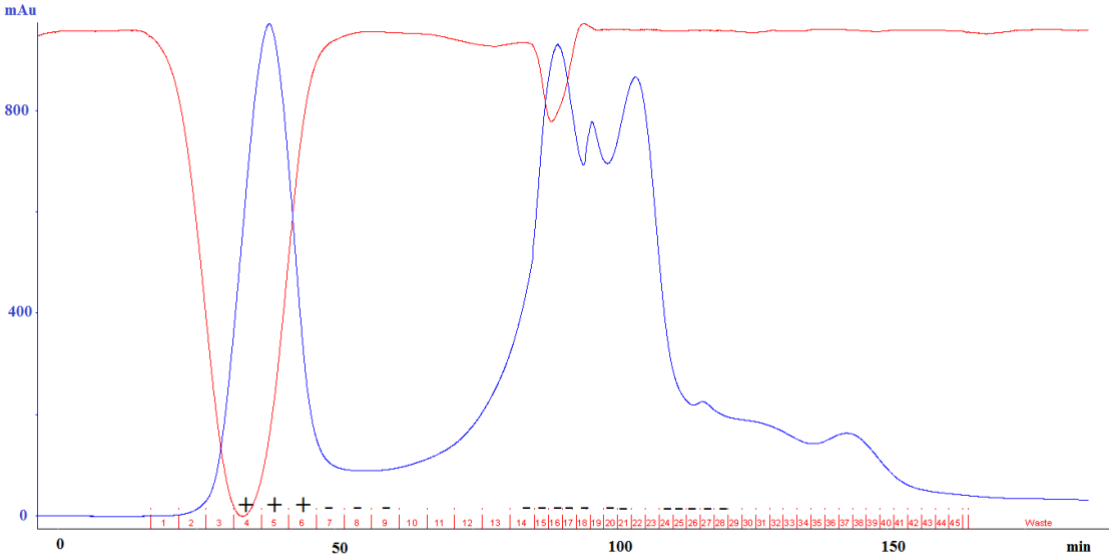


Figure 5.10 Size exclusion purification of *C. macra* retained fraction of 100 kDa MWCO centrifugal device by Superdex 75. INA (~-4.5 °C) was only found in fraction 4-6. +, fraction with INA; -, fraction with low or no INA.

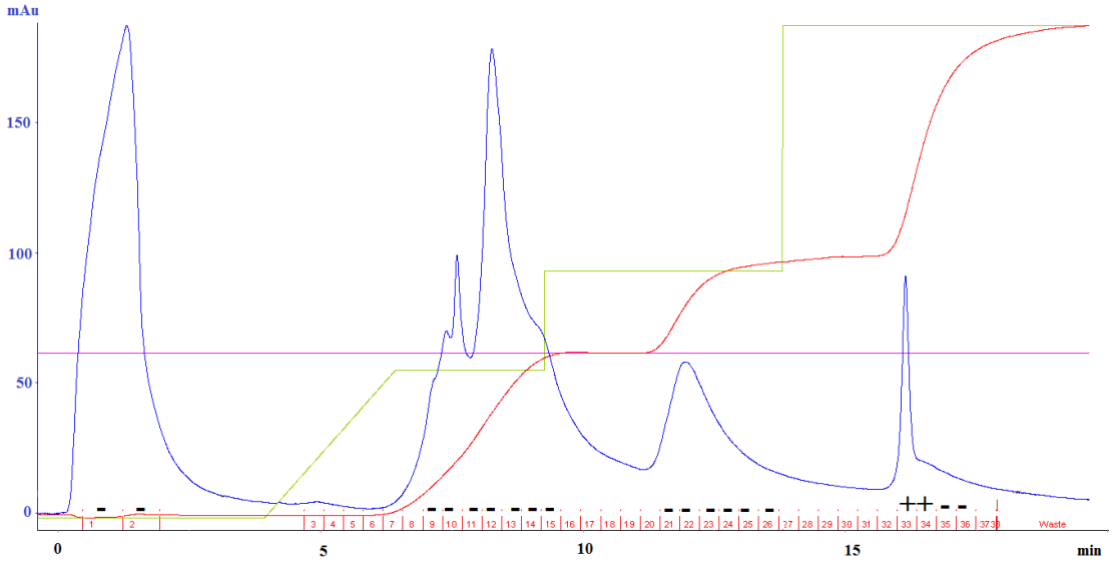


Figure 5.11 Ion exchange purification (Resource Q) of *C. macra* fractions 4-6 from size exclusion purification. Green line, step elution with 1 M NaCl elution buffer. Red line, ion conductivity. Blue, UV absorbance (mAu). +, fraction with INA; -, fraction with low or no INA.

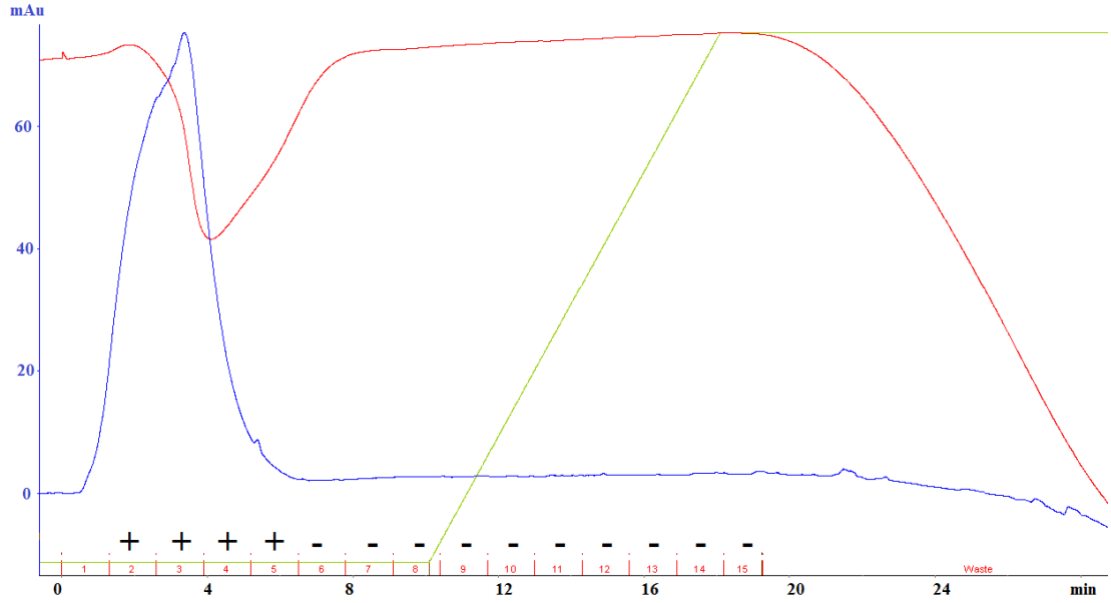


Figure 5.12 Hydrophobic interaction purification of *C. macra* fractions 4-6 from size exclusion purification. +, fraction with INA; -, fraction with low or no INA.

5.3.3.1. SDS-PAGE of INA positive fraction

Fraction 11 to 13, fraction 19 to 21 and fraction 33-34 in the ion exchange column (Resource Q) (section 5.3.3) were run on 1.2% SDS-PAGE then stained with Colloidal Coomassie Blue stain (Figure 5.13). Multiple protein bands (~28 kDa to 90 kDa) were found in fractions 11-13, 19-21 and 33-34, with a ~45 kDa band highly enriched in the fractions 11-13.

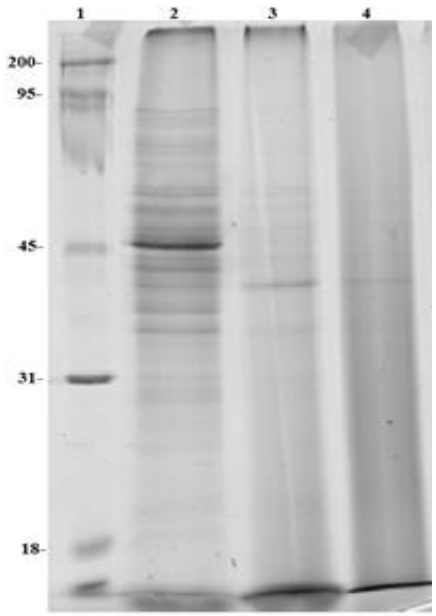


Figure 5.13 Gel electrophoresis of fractions in the ion exchange purification by SDS-PAGE. Lane 1, PAGEmark^{TM} unstained marker (kDa); lane 2, fraction 11-13; lane 3, fraction 19-21; lane 4, fraction 33-34.

5.3.3.2. Mass spectrometry

INA-positive fraction 33-34 from the ion exchange purification was assessed by MALDI-TOF mass spectrometry. Four peptide fragments (MIQIDFNK, SGETITFK, FDYPGLADK and RTQDGGTEVVEAK) matched four proteins respectively by searching against SwissProt: phosphoribosyl-AMP cyclohydrolase (HIS3_PELPD), plastocyanin (PLAS_ORYSI), serine/threonine-protein kinase TEL1 (ATM_CANAL) and malate dehydrogenase (MDHM_CAPAA). Four peptide fragments (KSGETITFKN, RAQSCPGLNLSNDLLSLKL, TQDGGTEVVEAK and RDISLNGMLEKH) matched four contigs from the *C. macra* transcriptome respectively (chapter 6): c367889_g1_i1 btb poz domain-containing protein at2g30600, c363414_g3_i1 curved dna-binding protein, c323939_g2_i1 malate dehydrogenase and c317154_g1_i2 ternary complex factor mip1-like.

5.3.4. Leaf surface proteins separation

The buffer in which moderate sonication was performed on the *C. macra* leaves had a high level of INA (Figure 5.14). Proteins in this buffer were concentrated by freeze-drying and resolved by gel electrophoresis. Two bands were detected in the range from 45 kDa to 65 kDa in the SDS-PAGE but mass spectrometry failed to identify any protein (Figure 5.14).

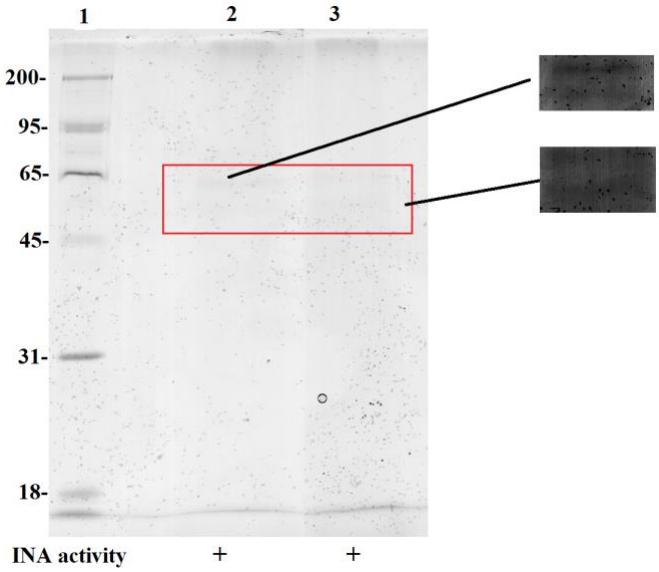


Figure 5.14 Identification of proteins from the *C. macra* leaf sonication. A, SDS-PAGE of proteins from the *C. macra* leaf sonication. Lane 1, PAGEmark™ unstained marker (10 µL), kDa; Lane 2 and 3, *C. macra* leaf surface proteins (15 µL). Only the contrast and the color of the original gel image (left) has been adjusted for better visualization of indicated bands (Right).

5.4. Discussion

Ice activities were confirmed in many *Chionochloa* species and protein components (ice active proteins) were essential in maintaining all three types of ice activities (chapter 3). The protein extraction methods for plants are different to those for animals and insects (Laing and Christeller 2004). Purification from TCA/acetone is the most common protein extraction method for plants, which includes suspending pulverized plant powder in cold acetone with 10% trichloroacetic acid to precipitate protein (with numerous downstream modifications) (Chatterjee *et al.* 2012; Cilia *et al.*

2009; Sheoran *et al.* 2009). However, TCA/acetone has strong effects on protein three-dimensional structure and the activity of ice active proteins might change, e.g. ice nucleation proteins (section 3.4, chapter 3). In this chapter, a liquid nitrogen based extraction method was used to maximize the ice activities in the raw protein extract.

There are not many methods to purify natural ice active proteins from the raw protein extraction. The most common method is based on the affinity between ice crystals and proteins (Kuiper *et al.* 2003). The most commonly used method is ice-finger purification whereas a higher efficiency ice-shell purification method was recently introduced (Marshall *et al.* 2016). The result in this chapter illustrated that ice-shell purification performed better than ice-finger purification in both *C. macra* and *C. spiralis*, as IAPs were identified in ice-shell purification other than ice-finger purification.

Not all proteins that have ice activities can be purified by ice affinity, e.g. in this study INPs from the *C. macra* extract. INA dropped from ~-4.8 °C in the raw extract to ~-6.0 °C in the 1st ice fraction (IF1), and to ~-10 °C in the 2nd ice fraction (IF2) from both ice-finger and ice-shell purification. Further dilution tests indicated that INA in the liquid fraction was higher than the ice fraction, suggesting a weak ice-binding ability of the substance responsible for INA or that other substances such as metal ion were essential in maintaining INA but these substances were not included in the ice fraction (Paschke 2017).

Thermal hysteresis was low in both the ice and liquid fractions from *C. spiralis* unless further concentrated, which may be a result of low concentrations of IAPs in the this species (Drori *et al.* 2015; Paschke 2017). Ice-shaping activity was observed in concentrated ice and liquid fractions from both *C. macra* and *C. spiralis*, with comparatively high-density faceting observed in the ice hemisphere from ice-shell purification than in ice-finger purification. Recrystallization inhibition activity was slightly reduced in the ice fraction and liquid fraction of *C. spiralis* compared with that of the raw extract but still significantly higher than the buffer control. One

hypothesis is that other substances such as metal ions are necessary to maintain the ice activity and these are excluded from the ice lattice when the IAPs are absorbed in the ice (Gilbert *et al.* 2005; Paschke 2017).

Fast performance liquid chromatography (FPLC) can be used to separate proteins according to the molecular size, ionic properties and hydrophobic interactions. Previous studies indicated INA in lichens was retained by a 100 kDa MWCO column but passed a 300 kDa MWCO column (Dreischmeier *et al.* 2017; Murray *et al.* 2015; Pummer *et al.* 2015). The combination of different FPLC columns and MWCO columns may help to get comparatively pure INPs. IAPs in *Chionochloa* failed to be purified by hydrophobic interaction column (Figure 5.12). Thus, a combination of ion exchange column and size exclusion column is ideal for separating IAPs in the *Chionochloa* species. Dialysis is needed to remove high salt levels and recover the protein function after high salt elution from the ion exchange column.

Class II endochitinase antifreeze protein, recrystallization inhibition protein and 1,3-ß glucanase were identified in *Chionochloa* with the ice-shell purification which might suggest a higher efficiency of ice-shell purification than the traditional ice-finger purification. In this chapter, the failure to detect peptide fragments in the INA positive fraction may due to interference of substances in the plant extract which have negative effect on tryptic digestion. RNA sequencing covered the entire profile of mRNAs in the plants and with the data of peptide fragments, related sequences could be blasted in the transcriptome to get the full genes of the target proteins.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), which is an essential enzyme in the initial step of photosynthetic CO_2 fixation, is found abundantly in the chloroplast of many plants. Rubisco comprises about 30% to 50% of plant total proteins, which reduced the resolution of 2DE and hindered the study of plant low abundant proteins. In this study, two rounds of ice-shell affinity purification failed to remove Rubisco suggested that Rubisco might interact with the ice lattices during the growth of the hollow ice hemisphere.

Yu *et al*. (2010) found the Rubisco large chain (~5 kDa) was present in the tryptic digestion of a single band (~120 kDa) on a SDS-PAGE gel. This band contained the AFP from the leaves of *Ammopiptanthus nanus* which lead to high thermal hysteresis and hexagonal shaped ice crystals (Yu *et al.* 2010). Lim *et al*. (2013) found that RuBisco was involved in the gelation of winter wheat leaf protein extracts and the removal of RuBisco led to no gel formation. Interestingly, abundant winter flounder AFP I in solution can also form a similar gel under freezing to thaw cycles (Lim *et al.* 2013). Thus, the ice activities and any antifreeze ability of RuBisco need to be further investigated. Rubisco removal might improve the resolution of either 2DE or mass spectrometry and methods have been developed including RuBisco IgY affinity, PEG fractionation, DTT and Ca^{2+}/phytate treatments (Sehrawat *et al*. 2013).

To conclude, a class II endochitinase antifreeze protein and recrystallization inhibition protein were identified in the ice fraction (IF) of ice-shell purification from both *C. macra* and *C. spiralis* whereas high levels of INA in *C. macra* could not be harvested after two rounds of ice-shell purification. This INA in *C. macra* remained after the FPLC purification procedures, and was presented in the high salt elution fraction in the ion exchange purification. Unfortunately, mass spectrometry failed to identify any potential INPs, which may due to either not many peptide fragments being present or there were compounds present that interfered with the tryptic digestion.

Chapter 6 De novo transcriptome assembly by RNA sequencing revealed featured biological features in the *Chionochloa macra*

6.1. Introduction

The genus, *Chionochloa* (tussocks), was fully described under the Tribe *Dantbonieae* from the subfamily *Arnndinoideae*, family *Poaceae* (Zotov 1963; Zotov 1970). This Australasian grass genus currently contains 25 species with 23 endemic to New Zealand. *Chionochloa* species are all interfertile, readily form hybrids and result in many subspecies and hybridized forms (Pirie *et al.* 2010). *Chionochloa* species are distributed from lowlands to alpine areas in both the North Island and South Island of New Zealand, with a greater taxon richness in the latter (Pirie *et al.* 2010).

More than 75% of the land in New Zealand is above an altitude of 200 m, with 19 mountain peaks higher than 3,000 m and permanent snow over 2,000 m (Wharton 2011). In the South Island, the Alpine Fault lays along nearly the entire island and in the North Island, the alpine area mainly lays between the Hikurangi Range and the Tararua Range, as well as Mount Taranaki and the Volcanic Plateau (Wilson 2007). Alpine species of *Chionochloa*, snow tussocks, dominate the majority of alpine grasslands, especially in the South Island (Chapin III *et al.* 1982).

Cold weather arrives in New Zealand at any time of the year, especially when the cold Antarctic Air from the south flows through the east of the South Island to the west because of the blocking anticyclones near the west of the country (Brenstrum 1998; Sturman *et al.* 1999). As a consequence, freezing and snow occur frequently in these alpine areas and alpine plants maintain high resistance to freezing to survive (Bannister *et al.* 2005; Mark *et al.* 2000). Freezing tolerance of snow tussocks has not been well characterized. Only *C. rigida* and *C. rubra* were reported to tolerate freezing below -20 ℃ in winter (Bannister *et al.* 2005; Reitsma 1994). A recent study by Wharton *et al*. (2010) found a high endogenous ice nucleation activity in *C. macra* and *C. rigida* (Wharton *et al.* 2010). The research presented in chapter 3 of this study

showed that ice nucleation activity existed in several species of *Chionochloa* with seasonal changes, except for the *C. macra*, which maintained high ice nucleation activity throughout the year. Moreover, a weak non-typical hexagonal shaped ice crystal was observed in the nanoliter osmometry of the winter *C. macra* extract (Chapter 3).

Most studies of *Chionochloa* have focused on its biology and ecology. Some of these investigated morphological differences in *Chionochloa* and concluded that a influence of both genetic and environmental variations resulted in the differences including various flowering times and growth periods (Mark 1965b). Later, stature and growth rates in two species, *C. rigida* and *C. macra*, were reported to be under strictly genetic control as species differences of these characters remained after reciprocal transplanting in places with various altitudes, and it was confirmed that flowering was influenced more by environmental factors such as the temperature (Greer 1979). Periodic flowering features was found in the *C. macra*, in which non-flowering seasons occurred after two or three flowering seasons (Campbell 1981). Another part of the research focused on adaptions to various soil fertilities. Many *Chionochloa* species followed soil fertility gradient or phosphate gradient distributions (Chapin III *et al.* 1982). Some *Chionochloa* species adapt to low fertility soil by reducing their growth rate to avoid nutrition shortage, others such as *C. rigida* and *C. rubra*, tolerate various soil types (Lee and Fenner 1989).

Variations in chemical composition among different species were another important field of interest in this genus. Minerals, nitrogen and carbohydrate composition varied in different *Chionochloa* species with the *C. macra* containing a higher concentration of phosphorus and silicon (Connor *et al.* 1970). Triterpene methyl ethers, a group of leaf waxes only found in the *Poaceae* or *Gramineae*, were identified in many *Chionochloa* species but not in the *C. macra* (Russell *et al.* 1976). A single gene was inferred controlled the synthesis of the leaf wax based on evidence from natural hybrids (Connor and Purdie 1981).

Rich carbohydrate content was another feature in *Chionochloa* species. A high level of polysaccharides (hemicellulose and cellulose) was reported in the *Chionochloa* species including the *C. macra* (Bailey and Connor 1972). Later, non-structural carbohydrate was found at a high level in the *C. macra*, with sucrose to be the main component, and small amount of starch, glucose, and fructose were also present in the *C. macra* (Payton and Brasch 1978). A cryoprotective function for these carbohydrates was suggested to protect tissues from winter freezing (Bannister and Ward 1981). A PhD thesis by Matthew Aaron Krna (2015) also discovered richer carbon sequestration in *Chionochloa* species with lower temperatures and increasing altitudes (Krna 2015).

Dispersal played an important role in the biography of southern hemisphere biota (Sanmartin and Ronquist 2004). Dispersal through Antarctica (Trans-Antarctica dispersal) was the main pattern in animal distribution between Australia and southern South America during the long period of connection of these three continents. Trans-Tasman dispersal was the main explanation of plant distribution between Australia and New Zealand (Sanmartin and Ronquist 2004). Modern New Zealand originated from the Zealandia, which was formed by its separation from the southern supercontinent Gondwana around 82 million year ago (Mya). Most part of Zealandia was gradually submerged under the sea during around 60 to 24 Mya, except for a few low altitude islands (Wallis and Trewick 2009). Mountain formation in New Zealand started at around 25 Mya because of the collision of the Indo-Australian and Pacific tectonic plates. Alpine habitat was formed starting at around 5 Mya with the elevation of the Alpine Fault. The origin of the New Zealand biota has been supposed to be either by descent from ancestors since the isolation of Zealandia or from Australian origins by long distance dispersal but Pire *et al.* (2010) suggested that *Chionochloa* did not fit either pattern (Pirie *et al.* 2010). Their phylogeny inferred from chloroplast DNA of *Chionochloa* species indicated that the single Australian species of

Chionochloa fell into the New Zealand *Chionochloa* clade and suggested a New Zealand origin of the genus.

Chionochloa, which arrived around 17-20 Mya but did not extend its diversity until the formation of the New Zealand high mountains, is considered to be one of the most important plants in the New Zealand alpine grassland habitat with their ability to conserve soil and water in this area, where is cold and relatively dry (Mark 1965a; Tanentzap *et al.* 2015). Alpine *Chionochloa*, the snow tussock, dated back to around only 7.6 Mya and did not have much diversity. *Chionochloa macra*, the slim snow tussock, was separated from its close relative, *C. rigida*, the narrow-leaved snow tussock, and was found mainly in the South Island of New Zealand (Chapin III *et al.* 1982; Zotov 1970). Dawson (1989) reported in the first chromosome study of *Chionochloa* that species in this genus had six sets of chromosomes, 2n=6x=42 (Dawson 1989; Murray *et al.* 2005). Since then, the genetics of *Chionochloa* species has not been well studied.

Past genetic studies of *Chionochloa* were mainly based on mitochondrial and/or rDNA markers (Barker *et al.* 2007; Hsiao *et al.* 1998; Pirie *et al.* 2010; Pirie *et al.* 2008). Phylogeny of *Chionochloa* has been explored by several studies to identify a Southern South Island origin of this genus (Barker *et al.* 2003; McCall *et al.* 2004; Pirie *et al.* 2010). Trees inferred by Pire *et al.* (2010) indicated the dispersal of ancestor to other parts of the South Island and to the North Island through the northwest South Island (Pirie *et al.* 2010). None of the *Chionochloa* species have yet been fully sequenced. There are only a few nucleotide, amino acid sequences and a full plastid genome from *C. macra* deposited in the NCBI GeneBank. The lack of genomic data hinders studies exploring the molecular genetics of key biological characters.

Transcriptome sequencing plays an important role in providing extensively useful gene resources for the study of non-model organisms. A well-characterized transcriptome provides a useful draft map of protein profiles and annotations about

which pathways are active under which conditions. In this study, Total RNA sequencing along with annotations of genes provides protein profiles and species specific metabolic pathways in the *C. macra*, and threw light on the water reservation and cold tolerance mechanism of this species.

6.2. Materials and methods

6.2.1. Sample collection

Leaves from four individuals of *C. macra* (labeled as macra1, macra2, macra3 and macra4) were collected at the Dunedin Botanic Garden on 8th July 2014, cut into 2 cm pieces with dry ice precooled autoclaved scissors, stored in separate clean sealed plastic sample bags and placed on dry ice immediately. Samples were transferred to an -80 °C freezer within 30 min and subsequently examined for various types of ice activities for each *C. macra* individual to be sequenced.

6.2.2. Total RNA extraction, total RNA sequencing and transcriptome analysis

The presence of large amount of polysaccharides in *C. macra* made RNA extraction and purification difficult. Commercial kits (Qiagen Rneasy plant mini kit, ISOLATE II RNA plant kit) both failed as columns were blocked. A modified Invitrogen plant RNA reagent - Qiagen Rneasy plant mini kit combined method was used instead (Supplementary experiment procedure for chapter 6). Four Illumina TruSeq stranded total RNA libraries were prepared following the Illumina TruSeq protocol, sequenced in a single lane (2×100 bp PE) on an Illumina HiSeq 2000 sequencer at New Zealand Genomics Limited (NZGL) (Supplementary experiment procedure for chapter 6).

A pool of ice-binding proteins was created for the further research on ice activities in the *Chionochloa*. Genetic explanations of several biological features in earlier studies of *C. macra* were given by investigating metabolic pathways of silicon, phosphorus, polysaccharide and leaf wax product.

A standard transcriptome analysis was performed and repeat elements, SNPs and alternative splicing events were studied in the *C. macra* for the purpose of better understanding the molecular genetics of this species (Supplementary experiment procedure).

6.3. Results

6.3.1. The *C. macra* transcriptome

Raw reads for each library was deposited in NCBI Sequence Read Archive (SRA) with accession numbers SRR3105314, SRR3105316, SRR3105317 and SRR3105318. Q20 reads from the four libraries were merged, which took 96.27% of the raw cluster. The trimmed pair-end Q20 reads and the trimmed singletons were processed to make a de novo assembly.

The Trinity assembler (v2.2.0) produced a total number of 901,869 contigs (Trinity gene isoforms) in 632,094 Trinity genes with an N50 length of 674, and this was defined as the raw *C. macra* transcriptome. The filter criteria of minimum FPKM 0.5 and minimum isoform 1.0% constrained the Trinity assembly to 297,695 contigs (Trinity gene isoforms) in 205,924 Trinity genes, with an N50 length of 1009. This *C. macra* transcriptome was deposited in NCBI Transcriptome Shotgun Assembly (TSA) database with an accession number GFMB00000000. There were 52.9% of contigs gave at least one Blastx/Blastn hit against the nr, UniProt, TrEMBL and nt database (e-value<10^{-5}). Blast2GO assigned GO annotations to 36.9% contigs in the *C. macra* transcriptome. Detail analysis of the *C. macra* transcriptome was described in supplementary information for chapter 6.

6.3.2. Featured KEGG pathways in the *C. macra* transcriptome

6.3.2.1. Metabolic pathways in the *C. macra* transcriptome related to biological features

6.3.2.1.1. Sugar metabolism

Starch and sucrose metabolism is the one of the KEGG pathways in the *C. macra* transcriptome that contained the highest number of enzymes (47) and also one of the pathways that contained the largest number of contigs (2,043). Seven contigs encoding eight enzymes had FPKM value larger than 50 in this pathway (Table 6.1). There were 720 contigs encoding 40 enzymes assigned to amino sugar and nucleotide sugar metabolism, with 13 contigs encoding 6 enzymes had FPKM value higher than 20 (Table 6.1).

KEGG pathway	FPKM range	Enzyme (EC)	Contig	Enzyme name
Starch and sucrose metabolism	>50	3.2.1.4	c295268_g1_i1	1,4-β-endoglucanases
		2.4.1.34	c350979_g1_i1	1,3-β-glucan synthase
		2.4.1.14	c363836_g1_i1	Sucrose-phosphate synthase
		2.4.1.13	c350240_g1_i1	Sucrose synthase
		2.4.1.11	c350991_g2_i8	Glycogen(starch) synthase
		2.4.1.10	c330227_g1_i1	Levansucrase
		3.6.1.21	c364118_g2_i1	ADP-sugar diphosphatase
		2.4.1.21	c350991_g2_i8	ADP-glucose synthase
Amino sugar and nucleotide sugar metabolism	>20	4.1.1.35	c336106_g4_i1	UDP-glucuronate decarboxylase
		5.4.99.30	c363296_g2_i1, c363296_g2_i7, c363296_g2_i5, c363296_g2_i4, c363296_g2_i3	UDP-arabinopyranose mutase
		3.2.1.55	c364279_g1_i3	Arabinofuranosidase
		2.7.7.27	c354520_g1_i2, c360332_g1_i1, c354520_g1_i3, c347923_g1_i3	ADP glucose pyrophosphorylase
		5.1.3.2	c347178_g2_i3	UDP-glucose 4-epimerase
		5.1.3.18	c346859_g1_i1	GDP-mannose 3,5-epimerase

Table 6.1 Highly expressed KEGG enzymes in sugar metabolisms in the *C. macra* transcriptome.

6.3.2.1.2. Triterpenoid metabolism

There were two important metabolic pathways involved in the biosynthesis of triterpenoid products in higher plants: the mevalonate (MVA) pathway in the cytoplasm and the methylerythritol phosphate pathway (MEP) pathway in the plastid (Vranová *et al*. 2013). Both pathways lead to two essential intermediates, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) catalyzed by the isopentenyl-diphosphate Delta-isomerase (EC: 5.3.3.2) (Henneman *et al*. 2008). IPP or DMAPP were then converted to geranyl-diphosphate (GPP) by geranyl-diphosphate synthase (EC: 2.5.5.1), which was then converted to (E, E)-Farnesyl diphosphate by farnesyl diphosphate synthase (EC: 2 .5.1.10) (Henneman *et al*. 2008). The (E, E)-Farnesyl diphosphate was later passed to sesquiterpenoid and triterpenoid biosynthesis, steroid biosynthesis, carotenoid biosynthesis and N-Glycan biosynthesis. The (E, E)-Farnesyl diphosphate could also be converted to geranylgeranyl diphosphate (GGPP) by geranylgeranyl diphosphate synthase (EC: 2.5.1.29), which were important in diterpenoid biosynthesis, carotenoid biosynthesis, indole diterpene alkaloid biosynthesis and ubiquinone and other terpenoid-quinone biosynthesis.

In the *C. macra* transcriptome, a comparatively complete terpenoid backbone biosynthesis pathway was identified (Figure 6.1). It was hypothesized that ß-amyrin and cycloartenol were synthesized in the *C. macra* as related enzymes were present in the transcriptome (Figure 6.1). The 3-hydroxy-3-methylglutaryl coenzyme A reductase (EC: 1.1.1.34), a rate limiting enzyme in the MVA pathway, and the 1-deoxy-D-xylulose 5-phosphate synthase (EC: 2.2.1.7) and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC: 1.1.1.267), two essential enzymes in the MEP pathway, were both found active in the *C. macra* transcriptome (Figure 6.1). Some of the contigs encoded isopentenyl-diphosphate Delta-isomerase (EC: 5.3.3.2) and the ß-amyrin synthase (EC: 5.4.99.39), were highly expressed (FPKM >20) in the terpenoid backbone biosynthesis (Table 6.2). Only the farnesyl diphosphatase (EC: 3.1.7.6) was absent in the main terpenoid backbone biosynthesis pathway in the *C. macra*

transcriptome whereas the downstream farnesol dehydrogenase (NADP+) (EC: 1.1.1.216) was present with a FPKM of 1.87.

Enzyme (EC)	Contig	FPKM
5.3.3.2	c328073_g1_i1	41.5
5.4.99.39	c343629_g1_i1	35.0
5.4.99.39	c333544_g2_i1	24.4
5.3.3.2	c328073_g2_i2	22.7
5.3.3.2	c328073_g2_i1	20.7
5.4.99.39	c279919_g1_i1	20.2
5.4.99.39	c333544_g4_i1	19.9
1.17.7.1	c356708_g1_i2	19.5
1.17.7.4	c356708_g1_i2	19.5
2.5.1.1	c358003_g3_i1	19.5
2.5.1.10	c358003_g3_i1	19.5

Table 6.1 Highly expressed contigs (FPKM >20) in the terpenoid backbone biosynthesis in the *C. macra* transcriptome.

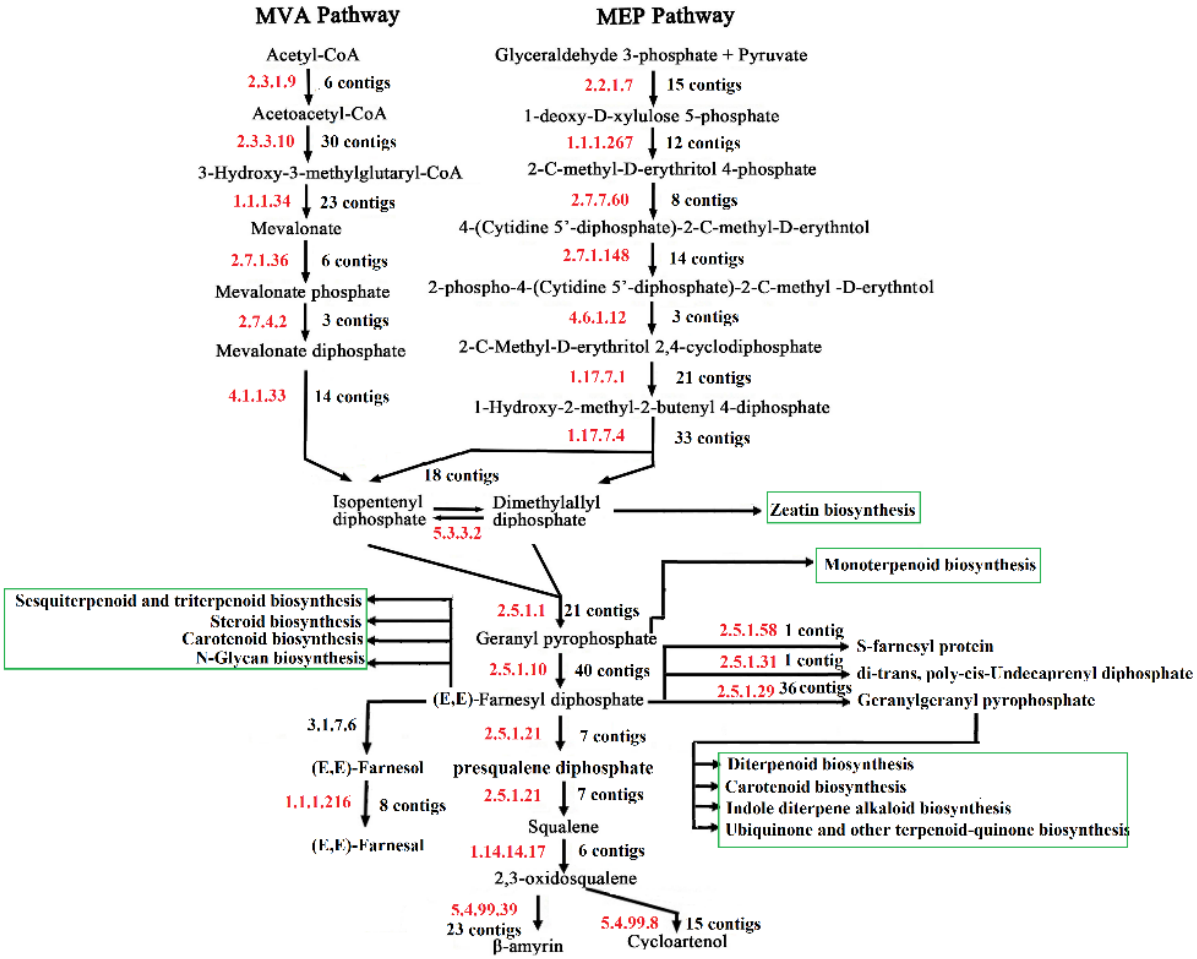


Figure 6.1 Putative triterpenoids biosynthesis pathways in the *C. macra* transcriptome. Numbers in red, KEGG enzymes that participated in the pathways. Color in green, downstream pathways.

6.3.2.1.3. Silicon and phosphorus metabolisms

Four groups of aquaporin were identified in the *C. macra* transcriptome including nodulin 26-like intrinsic protein (NIP), plasma membrane intrinsic proteins (PIP), tonoplast intrinsic proteins (TIP) and small basic intrinsic protein (SIP) as well as uncharacterized intrinsic proteins (XIP) (Lopez *et al*. 2013). There were 76 contigs gave Blastx hits to aquaporin proteins and 3 contigs were from nodulin 26-like intrinsic proteins-2 (NIP2), a subclass of aquaporin, which is involved in silicon absorption. The gene, c358634, encoding aquaporin NIP2 with a GSGR silicon filters and two asparagine-proline-alanine (NPA) motifs (108 amino acid length distance between the two NPA motifs) (Deshmukh *et al*. 2015).

Phosphate transporter proteins including glycerol-3-phosphate transporter, inorganic phosphate transporter, low-affinity inorganic phosphate transporter, phol phosphate transporter and sodium-dependent phosphate transporter were identified in 72 contigs from this transcriptome. BHLH transcription factor PTF1, which enhanced phosphorus accumulation, was found in seven contigs (c359053_g4_i1, c359053_g4_i5, c220120_g1_i1, c307998_g1_i2, c370579_g1_i1, c346134_g1_i1 and c346134_g1_i3) from this transcriptome. Phosphate translocators, sodium phosphate symporter, phytase and acid phosphatase (phosphohydrolase) were also reported in this transcriptome.

6.3.2.1.4. Ice active proteins and ice-binding proteins mapping

There was no protein in the raw transcriptome that had a significant match to the bacterial INPs (data not shown). The contig c90000_g1_i1 contained the whole length ORF that matched the *Lolium perenne* ice recrystallization inhibition protein-like protein (IRI1) gene (EU680848.1), with a FPKM value of 69.37. The contig c286849_g3_i3 and c82162_g1_i1 matched the *Lolium perenne* ice recrystallization inhibition protein-like protein (IRI3) gene (EU680850.1). These RIPs showed

similarities with RIPs identified in species including *Triticum aestivum* (AAX81542.1) and *Deschampsia antarctica* (ACN38302.1). This group of RIPs contained N-terminal secretion signal sequences, leucine-rich repeat domains (LRR) and ice-binding repeats (NxVxxG and NxVxG) in the C-terminal, which were common features of recrystallization inhibition protein in *Pooideae* grasses. The FPKM values of these two contigs were 17.9 and 7.4 respectively.

The contigs c295787_g1_i1 and c276124_g1_i1 match the antifreeze protein (XM_008650641). The contigs c259538_g1_1 and c359349_g1_i1 match the antifreeze protein (XM_003567816). The c352250_g2_i7 and c352250_g2_i4 gave Blastx hits to the *Secale cereale* 24.8 kDa class II endochitinase-antifreeze protein precursor (AF280438) and 31.7 kDa class I endochitinase-antifreeze protein (AF280437), with FPKM values of 5.6 and 2.4 respectively.

Peptide fragments from tryptic digestion of proteins in the *C. macra* ice fraction from ice-shell purification were identified with hits matched proteins from this *C. macra* transcriptome. There were 18 contigs encoding enzymes mapped to KEGG pathways with carbon fixation in photosynthetic organisms containing the most enzymes. There were nine contigs that had FPKM more than 20 (Table 6.3) whereas 19 contigs had FPKM less than 20 (Table 6.4).

Contig name	FPKM	Blastx description
c345230_g1_i1	124.9	Lactoylglutathione lyase
c90000_g1_i1	69.4	Recrystallization inhibition protein
c341086_g2_i2	60.1	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase
c361431_g1_i2	54.7	Cathepsin b-like isoform x1
c327888_g3_i1	46.3	5-methyltetrahydropteroyltriglutamate-homocysteine
c328408_g1_i2	42.6	Polyubiquitin 2
c336695_g1_i1	42.5	5-methyltetrahydropteroyltriglutamate-homocysteine
c369693_g1_i1	33.1	Ferredoxin-dependent glutamate chloroplastic
c368637_g2_i2	27.2	5-methyltetrahydropteroyltriglutamate-homocysteine

Table 6.2 Abundant ice-binding proteins in the *C. macra* protein extract. Contigs with FPKM>20 were shown.

Contig	Blastx description
c369335_g1_i1	Phosphoglycolate phosphatase chloroplastic-like
c368245_g3_i1	Tpa: exhydrolase ii
c367589_g2_i2	β-d-xylosidase 4-like
c362157_g2_i3	Superoxide dismutase
c361554_g1_i1	Peroxidase 15-like
c360024_g1_i2	Helicase-like protein
c359794_g2_i1	2-cys peroxiredoxin bas1
c359742_g2_i1	Ribose-5-phosphate isomerase
c354621_g1_i2	Sedoheptulose-chloroplastic
c350973_g8_i2	Chitinase 2-like
c350076_g1_i3	Aldo-keto reductase-like
c346615_g1_i1	Probable β-d-xylosidase 6
c339856_g1_i1	Peroxidase 17-like
c336970_g1_i2	β-d-xylosidase 7
c332544_g1_i3	O-glycosyl hydrolase superfamily protein
c323939_g2_i1	Malate dehydrogenase
c318336_g1_i2	Ubiquitin-nedd8-like protein rub2
c352250_g2_i7	Class II endochitinase protein
c24970_g1_i1	Ubiquitin-60s ribosomal protein l40

Table 6.3 Ice-binding proteins from the *C. macra* extract. Proteins were identified by mapping mass spectrometry data to the *C. macra* transcriptome. Contigs with 0.5<FPKM<20 were shown.

6.4. Discussion

Chionochloa species are New Zealand endemic plants and play an essential role in grassland and its water resource reservation (Mugambi 1971). The tussock groups are also important in studying the evolution of the New Zealand flora but genetic markers required have not been well investigated (Pirie *et al.* 2010). Prior to this study, there were only 309 nucleotide sequences and 349 protein sequences of *Chionochloa* deposited in NCBI GeneBank (11/04/2016), most of which were from mitochondria and ribosomes. The low cost and high coverage of next generation sequencing provides a comparatively complete transcriptome for developing genetic markers and analyzing the expression of genes (Duarte *et al.* 2014). This study provided the first transcriptome of *C. macra*, and the first transcriptome in the *Chionochloa* genus and *Danthonioideae* subfamily, and included a comprehensive characterization of *C. macra* transcripts. The Hiseq2000 platform generated approximately 171 million pairs of Q20 reads and finally proceeded to 297,695 Trinity contigs of high quality with an acceptable N50 value of 1009, compared with published transcriptomes of other plants (Lang *et al.* 2015). As rRNAs were not removed during the library construction process, 33.4% of the reads were remapped to rRNAs by the SAMtool whereas completeness analysis by CEGMA showed that the coverage of the *C. macra* transcriptome was still high. There were 46.0% of the *C. macra* transcriptome gaving at least one Blastx hit against nr database and 36.9% of the *C. macra* transcriptome were assigned GO annotations by Blast2GO. The remaining transcriptome may include genes of novel functions, which are not identified across known plant species. The unannotated transcriptome was assessed to contain comparatively high level of retroelements and DNA transposons by RepeatMasker with Repbase database.

Chionochloa macra collected in winter at Dunedin Botanic Garden had a high level of INA and weak ice-shaping ability in its protein extract (chapter 3). Previous studies showed that this INA was caused by plant endogenous proteins other than

bacterial INPs (Wharton *et al.* 2010). RNAs encoding bacterial ice active proteins (IAP) were not identified with significant Blastx hit in this transcriptome suggesting either there was no high level of bacterial INA on *C. macra* leaves or the library preparation method could not enrich RNAs for bacteria INPs. In chapter 3, a weak hexagonal shaped ice crystal was found from the *C. macra* extract in the nanoliter osmometry and ice active proteins (IAP) identified in this transcriptome might contribute.

Different types of putative ice active proteins were identified in the *C. macra* transcriptome. Comparatively high expression of putative recrystallization inhibition protein, which showed high similarity to known plant RIPs, may indicate similar RI functions. Putative antifreeze proteins, which showed high amino acid similarity to antifreeze proteins from *Secale cereale*, were also identified. The contigs, c341067_g1_i1, c341067_g1_i2, c341067_g1_i3, c366446_g2_i2 and c317088_g1_i1 did not show Blastx hits to known plant ice active proteins but they were assigned with GO terms of ice binding (GO: 0050825), which indicated the encoding proteins interact selectively and non-covalently with ice crystals.

There were 28 proteins identified in the ice fraction from two rounds of ice affinity purification of the *C. macra* extract, which suggested that these proteins were included in the ice lattices during the growth of ice hemisphere. A literature search indicated no ice-binding activity reported in these proteins from other plant species. There were 18 of the 28 proteins annotated to enzymes mapped to 20 KEGG pathways. The top three significantly enriched GO terms of the ice-binding proteins were 5-methyltetrahydropteroyltri-L-glutamate-dependent methyltransferase activity, 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity and S-methyltransferase activity whereas there is no study showing a close relationship of these activities with ice-binding ability.

There were 25,338 orthologous clusters generated based on the predicted protein sequences from the transcriptomes of *Chionochloa macra*, *Eragrostis tef*, *Oryza*

sativa Japonica, Seteria italica, Sorghum bicolor and *Brachypodium distachyon*. The 8,504 shared orthologous clusters indicated conserved protein families and similar biological functions among these species. Compared with the other five species, the specific clusters in the *C. macra* were larger and may indicate evolution and novel functions in the Biological process, Molecular function and Cellular component GO catalogs since its speciation. The 2,616 single copy gene cluster shared by the six species may suggest that those genes maintained a single copy during their evolution, which is quite useful in phylogenetic research (Müller *et al.* 2015; Simão *et al.* 2015).

Ethylene-responsive transcription factor-related families were comparatively highly expressed in the *C. macra* transcriptome. These group transcription factors are induced by high salt, extreme temperature and drought stress and play a function in disease resistance (Müller and Munné-Bosch 2015). It is reported that the ERF1 is sensitive to salt and enhanced expression of ERF leads to reduced water loss from the leaves by an increasing level of proline, an osmolyte which is essential in drought stress (Mawlong *et al.* 2015). The expression of dehydrin, which was synthesized in response to cold and drought stress, was quite high with FPKM >50 (Chiappetta *et al.* 2015). The NAC, AP2, MYB, bZIP, and zinc-finger families, which are important in the regulation of plant defense and stress responses, are all present in this *C. macra* transcriptome (An *et al.* 2015).

KEGG pathways related to biological features in the *C. macra* were mined in this study. Extremely high FPKM value of cytochrome P450 monooxygenase in the *C. macra* transcriptome suggested the abundance of P450 monooxygenase in the *C. macra*. Cytochrome P450 monooxygenases, one of the largest gene families in the most plant species, are essential in plant growth, development and response to stress (Tan *et al.* 2015; Zhang *et al.* 2011). P450 monooxygenase, the function of which is to oxidize hydrophobic compounds, have been reported to be involved in the majority of secondary metabolite biosynthesis pathways including terpenoids, alkaloids and phenylpropanoids.

Cell walls of *Poaceae* grasses are rich in mixed-linked (1,3; 1,4)-β-D-glucans, and the production of polysaccharide synthase and glycosyltransferases are involved in the synthesis of glucans (Lockhart 2015). Genes encoding 1,3-β-D-glucan synase were identified in this transcriptome but genes encoding 1,4-β-D-glucan synthase were not found. There were four contigs encoding three enzymes of callose synthase with FPKM no less than 20. Enzymes in the glycosyl transferase family (EC: 2.4.x.y), UDP-glycosyltransferase family (EC: 2.4.1.x) and glycoside hydrolase family (EC: 3.2.1.x), which are essential for the synthesis of polysaccharides, were highly expressed (FPKM>50) in this transcriptome. It was confirmed that the UDP-glucose dehydrogenase (EC: 1.1.1.22), which was essential for cell wall polysaccharide synthesis, is found in the *C. macra* leaves (Gao *et al.* 2015). These data suggested that sugar metabolism was active in the winter *C. macra* and there were comparatively high-levels of expression of several essential enzymes involved in the starch and sucrose metabolism. High levels of glucose, fructose, and sucrose have been found in the grasses *Melinis minutiflora* and *Echinolaena inflexa* from the *Poaceae*, and related enzymes were also highly expressed in the *C. macra* (Souza *et al.* 2010; Souza *et al.* 2005). These studies showed that high levels of non-structural carbohydrates played a function in the cold tolerance mechanisms of these plants.

Triterpene methyl ethers are leaf wax components in *Poaceae* grasses, but not abundant in other plant families (Russell *et al.* 1976). The presence of triterpene methyl ethers in *Chionochloa* species had been well studied. Not all the *Chionochloa* species have been reported to synthesize triterpene methyl ethers and triterpene methyl ethers in different *Chionochloa* species were different (Russell *et al.* 1976). Arundoin and miliacin were common in many *Chionochloa* species whereas other triterpene methyl ethers were less common and were used as chemotaxonomic markers (Connor and Purdie 1981). ß-amyrin methyl ethers were only found in the *C. rigida* but there was no triterpene methyl ethers reported in the *C. macra*. Russell *et al*. (1976) inferred that methyl ether was only produced by genes controlling

methylation during the synthesis of triterpenes, as no other sterol methyl ethers were found in the *Chionochloa* species. The synthesis of triterpene methyl ethers was proved to follow Mendelian inheritance with dominance and was directed by independent genes (Connor and Purdie 1976; 1981; Orians 2000; Russell *et al.* 1976). Interestingly, genes involved in the synthesis of cycloartenol and ß-amyrin, including ß-amyrin synthase and cycloartenol synthases, which are essential for the synthesis of leaf wax cycloartenol methyl ether and ß-amyrin methyl ether, were moderately abundant in the *C. macra* transcriptome. Meanwhile, nearly complete cycloartenol synthesis and ß-amyrin synthesis pathways were identified downstream a complete triterpenoids backbone synthesis pathway in the *C. macra* transcriptome. The function of ß-amyrin is not fully studied and it is an intermediate for more complex triterpenes. Studies showed that the accumulation of ß-amyrin in the cuticular wax lead to a reduction in the efficiency of water transport barrier of the intracuticular wax and other triterpenoids may act in a similar role (Jetter and Riederer 1996; 2016). Why plant species produce triterpenoids is still not fully understood and two hypothesises were given: either the water transport barrier in the cuticle structure was not compromised with the presence of triterpenoids, or the accumulation of triterpenoids played a balanced role to the water transport barrier which means it maintained the cuticular structure intergrity (Jetter and Riederer 1996; Jetter and Riederer 2016).

Silicon is important in plant health under various biotic and abiotic stresses and in toxic metal reduction (Sahebi *et al.* 2015). Generally, monocots can absorb comparatively high level of silicon (>1.0% dry weight) (Deshmukh *et al.* 2015). The intake of silicon is dependent on silicon influx transporters. Plant nodulin 26-like proteins (NIP), which belong to a subfamily of aquaporins and contain a NPA repeat motif and a selective motif that is specific in silicon absorption, are important in silicon accumulation. Silicon is absorbed in the form of silicic acid [$Si(OH)_4$] and stored in leaf cells in the forms of amorphous silica (Deshmukh *et al.* 2015). There are different sub-classes of NIPs, each of which performs different transport functions.

NIP1 is related to H_2O and glycerol transport and NIP3 is mainly involved with large molecules whereas NIP2 transports various metalloids. There are two types of silicon transporters (NIP2-1, NIP2-2) in the *C. macra* transcriptome, which shared the common features that 108 amino acids were found between two NPA motifs and a silicon-selective GSGR motif was present. It is reported that each plant species has at least one putative silicon transporter and contains at least 1.0% silicon from the total dry weight, whereas the *C. macra* in the wild has 2.0% of silicon dry weight (Connor *et al.* 1970; Deshmukh *et al.* 2015).

Phosphorus, which is a key component of many essential compounds including nucleic acids, phospholipids and ATP, is important in plants' growth (Versaw and Harrison 2002). More than 90% of the phosphorus that plants need comes from the soil and the accumulation of phosphorus in plants requires highly efficient phosphorus transporters (Li *et al.* 2011a; Versaw and Harrison 2002). The identification of several high affinity phosphorus transporters, which enhance the transport and intake of phosphorus may account for the relatively high concentration of phosphorus in the *C. macra*. Moreover, studies showed that the enhanced expression of efficient phosphorus transcription factors such as those belonging to MYB, bHLH and WRKY families increased the dry weight of phosphorus in several plants. The bHLH transcription factor PTF1 identified in the *C. macra* transcriptome was reported to improve phosphorus dry weight by 30% in transformed *Oryza sativa* than in the wild type (Yi *et al.* 2005). It is also reported that maize with the transformation of ZmPTF1 showed better adaption in phosphorus deficient culture media (Li *et al.* 2011b).

Genetic studies of *C. macra* are limited by there are not being many sequences identified, especially for related molecular markers. Simple sequence repeats (SSRs), short tandem repeats with a repeat unit size up to 6 bp, are useful markers in the population study of close species, including genetic mapping, genetic diversity analysis, DNA fingerprinting, and marker-assisted selection (Gamar *et al.* 2013). SSRs were present in ~15% of the *C. macra* transcriptome, which was higher than

wheat (~8%) and sweet potato (~8%) but lower than tea plant (~17%) (Wang *et al.* 2017). The *C. macra* transcriptome had a SSR frequency (repeat size no less than 2) of one in 5.0 kb, which was higher than sweet potato (1/7.1 kb) and *Arabidopsis* (1/6.0 kb) but lower than those in tea (1/2.4 kb) and radish (1/3.5 kb) (Zhou *et al.* 2016). Mono nucleotide unit repeat and four nucleotide unit repeat were the majority of SSRs in this study, with percentages of 25.6% and 29.5% respectively. You *et al.* (2015) suggested that trinucleotide CCG/CGG repeats were common in monocot but rare in dicotyledonous plants, and CCG/CGG repeats in this transcriptome were not abundant, with a proportion of 5.7% (You *et al.* 2015). High throughput sequencing of transcriptomes enabled full SNP discovery in the coding genes in many species (Yu *et al.* 2014). The *C. macra* transcriptome showed high levels of SNPs, with a total number of 709,367 SNPs and a variant rate of 1 in 139 bases on average. The miRNAs are short non-coding RNAs that regulate gene expression by binding to mRNAs which lead to the degradation of mRNAs or translation blocking (Hadjiargyrou *et al.* 2016). The contig c470538_g1_i1 (miRNA MIR5079B), which matched the sequences of gene c364052 that coded the dehydrogenase (EC: 1.6.99.3) in oxidative phosphorylation, might suggest a regulatory role of miRNA (MIR5079B) to the enzyme (EC: 1.6.99.3) in the *C. macra* transcriptome. Other miRNA-like sequences (FPKM>10) did not have any annotations and may indicate novel regulative functions.

Alternative splicing (AS), by which various mature mRNAs are formed from a single precursor mRNA and result in high protein complexity, is quite common in eukaryotes (Marquez *et al.* 2015). Up to 95% of human genes, 48% of rice genes and more than 61% of *Arabidopsis* genes underwent splicing events (Marquez *et al.* 2015). AS is important in plant growth and the response to various biotic and abiotic stresses including flowering, seeding and pathogen resistance (Zhang *et al.* 2017). Less than 15% of genes or gene fragments in the *C. macra* transcriptome underwent alternative splicing events, which was lower than expected. Pfam domain analysis of

AS contigs indicated that leucine rich repeats (LRR), which were common in plant-pathogen interactions, and pentatricopeptide repeats (PPR), which were comparatively abundance in proteins regulating gene expression and RNA processing (editing, splicing and translation), had a high frequency of alternative splicing in the *C. macra* transcriptome (Manna 2015; Shanmugam 2005). GO enrichment analysis of the AS genes against gene sets in the entire transcriptome as background suggested that most GO terms belonged to the Biological process GO catalog. GO terms related to isopentenyl diphosphate synthesis, which is essential in the terpenoid backbone biosynthesis was enriched. GO terms of polysaccharide biosynthesis and RNA splicing were also enriched in this catalog. The enrichment of isomerase activity and phosphatase activity in the molecular function GO catalog suggests significant AS events in these enzymes.

In the present study, the protein expression profile of *C. macra* was present and the first transcriptome of *C. macra* was annotated. Genetic explanations of several biological features in earlier studies of *C. macra* were given by investigating metabolic pathways of silicon, phosphorus, polysaccharide and triterpene products. Repeat elements, SNPs and alternative splicing events were studied in the *C. macra* for the purpose of better understanding the molecular genetics of this species. A pool of ice-binding proteins was created for the further research on ice activities in the *Chionochloa*.

Chapter 7 Expression of ice active proteins and phylogenetic study of cold response-related genes from the *C. macra* transcriptome

7.1. Introduction

7.1.1. Plant ice active proteins

Three types of ice activities, ice nucleation activity (INA), thermal hysteresis (TH) and recrystallization inhibition activity (RI), have been reported in many plant species whereas only endogenous proteins responsible for TH and/or RI activity have been identified and well characterized (Griffith and Yaish 2004; John *et al.* 2009; Tremblay *et al.* 2005; Wharton *et al.* 2010). Generally, TH values of plant thermal hysteresis proteins (THP) are low (~0.1 °C to 0.5 °C), compared with THPs in fishes and insects (Barrett 2001). Low levels of TH values were first reported in plant species including bittersweet nightshade (*Solanum dulcamara* L.), narrow leaved plantain (*Plantago lanceolata* L.) and wood aster (*Aster cordifolius* L.) and it was indicated that there were similar epitopes between plant and insect THPs (Urrutia *et al.* 1992). Later, the thermal hysteresis glyco-protein (THGP) in the bittersweet nightshade, was purified, which was 67 kDa in size with galactose carbohydrate components and high glycine components (Duman 1994). Other plant THPs were reported to be pathogenesis-related proteins including endochitinase-like proteins, endo-β-1,3-gluconase, thaumatin-like protein and polygalacturonase inhibitor proteins (Griffith and Yaish 2004; Hon *et al.* 1995).

As the ability to depress freezing points is low, the main strategy to overcome freezing in plants is the inhibition of ice recrystallization (Tremblay *et al.* 2005). Leucine rich repeats (LRR) were reported in many recrystallization inhibition proteins (RIP) including carrot, wheat, Antarctic hair grass and winter ryegrass, which showed weak TH values and high resistance to heat (Sidebottom *et al.* 2000; Zhang *et al.* 2004). RIPs were only well characterized in the *Pooideae* grass. LRR repeats in the N-terminus, which were similar to receptor-like kinases, and ice-binding repeats in

the C-terminus, which showed no sequence similarity to any other proteins, were identified in many RIPs (John *et al.* 2009). These RIPs may have originated from phytosulfokine LRR receptor kinase around 36 Mya (John *et al.* 2009).

Endochitinase-like proteins, thaumatin-like protein, polygalacturonase inhibitor proteins and phytosulfokine LRR receptor kinase-like RIPs were identified in the *C. macra* transcriptome (Table S1). The ice-shell purification approach (chapter 5) identified class II endochitinase-like proteins and RIP in the ice fraction of both the *C. macra* and the *C. spiralis* extracts, which indicated ice-binding ability of these proteins to the ice lattice. Whether these proteins identified in the *C. macra* transcriptome or ice fraction had ice activities or not was unknown.

7.1.2. Expression of ice active proteins in *E. coli*

Further study of the chemical and physical characters of IAPs requires large amounts of purified proteins. Protein isolation from plant tissues is costly, cumbersome and lengthy, as large amount of plant leaves are needed (Rosano and Ceccarelli 2014; Yesilirmak and Sayers 2009). The expression of identified genes with affinity tags in the *E. coli* system has been a convenient way to produce high levels of pure proteins (Yesilirmak and Sayers 2009).

The most commonly used tags are His-tag, which is cheap and can be absorbed to a metal ion on the resin such as Nickle and Cobalt, and GST-tag, which is captured by the reduced glutathione (GSH) on the GST resin (Franken *et al.* 2000; Lee *et al.* 2011). One major issue in the recombinant protein purification is that the target protein sometimes is expressed in the form of inclusion bodies (Rosano and Ceccarelli 2014). Another issue is that the His-tag often is not exposed and protein expressed fails to bind to the metal resin (Debeljak *et al.* 2006). Expression of both N-terminal and C-terminal His-tag may increase the chances of exposed His-tag. The other method is to use denatured reagents such as 8 M urea or 6 M Guanidine hydrochloride to denature proteins and expose the hidden His-tag (Wingfield 2002). After

purification, a refolding process is needed to regain the correct bonds and structures for protein activities. Drawbacks are that hidden multiple histidines in the proteins from the *E. coli* system are also exposed, which causes contamination (Robichon *et al.* 2011). GST-tag is also frequently used as it folds rapidly into high soluble protein during translation, which increases the solubility of recombinant proteins and promotes the expression of recombinant proteins (Costa *et al.* 2014).

In this study, expression of IRI2 and EAF2 both failed in the pET-28a(+) system with various strains of *E. coli* and different IPTG concentrations. Then, both IRI2 and EAF2 were inserted in the pGEX-6P-3 system and both genes were expressed.

7.1.3. Phylogenetic analysis of cold-related genes

The subfamily of *Pooideae* includes the majority of grass species and its evolutionary history has been well studied since the early 1990s (McKeown *et al.* 2016; Soreng *et al.* 1990). The most recent common ancestor (MRCA) of *Pooideae* was adapted to a warm climate whereas most species from the *Pooideae* are adapted to cold environments and are associated with the synthesis of ice active proteins (IAP) (McKeown *et al.* 2016). This makes these grasses ideal models to study adaption to climate change (Li *et al.* 2012a; Sandve *et al.* 2008). Cold adaption in *Pooideae* likely evolved after the split with *Brachypodium distachyon* at approximately 35 Mya (Wu and Ge 2012). Li *et al.* (2012) established phylogenies with three cold-related genes, recrystallization inhibition protein (RIP), C-repeat binding factor (CBF) and fructosyltransferase (FST), from *B. distachyon* and core *Pooideae* species (Li *et al.* 2012a). They proved that *B. distachyon* evolved differently to the core *Pooideae* group with regards to cold stress adaption. The genus *Chionochloa*, within the *Pooideae*, has evolved significant ice activities in several of its species however the evolutionary history of cold response-related genes has not been studied in this genus. In this study, the phylogenetic relationships of a recrystallization inhibition protein, a class II endochitinase antifreeze protein, a C-repeat binding factor and a

fructosyltransferase from the *C. macra* transcriptome were inferred and compared with those in Li *et al.* (2012).

7.1.4. Aims

This study aims to confirm ice activities of putative IAP genes identified in the *C. macra* transcriptome. In particular, the phylogenetic relationships of these genes together with *Pooideae* subfamily-specific cold response-related genes were inferred to study the evolutionary history of cold stress adaptation in *C. macra*.

7.2. Methods

7.2.1. Putative IAP gene identification

Open reading frames (ORF) of plant IAPs were downloaded from the NCBI GeneBank and then searched against the *C. macra* transcriptome in CLC main workbench 6 (Qiagen). Those contigs having less than 60% sequence similarities to the corresponding plant IAPs and with e-value higher than e^{-5} were not selected.

7.2.2. Gene synthesis

The open reading frame (ORF) of c90000_g1_i (named as IRI2) and c352250_g2_i5 (named as EAF2) were synthesized by gBlock Gene Fragments (Integrated DNA Technologies), with the restriction enzyme site *BamHI* and *NdeI* at the 5' end and *XhoI* at the 3' end. The nucleotide sequences were optimized according to the degeneracy of the genetic code to amino acids synthesized in the *E. coli* system to reduce the proportion of repeat nucleotide sequences and rare codon usages (Figure S5-S6).

7.2.3. Preparation of recombinant expression vectors

Plasmid vectors, pET-28a(+) and pGEX-6P-3, were extracted from glycerol stock *E. cloni* 10G-pET-28a(+) and *E. cloni* 10G-pGEX-6P-3 with the Roche plasmid

miniprep kit (Roche) respectively (Figure 7.1). Both gBlock gene fragments (~500 ng) and the pET-28a(+) vector (~1,000 ng) were double digested with *NdeI* and *XhoI* (New England BioLabs) following the manufacturer's protocol (Figure 7.2). Another aliquot of gBlock gene fragments (~500 ng) and pGEX-6P-3 vector (~1,000 ng) were double digested with *BamHI* and *XhoI* (Figure 7.2). Shrimp alkaline phosphatase (New England BioLabs) was only added to the digestion of the plasmid vector to dephosphorylate the 5' phosphate group of the linear vector. The digested gBlock gene fragments were purified by Isolate II PCR and Gel purification kit (Bioline) and the digested vector was processed by gel purification to harvest the linear vector with sticky ends.

NdeI-XhoI or *BamHI-XhoI* digested vector and *NdeI-XhoI* or *BamHI-XhoI* digested gBlock gene fragments were subjected to overnight ligation at 16 ℃ with T4 DNA ligase (Roche) (molar ratio of 1:4) in a 30 µL system respectively (Figure 7.2). T4 DNA ligase in the ligation reaction was then inactivated at 65 ℃ for 10 min.

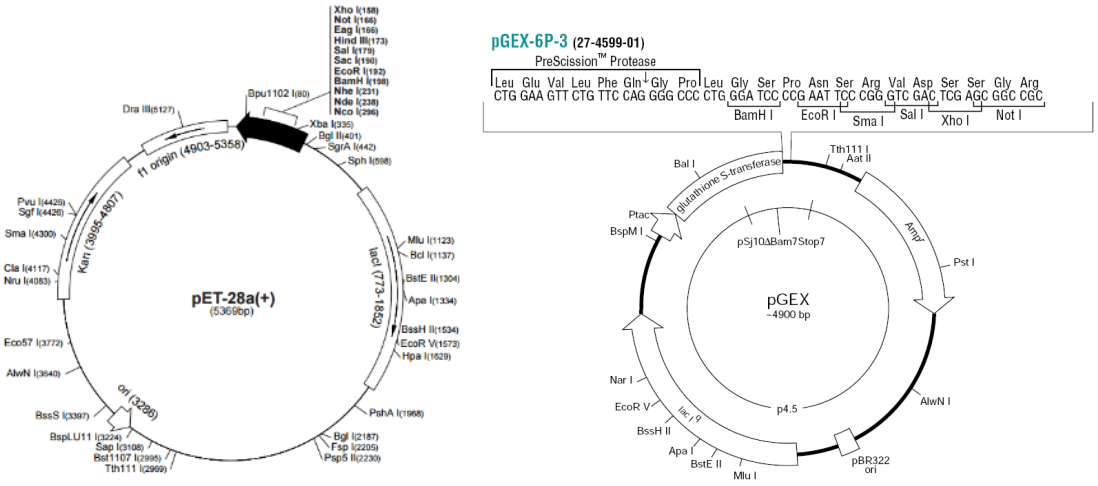


Figure 7.1 pET-28a(+) vector (Novagen) and pGEX-6P-3 vector (GE Healthcare). The antibiotic resistances were Kanamycin in pET-28a(+) and Ampicillin in pGEX-6P-3.

Figure 7.2 Transformation of putative IAP genes in pET-28a(+) and pGEX-6P-3 vector. Transformation was done by electroporation. Recombinant plasmids were selected on agar plates with corresponding antibiotic resistance.

7.2.4. Preparation of competent *E. coli*

Five strains of *E. coli* (E. cloni 10G, BL21(DE3)GOLD, C41(DE3)pLysS, C43(DE3) pLysS and SHuffle®T7) were achieved from glycerol stocks and streaked on agar plates with appropriate antibiotic resistance for overnight growth (no antibiotics for E. cloni 10G, BL21(DE3)GOLD and SHuffle®T7, 34 µg/mL Chloramphenicol for C41(DE3)pLysS, C43(DE3)pLysS). A single colony was picked up and grown in 5 mL LB medium with appropriate antibiotics resistance overnight. 3 mL LB medium from overnight culture was added to 500 mL new LB medium with appropriate antibiotics resistance and placed on 200 rpm shaker at 37 ℃ to grow to an absorbance of 0.35 to 0.4. Competent cells were prepared through the method described in Sambrook and Russell (2001) and 50 µL aliquots were stored at -80 ℃ (Sambrook and Russell 2001).

7.2.5. Transformation by electroporation

DNA ligation (2 µL) was added in 50 µL E. cloni 10G competent cell and transferred to a 2 mm gap width electroporation cuvette then pulsed at 2.5 kV. SOC medium (500 µL) was added for cell recovery immediately. Cells were incubated at 200 rpm shaker, 37 ℃ for 1 h and plated on a pre-warmed LB agar plate with appropriate antibiotic resistance (50 µg/mL Kanamycin for pET-28a(+); 100 µg/mL Ampicillin for pGEX-6P-3) (Figure 7.2). A single colony was selected and cultured in 2 mL LB medium with appropriate antibiotic resistance overnight. Glycerol stock was made for long-term storage of the plasmid in E. cloni 10G. Plasmids with insertions were extracted with Roche plasmid miniprep kit (Roche). Recombinant plasmid (1 ng) was transformed into expression *E. coli* strains with the same electroporation method.

7.2.6. Expression and purification of IAPs

A single colony was selected from LB agar plate with appropriate antibiotic resistance and transferred to 5 mL LB medium with appropriate antibiotic resistance for

overnight growth at 37 °C, 200 rpm shaker. Cultured cells (2 mL) were then harvested for plasmid extraction and double digestion with *NdeI-XhoI* for pET-28a(+) or *BamHI-XhoI* for pGEX-6P-3 to confirm the correct insertion. After confirmation of the recombinant plasmid, 3 mL cells were transferred to 500 mL LB medium with appropriate antibiotic resistance and grew at 37 °C to an OD of 0.4-0.8. IPTG (500 mM) was added to a final concentration of 0.6 mM and cells were induced at 18 °C overnight for the purpose of maximizing the soluble protein production. Cell pellets were harvested by centrifugation at 8000 rpm, 4 °C for 30 min and resuspended in lysis buffer at a ratio of 4 mL per gram wet cell pellet. Proteins were extracted by three cycles of sonication (30% amplitude, 0.5 s on/0.5 s off for 1 min), after being treated with lysozyme (a final concentration of 1 mg/mL) at 4 °C for 1 h.

7.2.7. Purification of His-tagged or GST-tagged proteins

Histrap FF crude (GE Healthcare) and Glutathione Sepharose 4B resin (GE Healthcare) were used to purify IAPs with His-tag and IAPs with GST-tag, respectively, following manufacturer's protocols.

7.2.8. Ice activity assays

INA, TH and RI activity were screened for the purified protein following ice activity assays described in chapter 2.

7.2.9. Phylogenetic analysis of ice active proteins in the *C. macra* transcriptome

Open reading frames for cold-related genes were downloaded from NCBI GeneBank and Li *et al*. (2012). Nucleotide sequence alignments of IAPs aligned by ClustalX were trimmed by PAL2NAL with default settings according to the corresponding amino acid sequence alignments aligned by ClustalX (Larkin *et al.* 2007; Suyama *et al.* 2006). The "best DNA/Protein model" of sequence alignments (tested by Jmodeltest and Protest in Phylemon 2.0) was used to infer Bayesian trees and

maximum likelihood (ML) trees, using MrBayes (v3.2.6) and GARLI (v2.01) respectively (Ronquist *et al.* 2012; Zwickl 2006). Bayesian posterior probability supports and bootstrap supports of tree nodes were also calculated respectively for the corresponding trees. The 50% majority consensus tree of each gene was generated by Sumtree (v3.3.1) in DendroPy program (Sukumaran and Holder 2010).

BEAST (v2.6.4), which utilized Markov chain Monte Carlo (MCMC) based Bayesian analysis, was used to estimate the divergence time of cold related genes (Drummond *et al.* 2012). The xml configuration files were generated in BEAUti with a combination of both nucleotide and amino acid alignments. The site model was unlinked and selected based on the results of Jmodeltest and Protest. The clock model and tree model of the nucleotide alignment and the amino acid alignment respectively were linked. The Bayesian consensus tree from MrBayes was used as a start tree to guide the BEAST analysis. Two independent runs were performed for each dataset, which ran for 10,000,000 generations with the tree sampled every 1,000 generations. Logcombiner was used to combine datasets for individual runs. Trees were interpreted by TreeAnnotator and viewed in Figtree (http://tree.bio.ed.ac.uk/). In the maximum clade credibility trees, *Oryza sativa japonica* was selected as the outgroup for *Pooideae*. Three sources of fossil evidence were used for the calibration of divergence times: *Oryza sativa japonica* and *Pooideae* (70-46 Mya) (Sandve *et al.* 2008), *Brachypodium* and *Triticum* (40-35 Mya) (Bossolini *et al.* 2007), and *Triticum* and *Aegilops* (4.5-2.5 Mya) (Grønvold 2013).

7.3. Results

7.3.1. Putative IAP gene selection for synthesis

Contigs in the *C. macra* transcriptome were found to match plant IAPs with full open reading frame (ORF). These included c90000_g1_i, c286849_g3_i1 and c82162_g1_i1 (mapped to the mRNAs for the recrystallization inhibition proteins, EU680848 and EU680850); c352250_g2_i3, c352250_g2_i4, c352250_g2_i5 and

c352250_g2_i7 (mapped to the mRNAs for the class II endochininase antifreeze proteins, AF280438 and AF280437). High similarities were reported in the corresponding amino acid sequences between c90000_g1_i and EU680848, with 74.8% sequence identity, and between c352250_g2_i5 and AF280437, with 70.8% sequence identity (Figure 7.3).

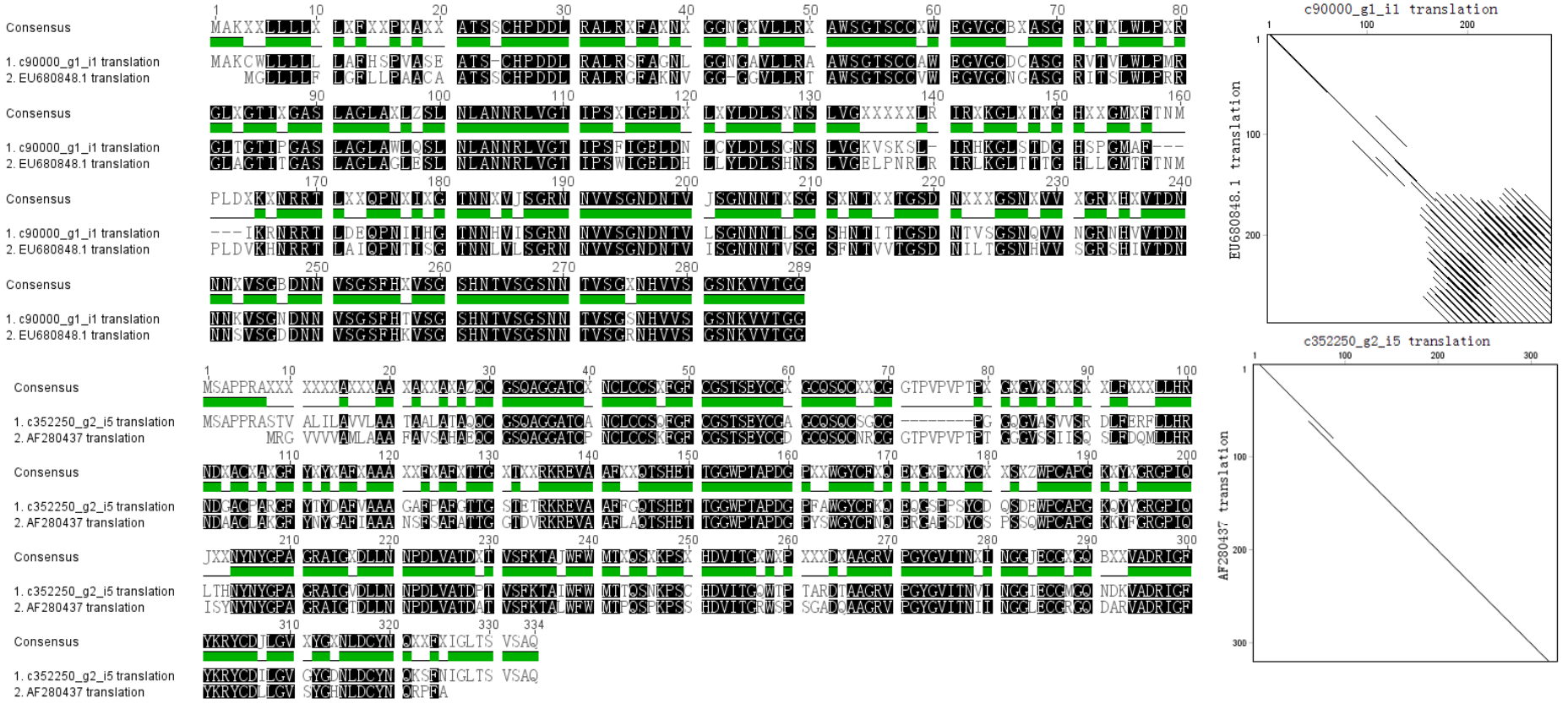


Figure 7.3 Amino acid sequence alignments of IAPs in the *C. macra* transcriptome. Alignments and dotplots were generated by Geneious 6.0 (Biomatters). Base in back stands for identical site. Dotplot of c90000_g1_i and EU680848 alignment indicates multiple repeats in the C-terminal.

7.3.2. Gene synthesis

Gene IRI2, which showed a high similarity to the mRNA of the recrystallization inhibition protein (RIP) (EU680848), and EAF2, which showed a high similarity to the mRNA of the class II endochitinase antifreeze protein (EAF) (AF280437), were synthesized by IDT gBlock Gene Fragment with restriction enzyme sites of *BamHI*, *NdeI* added at the 5' end, and *XhoI* at 3' end (Figure S5-S6).

7.3.3. Plasmid digestion, ligation and transformation

The pET-28a(+)-IRI2 and pET-28a(+)-EAF2 were confirmed by visualizing the recombinant plasmids with the *NdeI-XhoI* double digestion on 1.0% agarose gels. Two bands, one at ~5,000 bp and the other at ~880 bp, were observed on the agarose gel from the double digestion of pET-28a(+)-IRI2 (Figure 7.4). A ~5,000 bp band and a ~1,050 bp band, were observed on the 1.0% agarose gel from the double digestion of pET-28a(+)-EAF2 (Figure 7.5). The recombinant plasmids were also confirmed by the Sanger sequencing with T7 primers (Figure S7-S8).

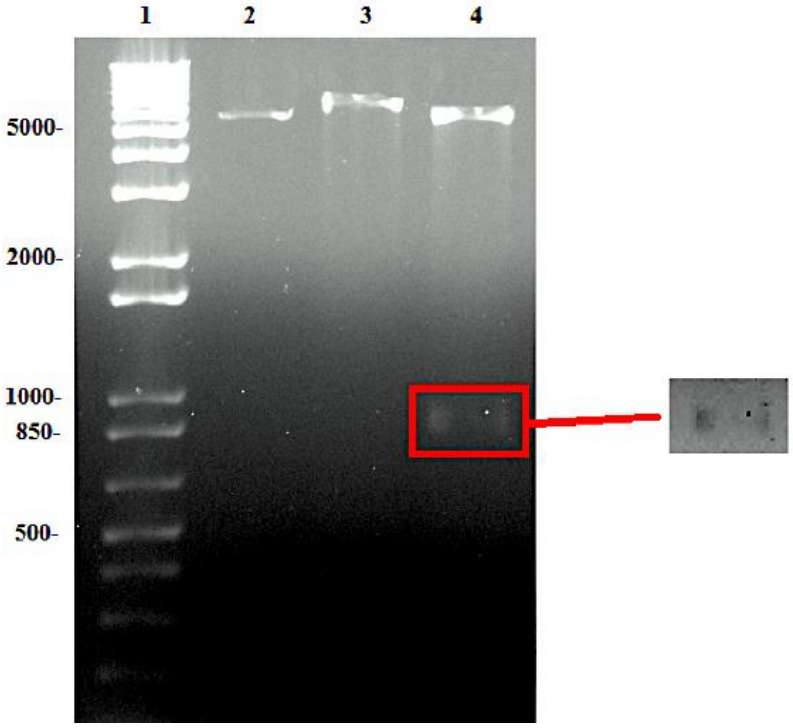


Figure 7.4 pET-28a(+)-IRI2 confirmation by restriction enzyme digestions.
Lane 1, 1Kb plus ladder (Invitrogen) (bp); Lane 2, Empty pET-28a(+) vector double digested by *NdeI- XhoI*, Lane 3, pET-28a(+)-IRI vector digested by *NdeI*; Lane 4, pET-28a(+)-IRI2 vector double digested by *NdeI-XhoI*. Only the contrast and the color of the original gel image (left) has been adjusted for better visualization of indicated bands (Right).

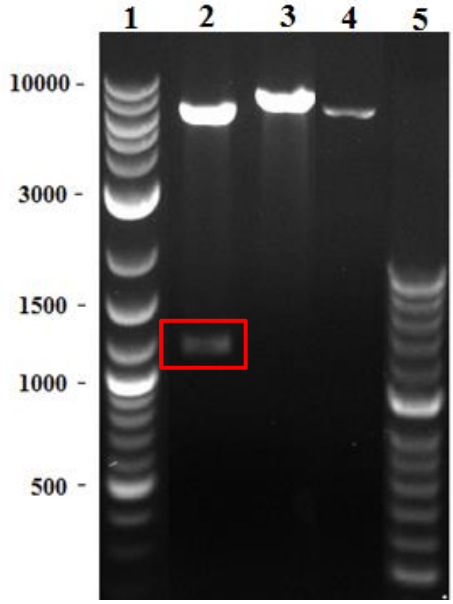


Figure 7.5 pET-28a(+)-EAF2 confirmation by restriction enzyme digestion. Lane 1, 2-log DNA ladder (New England Biolabs) (bp); Lane 2, pET-28a(+)-EAF2 double digested by *NdeI-XhoI*; Lane 3, pET-28a(+)-EAF2 digested by *NdeI*; Lane 4, pET-28a(+) digested by *NdeI*.

The pGEX-6P-3-IRI2 and pGEX-6P-3-EAF2 were confirmed by visualizing the double digestion of recombinant plasmids with *BamHI* and *XhoI* on 1.0% agarose gel. Two bands, one for the pGEX-6P-3 vector and the other for the inserted gene, were observed on the agarose gel (Figure 7.6).

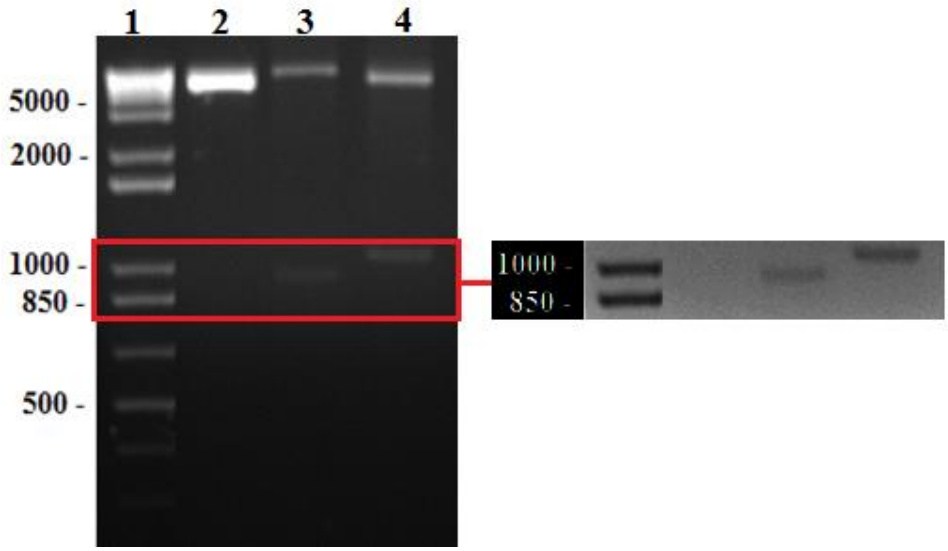


Figure 7.6 pGEX-6P-3-IRI2 and pGEX-6P-3-EAF2 confirmation by restriction enzyme digestions. Lane 1, 1 Kb plus ladder (Invitrogen) (bp); Lane 2, pGEX-6P-3 double digested by *BamHI-XhoI*; Lane 3, pGEX-6P-3-IRI2 double digested by *BamHI-XhoI*; Lane 4, pGEX-6P-3-EAF2 double digested by *BamHI-XhoI*. Only the contrast and the color of the original gel image (left) has been adjusted for better visualization of indicated bands (Right).

7.3.4. Protein expression and purification

IRI2/EAF2 insertion in pET-28a(+) vector failed to express in the BL21(DE3)GOLD cell with the 0.6 mM IPTG induction at 18 °C for 16 h and there was no His-tag IRI2/EAF2 protein after His-tag purification. No ice-shaping activity was observed in the raw protein extracts. Overexpressed proteins (~37 kDa) were found on the SDS-PAGE, which were not IRI2/EAF2, but *E. coli* ribosome proteins identified by mass spectrometry. IRI2/EAF2 proteins also failed to express in the application of IPTG gradient (Figure 7.7), and various cell strains including C41(DE3)pLysS, C43(DE3) pLysS and ShuffleT7 strains.

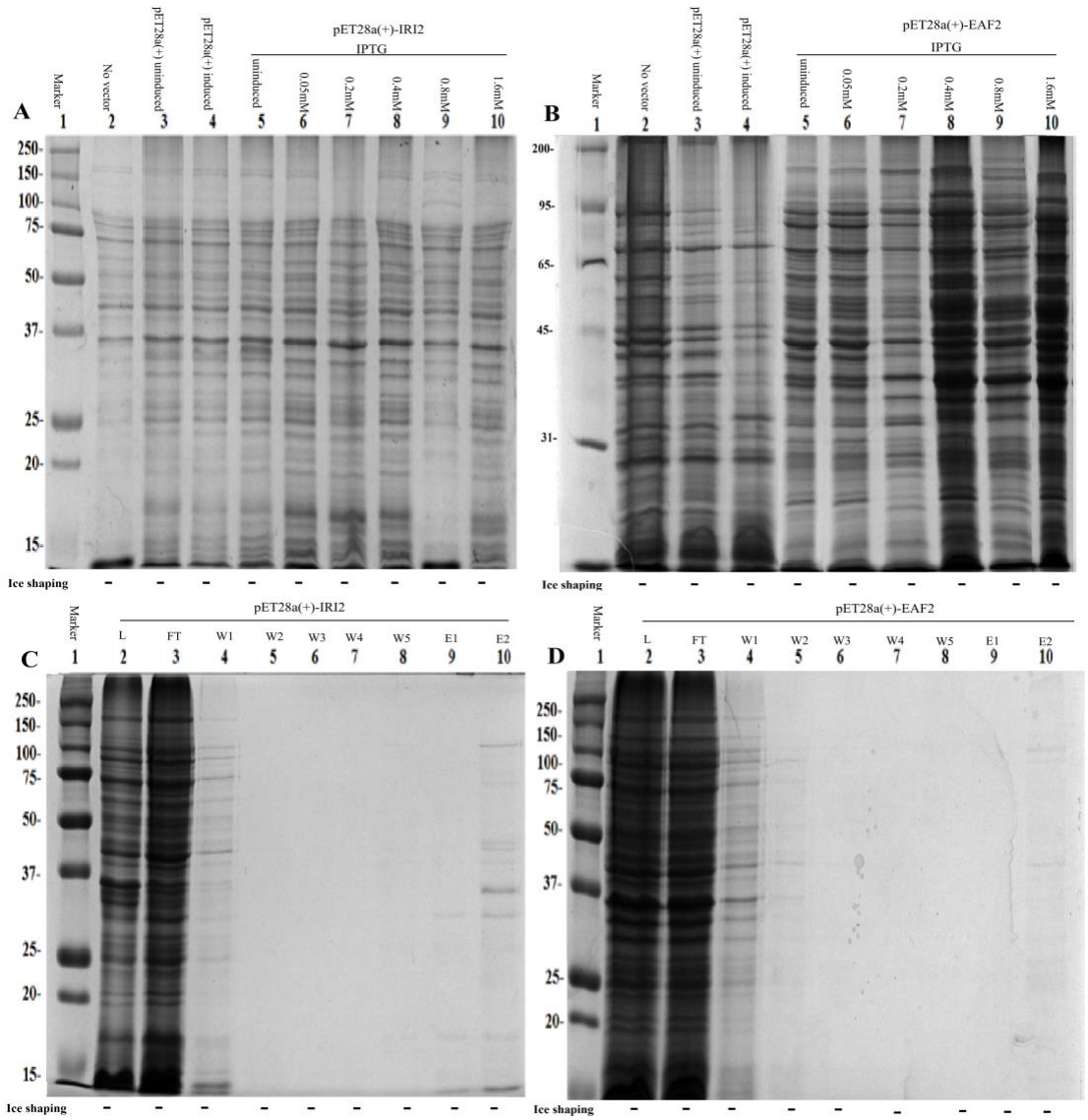


Figure 7.7 Expression failure of His-tag IRI2 and His-tag EAF2 in BL21(DE3)GOLD. A, B: Lane 1, marker (kDa); 2, BL21(DE3)GOLD only; 3 to 4, pET-28a(+), uninduced and induced with 0.6mM IPTG respectively; 5 to 10, pET-28a(+)-IRI2/EAF2, uninduced, induced with 0.05 mM, 0.2 mM, 0.4 mM, 0.8 mM, 1.6 mM IPTG respectively. C, D: Lane 1, all blue marker (kDa) (Bio-Rad); 2, cell lysate of pET-28a(+)-IRI2/EAF2; 3, flow through from His tag resin; 4-8, 1st to 5th wash of resin; 9-10, elution of His-tag IRI2/EAF2. "-", no ice shaping ability, round ice crystal under nanoliter osmometry.

IRI2/EAF2 insertion in pGEX-6P-3 vector were expressed in the BL21(DE3)GOLD cell, C41(DE3)pLysS, C43(DE3) pLysS and ShuffleT7 strains. C41(DE3)pLysS was selected for further protein expression. Glutathione Sepharose 4B resin was applied to purify the GST-tagged IRI2 and EAF2. In the elution, two bands (one at ~58 kDa and the other at ~26 kDa) were observed in elution from GST-IRI2 and GST-EAF2 (Figure 7.8). GST-IRI2, GST-EAF2 and the 26 kDa GST-tag bands were confirmed by mass spectrometry (Figure S8-S10).

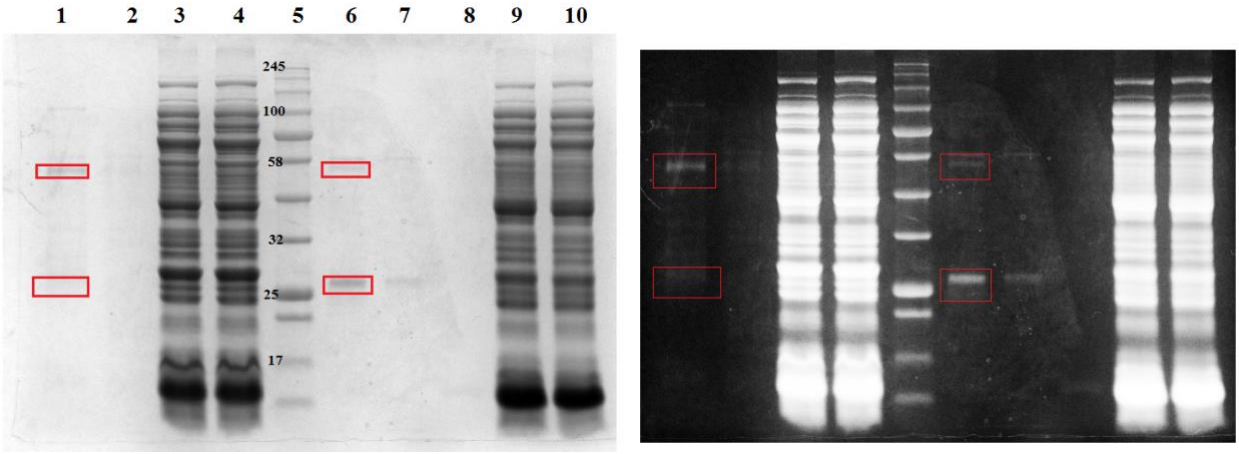


Figure 7.8 GST-IRI2 and GST-EAF2 purification. Lane 1, GST-EAF2 elution; Lane 2, 3rd wash of GST resin; Lane 3, flow through; Lane 4, cell lysate of pGEX-6P-3-EAF2; Lane 5, marker (kDa) (New England Biolabs); Lane 6-7, GST-IRI2 elution; Lane 8, 3rd wash GST resin; Lane 9, flow through; Lane 10, cell lysate of pGEX-6P-3-IRI2. Bands in red were processed to mass spectrometry. Only the contrast and the color of the original gel image (left) has been adjusted for better visualization of indicated bands (Right).

GST-tagged human rhinovirus 3C proteinase (GST-3C proteinase) was applied to cleave the GST tag from the GST-IRI2 (Figure 7.9). Mass spectrometry identified peptides after tryptic digestion matched IRI2 (Figure 7.10). However, the 3C proteinase failed to cleave the GST-tag from the GST-EAF2 so the GST-EAF2 was proceeded to ice activities analysis (Figure 7.9).

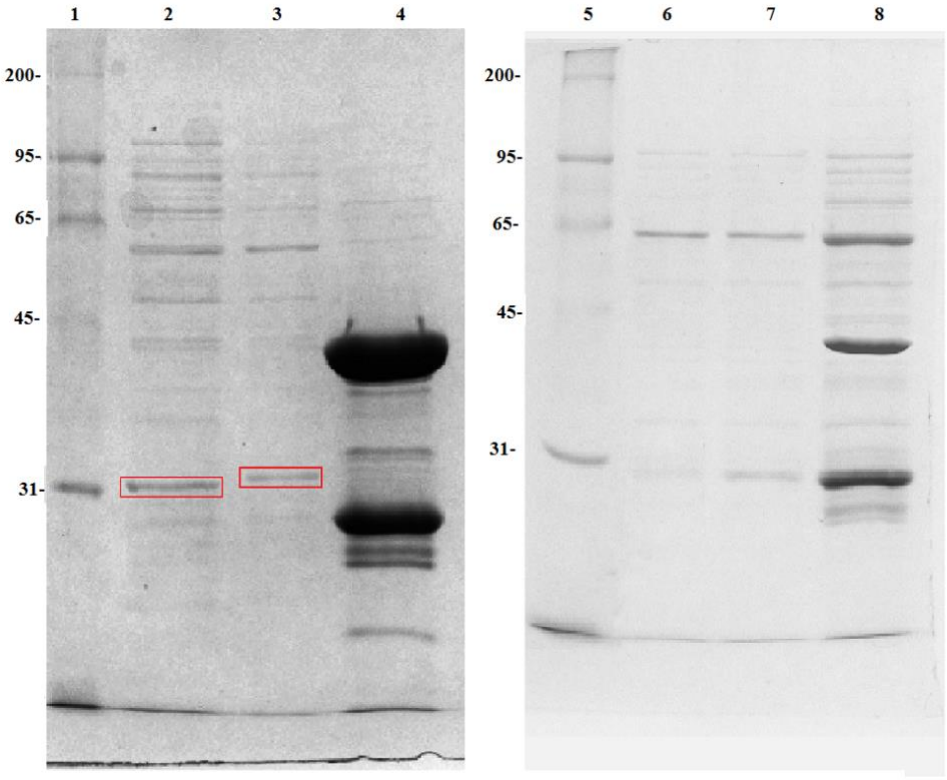


Figure 7.9 GST tag cleavage from GST-IRI2 (left)/GST-EAF2 (right). Lane 1 and 5, PAGEmarker^{TM} Unstained Protein Marker (G-Biosciences) (kDa); Lane 2-3, Cleavage of GST tag in GST-IRI2; Lane 4 and 8, wash of GST tag and GST-3C bound to the GST resin; Lane 6-7, Cleavage of GST tag in GST-EAF2 (unsuccessful). IRI2 was found at ~31 kDa. GST tag failed to be cleaved from GST-EAF2.

1 GPLGSPGIPG VHMAKCWLLL LLAFHSPVA SEATSCHPDD LRALRSFAGN

51 LGGNGAVLLR AAWSGTSCCA WEGVGCDCAS GRVTVLWLPM RGLTGTIPGA

101 SLAGLAWLQS LNLANNRLVG TIPSFIGELD NLCYLDLSGN SLVGKVSKSL

151 IRHKGLSTDG HSPGMAFIKR NRRTLDEQPN IIHGTNNHVI SGRNNVVSGN

201 DNTVLSGNNN TLSGSHNTIT TGSDNTVSGS NQVVNGRNHV VTDNNNKVSG

251 NDNNVSGSFH TVSGSHNTVS GSNNTVSGSN HVVSGSNKVV TGGLVPRGSH

301 HHHHH

Figure 7.10 Peptide fragments identification of IRI2 after tryptic digestion. Matched peptides were shown in bold red. Protein sequence coverage was 74%.

7.3.5. Ice activities in synthesized IAPs

INA in the purified IRI2 protein and GST-EAF2 protein were both significantly lower than that of the raw *C. macra* protein extract (Table 7.1). A typical hexagonal shaped ice crystal was observed in the nanoliter osmometry of purified IRI2 protein and a weak hexagonal shaped ice crystal was observed in the nanoliter osmometry of purified GST-EAF2 protein (Table 7.1). Thermal hysteresis of IRI2 and GST-EAF2 were 0.04 ℃ and 0.03 ℃ respectively. Comparatively high RI activity was found in the splat cooling assay of purified IRI2 protein, with a mean ice crystal diameter of 18.3 μm at t_{30min}, whereas no RI activity was detected in purified GST-EAF2 protein (mean ice crystal diameter of 60.4 μm at t_{30min}) (Table 7.1).

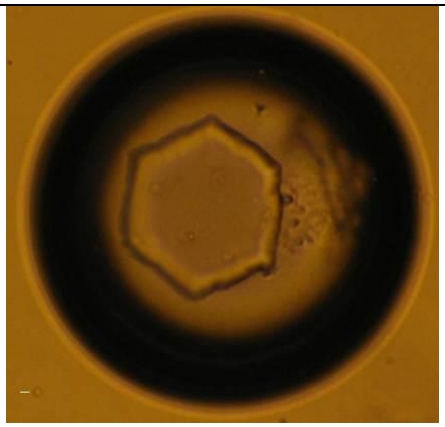
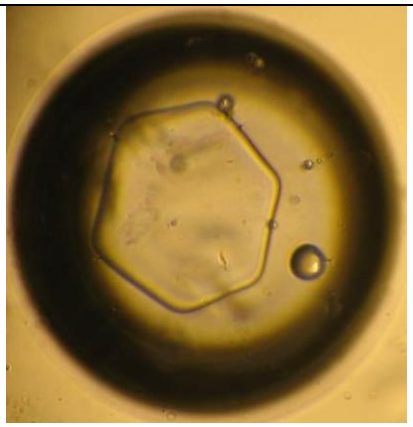
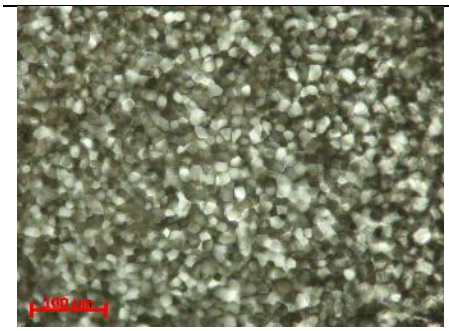
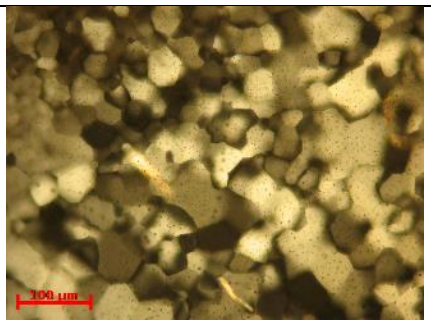
	IRI2 protein	GST-EAF2 protein
Ice shaping		
RI assay		
INA	14.8±1.4 ℃	16.9±0.9 ℃
TH value	0.03±0.01 ℃	0.04±0.01 ℃
$RI^{30\ min}$	18.3±2.0 μm	60.4±9.7 μm

Table 7.1 Ice activities of purified IRI2 protein and GST-EAF2 protein. Three measurements were performed for each activity. Images given are representative.

7.3.6. Phylogenetic analysis of cold-related genes

In the BEAST analysis of IRI2 genes, most parameters obtained effective sample size (ESS) values of not less than 200. In the maximum clade credibility tree, the majority of branches obtained a posterior probability of no less than 0.8, with bootstrap supports and Bayesian supports higher than 80 and 0.8 respectively (Figure 7.11). This tree used *Oryza sativa japonica* as the outgroup, and consisted of the *Brachypodium distachyon* only clade, and the core *Pooideae* clade (*Triticeae* and *Poeae*), with high bootstrap support (100/100) and Bayesian support (1/1). There were two groups of IRI2 genes in the *Brachypodium distachyon* only clade with high bootstrap and Bayesian support. The *Triticeae* sub-clade contained four groups: a *Triticum* only group, a *Triticum-Aeglops-Elymus-Secale-Thinopyrum* mixed group

and two *Triticum-Hordeum* groups, with good bootstrap support for nucleotide sequences (>60). The bootstrap support for amino acid sequences was low (<50) for the *Triticum* only group and the *Triticum-Hordeum* groups. The *Poeae* sub-clade contained three genera in this tree: *Lolium*, *Deschampsia* and *Chionochloa*. Divergence in the *Poeae* sub-clade was complex and tree branches were not well resolved or supported.

The divergence times based on the IRI2 gene were estimated to be 50.0±2.0 Mya (between *Oryza sativa japonic*a and *Pooideae* species), 36.0±1.0 Mya (between *Brachypodium distachyon* and core *Pooideae* species), and 5.8±0.7 Mya (among *Triticeae* species) and 5.9-27.1 Mya (95% HPD, node height highest posterior density intervals) (among *Poeae* species). The divergence time between the two groups of *Brachypodium distachyon* was estimated to be 5.9-27.5 Mya (95% HPD). The divergence time between *Chionochloa macra* and *Lolium* was a recent event based on the IRI2 gene (1.3-7.6 Mya, 95% HPD). The divergence time between *Chionochloa macra* and *Deschampsia* was 2.8-9.0 Mya (95% HPD) based on the IRI2 gene. Unlike *Brachypodium distachyon*, the IRI2 genes from *C. macra* had not diverged from those of the core *Pooideae*.

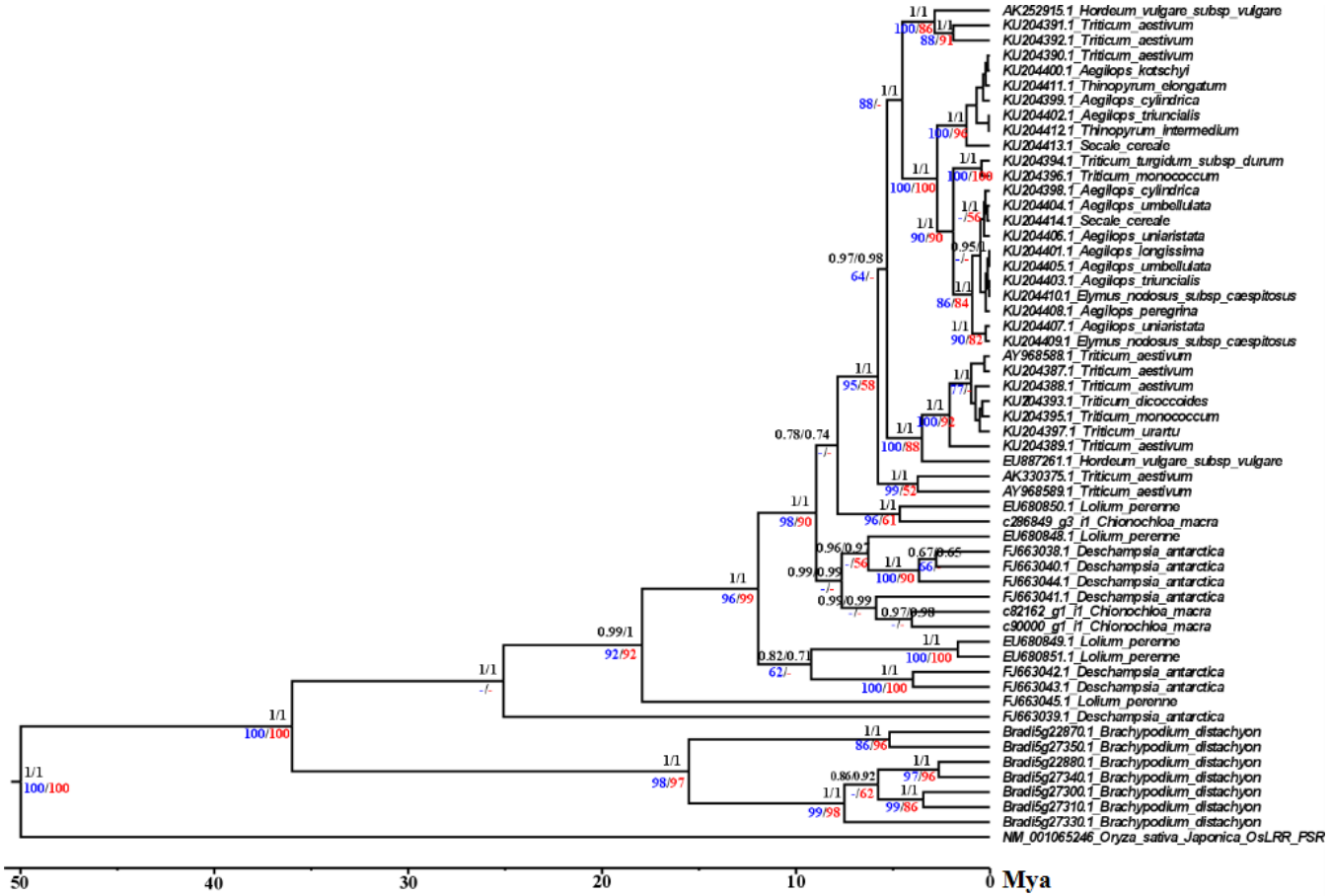


Figure 7.11 Evolutionary relationship of *Chionochloa* species to *Pooideae* species based on the IRI2 gene. Number above the branch in black, Bayesian posterior probability (BEAST/MrBayes). Number below the branch in blue/red, bootstrap support values inferred by GARLI (nucleotide/amino acid). Support values less than 50 are shown as "-". *Oryza sativa japonica* was set as the outgroup. Scale bar, millions of years ago (Mya).

Endochitinase proteins were identified in many plants showing high amino acid similarity with antifreeze protein AF280437 and AF280438 from winter rye (*Secale cereal*). BEAST analysis of EAF2 genes showed a slightly different tree to the IRI2 tree (Figure 7.12). There was full Bayesian and bootstrap support (1/1) for *Oryza sativa japonica* as the outgroup in this tree and *Brachypodium distachyon* as the sister group of the core *Pooideae*. The EAF2 genes in the core *Pooideae* mainly fell into two clades (bootstrap support for amino acid sequences, 75) in this tree, with genes in *Triticum* only found in one clade with AF280437, genes in *Chionochloa* only found in the other clade with AF280438, and genes in *Seceale*, *Aegilops* and *Hordeum* found in both clades. EAF2 (c352250_g2_i5) and c352250_g2_i7 from the *C. macra* transcriptome showed more than 70% amino acid similarity with proteins AF280437 and AF280438 respectively. These two proteins, together with c352250_g2_i3 and c352250_g2_i4, were clustered with AF280438 in a separate sub-clade, which might indicate similar ice-shaping function. The estimated divergence times were 26.8-38.6 Mya (95% HPD) of the two clades in *Pooideae* and 9.2-25.3 Mya (95% HPD) of the *Chionochloa*-AF280438 clade.

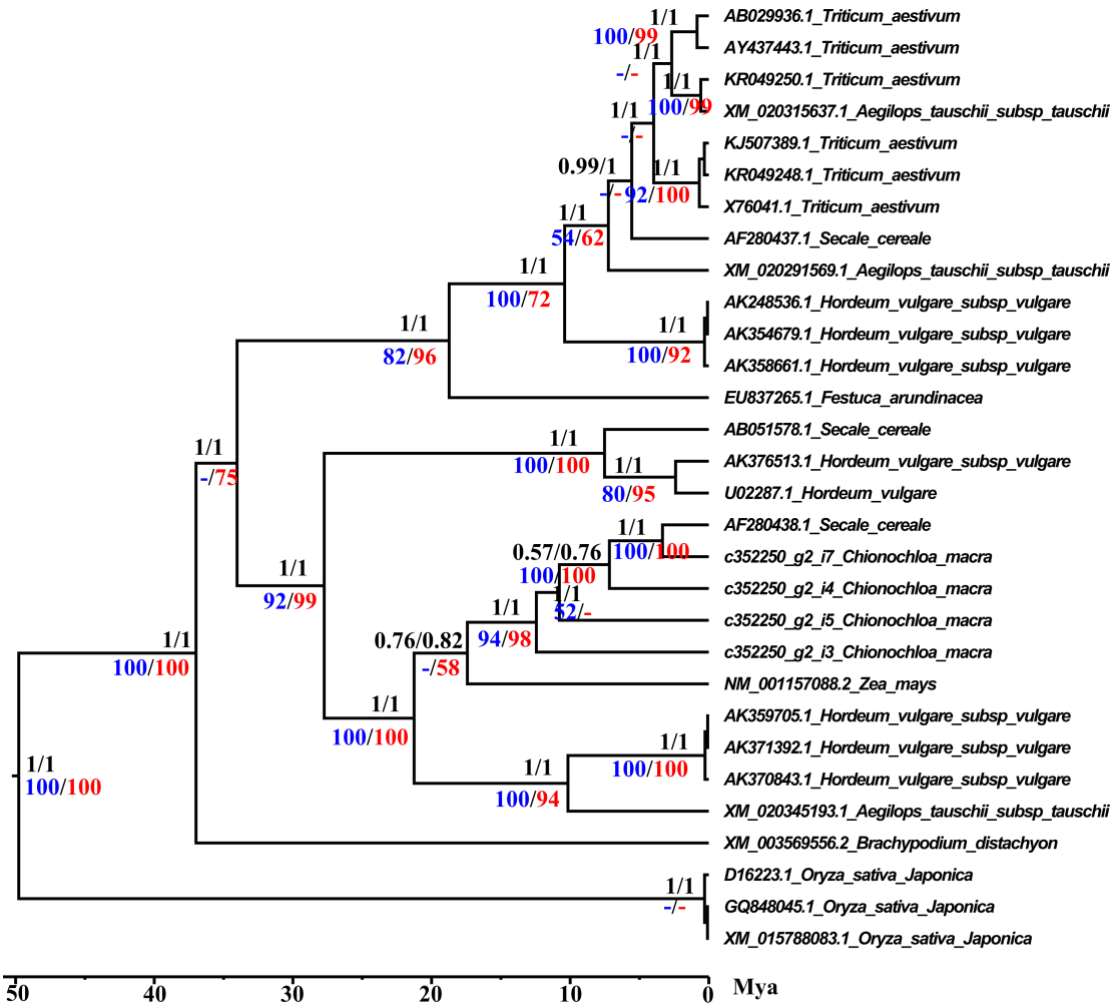


Figure 7.12 Evolutionary relationship of *Chionochloa* species to *Pooideae* species based on the EAF2 gene. Number above the branch in black, Bayesian posterior probability (BEAST/MrBayes). Number below the branch in blue/red, bootstrap support values inferred by GARLI (nucleotide/amino acid). Support values less than 50 are shown as "-". *Oryza sativa Japonica* was set as the outgroup. Scale bar, millions of years ago (Mya).

There were two main clades with good branch support in the tree inferred from the FST gene: the fructosyl transferase-like clade and the invertase-like clade (Figure 7.13) (Li *et al*. 2012a). Species in the fructosyl transferase-like clade were from the core *Pooideae*, which shared the common feature that all FST genes contained a h(A/G)Y/F motif, whereas species in the invertase-like clade instead contained the WGW motif (Figure 7.13) (Li *et al*. 2012a). Three FST-genes from *Chionochloa macra* fell into the fructosyl transferase-like clade with the presence of the h(A/G)Y/F motif and one *C. macra* gene fell into the invertase-like clade. The split of these two clades was estimated to be 21.3-30.9 Mya (95% HPD).

The tree inferred from the CBF3 gene showed that there were four main clades: a *Brachypodium distachyon* only clade, two core *Pooideae* clades and a *Chionochloa-Setaria* clade (Figure 7.14). The *Oryza sativa japonica* outgroup and the *Brachypodium distachyon* only clade were well supported, whereas the divergence of core *Pooideae* and *Chionochloa* did not have good support. The *Brachypodium distachyon* only clade and the upper core *Pooideae* clades in the CBF tree equated to the CBF3d subgroup of Li *et al*. (2012). The other core *Pooideae* clades and Bradi4g35630.01 equated to their CBF3c subgroup. The estimated divergence times in this tree were 29.6-31.8 Mya (95% HPD) between the core *Pooideae* CBF3d clade and the *Chionochloa-Setaria* clade, and 32.1-38.8 Mya (95% HPD) between the core *Pooideae* and the *Brachypodium distachyon* only clade.

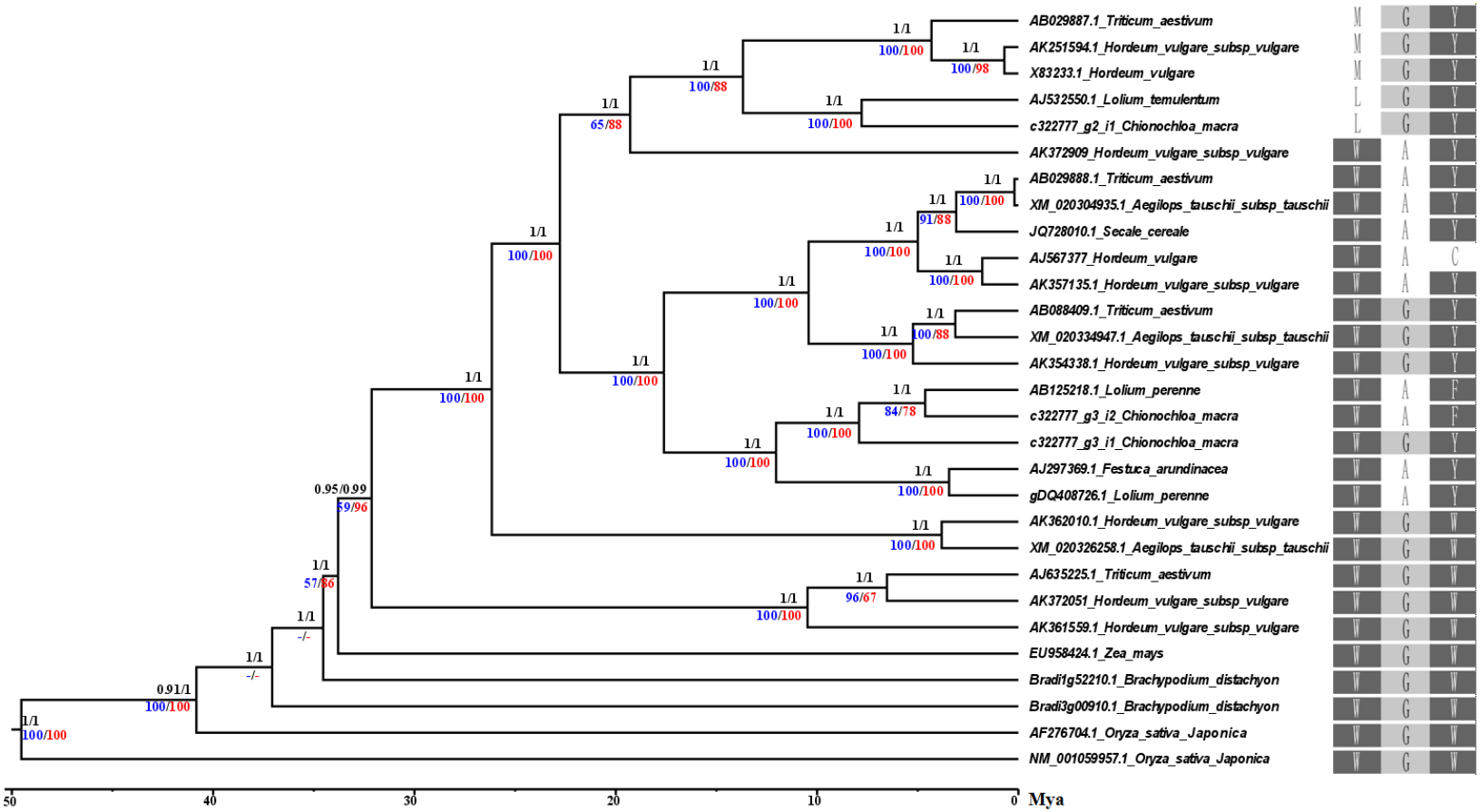


Figure 7.13 Evolutionary relationship of *Chionochloa* species to *Pooideae* species based on the FST gene. Number above the branch in black, Bayesian posterior probability (BEAST/MrBayes). Number below the branch in blue/red, bootstrap support values inferred by GARLI (nucleotide/amino acid). Support values less than 50 are shown as "-". *Oryza sativa japonica* was set as the outgroup. FST specific motif, h(A/G)Y/F motif was shown on the right. Scale bar, millions of years ago (Mya).

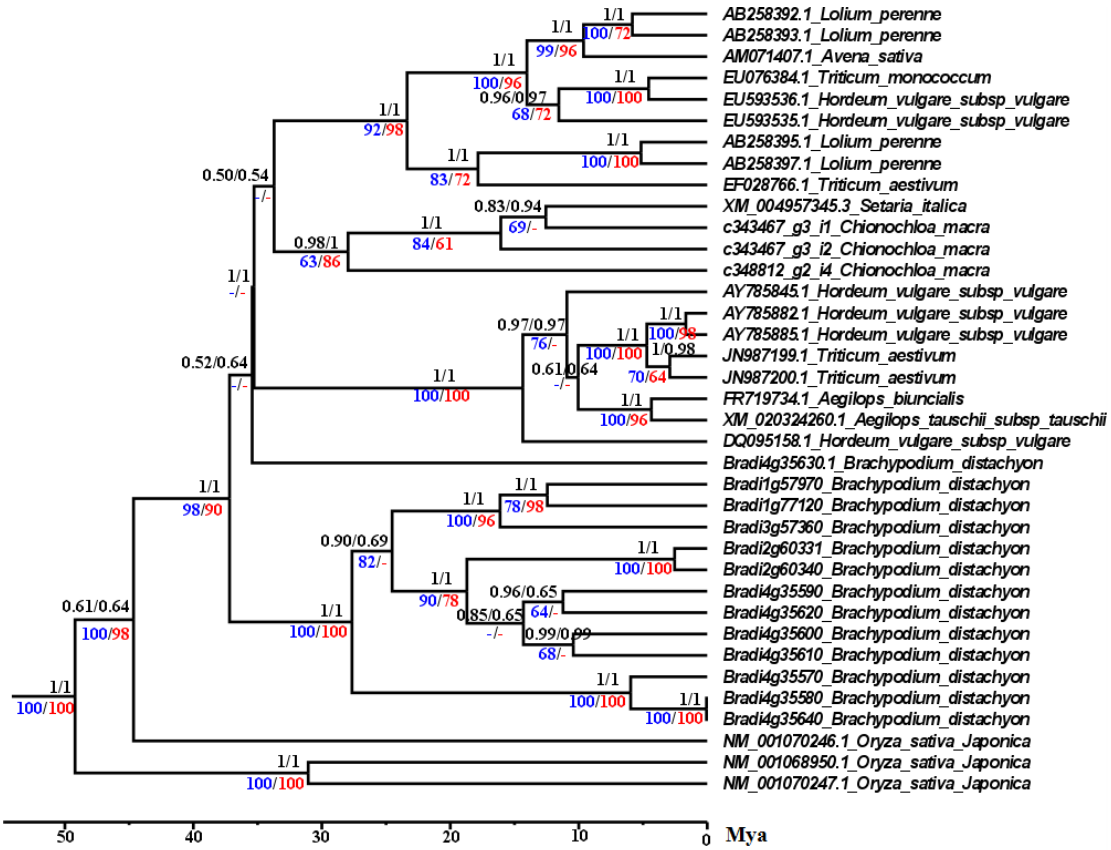


Figure 7.14 Evolutionary relationship of *Chionochloa* species to *Pooideae* species based on the CBF3 gene. Number above the branch in black, Bayesian posterior probability (BEAST/MrBayes). Number below the branch in blue/red, bootstrap support values inferred by GARLI (nucleotide/amino acid). Support values less than 50 are shown as "-". *Oryza sativa japonica* was set as the outgroup. Scale bar, millions of years ago (Mya).

7.3.7. Expression levels of cold response-related genes in the *C. macra* transcriptome.

The expression levels of RIP (IRI2), endochitinase AFP (EAF2), CBF and FST genes in the *C. macra* transcriptome are shown in Table 7.2.

Gene name	*C. macra* gene	FPKM
Endochitinase AFP	c352250_g2	27.4
FST	c322777_g3	58.1
CBF	c343467_g3	1.9
	c348812_g2	3.4
RIP	c90000_g1	69.4
	c82162_g1	19.4
	c286849_g3	9.4

Table 7.2 Cold response-related gene expression in the *C. macra* transcriptome. FPKM values indicate expression level of genes in the transcriptome.

7.4. Discussion

The grass family *Poaceae*, is the fifth largest flowering plant family, containing more than ten thousand species, and is widely distributed from the warmest to the coldest zones worldwide (Humphreys and Linder 2013). The *Pooideae* grasses, the largest subfamily of *Poaceae* are strictly C3 grass mainly found in Northern Hemisphere temperate areas, and have been an interesting model to study climate adaptation (Strömberg 2011). Generally, the *Poaceae* family are adapted to tropical environments whereas most *Pooideae* tolerate cold climates (Taylor *et al.* 2012; Taylor *et al.* 2011). The *Danthonioideae*, the Southern Hemisphere grass subfamily, consists of both C3 and C4 photosynthetic types and belongs to the PACMAD tropical clade (Taylor *et al.* 2012; Taylor *et al.* 2011). The distribution of *Danthonioideae* is not considered to be limited by temperature, and different responses to cold have been observed in several lineages of *Danthonioideae* (Humphreys and Linder 2013). Cold tolerance in *Danthonioideae* has evolved more recently (~8 Mya) than the evolution of a novel gene family which expanded in the *Pooideae* during global cooling (Humphreys and Linder 2013). *Danthonioideae* experienced dramatically increased rates of cold tolerance only in of the genera *Chionochloa*, *Danthonia* and *Rytidosperma* (Humphreys and Linder 2013). Cold tolerance in *Danthonia* and *Rytidosperma* has been correlated with geographical expansion into colder climates, whereas in *Chionochloa* it was thought to have evolved *in situ* (Pirie *et al.* 2010).

Chionochloa, a genus in the *Danthonioideae* and endemic in New Zealand, is widespread from low land to alpine areas and dominates grassland in the South Island (Zotov 1963). Several *Chionochloa* species have developed extreme cold tolerance which is thought to have evolve *in situ*, however the cold tolerance mechanism involved was not clear (Humphreys and Linder 2013). Wharton *et al*. (2010) reported high levels of INA in two *Chionochloa* species but there have been no other studies characterizing and identifying IAPs in the *Chionochloa* genus since then (Wharton *et al.* 2010). Several potential IAPs were identified in the *C. macra* transcriptome through Blast searches against both nucleotide sequence and corresponding amino acid sequence data. In this chapter, protein products from two IAP-like genes showed reasonable levels of ice activities when expressed in the *E. coli* system. Phylogenetic analysis also illustrated the evolution of these genes together with two cold response-

related genes (CBF and FST) in *C. macra* compared with *Brachypodium distachyon* and the core *Pooideae* family.

The Blast searches identified proteins in *C. macra* with high similarities in both nucleotide sequence and the corresponding amino acid sequence to IAPs in the *Pooideae* family, which may indicate similar ice-binding functions. The RIP-like proteins from *C. macra* featured LLR repeat domains and RIP repeat ice-binding domains (John *et al.* 2009). Genes encoding the EAF-like protein showed high similarity to those of endochitinase antifreeze proteins, AF280437 and AF280438 (Yeh *et al.* 2000).

In this study, protein expression in the pET-28a(+) system failed even though the insertions were correct, rare codon usage was adjusted and expression conditions were optimized in IPTG gradients, induction temperatures, medium types and cell strains. Some studies have showed that proteins failed to express in pET plasmid even with various optimizations (Gou *et al*. 2004; Hire *et al*. 2008; Jia *et al*. 2011; Li *et al*; 2015; Pereira *et al*. 2017). The expression of IRI2 and EAF2 was shifted to a GST fusion system as Pereira *et al*. (2017) indicated some proteins that failed in a pET system were expressed in a pGEX system. Though GST-IRI2 and GST-EAF2 were detected, a significant amount of GST-tag only protein (~26 kDa) was always formed. It is suggested that several different plasmid vectors should be used in case one fails to work.

Zheng and Wiens (2015) concluded that there was little negative effect of missing taxa in the estimation of divergence times by BEAST (Zheng and Wiens 2015). Though lacking some taxa from the *Pooideae* in the maximum clade credibility tree construction based on plant IAP-like genes, the estimation of divergence times for species in *Poaceae* was in agreement with published studies. Phylogenetic analysis of IRI2, EAF2, CBF3 and FST showed different patterns. Trees inferred from IRI2, EAF2 and FST all supported *Chionochloa* being nested within core *Pooideae* based on these cold-related genes, whereas the tree inferred from CBF3 indicated that *Chionochloa* had not diverged from the *Pooideae* at 32-38 Mya (95% HPD).

The ability to respond to cold is important and *Pooideae* grasses have adapted well to cold stress (Vigeland *et al.* 2013). Several cold-response gene families showed specific features in the *Pooideae* group: the C-repeat binding family was extensively distributed; fructosyltransferase enzymes and recrystallization inhibition proteins have evolved (Vigeland *et al.* 2013). Cold-response genes investigated in this study

obtained high levels of FPKM values except CBF3, indicating high activities of these genes in the *C. macra* transcriptome (Table 7.2). A literature search and a Blast search did not reveal that there were RIP and EAF proteins found in species from the *Panicoideae* and *Chloridoideae*, which may suggest either that there were no available sequences for species from these groups deposited in the database or these two genes were lost in the *Panicoideae* and *Chloridoideae* but regained in the *Chionochloa* due to environmental selection pressures. There were sequences in *Setaria italica*, *Zea Mays* and *Sorghum bicolor* that showed high similarity to the rice phytosulfokine LRR receptor kinase gene (OsLRR_PSR) (NM_001065246, *Oryza sativa japonica*), a possible ancestor of RIP, and the endochitinase-like antifreeze protein (AF280438).

In summary, this study investigated the ice activities of two putative IAPs and presented the first phylogenetic analysis of these IAP genes, together with two important cold response-related genes (CBF3 and FST) identified in the *C. macra* transcriptome. These showed different evolutionary relationships compared with genes from the core *Pooideae* family and the *Brachypodium distachyon*. It was inferred that IRI2 from *C. macra* had only recently diverged from other genera during the expansion of IRI2 throughout the *Pooideae* family (Bredow *et al.* 2017; Sandve *et al.* 2008). Likewise EAF2 from *C. macra* showed more recent diversification compared with the core *Pooideae.* In contrast the CBF3 and FST trees indicated that these genes had diversified in *Chionochloa* within the same timeframe as diversification had occurred within the core *Pooideae* around ~30 Mya.

Chapter 8 General discussion

8.1. Summary and conclusions

The genus *Chionochloa*, which is mainly endemic in New Zealand, has been widely studied for its biology and ecology, and is frequently used in the phylogenetic analysis of the *Danthonioideae* grasses (McGlone *et al.* 2014; Barker *et al.* 2007; Antonelli *et al.* 2011). This thesis found that *Chionochloa* had high levels of ice activities in several of its species. It was also found that ice-shell purification, general methods of protein purification and high throughput transcriptome sequencing helped to identify proteins and corresponding genes related to these activities. The phylogenies of ice active protein coding genes (IAP genes) suggested that the *C. macra* retained or regained the ice shaping ability and recrystallization inhibition activity similar to those from the *Pooideae* grass. Meanwhile, the phylogenies of C-repeat binding factor and fructosyltransferase, two *Pooideae* specific cold response-related genes found in the *C. macra* transcriptome, indicated indistinguishable relationship of the *Chionochloa* species with the *Pooideae* grass. Moreover, the ice nucleation activity in the *Chionochloa* species may be developed locally to help with the cold tolerance (Humphreys and Linder 2013).

Chionochloa species are distributed from lowland to alpine grassland in the South Island of New Zealand where cold weather occurs at any time of the year because of the incursions of the polar air masses (Connor 1991). There is the possibility that these species developed abilities to deal with freezing and showed at least one type of ice activity. Chapter 3 aimed to investigate by ice activity assays whether there were ice activities and whether there were certain patterns of the ice activities in the *Chionochloa* species.

Since ice active proteins (IAP) in some plants were mostly cold acclimated, it was hypothesized that ice activities in the *Chionochloa* species may also have followed similar patterns. The results showed that seasonal ice activities in the *Chionochloa* genus were present, with significant differences between species. Dilution screenings indirectly indicated different expression levels of ice active agents between *Chionochloa* species or in the same species in different seasons. However, other properties of ice activities in the *Chionochloa* species including origins, physical and chemical characters were unknown. Ice nucleation activity has been

reported in bacteria, so whether INA in the *Chionochloa* species was caused by INA bacteria or was intrinsic to the plant INA agent was examined.

Chapter 4 characterized ice activities in *Chionochloa* species and tested whether intrinsic protein components were essential in maintaining these activities. Protein extract from *Chionochloa* species collected in winter was received with various treatments including heat, salt, reducing agent, oxidizing agent, proteinase K and lysozyme, and the effects on ice activities were examined. Meanwhile, bacteria from leaf surfaces were separated and cultured on NAG agar. It was found that ice activities from different *Chionochloa* species behaved differently after exposure to these treatments. Proteins were essential in maintaining all three types of ice activities. Both intrinsic agents and INA bacteria are involved in the INA of the *Chionochloa* species.

As *C. macra* and *C. spiralis* obtained the highest INA and RI activity respectively, chapter 5 aimed to identify IAPs in *C. macra* and *C. spiralis*. Ice-shell purification and various chromatographic techniques were applied. Peptide fragments matched the class II endochitinase antifreeze-like protein, the recrystallization inhibition protein, and other proteins with unconfirmed ice activities in the ice fraction from ice-shell purification of both *C. macra* and *C. spiralis*. INA in the *C. macra* decreased dramatically in the ice fraction of ice affinity purification. INPs in the *C. macra* showed high affinity to an ion exchange column and could only be washed off under high salt elution. Due to fewer peptide fragments being present in mass spectrometry, INPs in the high salt elution was not identified.

In the grass family, especially the *Pooideae* grasses, IAPs including THPs and RIPs have been identified in several genera. To screen whether there are known IAP homologs in *Chionochloa*, chapter 6 aimed to establish protein expression profiles in the winter *C. macra* by sequencing its total RNAs. The first transcriptome of *C. macra*, also the first transcriptome in the *Chionochloa* genus, was assembled. As a sub-project of this thesis, this transcriptome was annotated and species-specific metabolic pathways were studied. Meanwhile, peptide fragments of ice fraction in chapter 5 were mapped to personalized MASCOT database created with the *C. macra* transcriptome.

Chapter 7 aimed to identify putative IAP genes in the *C. macra* transcriptome through a Blast search against known plant IAPs and express functional IAP genes in *E. coli* system. Meanwhile, the evolutionary history of these genes together with two

cold response-related genes were inferred and compared with those of species in the *Pooideae* subfamily. IAP genes were discovered in the *C. macra* transcriptome with high sequence similarities to known plant IAPs. The two selected genes (IAPs, IRI2 and EAF2) were expressed and ice activities were confirmed. Phylogenetic studies showed that these genes were not split from the core *Pooideae* subfamily and evolved differently compared to the *Brachypodium distachyon*.

The overall aim of the thesis was to screen ice activities and investigate ice active proteins (IAPs) in the *Chionochloa* genus, the approaches used in this study would be helpful in screening IAPs in other plant species. The first part of this thesis presented seasonal variations and confirmed protein components were essential in maintaining INA, TH and RI activity in species from this genus. The second part of this thesis identified IAPs in the *Chionochloa* species, annotated the *C. macra* transcriptome, expressed IAPs in the *E. coli* system and studied their ice activities, and presented a phylogenetic study of related genes.

8.2. Suggestions for further research

This study has added new knowledge of the ice activities and ice active proteins (IAPs) in New Zealand *Chionochloa* species. However, further research is needed to improve our understanding of ice activities in this genus.

Only 17 of the 36 endemic species and subspecies were included in this study and further research should investigate ice activities in the other 19 *Chionochloa* species and subspecies to obtain a complete view of whether there was phylogentic distribution and/or geographic distribution of ice activities in the *Chionochloa* when mapping their ice activities to the phylogentic trees (figure 1 and figure 3) in Pirie *et al*. (2010). This is because in the current study ice shaping activity was only observed in *C. spiralis* and *C. teretifolia*, which were found in the south South Island of New Zealand in the figure 1 from Pirie *et al*. (2010), and in *C. macra*, which was found in the middle South Island whereas no high level of INA was observed in *C. debbiei*, *C. flavescens* and *C. flavicans*, which was found in the north South Island and the North Island. Meanwhile, physical and chemical properties of ice activities in the other 19 *Chionochloa* species and subspecies are worth to be characterized and compared with the results in chapter 4.

Low level of INA (~-8 °C) was found in bacterial colonies derived from *C. macra* leaves. Which strain of bacteria contributed to this bacterial INA need to be determined. Whether there is any bacterial INA in other *Chionochloa* species worths further investigations. As not all bacteria are culturable on the NAG agar, a quantitative PCR based method will show more accurate identification of potential INA bacteria strain on *Chionochloa* leaves (Hill *et al*. 2013).

Ice-shell purification showed better enrichments of current known IAPs (*Pooideae* grass specific recrystallization inhibition protein and class II endochitinase antifreeze protein) in the *C. macra* and *C. spiralis* extract than those of the ice-finger method. Therefore, ice-shell purification is recommended to enrich IAPs in other *Chionochloa* species other than ice-finger. Rubisco was also largely enriched in the ice fractions of both *C. macra* and *C. spiralis*. Further study may include an additional Rubisco removal step to enhance the resolution of mass spectrometry (Sehrawat *et al*. 2013). Meanwhile, whether Rubisco play any role in the ice activities or antifreeze of the *Chionochloa* species or not need to be further investigated.

INPs from *C. macra* were enriched in the non ice-binding fraction after ice-shell purification. Whether INPs in other *Chionochloa* species are also in the non ice-binding fraction worths further investigations. A protocol of a 100 kDa MWCO centrifugal device, a size exclusion chromatography and an ion exchange chromatography were found to efficiently separate INPs in the *C. macra* extract, which is highly suggested for the purification of INPs in other *Chionochloa* species. INPs in *Chionochloa* may retained their secondary and/or tertiary structure under various physical and chemicdal treatments in the chapter 4 which proved they were comparative structure stable. Mass spectrometry failed to generate enough high quality peptide fragments to detect potential INPs in *Chionochloa* might due to this character and their resistance to the tryptic digestion. The choice of other single protease digestion or multiple-protease digestion will be tested which may generate more digested peptide fragments and improve the resolution of mass spectrometry (Tsiatsiani *et al*. 2015).

This study provided a first transcriptome of *C. macra* as a basis for studying the molecular biology of this genus whereas too many gene fragments generated by Trinity mapped to the same real gene (e.g. the IRI gene), which may because of the hexaploid of cells in *C. macra* and polymorphisms of allele genes. Though the long-read sequencing (e.g. PacBio, Nanopore and MOLECULO) generates sequences with

comparative high error rates, it can generate sequences long enough to serve as scaffold references, to which the contigs or paired-end reads from *C. macra* libraries can map to increase both gene completeness and base calling accuracy (Sharon *et al.* 2013; Tilgner *et al.* 2015). This method could to some extent reduce the number of gene fragments and increase the number of full genes in the current *C. macra* transcriptome. Further study is needed to improve the understanding of molecular genetics and the regulation of ice activities in this genus. The *C. macra* transcriptome presented a draft map of gene expression in the winter *C. macra* but much is still unknown. A comparative transcriptome analysis of *C. macra* and *C. spiralis* after exposure to various conditions (e.g. cold and warmth) could help identify cold-induced genes and their expression levels. Meanwhile, quantitative PCR needs to be done to verify the expression levels of ice active genes (Jozefczuk and Adjaye 2011).

Schubert *et al.* (2017) studied another three cold response related genes which is important in the *Pooideae,* and their results indicated lineage-specific evolution and gene expansion in the *Pooideae*. Further study in *Chionochloa* could also include these genes and comare with the *Pooideae* grass for the purpose of generating a more complete of evolutionary map of these genes in the *Chionochloa*.

Two of the ice active genes from *C. macra* could not be expressed in pET-28a(+) system with various prokaryote *E. coli* stains. Meanwhile, these two genes were expressed at low levels in the pGEX-6P-3 system with N-terminal GST-tag. Considerable high level of protein expression is essential for protein crystallization and protein structure analysis. Recent studies showed successful expression of IRI genes from *Brachypodium distachyon* in *E.coli* with different vectors, which is worth to try with the IRI2 gene in *Chionochloa* (Bredow *et al.* 2016; Bredow *et al.* 2017). In addition, other vectors (e.g. with high soluble tag, Trx-tag) are recommended to enhance expression. Furthermore, model plants as expression systems are more beneficial for the expression of eukaryotic recombinant protein and the study of plant genes (Thuenemann *et al.* 2013). The expression of ice active genes from *Chionochloa* in model plant system will also allow to study their in vivo function.

Homologs of ice active proteins may exhibit different level of ice activities compared with the origin ones (Bredow *et al.* 2017). There are homologs of IRI2 and EAF2 respectively in the *C. macra* transcriptome. Further studies could investigate whether and to what extent products of these genes had ice activities and/or ice-shaping ability.

References

Abraham, S. Keillor, K. Capicciotti, C.J. Perley-Robertson, G.E. Keillor, J.W. *et al*. 2015. Quantitative analysis of the efficacy and potency of novel small molecule ice recrystallization inhibitors. *Crystal Growth & Design*. **15**(10): 5034-5039.

An, X. Chen, J. Zhang, J.Y. Liao, Y.W. Dai, L.J. *et al*. 2015. Transcriptome profiling and identification of transcription factors in ramie (*Boehmeria nivea L. Gaud*) in response to PEG treatment, using illumina paired-end sequencing technology. *International Journal of Molecular Sciences*. **16**(2): 3493-3511.

Andrade, M.A. Perez-Iratxeta, C. Ponting, C.P. 2001. Protein repeats: structures, functions, and evolution. *Journal of Structural Biology*. **134**(2-3): 117-131.

Andrews, S. 2010. FastQC-A quality control tool for high throughput sequence data [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Antikainen, M. Griffith, M. 1997. Antifreeze protein accumulation in freezing-tolerant cereals. *Physiologia Plantarum*. **99**(3): 423-432.

Antikainen, M. Griffith, M. Zhang, J. Hon, W.C. Yang, D.S. *et al*. 1996. Immunolocalization of antifreeze proteins in winter rye leaves, crowns, and roots by tissue printing. *Plant Physiology*. **110**(3): 845-857.

Antonelli, A. Humphreys, A.M. Lee, W.G. Linder, H.P. 2011. Absence of mammals and the evolution of New Zealand grasses. *Proceedings of the Royal Society B: Biological Sciences*. **278**(1706): 695-701.

Ashworth, E.N. Kieft, T.L. 1992. Measurement of ice nucleation in lichens using thermal analysis. *Cryobiology*. **29**(3): 400-406.

Ashworth, E.N. Kieft, T.L. 1995. Ice nucleation activity associated with plants and fungi. *Biological Ice Nucleation and Its Applications*.137-162.

Atici, O. Nalbantoglu, B. 2003. Antifreeze proteins in higher plants. *Phytochemistry*. **64**(7): 1187-1196.

Baardsnes, J. Kondejewski, L.H. Hodges, R.S. Chao, H. Kay, C. *et al*. 1999. New ice-binding face for type I antifreeze protein. *FEBS Letters*. **463**(1): 87-91.

Baardsnes, J. Davies, P.L. 2001. Sialic acid synthase: the origin of fish type III antifreeze protein? *Trends in biochemical sciences*. **26**(8): 468-469.

Bailey, R.W. Connor, H.E. 1972. Structural polysaccharides in leaf blades and sheaths in the arundinoid grass *Chionochloa*. *New Zealand Journal of Botany*. **10**(4): 533-544.

Banin, A. Anderson, D.M. 1974. Effects of salt concentration changes during freezing on the unfrozen water content of porous materials. *Water Resources Research*. **10**(1): 124-128.

Bannister, P. Maegli, T. Dickinson, K.J. Halloy, S.R. Knight, A. *et al*. 2005. Will loss of snow cover during climatic warming expose New Zealand alpine plants to increased frost damage? *Oecologia*. **144**(2): 245-256.

Bannister, P. Ward, G. 1981. Changes in non-structural carbohydrate contents of *Chionochloa rigida* and *C. macra* during winter. *New Zealand Journal of Botany*. **19**(2): 233-239.

Bar, D.M. Braslavsky, I. Davies, P.L. 2016. Ice-binding proteins and their function. *Annual Review of Biochemistry*. **85**: 515-542.

Bar-Dolev, M. Braslavsky, I. Davies, P.L. 2016. Ice-Binding Proteins and Their Function. *Annual Review of Biochemistry*. **85**: 515-542.

Barker, N.P. Galley, C. Verboom, G.A. Mafa, P. Gilbert, M. *et al*. 2007. The phylogeny of the austral grass subfamily *Danthonioideae*: evidence from multiple data sets. *Plant Systematics and Evolution*. **264**(3-4): 135-156.

Barker, N.P. Linder, H.P. Morton, C.M. Lyle, M. 2003. The paraphyly of *Cortaderia* (*Danthonioideae*; *Poaceae*): evidence from morphology and chloroplast and nuclear DNA sequence data. *Annals of the Missouri Botanical Garden*. 1-24.

Barrett, J. 2001. Thermal hysteresis proteins. *The International Journal of Biochemistry & Cell Biology*. **33**(2): 105-117.

Beck, A.T. Steer, R.A. 1984. Internal consistencies of the original and revised beck depression inventory. *Journal of Clinical Psychology*. **40**(6): 1365-1367.

Beck, E. 1994. Cold tolerance in tropical alpine plants. Tropical alpine environments. Cambridge: *Cambridge University Press*.

Bigg, E.K. 1953. The supercooling of water. *Proceedings of the Physical Society. Section B*. **66**(8): 688.

Bossolini, E. Wicker, T. Knobel, P.A. Keller, B. 2007. Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. *The Plant Journal*. **49**(4): 704-717.

Braslavsky, I. Drori, R. 2013. LabVIEW-operated novel nanoliter osmometer for ice binding protein investigations. *Journal of Visualized Experiments: JoVE*. (72): e4189.

Bravo, L.A. Griffith, M. 2005. Characterization of antifreeze activity in Antarctic plants. *Journal of Experimental Botany*. **56**(414): 1189-1196.

Bredow, M. Vanderbeld, B. Walker, V.K. 2016. Knockdown of ice-binding proteins in *Brachypodium distachyon* demonstrates their role in freeze protection. *PloS one*. **11**(12): e0167941.

Bredow, M. Tomalty, H.E. Smith, L. Walker, V.K. 2017. Ice and anti-nucleating activities of an ice-binding protein from the annual grass, *Brachypodium distachyon*. *Plant, Cell & Environment*. **41**: 983-992. DOI: 10.1111/pce.12889.

Brenstrum, E. 1998. Cook and the weather. *Marine Observer*.120-124.

Briard, J.G. Poisson, J.S. Turner, T.R. Capicciotti, C.J. Acker, J.P. *et al*. 2016. Small molecule ice recrystallization inhibitors mitigate red blood cell lysis during freezing, transient warming and thawing. *Scientific Reports*. **6**: 23619.

Brown, C.S. Lee, M.S. Leung, D.W. Wang, T. Xu, W. *et al*. 2014. In silico derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity. *Journal of Molecular Biology*. **426**(10): 2045-2058.

Brown, M.B. Forsythe, A.B. 1974a. Robust tests for the equality of variances. *Journal of the American Statistical Association*. **69**: 364-367.

Brown, M.B. Forsythe, A.B. 1974b. The small sample behavior of some statistics which test the equality of several means. *Technometrics*. **16**: 129-132.

Brush, R.A. Griffith, M. Mlynarz, A. 1994. Characterization and quantification of intrinsic ice nucleators in winter rye (*Secale cereale*) leaves. *Plant Physiology*. **104**(2): 725-735.

Byard, S. Wisniewski, M. Li, J.H. Karlson, D. 2010. Interspecific Analysis of Xylem Freezing Responses in *Acer* and *Betula*. *HortScience*. **45**(1): 165-168.

Campbell, A.D. 1981. Flowering records for *Chionochloa*, *Aciphylla*, and *Celmisia* species in the Craigieburn Range, South Island, New Zealand. *New Zealand Journal of Botany*. **19**(1): 97-103.

Cannarozzi, G. Plaza-Wüthrich, S. Esfeld, K. Larti, S. Wilson, Y.S. *et al*. 2014. Genome and transcriptome sequencing identifies breeding targets in the orphan crop tef (*Eragrostis tef*). *BMC Genomics*. **15**(1): 581.

Capicciotti, C.J. Doshi, M. Ben, R.N. 2013. Ice recrystallization inhibitors: from biological antifreezes to small molecules. Recent Developments in the Study of Recrystallization. *InTech*. DOI: 10.5772/54992.

Carpenter, J.F. Hansen, T.N. 1992. Antifreeze protein modulates cell survival during cryopreservation: mediation through influence on ice crystal growth. *Proceedings of the National Academy of Sciences*. **89**(19): 8953-8957.

Chao, H. Davies, P.L. Carpenter, J.F. 1996. Effects of antifreeze proteins on red blood cell survival during cryopreservation. *Journal of Experimental Biology*. **199**(9): 2071-2076.

Chao, H. DeLuca, C.I. Davies, P.L. 1995. Mixing antifreeze protein types changes ice crystal morphology without affecting antifreeze activity. *FEBS Letters*. **357**(2): 183-186.

Chapin III, F.S. Follett, J.M. O'Connor, K.F. 1982. Growth, phosphate absorption, and phosphorus chemical fractions in two *Chionochloa* species. *The Journal of Ecology*. 305-321.

Chatterjee, M. Gupta, S. Bhar, A. Das, S. 2012. Optimization of an efficient protein extraction protocol compatible with two-dimensional electrophoresis and mass spectrometry from recalcitrant phenolic rich roots of chickpea (*Cicer arietinum L.*). *International Journal of Proteomics*. DOI: 10.1155/2012/536963.

Chen, K. Rajewsky, N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews. Genetics*. **8**(2): 93.

Chen, L. DeVries, A.L. Cheng, C.H. 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences*. **94**(8): 3811-3816.

Cheng, C.H. 1998. Evolution of the diverse antifreeze proteins. *Current Opinion in Genetics & Development*. **8**(6): 715-720.

Chiappetta, A. Muto, A. Bruno, L. Woloszynska, M. Van Lijsebettens, M. *et al*. 2015. A dehydrin gene isolated from feral olive enhances drought tolerance in *Arabidopsis* transgenic plants. *Frontiers in Plant Science*. **6**. DOI: 10.3389/fpls.2015.00392.

Christner, B.C. Morris, C.E. Foreman, C.M. Cai, R. Sands, D.C. 2008. Ubiquity of biological ice nucleators in snowfall. *Science*. **319**(5867): 1214-1214.

Chun, J.U. Yu, X.M. Griffith, M. 1998. Genetic studies of antifreeze proteins and their correlation with winter survival in wheat. *Euphytica*. **102**(2): 219-226.

Cilia, M. Fish, T. Yang, X. Mclaughlin, M. Thannhauser, T. *et al*. 2009. A comparison of protein extraction methods suitable for gel-based proteomic studies of aphid proteins. *Journal of biomolecular techniques: JBT*. **20**(4): 201.

Cingolani, P. Platts, A. Wang, L.L. Coon, M. Nguyen, T. *et al*. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila* melanogaster strain w1118; iso-2; iso-3. *Fly*. **6**(2): 80-92.

Clark, M.S. Thorne, M.A. Purac, J. Burns, G. Hillyard, G. *et al*. 2009. Surviving the cold: molecular analyses of insect cryoprotective dehydration in the Arctic springtail *Megaphorura arctica* (Tullberg). *BMC Genomics*. **10**(328): 1471-2164.

Clarke, W.E. Higgins, E.E. Plieske, J. Wieseke, R. Sidebottom, C. *et al*. 2016. A high-density SNP genotyping array for Brassica napus and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theoretical and Applied Genetics*. **129**(10): 1887-1899.

Coleman, N.J. Craig, D.Q.M. 1996. Modulated temperature differential scanning calorimetry: a novel approach to pharmaceutical thermal analysis. *International Journal of Pharmaceutics*. **135**(1-2): 13-29.

Conesa, A. Götz, S. García-Gómez, J.M. Terol, J. Talón, M. *et al*. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. **21**(18): 3674-3676.

Connor, H.E. 1991. *Chionochloa* Zotov (*Gramineae*) in New Zealand. *New Zealand Journal of Botany*. **29**: 219-282.

Connor, H.E. Bailey, R. O'Connor, K.F. 1970. Chemical composition of New Zealand tall-tussocks (*Chionochloa*). *New Zealand Journal of Agricultural Research*. **13**(3): 534-554.

Connor, H.E. Purdie, A.W. 1976. Triterpene methyl ether differentiation in *Chionochloa* (*Gramineae*). *New Zealand Journal of Botany*. **14**(4): 315-326.

Connor, H.E. Purdie, A.W. 1981. Triterpene methyl ethers in *Chionochloa* (*Gramineae*): Distribution in western South Island, New Zealand. *New Zealand Journal of Botany*. **19**(2): 161-170.

Constantinidou, H.A. Menkissoglu, O. 1992. Characteristics and importance of heterogeneous ice nuclei associated with citrus fruits. *Journal of Experimental Botany*. **43**(4): 585-591.

Constantinidou, H.A. Menkissoglu, O. Stergiadou, H.C. 1991. The role of ice nucleation activcteria in supercooling of citrus tissues. *Plant Physiology*. **81**: 548-554

Costa, S. Almeida, A. Castro, A. Domingues, L. 2014. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Frontiers in Microbiology*. **5**. DOI: 10.3389/fmicb.2014.00063.

Costanzo, J.P. Lee, R.E. 1996. Mini-review: ice nucleation in freeze-tolerant vertebrates. *CryoLetters*. **17**(2): 111-118.

Cox, M.P. Peterson, D.A. Biggs, P.J. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*. **11**(1): 485.

Dave, R.S. Mitra, R.K. 1998. A low temperature induced apoplastic protein isolated from *Arachis hypogaea*. *Phytochemistry*. **49**(8): 2207-2213.

Davies, P.L. 2014. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends in Biochemical Sciences*. **39**(11): 548-555.

Davies, P.L. Hew, C.L. 1990. Biochemistry of fish antifreeze proteins. *The FASEB Journal*. **4**(8): 2460-2468.

DeVries, A.L. Cheng, C.H. 2005. Antifreeze proteins and organismal freezing avoidance in polar fishes. *Fish Physiology*. **22**: 155-201.

DeVries, A.L. Komatsu, S.K. Feeney, R.E. 1970. Chemical and physical properties of freezing point-depressing glycoproteins from Antarctic fishes. *The Journal of Biological Chemistry*. **245**(11): 2901-2908.

Deng, G. Laursen, R.A. 1998. Isolation and characterization of an antifreeze protein from the longhorn sculpin, *Myoxocephalus octodecimspinosis*. *Biochimica et Biophysica Acta*. **1388**(2): 305-314.

Dawson, M.I. 1989. Contributions to a chromosome atlas of the New Zealand flora-30 miscellaneous species. *New Zealand Journal of Botany*. **27**(2): 163-165.

Debeljak, N. Feldman, L. Davis, K.L. Komel, R. Sytkowski, A.J. 2006. Variability in the immunodetection of His-tagged recombinant proteins. *Analytical Biochemistry*. **359**(2): 216-223.

Deininger, C.A. Mueller, G.M. Wolber, P.K. 1988. Immunological characterization of ice nucleation proteins from *Pseudomonas syringae*, *Pseudomonas fluorescens*, and *Erwinia herbicola*. *Journal of Bacteriology*. **170**(2): 669-675.

Deshmukh, R.K. Vivancos, J. Ramakrishnan, G. Guérin, V. Carpentier, G. *et al*. 2015. A precise spacing between the NPA domains of aquaporins is essential for silicon permeability in plants. *The Plant Journal*. **83**(3): 489-500.

Deswal, D. Gupta, R. Nandal, P. Kuhad, R.C. 2014. Fungal pretreatment improves amenability of lignocellulosic material for its saccharification to sugars. *Carbohydrate Polymers*. **99**: 264-269.

DeVries, A.L. Wohlschlag, D.E. 1969. Freezing resistance in some Antarctic fishes. *Science*. **163**(3871): 1073-1075.

Doucet, C.J. Byass, L. Elias, L. Worrall, D. Smallwood, M. *et al*. 2000. Distribution and characterization of recrystallization inhibitor activity in plant and lichen species from the UK and maritime Antarctic. *Cryobiology*. **40**(3): 218-227.

Dreischmeier, K. Budke, C. Wiehemeier, L. Kottke, T. Koop, T. 2017. Boreal pollen contain ice-nucleating as well as ice-binding 'antifreeze'polysaccharides. *Scientific Reports*. **7**: 41890.

Drori, R. Davies, P.L. Braslavsky, I. 2015. When are antifreeze proteins in solution essential for ice growth inhibition? *Langmuir*. **31**(21): 5805-5811.

Drummond, A.J. Suchard, M.A. Xie, D. Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*. **29**(8): 1969-1973.

Duarte, J. Rivière, N. Baranger, A. Aubert, G. Burstin, J. *et al*. 2014. Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. *BMC Genomics*. **15**(1): 126.

Duman, J.G. 1994. Purification and characterization of a thermal hysteresis protein from a plant, the bittersweet nightshade *Solanum dulcamara*. *Biochimica et Biophysica Acta*. **1206**(1): 129-135.

Duman, J.G. 2001. Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annual Review of Physiology*. **63**: 327-357.

Duman, J.G. Bennett, V. Sformo, T. Hochstrasser, R. Barnes, B.M. 2004. Antifreeze proteins in Alaskan insects and spiders. *Journal of Insect Physiology*. **50**(4): 259-266.

Duman, J.G. DeVries, A.L. 1976. Isolation, characterization, and physical properties of protein antifreezes from the winter flounder, *Pseudopleuronectes americanus*. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*. **54**(3): 375-380.

Duman, J.G. Olsen, T.M. 1993. Thermal hysteresis protein activity in bacteria, fungi, and phylogenetically diverse plants. *Cryobiology*. **30**(3): 322-328.

Earl, D. Bradnam, K. John, J.S. Darling, A. Lin, D. *et al*. 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Research*. **21**(12): 2224-2241.

Embuscado, M.E. BeMiller, J.N. Knox, E.B. 1996. A survey and partial characterization of iceucleating fluids secreted by giant-rosette (*Lobelia* and *Dendrosenecio*) plants of the mountains of eastern Africa. *Carbohydrate Polymers*. **31**(1-2): 1-9.

Eniade, A. Murphy, A.V. Landreau, G. Ben, R.N. 2001. A general synthesis of structurally diverse building blocks for preparing analogues of C-linked antifreeze glycoproteins. *Bioconjugate Chemistry*. **12**(5): 817-823.

Fei, Y.B. Wei, L.B. Gao, S.Q. Lu, M.C. Wang, B.H. *et al*. 2001. Isolation, purification and characterization of secondary structure of antifreeze protein from *Ammopiptanthus mongolicus*. *Chinese Science Bulletin*. **46**(6): 495-498.

Fletcher, G.L. Hew, C.L. Davies, P.L. 2001. Antifreeze proteins of teleost fishes. *Annual Review of Physiology*. **63**: 359-390.

Forests, N.I.I. 1976. A revised classification of the North Island indigenous forests. *The New Zealand Journal of Forestry*. **21**: 105.

Fröhlich-Nowoisky, J. Hill, T.C. Pummer, B.G. Yordanova, P. Franc, G.D. *et al*. 2015. Ice nucleation activity in the widespread soil fungus *Mortierella alpina*. *Biogeosciences*. **12**(4): 1057-1071.

Franken, K.L.M.C. Hiemstra, H.S. van Meijgaarden, K.E. Subronto, Y. Den Hartigh, J. *et al*. 2000. Purification of his-tagged proteins by immobilized chelate affinity chromatography: the benefits from the use of organic solvent. *Protein Expression and Purification*. **18**(1): 95-99.

Franks, E. 1985. Biophysics and biochemistry at low temperatures. Cambridge: *Cambridge University Press*.

Freire, E. 1995. Differential scanning calorimetry. *Protein Stability and Folding: Theory and Practice*. 191-218.

Friis, D.S. Johnsen, J.L. Kristiansen, E. Westh, P. Ramlov, H. 2014. Low thermodynamic but high kinetic stability of an antifreeze protein from *Rhagium mordax*. *Protein Science*. **23**(6): 760-768.

Jozefczuk, J. Adjaye, J. 2011. Quantitative real-time PCR-based analysis of gene expression. *Methods Enzymol*ogy. **500**: 99-109.

Gamar, Y.A. Kiambi, D. Kairichi, M. Kyallo, M. Elgada, M.H. 2013. Assessment of genetic diversity and structure of *sudanese sorghum* accessions using Simple Sequence Repeat (SSRs) Markers. *Greener Journal of Plant Breeding and Crop Sciences*. **1**: 16-24.

Gao, J.P. Wang, D. Cao, L.Y. Sun, H.F. 2015. Transcriptome sequencing of *Codonopsis pilosula* and identification of candidate genes involved in polysaccharide biosynthesis. *PloS One*. **10**(2): e0117342.

Garcia-Bañuelos, M. Gardea, A. Winzerling, J. Vazquez-Moreno, L. 2009. Characterization of a midwinter-expressed dehydrin (DHN) gene from apple trees *(Malus domestica*). *Plant Molecular Biology Reporter*. **27**(4): 476-487.

Garnham, C.P. Campbell, R.L. Davies, P.L. 2011a. Anchored clathrate waters bind antifreeze proteins to ice. *Proceedings of the National Academy of Sciences*. **108**(18): 7363-7367.

Garnham, C.P. Campbell, R.L. Walker, V.K. Davies, P.L. 2011b. Novel dimeric β-helical model of an ice nucleation protein with bridged active sites. *BMC Structural Biology*. **11**(1): 36.

Gilbert, J.A. Davies, P.L. Laybourn-Parry, J. 2005. A hyperactive, Ca^{2+}-dependent antifreeze protein in an Antarctic bacterium. *FEMS Microbiology Letters*. **245**(1): 67-72.

Gilbert, J.A. Hill, P.J. Dodd, C.E. Laybourn-Parry, J. 2004. Demonstration of antifreeze protein activity in Antarctic lake bacteria. *Microbiology*. **150**(1): 171-180.

Gillard, G.B. Garama, D.J. Brown, C.M. 2014. The transcriptome of the NZ endemic sea urchin Kina (*Evechinus chloroticus*). *BMC Genomics*. **15**(1): 45.

Goldstein, G. Nobel, P.S. 1994. Water relations and low-temperature acclimation for cactus species varying in freezing tolerance. *Plant Physiology*. **104**(2): 675-681.

Goto, M. Komaba, M. Horikawa, T. Nakamura, N. 1993. The role of ice nucleation-active bacteria on frost damage of tea plants. *Japanese Journal of Phytopathology*. **59**(5): 535-543.

Grønvold, L. 2013. Analysis of homeolog specific transcription in hexaploid wheat. Master thesis, Norwegian University of Life Sciences.

Graether, S.P. Jia, Z.C. 2001. Modeling *Pseudomonas syringae* ice-nucleation protein as a β-helical protein. *Biophysical Journal*. **80**(3): 1169-1173.

Graether, S.P. Kuiper, M.J. Gagne, S.M. Walker, V.K. Jia, Z. *et al*. 2000. Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature*. **406**(6793): 325-328.

Graether, S.P. Sykes, B.D. 2004. Cold survival in freeze-intolerant insects: the structure and function of beta-helical antifreeze proteins. *The FEBS Journal*. **271**(16): 3285-3296.

Graham, I.J. 2008. A continent on the move: New Zealand geoscience into the 21st century. Wellington: *Geological Society of New Zealand*.

Graham, L.A. Hobbs, R.S. Fletcher, G.L. Davies, P.L. 2013. Helical antifreeze proteins have independently evolved in fishes on four occasions. *PLoS One*. **8**(12): e81285.

Graham, L.A. Marshall, C.B. Lin, F.H. Campbell, R.L. Davies, P.L. 2008b. Hyperactive antifreeze protein from fish contains multiple ice-binding sites. *Biochemistry*. **47**(7): 2051-2063.

Greeff, C. Roux, M. Mundy, J. Petersen, M. 2012. Receptor-like kinase complexes in plant innate immunity. *Frontiers in Plant Science*. **3**. DOI: 10.3389/fpls.2012.00209.

Greer, D.H. 1979. Effects of long term pre conditioning on growth and flowering of some snow tussock *Chionochloa* spp populations in otago new zealand. *Australian Journal of Botany*. 617-630.

Griffith, M. Ala, P. Yang, D.S. Hon, W.C. Moffatt, B.A. 1992. Antifreeze protein produced endogenously in winter rye leaves. *Plant Physiology*. **100**(2): 593-596.

Griffith, M. Antikainen, M. 1996. Extracellular ice formation in freezing-tolerant plants. *Advances in Low-Temperature Biology.* London: *JAI Press, Inc.*

Griffith, M. Lumb, C. Wiseman, S.B. Wisniewski, M. Johnson, R.W. *et al*. 2005. Antifreeze proteins modify the freezing process in planta. *Plant Physiology*. **138**(1): 330-340.

Griffith, M. Yaish, M.W. 2004. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends in Plant Science*. **9**(8): 399-405.

Groen, S.C. Purugganan, M.D. 2016. Systems genetics of plant adaptation to environmental stresses. *American Journal of Botany*. **103**(12): 2019-2021.

Gronwald, W. Loewen, M.C. Lix, B. Daugulis, A.J. Sonnichsen, F.D. *et al*. 1998. The solution structure of type II antifreeze protein reveals a new member of the lectin family. *Biochemistry*. **37**(14): 4712-4721.

Gross, D.C. Proebsting, E.J. MacCrindale-Zimmerman, H. 1988. Development, distribution, and characteristics of intrinsic, nonbacterial ice nuclei in *Prunus* wood. *Plant Physiology*. **88**: 915-922

Guo, S. Garnham, C.P. Whitney, J.C. Graham, L.A. Davies, P.L. 2012. Re-evaluation of a bacterial antifreeze protein as an adhesin with ice-binding activity. *PLoS One*. **7**(11): e48805. DOI:10.1371/journal.pone.0048805.

Guo, S. Stevens, C.A. Vance, T.D. Olijve, L.L. Graham, L.A. *et al*. 2017. Structure of a 1.5-MDa adhesin that binds its Antarctic bacterium to diatoms and ice. *Science advances*. **3**(8): e1701440.

Gupta, R. Deswal, R. 2014a. Antifreeze proteins enable plants to survive in freezing conditions. *Journal of Biosciences*. **39**(5): 931-944.

Gupta, R. Deswal, R. 2014b. Refolding of β-stranded class I chitinases of *Hippophae rhamnoides* enhances the antifreeze activity during cold acclimation. *PLoS One*. **9**(3): e91723.

Gurian-Sherman, D. Lindow, S.E. 1993. Bacterial ice nucleation: significance and molecular basis. *The FASEB Journal*. **7**(14): 1338-1343.

Haas, B.J. Papanicolaou, A. Yassour, M. Grabherr, M. Blood, P.D. *et al*. 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*. **8**(8). DOI: 10.1038/nprot.2013.084.

Hacker, J. Ladinig, U. Wagner, J. Neuner, G. 2011. Inflorescences of alpine cushion plants freeze autonomously and may survive sub-zero temperatures by supercooling. *Plant Science*. **180**(1): 149-156.

Hacker, J. Neuner, G. 2008. Ice propagation in dehardened alpine plant species studied by infrared differential thermal analysis (IDTA). *Arctic, Antarctic, and Alpine Research*. **40**(4): 660-670.

Hadjiargyrou, M. Zhi, J. Komatsu, D.E. 2016. Identification of the microRNA transcriptome during the early phases of mammalian fracture repair. *Bone*. **87**: 78-88.

Hansen, T.N. Baust, J. 1988. Serial dilution of *Tenebrio molitor* haemolymph: analysis of antifreeze activity by differential scanning calorimetry. *CryoLetters*. **9**: 386-391.

Hartmann, S. Augustin, S. Clauss, T. Wex, H. Šantl-Temkiv, T. *et al*. 2013. Immersion freezing of ice nucleation active protein complexes. *Atmospheric Chemistry and Physics*. **13**(11): 5751-5766.

Hassas-Roudsar, M. 2011. Extraction, purification and study of mechanism of action of apoplastic ice structuring proteins from cold acclimated winter wheat leaves. PhD thesis, The University of Guelph

Hassas-Roudsari, M. Goff, H.D. 2012. A new quantitative method to measure activity of ice structuring proteins using differential scanning calorimetry. *Cryoletters*. **33**(2): 117-124.

Heenan, P.B. McGlone, M.S. 2013. Evolution of New Zealand alpine and open-habitat plant species during the late Cenozoic. *New Zealand Journal of Ecology*. **37**(1): 105-113.

Hengherr, S. Worland, M.R. Reuner, A. Brümmer, F. Schill, R.O. 2009. Freeze tolerance, supercooling points and ice formation: comparative studies on the sub-zero temperature survival of limno-terrestrial tardigrades. *Journal of Experimental Biology*. **212**(6): 802-807.

Henneman, L. van Cruchten, A.G. Denis, S.W. Amolins, M.W. Placzek, A.T. *et al*. 2008. Detection of non-sterol isoprenoids by HPLC-MS/MS. *Analytical Biochemistry*. **383**(1): 18-24.

Hew, C.L. Wang, N.C. Joshi, S. Fletcher, G.L. Scott, G.K. *et al*. 1988. Multiple genes provide the basis for antifreeze protein diversity and dosage in the ocean pout, *Macrozoarces americanus*. *Journal of Biological Chemistry*. **263**(24): 12049-12055.

Hew, C.L. Yang, D.S.C. 1992. Protein interaction with ice. *The FEBS Journal*. **203**(1-2): 33-42.

Hiilovaara-Teijo, M. Hannukkala, A. Griffith, M. Yu, X.-M. Pihakaski-Maunsbach, K. 1999. Snow-mold-induced apoplastic proteins in winter rye leaves lack antifreeze activity. *Plant Physiology*. **121**(2): 665-674.

Hiilovaara-Teijo, M. Palva, E.T. 1999. Molecular responses in cold-adapted plants. *Cold-Adapted Organisms.* Berlin: *Springer*.

Hill, T.C.J. Moffett, B.F. DeMott, P.J. Georgakopoulos, D.G. Stump, W.L. *et al*. 2014. Measurement of ice nucleation-active bacteria on plants and in precipitation by quantitative PCR. *Applied and Environmental Microbiology*. **80**(4): 1256-1267.

Hoch, G. Richter, A. Körner, C. 2003. Non-structural carbon compounds in temperate forest trees. *Plant, Cell & Environment*. **26**(7): 1067-1081.

Hodel, R.G.J. Segovia-Salcedo, M.C. Landis, J.B. Crowl, A.A. Sun, M. *et al*. 2016. The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*. **4**(6): 1600025.

Hon, W.C. Griffith, M. Mlynarz, A. Kwok, Y.C. Yang, D.S.C. 1995. Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiology*. **109**(3): 879-889.

Hsiao, C. Jacobs, S.W.L. Barker, N.P. Chatterton, N.J. 1998. A molecular phylogeny of the subfamily *Arundinoideae* (*Poaceae*) based on sequences of rDNA. *Australian Systematic Botany*. **11**(1): 41-52.

Huang, T. 1995. Purification and characterization of thermal hysteresis protein from cold-acclimated kale, *Brasica oleracea*. *Cryobiology*. **32**: 577.

Huang, T. Duman, J.G. 2002. Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamar*a. *Plant Molecular Biology*. **48**(4): 339-350.

Hugh, W. 2012. Alpine plant. *Te Ara - the Encyclopedia of New Zealand.* Available: https://teara.govt.nz/en/alpine-plants

Humphreys, A.M. Linder, H.P. 2013. Evidence for recent evolution of cold tolerance in grasses suggests current distribution is not limited by (low) temperature. *New Phytologist*. **198**(4): 1261-1273.

Ingraham, N.L. Mark, A.F. Frew, R.D. 2008. Fog deposition by snow tussock grassland on the Otago uplands: response to a recent review of the evidence. *Journal of Hydrology (New Zealand)*. **47**(2): 107.

Ishikawa, M. Ishikawa, M. Toyomasu, T. Aoki, T. Price, W.S. 2015. Ice nucleation activity in various tissues of *Rhododendron* flower buds: their relevance to extraorgan freezing. *Frontiers in Plant Science*. **6**. DOI: 10.3389/fpls.2015.00149.

Jetter, R. Riederer, M. 1996. Cuticular waxes from the leaves and fruit capsules of eight *Papaveraceae* species. *Canadian Journal of Botany*. **74**(3): 419-430.

Jetter, R. Riederer, M. 2016. Localization of the transpiration barrier in the epi and intracuticular waxes of eight plant species: water transport resistances are associated with fatty acyl rather than alicyclic components. *Plant Physiology*. **170**(2): 921-934.

Jia, Z. Davies, P.L. 2002. Antifreeze proteins: an unusual receptor-ligand interaction. *Trends in Biochemistry Scinence*. **27**(2): 101-106.

Jin, J.P. Zhang, H. Kong, L. Gao, G. Luo, J.C. 2013. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research*. **42**(D1): 1182-1187.

John, U.P. Polotnianka, R.M. Sivakumaran, K.A. Chew, O. Mackin, L. *et al*. 2009. Ice recrystallization inhibition proteins (IRIPs) and freeze tolerance in the cryophilic Antarctic hair grass *Deschampsia antarctica E. Desv*. *Plant, Cell & Environment*. **32**(4): 336-348.

Jurka, J. Kapitonov, V.V. Pavlicek, A. Klonowski, P. Kohany, O. *et al*. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*. **110**(1-4): 462-467.

Kaida, R. Serada, S. Norioka, N. Norioka, S. Neumetzler, L. *et al*. 2010. Potential role for purple acid phosphatase in the dephosphorylation of wall proteins in tobacco cells. *Plant Physiology*. **153**(2): 603-610.

Kanehisa, M. Goto, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. **28**(1): 27-30.

Kanehisa, M. Sato, Y. Kawashima, M. Furumichi, M. Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*. **44**(D1): 457-462.

Karlson, D.T. Xiang, Q.Y. Stirm, V.E. Shirazi, A.M. Ashworth, E.N. 2004. Phylogenetic analyses in cornus substantiate ancestry of xylem supercooling freezing behavior and reveal lineage of desiccation related proteins. *Plant Physiology*. **135**(3): 1654-1665.

Kawahara, H. Mano, Y. Hamada, R. Obata, H. 1994. Role of spermidine in the ice-nucleating activity of the EIM from *Erwinia uredovora* KUIN-3. *Bioscience, Biotechnology, and Biochemistry*. **58**(12): 2201-2206.

Kawahara, H. Tanaka, Y. Obata, H. 1995. Isolation and characterization of a novel ice-nucleating bacterium, *Pseudomonas* sp. KUIN-4, which has stable activity in acidic solution. *Bioscience, Biotechnology, and Biochemistry*. **59**(8): 1528-1532.

Kellogg, E.A. 2014. Flowering Plants: Monocots - *Poaceae*. *Springer*.

Kent, W.J. 2002. BLAT - the BLAST-like alignment tool. *Genome Research*. **12**(4): 656-664.

Kieft, T.L. 1988. Ice nucleation activity in lichens. *Applied and Environmental Microbiology*. **54**(7): 1678-1681.

Kieft, T.L. Ruscetti, T. 1990. Characterization of biological ice nuclei from a lichen. . *Journal of Bacteriology*. **72**: 3519-3523.

Kishimoto, T. Yamazaki, H. Saruwatari, A. Murakawa, H. Sekozawa, Y. *et al*. 2014. High ice nucleation activity located in blueberry stem bark is linked to primary freeze initiation and adaptive freezing behaviour of the bark. *AoB Plants*. **6**: plu044. DOI: 10.1093/aobpla/plu044.

Knight, C.A. DeVries, A.L. Oolman, L.D. 1984. Fish antifreeze protein and the freezing and recrystallization of ice. *Nature*. **308**(5956): 295-296.

Knight, C.A. Hallett, J. DeVries, A.L. 1988. Solute effects on ice recrystallization: An assessment technique. *Cryobiology*. **25**(1): 55-60.

Knight, C.A. Wen, D.Y. Laursen, R.A. 1995. Nonequilibrium antifreeze peptides and the recrystallization of ice. *Cryobiology*. **32**(1): 23-34.

Kobashigawa, Y. Nishimiya, Y. Miura, K. Ohgiya, S. Miura, A. *et al*. 2005. A part of ice nucleation protein exhibits the ice-binding ability. *FEBS letters*. **579**(6): 1493-1497.

Kothapalli, K.S.D. Ye, Kaixiong Gadgil, M.S. Carlson, S.E. O'Brien, K.O. *et al*. 2016. Positive selection on a regulatory insertion–deletion polymorphism in FADS2 influences apparent endogenous synthesis of arachidonic acid. *Molecular Biology and Evolution*. **33**(7): 1726-1739.

Kozloff, L.M. Schofield, M.A. Lute, M. 1983. Ice nucleating activity of *Pseudomonas syringae* and *Erwinia herbicola*. *Journal of Bacteriology*. **153**(1): 222-231.

Krna, M.A. 2015. Productivity, decomposition and carbon sequestration of *Chionochloa* species across altitudinal gradients in montane tussock grasslands of New Zealand. Doctoral dissertation, Massey University.

Krog, J. Zachariassen, K. Larsen, B. Smidsrod, O. 1979. Thermal buffering in Afroalpine plants due to nucleating agent-induced water freezing. *Nature*. **282**: 300-301

Kruip, J. Boekema, E.J. Bald, D. Boonstra, A.F. Rögner, M. 1993. Isolation and structural characterization of monomeric and trimeric photosystem I complexes (P700. FA/FB and P700. FX) from the cyanobacterium Synechocystis PCC 6803. *Journal of Biological Chemistry*. **268**(31): 23353-23360.

Kuiper, H.A. Kleter, G.A. Noteborn, H.P. Kok, E.J. 2001. Assessment of the food safety issues related to genetically modified foods. *The Plant Journal*. **27**(6): 503-528.

Kuiper, M.J. Lankin, C. Gauthier, S.Y. Walker, V.K. Davies, P.L. 2003. Purification of antifreeze proteins by adsorption to ice. *Biochemical and Biophysical Research Communications*. **300**(3): 645-648.

Kukal, O. Duman, J.G. Serianni, A.S. 1989. Cold-induced mitochondrial degradation and cryoprotectant synthesis in freeze-tolerant arctic caterpillars. *Journal of Comparative Physiology*. **158**(6): 661-671.

Kumble, K.D. Demmer, J. Fish, S. Hall, C. Corrales, S. *et al*. 2008. Characterization of a family of ice-active proteins from the Ryegrass, *Lolium perenne*. *Cryobiology*. **57**(3): 263-268.

Kuwabara, C. Imai, R. 2009. Molecular basis of disease resistance acquired through cold acclimation in overwintering plants. *Journal of Plant Biology*. **52**(1): 19-26.

Laing, W. Christeller, J. 2004. Extraction of proteins from plant tissues. *Current Protocols in Protein Science*. **38**(4.7): 4.7.1-4.7.7.

Landis, C.A. Campbell, H.J. Begg, J.G. Mildenhall, D.C. Paterson, A.M. *et al*. 2008. The Waipounamu Erosion Surface: questioning the antiquity of the New Zealand land surface and terrestrial fauna and flora. *Geological Magazine*. **145**(02): 173-197.

Lang, V. Usadel, B. Obermeyer, G. 2015. De novo sequencing and analysis of the lily pollen transcriptome: an open access data source for an orphan plant species. *Plant Molecular Biology*. **87**(1-2): 69-80.

Langmead, B. Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. **9**(4): 357-359.

Larkin, M.A. Blackshields, G. Brown, N.P. Chenna, R. McGettigan, P.A. *et al*. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**(21): 2947-2948.

Lauersen, K.J. Brown, A. Middleton, A. Daviesa, P.L. Walker, V.K. 2011. Expression and characterization of an antifreeze protein from the perennial rye grass, *Lolium perenne*. *Cryobiology*. **62**(3): 194-201.

Lee, M.S. Hseu, Y.C. Lai, G.H. Chang, W.T. Chen, H.J. *et al*. 2011. High yield expression in a recombinant *E. coli* of a codon optimized chicken anemia virus capsid protein VP1 useful for vaccine development. *Microbial Cell Factories*. **10**(1): 56.

Lee, W.G. Fenner, M. 1989. Mineral nutrient allocation in seeds and shoots of twelve *Chionochloa* species in relation to soil fertility. *Journal of Ecology*. **77**(3): 704-716.

Levitt, J. 1980. Responses of plants to environmental stress. Chilling, freezing, and high temperature stress. New York: *Academic Press*.

Li, C. Rudi, H. Stockinger, E.J. Cheng, H. Cao, M. *et al*. 2012a. Comparative analyses reveal potential uses of *Brachypodium distachyon* as a model for cold stress responses in temperate grasses. *BMC Plant Biology*. **12**(65): 1471-2229.

Li, H. Handsaker, B. Wysoker, A. Fennell, T. Ruan, J. *et al*. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. **25**(16): 2078-2079.

Li, H. Huang, G. Meng, Q. Ma, L. Yuan, L. *et al*. 2011a. Integrated soil and plant phosphorus management for crop and environment in China. A review. *Plant and Soil*. **349**(1-2): 157-167.

Li, L. Gyun Kang, D. Joon Cha, H. 2004. Functional display of foreign protein on surface of *Escherichia coli* using N-terminal domain of ice nucleation protein. *Biotechnology and Bioengineering*. **85**(2): 214-221.

Li, Q. Yan, Q. Chen, J. He, Y. Wang, J. *et al*. 2012b. Molecular characterization of an ice nucleation protein variant (InaQ) from *Pseudomonas syringae* and the analysis of its transmembrane transport activity in *Escherichia coli*. *International Journal of Biological Sciences*. **8**(8): 1097-1108.

Li, Z.X. Gao, Q. Liu, Y.Z. He, C.M. Zhang, X.R. *et al*. 2011b. Overexpression of transcription factor ZmPTF1 improves low phosphate tolerance of maize by regulating carbon metabolism and root growth. *Planta*. **233**(6): 1129-1143.

Lim, Z.L. Low, N.H. Moffatt, B.A. Gray, G.R. 2013. Gelation in protein extracts from cold acclimated and non-acclimated winter rye (*Secale cereale L.* cv Musketeer). *Cryobiology*. **66**(2): 156-166.

Lindow, S.E. 1982. Population dynamics of epiphytic ice nucleation active bacteria on frost sensitive plants and frost control by means of antagonistic bacteria. *Plant Cold Hardiness and Freezing Stress. Academic Press*.

Lindow, S.E. 1983. The role of bacterial ice nucleation in frost injury to plants. *Annual Review of Phytopathology*. **21**(1): 363-384.

Lindow, S.E. Arny, D.C. Upper, C.D. 1978. Distribution of ice nucleation-active bacteria on plants in nature. *Applied and Environmental Microbiology*. **36**(6): 831-838.

Liou, Y.C. Tocilj, A. Davies, P.L. Jia, Z.C. 2000. Mimicry of ice structure by surface hydroxyls and water of a β-helix antifreeze protein. *Nature*. **406**(6793): 322-324.

Lockhart, J. 2015. Uncovering the unexpected site of biosynthesis of a major cell wall component in grasses. *The Plant Cell*. **27**: 483-483.

Lorv, J.S.H. Rose, D.R. Glick, B.R. 2014. Bacterial ice crystal controlling proteins. *Scientifica (Cairo)*. DOI: 10.1155/2014/976895.

Lu, M. Wang, B. Li, Z. Fei, Y. Wei, L. *et al*. 2002. Differential scanning calorimetric and circular dichroistic studies on plant antifreeze proteins. *Journal of Thermal Analysis and Calorimetry*. **67**(3): 689-698.

Lu, X. Zhou, H. Pan, Y.B. Chen, C.Y. Zhu, J.R. *et al*. 2015. Segregation analysis of microsatellite (SSR) markers in sugarcane polyploids. *Genetics and Molecular Research*. **14**: 18384-18395.

Müller, M. Munné-Bosch, S. 2015. Ethylene response factors: a key regulatory hub in hormone and stress signaling. *Plant Physiology*. **169**(1): 32-41.

Müller, S. Salomo, K. Salazar, J. Naumann, J. Jaramillo, M.A. *et al*. 2015. Intercontinental long-distance dispersal of Canellaceae from the New to the Old World revealed by a nuclear single copy gene and chloroplast loci. *Molecular Phylogenetics and Evolution*. **84**: 205-219.

Maki, L.R. Galyan, E.L. Chang-Chien, M.M. Caldwell, D.R. 1974. Ice nucleation induced by *Pseudomonas syringae*. *Applied microbiology*. **28**(3): 456-459.

Manna, S. 2015. An overview of pentatricopeptide repeat proteins and their applications. *Biochimie*. **113**: 93-99.

Mark, A.F. 1965a. The environment and growth rate of narrow-leaved snow tussock, *Chionochloa rigida*, in Otago. *New Zealand Journal of Botany*. **3**(2): 73-103.

Mark, A.F. 1965b. Flowering, seeding, and seedling establishment of narrow-leaved snow tussock, *Chionochloa rigida*. *New Zealand Journal of Botany*. **3**(3): 180-193.

Mark, A.F. 2012. Grasslands - Types of tussock. *Te Ara - the Encyclopedia of New Zealand.* Available: https://teara.govt.nz/en/grasslands.

Mark, A.F. Adams, N. 1996. Alpine plants of New Zealand. Wellington *A.H. & A.W. Reed*.

Mark, A.F. Dickinson, K.J.M. Hofstede, R.G.M. 2000. Alpine vegetation, plant distribution, life forms, and environments in a perhumid New Zealand region: oceanic and tropical high mountain affinities. *Arctic, Antarctic, and Alpine Research*. 240-254.

Marquez, Y. Höpfler, M. Ayatollahi, Z. Barta, A. Kalyna, M. 2015. Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Research*. **25**(7): 995-1007.

Marshall, C.J. Basu, K. Davies, P.L. 2016. Ice-shell purification of ice-binding proteins. *Cryobiology*. **72**(3): 258-263.

Martinez, N.J. Gregory, R.I. 2010. MicroRNA gene regulatory pathways in the establishment and maintenance of ESC identity. *Cell Stem Cell*. **7**(1): 31-35.

Mawlong, I. Ali, K. Srinivasan, R. Rai, R.D. Tyagi, A. 2015. Functional validation of a drought-responsive AP2/ERF family transcription factor-encoding gene from rice in *Arabidopsis*. *Molecular Breeding*. **35**(8): 163.

Mazarei, M. Kerr, A. 1987. Identification and seasonal population changes of ice nucleation active bacteria [*Pseudomonas*] in a pea field in South Australia. *Australasian Plant Pathology*. **16**(1): 13-15.

McCall, A.C. Kelly, D. Chapman, H.M. 2004. Little geographic or host plant genetic variation in a *Chionochloa* (*Poaceae*) seed predator (*Cecidomyiidae*: undescribed species). *New Zealand Journal of Ecology*. 215-224.

McGlone, M.S. Perry, G.L.W. Houliston, G.J. Connor, H.E. 2014. Fire, grazing and the evolution of New Zealand grasses. *New Zealand Journal of Ecology*. 1-11.

McKeown, M. Schubert, M. Marcussen, T. Fjellheim, S. Preston, J.C. 2016. Evidence for an early origin of vernalization responsiveness in temperate *Pooideae* grasses. *Plant Physiology*. **172**(1): 416-426.

Meijenfeldt, N. 2010. Unraveling the cold response in *Draba*. PhD thesis, Universiteit van Amsterdam.

Meyer, K. Keil, M. Naldrett, M.J. 1999. A leucine-rich repeat protein of carrot that exhibits antifreeze activity. *FEBS letters*. **447**(2-3): 171-178.

Middleton, A.J. Brown, A.M. Davies, P.L. Walker, V.K. 2009. Identification of the ice-binding face of a plant antifreeze protein. *FEBS letters*. **583**(4): 815-819.

Minic, Z. Jouanin, L. 2006. Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiology and Biochemistry*. **44**(7): 435-449.

Mizrahy, O. Bar-Dolev, M. Guy, S. Braslavsky, I. 2013. Inhibition of ice growth and recrystallization by zirconium acetate and zirconium acetate hydroxide. *PLoS One*. **8**(3): e59540.

Moffet, B. Getti, G. Henderson-Begg, S. Hill, T. 2015. Ubiquity of ice nucleation in lichen–possible atmospheric implications. *Lindbergia*. **38**: 39-43.

Mohr, G. Perlman, P.S. Lambowitz, A.M. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Research*. **21**(22): 4991-4997.

Morris, C.E. Georgakopoulos, D.G. Sands, D.C. 2004. Ice nucleation active bacteria and their potential role in precipitation. *Journal de Physique IV (Proceedings)*. 87-103.

Mugambi, S.M. 1971. Edaphic adaptation in species of *Chionochloa*. Doctoral dissertation, Lincoln College, University of Canterbury.

Muldrew, K. McGann, L.E. 1994. The osmotic rupture hypothesis of intracellular freezing injury. *Biophysical Journal*. **66**(2): 532.

Mullet, J.E. Klein, P.G. Klein, R.R. 1990. Chlorophyll regulates accumulation of the plastid-encoded chlorophyll apoproteins CP43 and D1 by increasing apoprotein stability. *Proceedings of the National Academy of Sciences*. **87**(11): 4038-4042.

Murray, B.G. De Lange, P.J. Ferguson, A.R. 2005. Nuclear DNA variation, chromosome numbers and polyploidy in the endemic and indigenous grass flora of New Zealand. *Annals of Botany*. **96**(7): 1293-1305.

Murray, B.J. Broadley, S.L. Wilson, T.W. Bull, S.J. Wills, R.H. *et al*. 2010. Kinetics of the homogeneous freezing of water. *Physical Chemistry Chemical Physics*. **12**(35): 10380-10387.

Murray, B.J. Ross, J.F. Whale, T.F. Price, H.C. Atkinson, J.D. *et al*. 2015. The relevance of nanoscale biological fragments for ice nucleation in clouds. *Scientific Reports*. **5**: 8082.

Muryoi, N. Sato, M. Kaneko, S. Kawahara, H. Obata, H. *et al*. 2004. Cloning and expression of afpA, a gene encoding an antifreeze protein from the arctic plant growth-promoting rhizobacterium *Pseudomonas putida* GR12-2. *Journal of Bacteriology*. **186**(17): 5661-5671.

Nakasugi, K. Crowhurst, R.N. Bally, J. Wood, C.C. Hellens, R.P. *et al*. 2013. De novo transcriptome sequence assembly and analysis of RNA silencing genes of *Nicotiana benthamiana*. *PLoS One*. **8**(3): e59534.

Nejad, P. 2005. Pathogenic and ice-nucleation active (INA) bacteria causing dieback of willows in short rotation forestry. Doctoral thesis, Swedish University of Agricultural Sciences.

Ng, N.F. Hew, C.L. 1992. Structure of an antifreeze polypeptide from the sea raven. Disulfide bonds and similarity to lectin-binding proteins. *Journal of Biological Chemistry*. **267**(23): 16069-16075.

Nishimiya, Y. Kondo, H. Takamichi, M. Sugimoto, H. Suzuki, M. *et al*. 2008. Crystal structure and mutational analysis of Ca2+-independent type II antifreeze protein from longsnout poacher, *Brachyopsis rostratus*. *Journal of molecular biology*. **382**(3): 734-746.

O'Brien, R.D. Lindow, S.E. 1988. Effect of plant species and environmental conditions on ice nucleation activity of *Pseudomonas syringae* on leaves. *Applied and Environmental Microbiology*. **54**(9): 2281-2286.

Olijve, L.L.C. Hendrix, M.M. Voets, I.K. 2016. Influence of polymer chain architecture of poly (vinyl alcohol) on the inhibition of ice recrystallization. *Macromolecular Chemistry and Physics*. **217**(8): 951-958.

Orians, C.M. 2000. The effects of hybridization in plants on secondary chemistry: implications for the ecology and evolution of plant–herbivore interactions. *American Journal of Botany*. **87**(12): 1749-1756.

Parody-Morreale, A. Bishop, G. Fall, R. Gill, S.J. 1986. A differential scanning calorimeter for ice nucleation distribution studies—Application to bacterial nucleators. *Analytical Biochemistry*. **154**(2): 682-690.

Parody-Morreale, A. Murphy, K.P. Di Cera, E. Fall, R. DeVries, A.L. Gill, S.J. 1988. Inhibition of bacterial ice nucleators by fish antifreeze glycoproteins. *Nature*. **333**(6175): 782.

Parra, G. Bradnam, K. Korf, I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. **23**(9): 1061-1067.

Paschke, D. 2017. Cold tolerance mechanisms of the New Zealand alpine wētā, *Hemideina maori*. PhD thesis, University of Otago.

Payton, I.J. Brasch, D.J. 1978. Growth and nonstructural carbohydrate reserves in *Chionochloa rigida* and *C. macra*, and their short-term response to fire. *New Zealand Journal of Botany*. **16**(4): 435-460.

Peng, C.Y. 2007. How mergers may affect the mass scaling relation between gravitationally bound systems. *The Astrophysical Journal*. **671**(2): 1098.

Persans, M.W. Wang, J. Schuler, M.A. 2001. Characterization of maize cytochrome P450 monooxygenases induced in response to safeners and bacterial pathogens. *Plant Physiology*. **125**(2): 1126-1138.

Pirie, M.D. Humphreys, A.M. Galley, C. Barker, N.P. Verboom, G.A. *et al*. 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Molecular Phylogenetics and Evolution*. **48**(3): 1106-1119.

Pirie, M.D. Lloyd, K.M. Lee, W.G. Linder, H.P. 2010. Diversification of *Chionochloa* (*Poaceae*) and biogeography of the New Zealand Southern Alps. *Journal of Biogeography*. **37**(2): 379-392.

Preston, J.C. Sandve, S.R. 2013. Adaptation to seasonality and the winter freeze. *Frontiers in Plant Science*. **4**(167). DOI: 10.3389/fpls.2013.00167.

Pudney, P.D.A. Buckley, S.L. Sidebottom, C.M. Twigg, S.N. Sevilla, M.P. *et al*. 2003. The physico-chemical characterization of a boiling stable antifreeze protein from a perennial grass (*Lolium perenne*). *Archives of Biochemistry and Biophysics*. **410**(2): 238-245.

Pummer, B.G. Budke, C. Augustin-Bauditz, S. Niedermeier, D. Felgitsch, L. *et al*. 2015. Ice nucleation by water-soluble macromolecules. *Atmospheric Chemistry and Physics*. **15**(8): 4077-4091.

Pummer, B.G. Budke, C. Augustin-Bauditz, S. Niedermeier, D. Felgitsch, L. *et al*. 2014. Ice nucleation by water-soluble macromolecules. *Atmospheric Chemistry and Physics*. **14**(17): 24273-24309.

Qian, H. Ice nucleation by water-soluble macromolecules. Zhang, Y.J. Zhang, J. Wang, X.L. 2013. Latitudinal gradients in phylogenetic relatedness of angiosperm trees in North America. *Global Ecology and Biogeography*. **22**(11): 1183-1191.

Ramløv, H. DeVries, A.L. Wilson, P.W. 2005. Antifreeze glycoproteins from the antarctic fish Dissostichus mawsoni studied by differential scanning calorimetry (DSC) in combination with nanoliter osmometry. *Cryoletters*. **26**(2): 73-84.

Ramløv, H. Wharton, D.A. Wilson, P.W. 1996. Recrystallization in a freezing tolerant Antarctic nematode, *Panagrolaimus davidi*, and an alpine weta, *Hemideina maori* (*Orthoptera*; *Stenopelmatidae*). *Cryobiology*. **33**(6): 607-613.

Ramsay, J.A. 1964. The rectal complex of the mealworm Tenebrio molitor, L.(*Coleoptera, Tenebrionidae*). *Philosophical Transactions of the Royal Society B: Biological Sciences*. **248**(748): 279-314.

Raymond, J.A. Christner, B.C. Schuster, S.C. 2008. A bacterial ice-binding protein from the Vostok ice core. *Extremophiles*. **12**(5): 713-717.

Raymond, J.A. Fritsen, C. Shen, K. 2007. An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiology Ecology*. **61**(2): 214-221.

Raymond, M. 1996. New Zealand alpine plants: A challenge for growers. *Journal of the Royal New Zealand Institute of Horticulture.* **1**(2): 9-11.

Regand, A. Goff, H.D. 2006. Ice recrystallization inhibition in ice cream as affected by ice structuring proteins from winter wheat grass. *Journal of Dairy Science*. **89**(1): 49-57.

Reitsma, L. 1994. The frost resistance of some native plants from the Central Volcanic Plateau, North Island, New Zealand, in relation to plant succession. *New Zealand Journal of Botany*. **32**(2): 217-226.

Rexer-Huber, K.M. Bishop, P.J. Wharton, D.A. 2011. Skin ice nucleators and glycerol in the freezing-tolerant frog *Litoria ewingii*. *Journal of Comparative Physiology*. **181**(6): 781-792.

Robichon, C. Luo, J. Causey, T.B. Benner, J.S. Samuelson, J.C. 2011. Engineering *Escherichia coli* BL21 (DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. *Applied and Environmental Microbiology*. **77**(13): 4634-4646.

Ronquist, F. Teslenko, M. Van Der Mark, P. Ayres, D.L. Darling, A. *et al*. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*. **61**(3): 539-542.

Rosano, G.L. Ceccarelli, E.A. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*. **5**. DOI: 10.3389/fmicb.2014.00172.

Russell, G.B. Connor, H.E. Purdie, A.W. 1976. Triterpene methyl ethers of *Chionochloa* (*Gramineae*). *Phytochemistry*. **15**(12): 1933-1935.

Sørensen, J.G. Holmstrup, M. 2011. Cryoprotective dehydration is widespread in Arctic springtails. *Journal of Insect Physiology*. **57**(8): 1147-1153.

Sacomoto, G.A.T. Kielbassa, J. Chikhi, R. Uricaru, R. Antoniou, P. *et al*. 2012. KISSPLICE: de-novo calling alternative splicing events from RNA-seq data. *BMC Bioinformatics*. **13**(6): S5.

Sahebi, M. Hanafi, M.M. Siti Nor Akmar, A. Rafii, M.Y. Azizi, P. *et al*. 2015. Importance of silicon and mechanisms of biosilica formation in plants. *BioMed Research International*. **2015**. DOI: 10.1155/2015/396010.

Sakai, A. Larcher, W. 1987. Frost survival of plants: responses and adaptation to freezing stress. Berlin: *Springer*.

Sambrook, J.R. Russell, D. 2001. DW. 2001 Molecular cloning: a laboratory manual. *Quarterly Review of Biology*. **76**(3): 348-349.

Sandve, S.R. Rudi, H. Asp, T. Rognli, O.A. 2008. Tracking the evolution of a cold stress associated gene family in cold tolerant grasses. *BMC Evolutionary Biology* **8** (245). DOI: 10.1186/1471-2148-8-245.

Sanmartin, I. Ronquist, F. 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology*. **53**(2): 216-243.

Sarhan, M.A.A. 2011. Ice nucleation protein as a bacterial surface display protein. *Archives of Biological Science*. **63**(4): 943-948.

Schubert, M. Groenvold, L. Sandve, S.R. Hvidsten, T.R. Fjellheim, S. 2017. Evolution of cold acclimation in temperate grasses *(Pooideae*). *bioRxiv*. 210021. DOI: http://dx.doi.org/10.1101/210021.

Sdebottom, C. Buckley, S. Pudney, P. 1997. Heat-stable antifreeze protein from grass. *Nature*. **388**(6): 644.

Sehrawat, A. Abat, J.K. Deswal, R. 2013. RuBisCO depletion improved proteome coverage of cold responsive S-nitrosylated targets in *Brassica juncea*. *Frontiers in Plant Science*. **4**: 342. DOI: 10.3389/fpls.2013.00342.

Sekozawa, Y. Sugaya, S. Gemma, H. Iwahori, S. Ishikawa, M. 2001. Seasonal changes in the ice nucleation activity of various tissues in Japanese pear (*Pyrus pyrifolia Nakai*) in relation to their freezing behavior and frost injury. *Acta Hortic*. **587**: 543-547.

Shanmugam, V. 2005. Role of extracytoplasmic leucine rich repeat proteins in plant defence mechanisms. *Microbiological Research*. **160**(1): 83-94.

Sharon, D. Tilgner, H. Grubert, F. Snyder, M. 2013. A single-molecule long-read survey of the human transcriptome. *Nature biotechnology*. **31**(11): 1009.

Sheoran, I.S. Ross, A.R.S. Olson, D.J.H. Sawhney, V.K. 2009. Compatibility of plant protein extraction methods with mass spectrometry for proteome analysis. *Plant Science*. **176**(1): 99-104.

Shevchenko, A. Wilm, M. Vorm, O. Mann, M. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical Chemistry*. **68**(5): 850-858.

Sicheri, F. Yang, D.S. 1995. Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nature*. **375**(6530): 427-431.

Sidebottom, C. Buckley, S. Pudney, P. Twigg, S. Jarman, C. *et al*. 2000. Phytochemistry: heat-stable antifreeze protein from grass. *Nature*. **406**(6793): 256.

Simão, F.A. Waterhouse, R.M. Ioannidis, P. Kriventseva, E.V. Zdobnov, E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. **31**(19): 3210-3212.

Sinclair, B.J. Addo-Bediako, A. Chown, S.L. 2003. Climatic variability and the evolution of insect freeze tolerance. *Biological Reviews*. **78**(2): 181-195.

Singh, P. Hanada, Y. Singh, S.M. Tsuda, S. 2014. Antifreeze protein activity in Arctic cryoconite bacteria. *FEMS Microbiology Letters*. **351**(1): 14-22.

Smallwood, M. Bowles, D.J. 2002. Plants in a cold climate. *Philosophical Transactions of the Royal Society of London*. **357**(1423): 831-847.

Smith, S.A. Beaulieu, J.M. Stamatakis, A. Donoghue, M.J. 2011. Understanding angiosperm diversification using small and large phylogenetic trees. *American Journal of Botany*. **98**(3): 404-414.

Soreng, R.J. Davis, J.I. Doyle, J.J. 1990. A phylogenetic analysis of chloroplast DNA restriction site variation in *Poaceae* subfamily. *Pooideae*. *Plant Systematics and Evolution*. **172**(1): 83-97.

Soreng, R.J. Peterson, P.M. Romschenko, K. Davidse, G. Zuloaga, F.O. *et al*. 2015. A worldwide phylogenetic classification of the *Poaceae* (*Gramineae*). *Journal of Systematics and Evolution*. **53**(2): 117–137.

Souza, A. Sandrin, C. Calió, M. Meirelles, S. Pivello, V. *et al*. 2010. Seasonal variation of soluble carbohydrates and starch in *Echinolaena inflexa*, a native grass species from the Brazilian savanna, and in the invasive grass *Melinis minutiflora*. *Brazilian Journal of Biology*. **70**(2): 395-404.

Souza, A.d. Sandrin, C.Z. Moraes, M.G. Figueiredo-Ribeiro, R.d.C.L. 2005. Diurnal variations of non-structural carbohydrates in vegetative tissues of Melinis minutiflora, Echinolaena inflexa and Lolium multiflorum (Poaceae). *Brazilian Journal of Botany*. **28**(4): 755-763.

Stan, C.A. Schneider, G.F. Shevkoplyas, S.S. Hashimoto, M. Ibanescu, M. *et al*. 2009. A microfluidic apparatus for the study of ice nucleation in supercooled water drops. *Lab on a Chip*. **9**(16): 2293-2305.

Storey, J.M. Storey, K.B. 2005. Cold hardiness and freeze tolerance. *John Wiley and Sons, Inc*.

Storey, K.B. Storey, J.M. 1988. Freeze tolerance in animals. *Physiological Reviews* **68**(1): 27-84.

Storey, K.B. Storey, J.M. 2013. Molecular biology of freezing tolerance. *Comprehensive Physiology*. **3**(3): 1283-1308.

Strömberg, C.A.E. 2011. Evolution of grasses and grassland ecosystems. *Annual Review of Earth and Planetary Sciences*. **39**: 517-544.

Stressmann, M. Kitao, S. Griffith, M. Moresoli, C. Bravo, L.A. *et al*. 2004. Calcium interacts with antifreeze proteins and chitinase from cold-acclimated winter rye. *Plant Physiology*. **135**(1): 364-376.

Sturman, A. McGowan, A. Spronken-Smith, A. 1999. Mesoscale and local climates in New Zealand. *Progress in Physical Geography*. **23**(4): 611-635.

Sukumaran, J. Holder, M.T. 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*. **26**(12): 1569-1571.

Sun, D.W. 2016. Handbook of frozen food processing and packaging. Florida: *CRC Press*.

Suyama, M. Torrents, D. Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*. **34**(suppl_2): 609-612.

Tam, R.Y. Ferreira, S.S. Czechura, P. Chaytor, J.L. Ben, R.N. 2008. Hydration index-A better parameter for explaining small molecule hydration in inhibition of ice recrystallization. *Journal of the American Chemical Society*. **130**(51): 17494-17501.

Tan, L.R. Lu, Y.C. Zhang, J.J. Luo, F. Yang, H. 2015. A collection of cytochrome P450 monooxygenase genes involved in modification and detoxification of herbicide atrazine in rice (*Oryza sativa*) plants. *Ecotoxicology and Environmental Safety*. **119**: 25-34.

Tanentzap, A.J. Brandt, A.J. Smissen, R.D. Heenan, P.B. Fukami, T. *et al*. 2015. When do plant radiations influence community assembly? The importance of historical contingency in the race for niche space. *New Phytologist*. **207**(2): 468-479.

Tarailo-Graovac, M. Chen, N.S. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*. 4.10.1-4.10.14.

Taylor, S.H. Franks, P.J. Hulme, S.P. Spriggs, E. Christin, P.A. *et al*. 2012. Photosynthetic pathway and ecological adaptation explain stomatal trait diversity amongst grasses. *New Phytologist*. **193**(2): 387-396.

Taylor, S.H. Ripley, B.S. Woodward, F.I. Osborne, C.P. 2011. Drought limitation of photosynthesis differs between C3 and C4 grass species in a comparative experiment. *Plant, Cell & Environment*. **34**(1): 65-75.

Tegos, G. Vargas, C. Perysinakis, A. Koukkou, A.I. Christogianni, A. *et al*. 2000. Release of cell-free ice nuclei from *Halomonas elongata* expressing the ice nucleation gene *inaZ* of *Pseudomonas syringae*. *Jounral of Applied Microbiology*. **89**(5): 785-792.

Thiel, T. Michalek, W. Varshney, R. Graner, A. 2002. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley *(Hordeum vulgare* L.). *Theoretical and Applied Genetics*. **106**(3): 411-422.

Theeuwes, F. Hussain, A. Higuchi, T. 1974. Quantitative analytical method for determination of drugs dispersed in polymers using differential scanning calorimetry. *Journal of Pharmaceutical Sciences*. **63**(3): 427-429.

Thomashow, M.F. 1998. Role of cold-responsive genes in plant freezing tolerance. *Plant Physiology*. **118**(1): 1-8.

Thomashow, M.F. 1999. Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annual Review of Plant Biology*. **50**(1): 571-599.

Thuenemann E.C. Lenzi P. Love A.J. Taliansky M. Becares M. Zuniga S. *et al*. 2013. The use of transient expression systems for the rapid production of virus-like particles in plants. *Current Pharmaceutical Design*. **19**: 5564–5573.

Tilgner, H. Jahanbani, F. Blauwkamp, T. Moshrefi, A. Jaeger, E. *et al*. 2015. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nature biotechnology*. **33**(7): 736.

Tremblay, K. Ouellet, F. Fournier, J. Danyluk, J. Sarhan, F. 2005. Molecular characterization and origin of novel bipartite cold-regulated ice recrystallization inhibition proteins from cereals. *Plant and Cell Physiology*. **46**(6): 884-891.

Tsiatsiani, L. Heck, A.J. 2015. Proteomics beyond trypsin. *The FEBS journal.* **282**(14): 2612-2626.

Turner, M.A. Arellano, F. Kpzloff, L.M. 1990. Three separate classes of bacterial ice nucleation structures. *Journal of Bacteriology*. **172**(5): 2521-2526.

Urrutia, M.E. Duman, J.G. Knight, C.A. 1992. Plant thermal hysteresis proteins. *Biochimica et Biophysica Acta*. **1121**(1-2): 199-206.

Vasques, G.M. Grunwald, S. Comerford, N.B. Sickman, J. O. (2010). Regional modelling of soil carbon at multiple depths within a subtropical watershed. *Geoderma*, **156**: 326-336.

Versaw, W.K. Harrison, M.J. 2002. A chloroplast phosphate transporter, PHT2; 1, influences allocation of phosphate within the plant and phosphate-starvation responses. *The plant cell*. **14**(8): 1751-1766.

Vigeland, M.D. Spannagl, M. Asp, T. Paina, C. Rudi, H. *et al*. 2013. Evidence for adaptive evolution of low-temperature stress response genes in a *Pooideae* grass ancestor. *New Phytologist*. **199**(4): 1060-1068.

Vranová, E. Coman, D. Gruissem, W. 2013. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology*. **64**: 665-700.

Walker, J.E. Saraste, M. Runswick, M.J. Gay, N.J. 1982. Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal*. **1**(8): 945.

Wallis, G.P. Trewick, S.A. 2009. New Zealand phylogeography: evolution on a small continent. *Molecular Ecology*. **18**(17): 3548-3580.

Wang, B.Y. Tan, H.W. Fang, W. Meinhardt, L.W. Mischke, S. *et al*. 2015a. Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. *Horticulture Research*. **2**: 201465. DOI: 10.1038/hortres.2014.65.

Wang, S. Zhou, G. Huang, X. Hu, J. Wang, B. *et al*. 2017. Transcriptome analysis of non-heading Chinese cabbage under heat stress by RNA-seq and marker identification. *Euphytica*. **213**(5): 109.

Wang, W.X. Wei, L.B. 2003. Purification of boiling-soluble antifreeze protein from the legume *Ammopiptanthus mongolicus*. *Preparative Biochemistry and Biotechnology*. **33**(1): 67-80.

Wang, Y. Coleman-Derr, D. Chen, G. Gu, Y.Q. 2015b. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*. **43**(W1): 78-84.

Warren, G.J. 1987. Bacterial ice nucleation: molecular biology and applications. *Biotechnology and Genetic Engineering Reviews*. **5**(1): 107-136.

Wegener, A. 1966. The origin of continents and oceans. New York: *Dover publications, Inc.*

Wellman, H.W. 1955. New Zealand quaternary tectonics. *Geologische Rundschau*. **34**(1): 248-257.

Wharton, D.A. 2011. Cold tolerance of New Zealand alpine insects. *Journal of Insect Physiology*. **57**(8): 1090-1095.

Wharton, D.A. 2012. Supercooling and Freezing Tolerant Animals. Rijeka: *InTech*

Wharton, D.A. Barrett, J. Goodall, G. Marshall, C.J. Ramlov, H. 2005. Ice-active proteins from the Antarctic nematode *Panagrolaimus davidi*. *Cryobiology*. **51**(2): 198-207.

Wharton, D.A. Block, W. 1997. Differential scanning calorimetry studies on an Antarctic nematode (*Panagrolaimus davidi*) which survives intracellular freezing. *Cryobiology*. **34**(2): 114-121.

Wharton, D.A. Mutch, J.S. Wilson, P.W. Marshall, C.J. Lim, M. 2004. A simple ice nucleation spectrometer. *CryoLetters*. **25**(5): 335-340.

Wharton, D.A. Selvanesan, L. Marshall, C.J. 2010. Ice-active proteins from New Zealand snow tussocks, *Chionochloa macra* and *C. rigida*. *CryoLetters*. **31**(3): 239-248.

Wharton, D.A. To, N.B. 1996. Osmotic stress effects on the freezing tolerance of the Antarctic nematode *Panagrolaimus davidi*. *Journal of Comparative Physiology*. **166**(5): 344-349.

Wharton, D.A. Wilson, P.W. Mutch, J.S. Marshall, C.J. Lim, M. 2007. Recrystallization inhibition assessed by splat cooling and optical recrystallometry. *CryoLetters*. **28**(1): 61-68.

Wilson, H. 2007. Alpine plants - Alpine New Zealand [Online]. Available: http://www.TeAra.govt.nz/en/map/11022/alpine-areas-of-new-zealand.

Wilson, P.W. 1993. Explaining thermal hysteresis by the Kelvin effect. *CryoLetters*. **14**: 31-36.

Wilson, P.W. Osterday, K.E. Heneghan, A.F. Haymet, A.D. 2010. Type I antifreeze proteins enhance ice nucleation above certain concentrations. *Journal of Biological Chemistry*. **285**(45): 34741-34745.

Wilson, S.L. Kelley, D.L. Walker, V.K. 2006. Ice-active characteristics of soil bacteria selected by ice-affinity. *Environmental Microbiology*. **8**(10): 1816-1824.

Winfield, M.O. Lu, C. Wilson, I.D. Coghill, J.A. Edwards, K.J. 2010. Plant responses to cold: transcriptome analysis of wheat. *Plant Biotechnology Journal*. **8**(7): 749-771.

Wingfield, P.T. 2002. Use of protein folding reagents. *Current Protocols in Protein Science*. A.3A.1-A.3A.4.

Winkworth, R.C. Wagstaff, S.J. Glenny, D. Lockhart, P.J. 2002. Plant dispersal news from New Zealand. *Trends in Ecology & Evolution*. **17**(11): 514-520.

Winkworth, R.C. Wagstaff, S.J. Glenny, D. Lockhart, P.J. 2005. Evolution of the New Zealand mountain flora: origins, diversification and dispersal. *Organisms Diversity & Evolution*. **5**(3): 237-247.

Wisniewski, M. 1997. Extraction of arsenic from sulphuric acid solutions by Cyanex 923. *Hydrometallurgy*. **46**(1-2): 235-241.

Wisniewski, M. Fuller, M. Palta, J. Carter, J. Arora, R. 2004. Ice nucleation, propagation, and deep supercooling in woody plants. *Journal of Crop Improvement*. **10**(1-2): 5-16.

Wisniewski, M. Neuner, G. Gusta, L.V. 2015. The use of high-resolution infrared thermography (HRIT) for the study of ice nucleation and ice propagation in plants. *Journal of Visualized Experiments: JoVE*. **99**: e52703-e52703.

Wisniewski, M. Webb, R. Balsamo, R. Close, T.J. Yu, X.M. *et al*. 1999. Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: a dehydrin from peach (*Prunus persica*). *Physiologia Plantarum*. **105**(4): 600-608.

Wolanczyk, J.P. Storey, K.B. Baust, J.G. 1990. Ice nucleating activity in the blood of the freeze-tolerant frog, *Rana sylvatica*. *Cryobiology*. **27**(3): 328-335.

Wolber, P.K. 1993. Bacterial ice formation. *Advances in Microbial Physiology*. **34**: 203-237.

Wolber, P.K. Deininger, C.A. Southworth, M.W. Vandekerckhove, J. van Montagu, M. *et al*. 1986. Identification and purification of a bacterial ice-nucleation protein. *Proceedings of the National Academy of Sciences*. **83**(19): 7256-7260.

Wolber, P.K. Green, R.L. 1990. Detection of bacteria by transduction of ice nucleation genes. *Trends Biotechnology*. **8**(10): 276-279.

Worrall, D. Elias, L. Ashford, D. Smallwood, M. Sidebottom, C. *et al*. 1998. A carrot leucine-rich-repeat protein that inhibits ice recrystallization. *Science*. **282**(5386): 115-117.

Wu, B. Suo, F.M. Lei, W.J. Gu, L.F. 2014. Comprehensive analysis of alternative splicing in *Digitalis purpurea* by strand-specific RNA-Seq. *PLoS One*. **9**(8): e106001.

Wu, S. Zhu, Z.W. Fu, L.M. Niu, B.F. Li, W.Z. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*. **12**(1): 444.

Wu, Z.Q. Ge, S. 2012. The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. *Molecular Phylogenetics and Evolution*. **62**(1): 573-578.

Xie, Y.L. Wu, G.X. Tang, J.B. Luo, R.B. Patterson, J. *et al*. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. **30**(12): 1660-1666.

Xu, H. Griffith, M. Patten, C.L. Glick, B.R. 1998. Isolation and characterization of an antifreeze protein with ice nucleation activity from the plant growth promoting rhizobacterium *Pseudomonas putida* GR12-2. *Canadian Journal of Microbiology*. **44**(1): 64-73.

Yeh, S. Moffatt, B.A. Griffith, M. Xiong, F. Yang, D.S.C. *et al*. 2000. Chitinase genes responsive to cold encode antifreeze proteins in winter cereals. *Plant Physiology*. **124**(3): 1251-1264.

Yesilirmak, F. Sayers, Z. 2009. Heterelogous expression of plant genes. *International Journal of Plant Genomics*. **2009**. DOI: 10.1155/2009/296482

Yi, K.K. Wu, Z.C. Zhou, J. Du, L.M. Guo, L. *et al*. 2005. OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiology*. **138**(4): 2087-2096.

You, Y.N. Liu, D.C. Liu, H.B. Zheng, X.F. Diao, Y. *et al*. 2015. Development and characterisation of EST-SSR markers by transcriptome sequencing in taro (*Colocasia esculenta (L.) Schoot*). *Molecular Breeding*. **35**(6): 134.

Yu, S.S. Yin, L.K. Mu, S.Y. 2010. Discovery of an antifreeze protein in the leaves of *Ammopiptanthus nanus*. *Canadian Journal of Plant Science*. **90**(1): 35-40.

Yu, X.M. Griffith, M. 1999. Antifreeze proteins in winter rye leaves form oligomeric complexes. *Plant Physiology*. **119**(4): 1361-1370.

Yu, X.M. Griffith, M. Wiseman, S.B. 2001. Ethylene induces antifreeze activity in winter rye leaves. *Plant Physiology*. **126**(3): 1232-1240.

Zachariassen, K.E. Kristiansen, E. 2000. Ice nucleation and antinucleation in nature. *Cryobiology*. **41**(4): 257-279.

Zhang, C. Fei, S.Z. Arora, R. Hannapel, D.J. 2010. Ice recrystallization inhibition proteins of perennial ryegrass enhance freezing tolerance. *Planta*. **232**(1): 155-164.

Zhang, D.Q. Liu, B. Feng, D.R. He, Y.M. Wang, S.Q. *et al*. 2004. Significance of conservative asparagine residues in the thermal hysteresis activity of carrot antifreeze protein. *Biochemical Journal*. **377**(3): 589-595.

Zhang, R. Calixto, C.P. Marquez, Y. Venhuizen, P. Tzioutziou, N.A. *et al*. 2017. A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*. **45**(9): 5061-5073.

Zhang, Y.Y. Zhang, B.C. Yan, D.W. Dong, W.X. Yang, W.B. *et al*. 2011. Two *Arabidopsis* cytochrome P450 monooxygenases, CYP714A1 and CYP714A2, function redundantly in plant development through gibberellin deactivation. *The Plant Journal*. **67**(2): 342-353.

Zheng, Y. Wiens, J.J. 2015. Do missing data influence the accuracy of divergence-time estimation with BEAST? *Molecular phylogenetics and evolution*. **85**: 41-49.

Zhou, Q. Luo, D. Ma, L.C. Xie, W.G. Wang, Y. *et al*. 2016. Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus L.*) using Illumina sequencing. *Scientific Reports*. **6**: 20549.

Zotov, V.D. 1963. Synopsis of the grass subfamily *Arundinoideae* in New Zealand. *New Zealand Journal of Botany*. **1**: 78-136.

Zotov, V.D. 1970. *Chionochloa macra* (Gramineae): A new species. *New Zealand Journal of Botany*. **8**(1): 91-93.

Zwickl, D.J. 2006. GARLI, vers. 0.951. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, University of Texas.

Appendix

Part I supplymentary tables and figures

Protein	Contig
Thaumatin-like protein	c228340_g1_i1; c350175_g4_i1; c395750_g1_i1; c361984_g4_i1; c87067_g2_i1; c93125_g1_i1; c221180_g1_i1; c321881_g3; c321881_g4; c355160_g1_i8; c433511_g1_il; c228367_g1_i1; c297937_g1; c314992_g1_i1; c349116_g1; c349116_g3; c353615_g1_i1; c353615_g3_i1; c354931_g2_i9; c57441_g1_i1; c360201_g1; c332254_g2_i1; c187636_g1_i1; c247915_g1_i1; c354931_g2; c332254_g2_i2; c156312_g1_i1; c349116_g2_i1; c350175_g1_i4; c354931_g2; c354931_g3_i2; c361054_g5_i1; c339513_g3_i1; c319582_g2_i2; c208_g1;
Polygalacturonase inhibitor proteins	c232915_g2_i1; c118301_g1_i1; c554961_g1_i1; c366516_g4_i1; c366516_g5_i2; c357658_g1; c366516_g5_i1; c280482_g1_i1; c282104_g1_i1; c335556_g2_i1; c580117_g1_i1; c604909_g1_i1; c348699_g1_i5; c348699_g2_i1; c381143_g1_i1; c499430_g1_i1; c356928_g3; c156944_g1_i2;

Table S1 Other possible IAPs in the *C. macra* transcriptome.

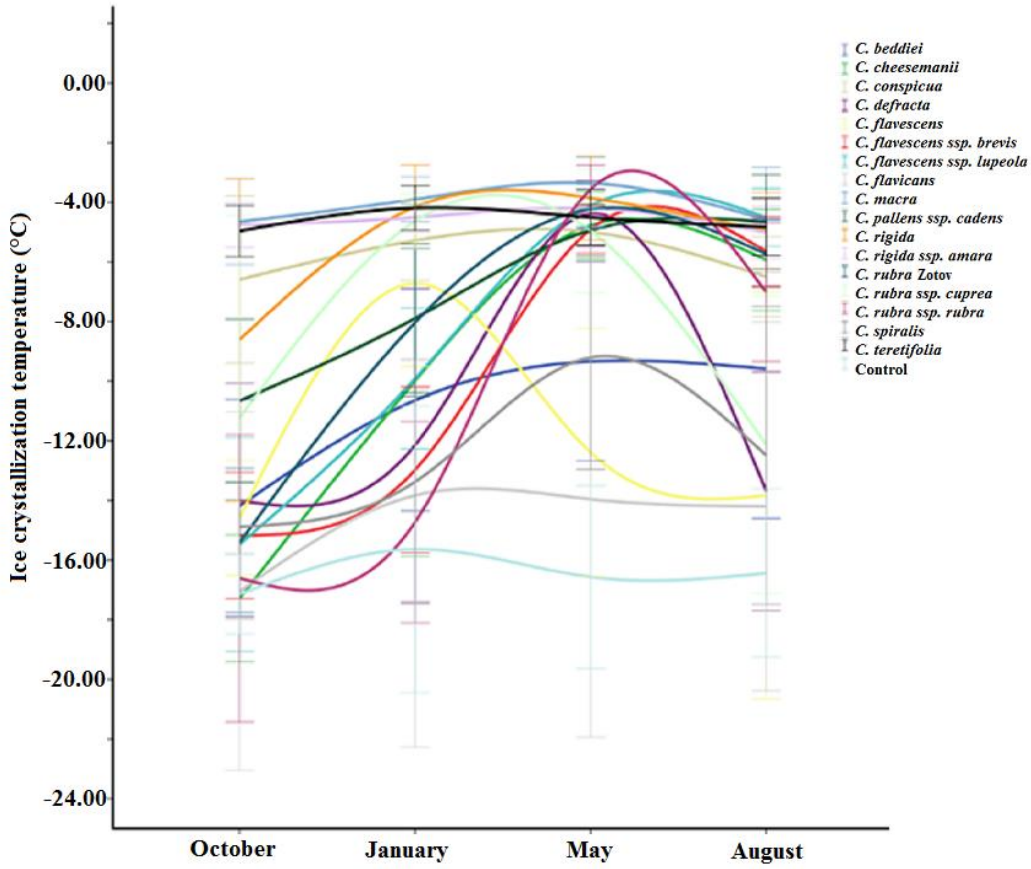


Figure S1 Seasonal pattern of INA in *Chionochloa* species. Mean T_cs±standard deviations, N=3.

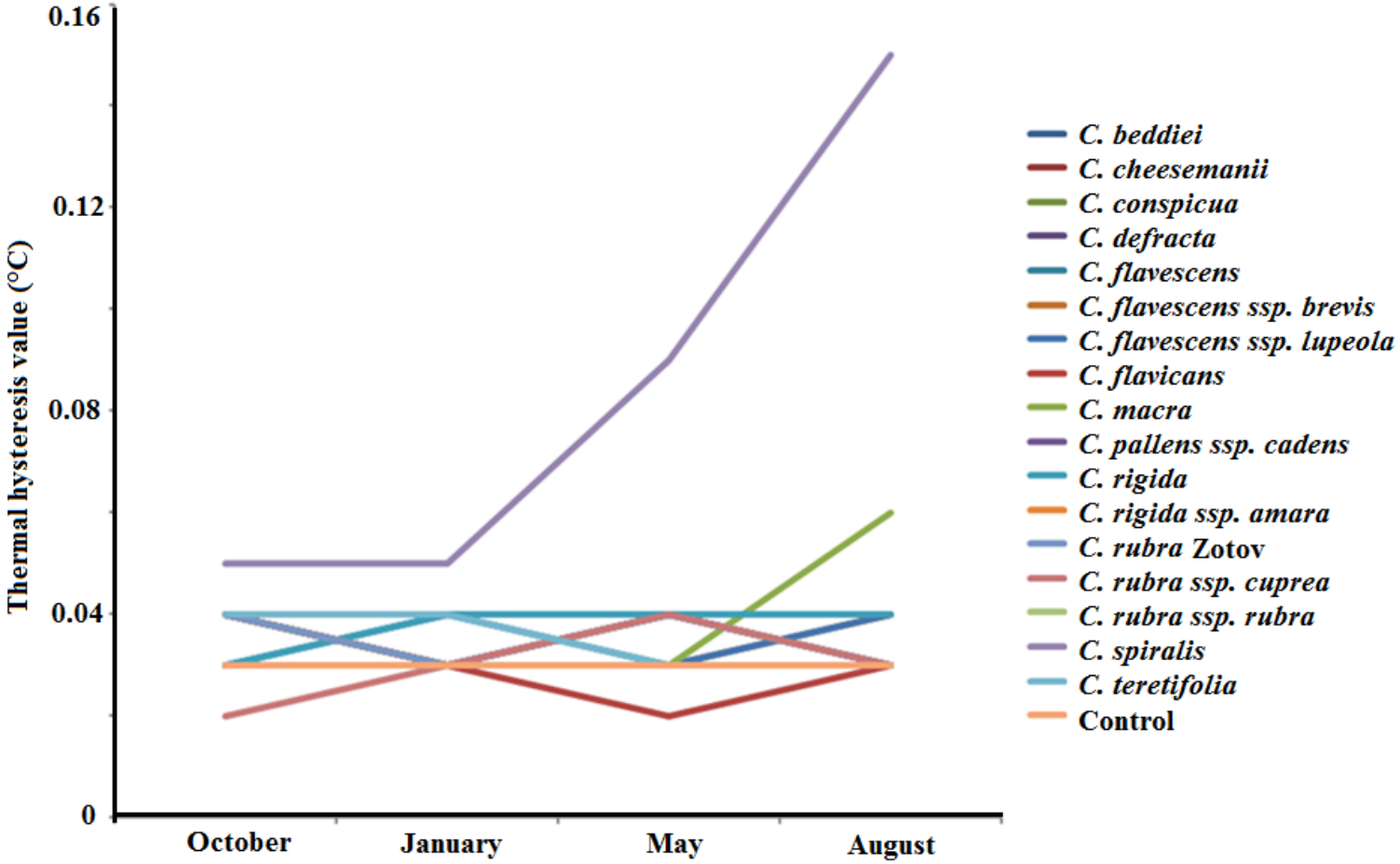


Figure S1 Seasonal pattern of TH value in the *Chionochloa* species. Mean TH, N=3.

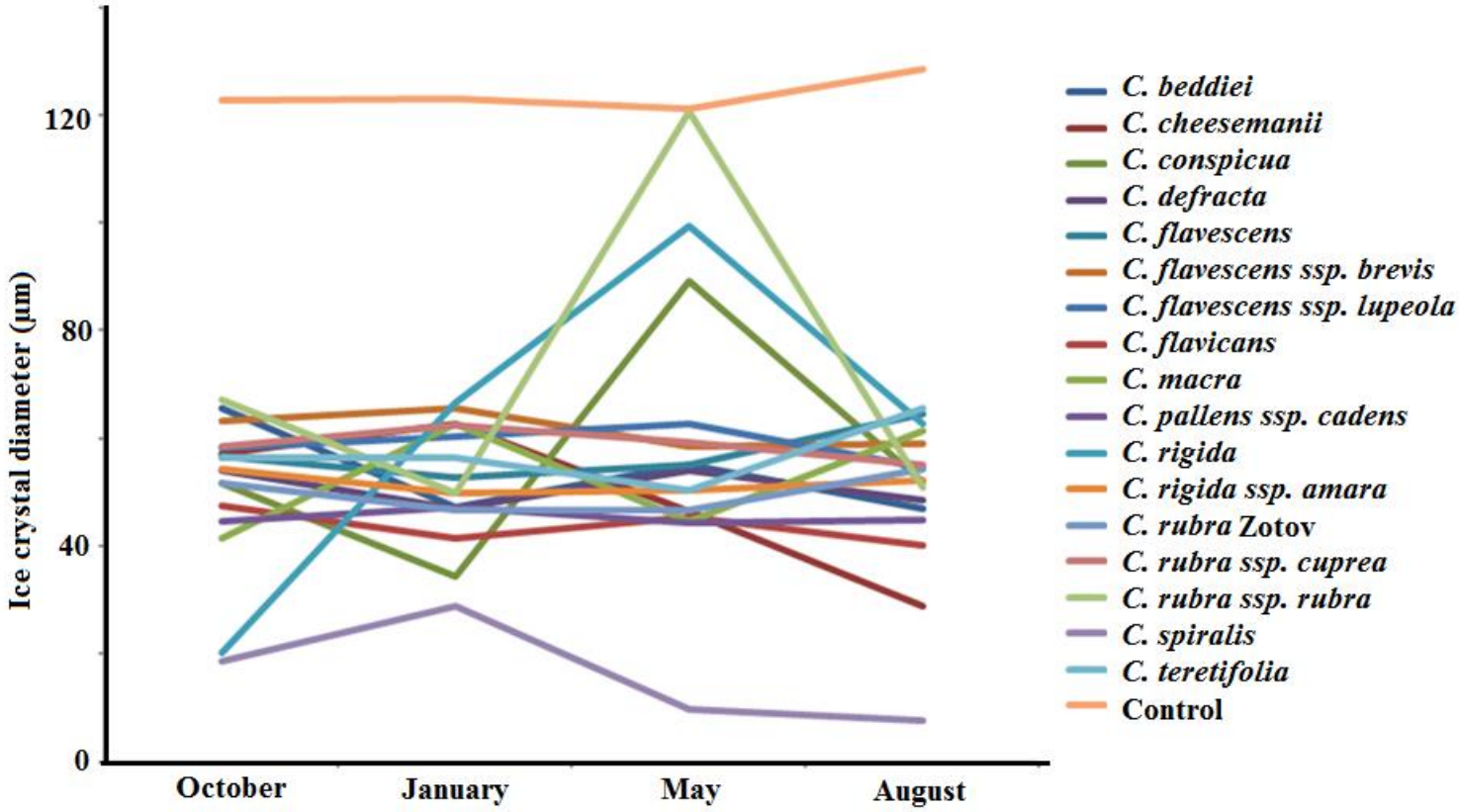


Figure S3 Seasonal pattern of RI activity in *Chionochloa* species. Mean ice crystal diameters (µm) at t_{30min} of splat cooling assay were present for 10 randomly selected largest ice crystals.

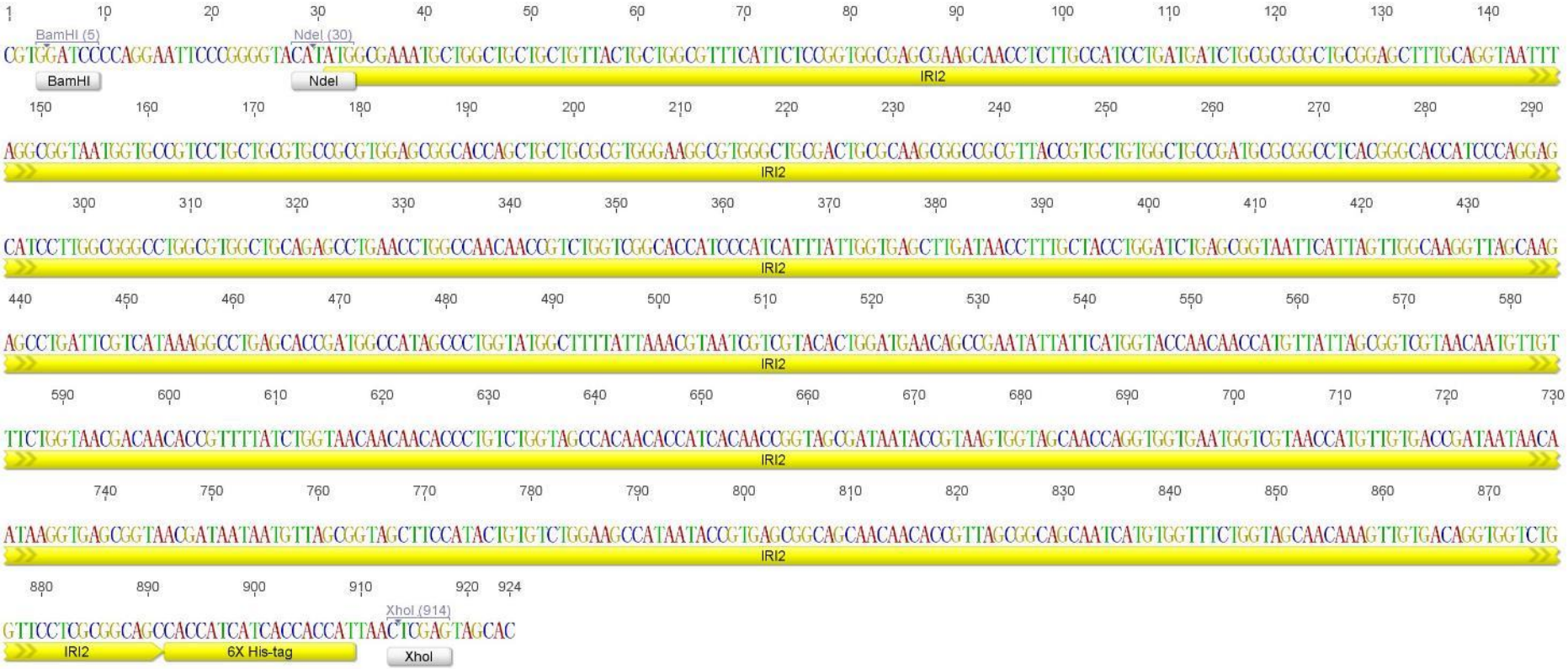


Figure S4 The nucleotide sequence for IRI2. Restriction enzyme sites were shown.

Figure S5 The nucleotide sequence for EAF2.

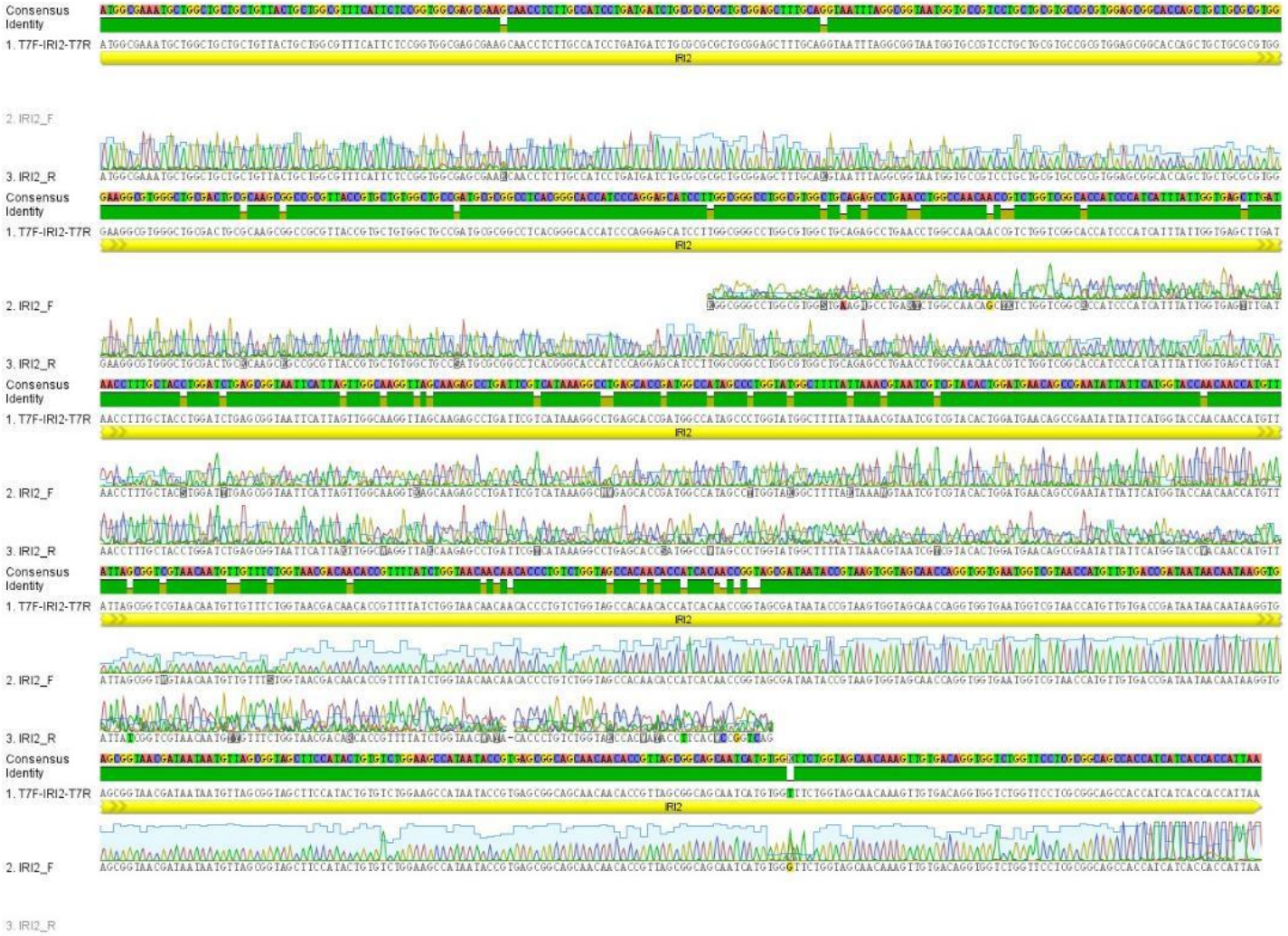


Figure S6 Identification of the nucleotide sequence of IRI2 by Sanger sequencing.

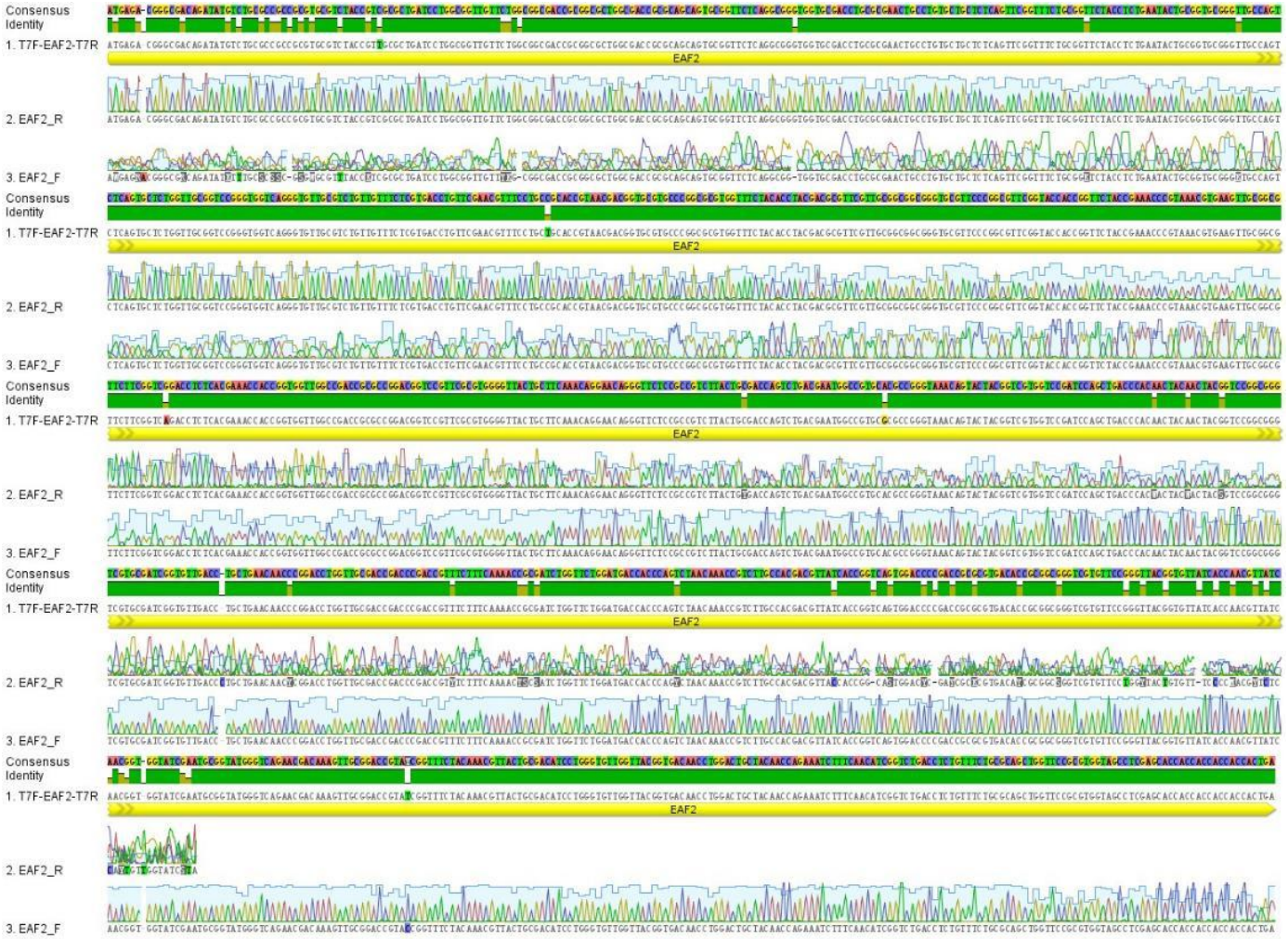


Figure S7 Identification of the nucleotide sequence of EAF2 by Sanger sequencing.

1 MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL
51 EFPNLPYYID GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL
101 DIRYGVSRIA YSKDFETLKV DFLSKLPEML KMFEDRLCHK TYLNGDHVTH
151 PDFMLYDALD VVLYMDPMCL DAFPKLVCFK KRIEAIPQID KYLKSSKYIA
201 WPLQGWQATF GGGDHPPKSD LEVLFQGPLG SPGIPGVHMA KCWLLLLLLA
251 FHSPVASEAT SCHPDDLRAL RSFAGNLGGN GAVLLRAAWS GTSCCAWEGV
301 GCDCASGRVT VLWLPMRGLT GTIPGASLAG LAWLQSLNLA NNRLVGTIPS
351 FIGELDNLCY LDLSGNSLVG KVSKSLIRHK GLSTDGHSPG MAFIKRNRRT
401 LDEQPNIIHG TNNHVISGRN NVVSGNDNTV LSGNNNTLSG SHNTITTGSD
451 NTVSGSNQVV NGRNHVVTDN NNKVSGNDNN VSGSFHTVSG SHNTVSGSNN
501 TVSGSNHVVS GSNKVVTGGL VPRGSHHHHH H

Figure S8 Identification of GST-IRI2. Matched peptides were shown in bold red. Protein sequence coverage was 59%.

1 MMSPILGYWK IKGLVQPTRL LLEYLEEKYE EHLYERDEGD KWRNKKFELG
51 LEFPNLPYYI DGDVKLTQSM AIIRYIADKH NMLGGCPKER AEISMLEGAV
101 LDIRYGVSRI AYSKDFETLK VDFLSKLPEM LKMFEDRLCH KTYLNGDHVT
151 HPDFMLYDAL DVVLYMDPMC LDAFPKLVCF KKRIEAIPQI DKYLKSSKYI
201 AWPLQGWQAT FGGGDHPPKS DLEVLFQGPL GSVHMRRATD MSAPPRASTV
251 ALILAVVLAA TAALATAQQC GSQAGGATCA NCLCCSQFGF CGSTSEYCGA
301 GCQSQCSGCG PGGQGVASVV SRDLFERFLL HRNDGACPAR GFYTYDAFVA
351 AAGAFPAFGT TGSTETRKRE VAAFFGQTSH ETTGGWPTAP DGPFAWGYCF
401 KQEQGSPPSY CDQSDEWPCA PGKQYYGRGP IQLTHNYNYG PAGRAIGVDL
451 LNNPDLVATD PTVSFKTAIW FWMTTQSNKP SCHDVITGQW TPTARDTAAG
501 RVPGYGVITN VINGGIECGM GQNDKVADRI GFYKRYCDIL GVGYGDNLDC
551 YNQKSFNIGL TSVSAQLVPR GSLE

Figure S9 Identification of GST-EAF2. Matched peptides were shown in bold red. Protein sequence coverage was 47%.

1 MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL
51 EFPNLPYYID GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL
101 DIRYGVSRIA YSKDFETLKV DFLSKLPEML KMFEDRLCHK TYLNGDHVTH
151 PDFMLYDALD VVLYMDPMCL DAFPKLVCFK KRIEAIPQID KYLKSSKYIA
201 WPLQGWQATF GGGDHPPKSD LEVLFQGPLG SPGIPGVHMA KCWLLLLLLA
251 FHSPVASEAT SCHPDDLRAL RSFAGNLGGN GAVLLRAAWS GTSCCAWEGV
301 GCDCASGRVT VLWLPMRGLT GTIPGASLAG LAWLQSLNLA NNRLVGTIPS
351 FIGELDNLCY LDLSGNSLVG KVSKSLIRHK GLSTDGHSPG MAFIKRNRRT
401 LDEQPNIIHG TNNHVISGRN NVVSGNDNTV LSGNNNTLSG SHNTITTGSD
451 NTVSGSNQVV NGRNHVVTDN NNKVSGNDNN VSGSFHTVSG SHNTVSGSNN
501 TVSGSNHVVS GSNKVVTGGL VPRGSHHHHH H

Figure S10 Identification of over-expressed GST-tag in the GST-IRI2 expression. The single GST-tag band (~26 kDa) from the gel electrophoresis was cut, tryptic digested and sent to mass spectrometry. Matched peptides were shown in bold red. Protein sequence coverage was 34%.

Part II Supplymentary information for chapter 6

S6. 1 Supplymentary experiment procedures

S6.1.1 Total RNA extraction and total RNA sequencing

The presence of large amount of polysaccharides in the *C. macra* made RNA extraction and purification difficult. Commercial kits (Qiagen Rneasy plant mini kit, Isolate II RNA plant kit) both failed as columns were blocked. A modified Invitrogen plant RNA reagent - Qiagen Rneasy plant mini kit combined method, was used instead. Starting material (~50 mg, two pieces of 2 cm leaves) was ground to a powder in an autoclaved mortar with liquid nitrogen and resuspended in 500 µL Invitrogen plant RNA reagent. After incubating at room temperature for 30 min, the lysate was centrifuged at 12,000 × g, room temperature for 2 min. The supernatant was transferred to a new 2 mL Eppendorf tube. 100 µL 5M NaCl and 300 µL chloroform were added and mixed vigorously. The upper supernatant was transferred to a RNeasy centrifugal device and an on-column Dnase I digestion step was performed for 15 min at room temperature. Finally, a RNA clean up procedure was done following the manufacturer's protocols (Qiagen) and Rnase-free water was used to elute the total RNAs. Total RNAs were initially checked by a Nanodrop ND-1000 (Thermal Scientific), a Qubit RNA assay, a 1.0% agarose gel and a Bioanalyser (Agilent Technologies) respectively.

Four Illumina TruSeq stranded total RNA libraries were prepared following the Illumina TruSeq protocol and sequenced in a single lane (2×100 bp PE) on an Illumina HiSeq 2000 sequencer at New Zealand Genomics Limited (NZGL). Ribosome RNA was not removed in order to search for the presence of ice nucleation bacteria. Briefly, the total RNAs were fragmented and single stranded cDNAs were synthesized with random primers in the kit by the reverse transcription. RNA templates were removed and the second strand cDNAs were synthesized to generate double-strand DNAs (dsDNA). A single adenine base added to the 3' end of dsDNA and indexing adapters were ligated to the dsDNA. DNA fragments with adaptors at both ends were selectively enriched through PCR with primers annealing to the end of adapters. The quality of libraries was checked on a Bioanalyser (Agilent Technologies) and the average library sizes were 314 bp, 331 bp, 598 bp and 596 bp

respectively. The corresponding average insertion sizes were 193 bp, 209 bp, 477 bp and 475 bp respectively, removing the 121 bp adapter.

S6.1.2 Sequence assembly

Raw sequence reads were delivered in fastq format. All raw sequence reads underwent quality control steps as a part of post-sequencing processing. First, adapter sequences were removed and FASTQC was applied to analyze the quality of the four libraries datasets (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) (Andrews 2010). Sequence reads were trimmed at Phred scores Q20 (1.0 % probability of false calling rate at a base) by dynamictrim.pl in SolexaQA++ 3.1.3 package (http://solexaqa.sourceforge.net) and sorted by lengthsort.pl in the same package with default setting, 25 bases (Cox *et al.* 2010). The four individual *C. macra* library was merged to generate a merged library.

Two assemblers (Trinity and SOAPdenovo-Trans) were used to de novo assemble the transcriptome with the Q20 reads from the merged library (Haas *et al.* 2013; Xie *et al.* 2014). Trinity was run on standard default settings with kmer 25 and the minimal contig length was set to a default value (201 bp long). SOAPdenovo-Trans was run with kmer 25 and the other default settings. The computing server at the Department of Biochemistry, University of Otago, was used to run the de novo assemblers. Transcriptome assembly statistics including the total number of contigs, contig lengths and N_{50} (a value where at least 50% of contigs in the assembly have a length no less than this value) were obtained with the help of the assemblathon.pl script (Earl *et al.* 2011). The raw *C. macra* transcriptome generated by Trinity was filtered by the criteria based on the expression levels that minimum FPKM (Fragments Per Kilobase of transcript per Million mapped reads) was 0.5 and minimum isoform was 1.0%. Under these criteria, isoforms (contigs) of a Trinity gene with isoform percentage less than 1.0% and FPKM value less than 0.5 were excluded. The filtered *C. macra* transcriptome was used for downstream annotation and named as the *C. macra* transcriptome. The filtered-out contigs by the mentioned criteria were named as the filtered-out transcriptome. The transcriptome assembled with the merged libraries method and individual library method were compared and assessed the impact of individual variations.

S6.1.3 Assembly completeness

To assess the completeness of this transcriptome assembly, three methods were applied: obtain coverage distribution of amino acid sequences from close relatives, compare with known *C. macra* sequences in the NCBI GeneBank and search against the representation of core eukaryotic genes (Gillard *et al.* 2014).

The Core Eukaryotic Genes Mapping Approach (CEGMA) (http://korflab.ucdavis.edu/datasets/cegma) (Parra *et al.* 2007) was previously used in assessing the completeness of assembled genomes. It was also used to evaluate the completeness of assembled transcriptomes (Nakasugi *et al.* 2013). The *C. macra* transcriptome was searched against the 458 core-conserved proteins present in all eukaryotic species and the 248 highly core-conserved eukaryotic proteins. The transcriptome, which matched the CEGMA best, was chosen for downstream analysis. The selected transcriptome was mapped against nucleotide sequences and protein sequences of all available *Chionochloa* species deposited in NCBI GeneBank to investigate the sequence coverage. Predicted open reading frames (ORFs) from the *C. macra* transcriptome was also blasted against predicted ORFs from known plant species (Table S6.1). As all plant species in the analysis were *Poaceae* grasses, high similarity in protein profiles was expected between these species and the *C. macra* transcriptome.

Species	Sequence reference
Brachypodium distachyon	Brachypodium_distachyon.v1.0.28.pep.abinitio.fa
Eragrostis tef	Cannarozzi *et al.* 2014
Oryza sativa	Oryza_sativa.IRGSP-1.0.28.pep.abinitio.fa
Setaria italic	Setaria_italica.JGIv2.0.28.pep.abinitio.fa
Sorghum bicolor	Sorghum_bicolor.Sorbi1.28.pep.abinitio.fa
Zea mays	Zea_mays.AGPv3.28.pep.abinitio.fa

Table S6.1 Predicted ORFs derived from transcriptomes from six *Poaceae* grass. All the ORFs were downloaded from *Ensemble* database except *Eragrostis tef*.

S6.1.4 Transcriptome annotation

The *C. macra* transcriptome was locally blasted against the non-redundant database nr (version 28/5/2015, 67,337,701 sequences) with a total length of 24,122,812,982 bp, the UniProt protein knowledgebase (version 27/5/2015 contains Swiss-Prot with 548,586 sequences and TrEMBL with 48,744,721 sequences). The nr database was constrained by the GI list of *viridiplantae*, by which only sequences from

viridiplantae were available for the Blastx search. The Blast search result file (.xml) generated was imported to Blast2GO and processed to assign annotations (Conesa *et al.* 2005). WebMGA web service was used for COG, KOG, KEGG pathway analysis and Pfam analysis (http://weizhong-lab.ucsd.edu/metagenomic-analysis/) (Wu *et al.* 2011).

S6.1.5 Unannotated transcriptome analysis

Contigs in the *C. macra* transcriptome that failed to give a Blastx hit with the databases in section 6.2.5 were locally blasted against the NCBI nt database to search for any non-protein components such as non-coding RNAs. The *C. macra* transcriptome and contigs from the *C. macra* transcriptome without any Blast hit were both masked with the RepeatMasker program (http://www.repeatmasker.org) with Repbase database (http://www.girinst.org/repbase) in a search for interspersed repeats and low complexity DNA sequences (Jurka *et al.* 2005; Tarailo-Graovac and Chen 2009). The masking process was constrained by only searching for repeats in *viridiplantae* deposited in the RepeatMasker and Repbase database. The redundancy of the *C. macra* transcriptome was assessed by CDHIT at the threshold of 90% similarity.

S6.1.6 Comparative analysis with model plants

To compare the *C. macra* to other related plant species, the *C. macra* transcriptome were processed with TransDecoder in the Trinity package to predict open reading frames (ORF) (Haas *et al.* 2013). The UniProt database and Pfam database were searched to capture ORFs that may have functional significances. The predicted ORFs were searched against OrthoVenn (http://probes.pw.usda.gov/OrthoVenn/), which utilized OrthoMCL for the identification of Ortholog groups, to compare clusters of orthologous genes among multiple species including *Eragrostis tef*, *Oryza sativa Japonica*, *Seteria italica*, *Sorghum bicolor* and *Brachypodium distachyon* (Wang *et al.* 2015b). The predicted coding regions of *Eragrostis tef* transcripts were derived from Cannarozzi *et al.* (2014) (Cannarozzi *et al.* 2014).

S6.1.7 Ice-binding protein mapping

Two rounds of ice-shell purification were performed on protein extracts from the winter *C. macra* (chapter 5) to isolate proteins interacting with ice lattices, which were named as the ice-binding proteins (Kuiper *et al.* 2003). Ice fraction containing ice-binding proteins were freeze-dried to concentrate and an in-solution tryptic digestion step was performed. LC-MS-based protein profiling was performed with a Nanospray Flex™ Ion Sources to the LTQ-Obrbitrap XL mass spectrometry inline coupled an Ultimate 3000 nano-flow U-FPLC. Predicted ORFs from the *C. macra* transcriptome were imported into MASCOT (Matrix Science) search engine as localized database. The dataset from Orbitrap was then search against the mentioned localized database to identify potential ice-binding proteins in the *C. macra* transcriptome.

S6.1.8 SNPs and SSR marker detection

Single nucleotide polymorphisms (SNP) are the most abundant form of DNA sequence variations. SNP makers are useful in fields of marker-assisted plant breeding under various pressures and plant germplasm management (Wang *et al.* 2015a). Development of next generation sequencing helped to identify SNPs in many species (Clarke *et al.* 2016; Duarte *et al.* 2014; Kothapalli *et al.* 2016). Q20 reads from each library were mapped to the *C. macra* transcriptome reference by Bowtie2 with four .bam files generated (Langmead and Salzberg 2012). SAMtools with pileup command was applied to search the SNPs and short INDELs (Li *et al.* 2009). Filtering threshold was set as follows: minimum read depth 10, maximum read depth 200, base quality scores no less than 30, variant minimum frequency 0.25. The generated .vcf files were processed to snpEff to analysis the effect on amino acid change (Cingolani *et al.* 2012).

Single sequence repeats (SSR) markers, which contain tandem repeats of short DNA fragments, are abundant, highly polymorphic and co-dominantly inherited in eukaryotic genomes (Lu *et al.* 2015). SSR markers are widely used in genetic diversity and genotyping thus they will be helpful in large-scale genotyping and analysis of the diversity of the *Chionochloa* genus (Hodel *et al.* 2016). MISA (MIcroSAtellite; http://pgrc.ipk-gatersleben.de/misa/) was used for simple sequence repeats (SSRs) marker identification (Thiel *et al.* 2002). The parameter for detecting SSR in this study were as follows: a minimum of 10, 6, 5, 3, 3, and 3 repeat units for

1-6 nucleotide motifs, respectively. Maximal number of bases interrupting 2 SSRs in a compound microsatellite was set at 10.

S6.1.9 Highly expressed proteins and KEGG pathways analysis

Variations in the protein expression of plant species induced by selective pressures present the main molecular adaptions to environment stress (Groen and Purugganan 2016). KEGG (Kyoto Encyclopedia of Genes and Genomes) database provides up to date knowledge of gene ortholog groups and their functions on metabolisms in sequenced genomes (Kanehisa and Goto 2000). Specific metabolic pathways and functions of genes can be predicted by mapping unannotated genomes to the KEGG database (Kanehisa *et al.* 2016). Transcript abundance was estimated by RSEM in the Trinity Package. Q20 reads were first remapped to the *C. macra* transcriptome by the Bowtie. The align_and_estimate_abundance.pl script was used to estimate the expression level of each contig with FPKM value. The *C. macra* transcriptome were further filtered at a FPKM threshold of 20 and mapped to the KEGG pathways.

S6.1.10 Transcription factors and miRNAs search

Transcription factors play important roles in the regulation of plant growth and development through binding to the *cis*-regulatory elements in the promoter region of target genes and regulate downstream transcription whereas miRNA, which consist of short length (~22 bp) non-coding RNAs, act as repressors to silent mRNA formed at the post-transcription level (Chen and Rajewsky 2007; Martinez and Gregory 2010). Transcription factor and miRNA are two of the most studied gene regulation methods in plants and animals (Chen and Rajewsky 2007). Transcription factor and miRNA in the *C. macra* transcriptome were detected by searching against plant transcription factor (TF) database (http://planttfdb.cbi.pku.edu.cn/) and Plant miRNA database (http://www.mirbase.org/) (Jin *et al.* 2013).

S6.1.11 Alternative splicing (AS) events detection

Alternative splicing of pre-mRNA increases the complexity of gene expression under various selective pressures. The method to detect alternative splicing by mapping a transcriptome to its genome reference is not available due to lack of either *C. macra* genome sequences or a genome from a close relative. Alternative splicing events in

this transcriptome were detected by two methods, the Blast like alignment tool, BLAT, and the de novo alternative splicing caller, Kissplice, which is more sensitive than Trinity in reporting alternative splicing events by utilizing de novo sequencing reads. Contigs in the *C. macra* transcriptome were BLAT against each other with the following setting: -titleSize=18, -minIdentity=96 and maxInton=10000 (Kent 2002; Wu *et al.* 2014). Kissplice (v2.3.1) was run with the default settings except -C=0.05 and -e=7 (minimum gap bases) (Sacomoto *et al.* 2012). The merged libraries were imported in the Kissplice to generate the de Bruijn graphs. The output fasta file contains alternative splicing events were blasted against the *C. macra* transcriptome for AS contig annotations.

S6.2 Supplymentary results

S6.2.1 Quality of total RNAs

S6.2.1.1 Total RNA quantification

All four total RNA samples obtained good $OD_{260/280}$ and $OD_{260/230}$, which indicated no serious contaminations (Table S6.2). The integrity of these total RNAs was quick checked on the 1.0% agarose gel. Both 28S and 18S rRNAs were presented without obvious degradation. Quality control from NZGL indicated all four samples had RNA integrity number (RIN) no less than 8 on a Bioanalyzer (Agilent Technologies).

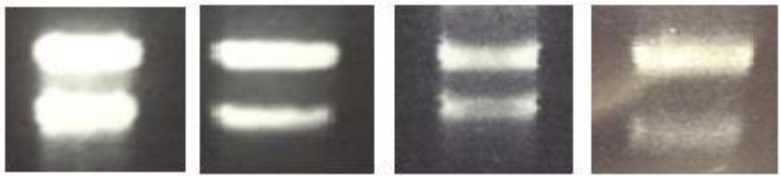
Total RNA sample	macra1	macra2	macra3	macra4
$OD_{260/280}$	2.11	2.13	2.09	2.19
$OD_{260/230}$	2.39	1.78	2.01	2.36
Nanodrop ng/μL	417	238	401	235
Qubit assay ng/μL	290	140	280	180
1.0% agarose gel				

Table S6.2 Total RNA quantification. $OD_{260/280}$ and $OD_{260/230}$ were measured by Nanodrop ND-1000 (Thermo Scientific). 2 μL samples were used in the Qubit RNA assays (Thermo Scientific). 2 μL samples were loaded on 1.0% agarose gels.

S6.2.1.2 Sequencing quality

S6.2.1.3 Summary of sequencing reads

The single sequencing lane of Hiseq2000 (Illumina) generated 36,829 Mbases data, which were 182,321,833 pair-end raw reads in total (Table S6.3). Raw reads in each library were filtered by Phred quality score Q20 (99% accuracy of correct base calling) for downstream analysis, which resulted in 97.02%, 96.50%, 95.27% and 96.35% of the raw reads remaining in the library macra1, macra2, macra3 and macra4 respectively. High proportion of library macra4 in the cluster was observed which might be due to its high RNA ratio during the library pooling process.

Sample	Index	Yield (Mbases)	# of Read pair	# of Q20 read (forward read / reverse read)
macra1	ATCACG	7,219	35,739,577	35,360,863/33,988,450
macra2	ACTTGA	7,350	36,385,655	35,918,606/34,303,478
macra3	TAGCTT	8,196	40,572,675	39,940,532/37,367,081
macra4	GGCTAC	14,064	69,623,926	68,701,214/65,468,601

Table S6.3 Summary of sequencing read quality. Raw reads from library macra1 to macra4 took 19.28%, 19.63%, 21.89% and 37.56% respectively of the raw cluster in the sequencing lane. #, Total number.

S6.2.2 Sequencing reads proceeding

Q20 reads from the four libraries were merged, which produced a total of 351,048,825 sequence reads. This took 96.27% of the reads in the raw cluster. There were 171,127,610 read pairs with quality scores no less than Q20, which were 93.86% of the raw reads and 97.50% of the Q20 reads. The trimmed pair-end reads and the trimmed singletons were processed to make a de novo assembly.

S6.2.3 De novo assembly

The Trinity assembler (v2.2.0) produced a total number of 901,869 contigs (Trinity gene isoforms) in 632,094 Trinity genes with an N50 length of 674, and this was defined as the raw *C. macra* transcriptome (Table S6.4). The filter criteria of minimum FPKM 0.5 and minimum isoform 1.0% constrained the Trinity assembly to 297,695 contigs (Trinity gene isoforms) in 205,924 Trinity genes, with an N50 length of 1009. There were 604,174 contigs (Trinity gene isoforms) in 480,591 Trinity genes filtered out by the criteria. A total of 179,921,215 reads, of which 171,127,610 (95.11%) were paired and 8,793,605 reads (4.89%) were unpaired, were aligned to the initial Trinity output (the raw *C. macra* transcriptome) by Bowtie2 with a 76.35% overall alignment rate. These reads were also aligned to the *C. macra* transcriptome with 72.21% overall alignment rate.

The alignment rates for Q20 reads from each library mapped to the *C. macra* transcriptome were 80.53%, 70.82%, 66.27% and 82.90% respectively. As the rRNA was not excluded in the library construction process in order to obtain any bacterial information, sequence reads that were contributed to the abundant rRNA was assessed that 33.38% of the total Q20 trimmed reads that were mapped to the *C. macra* plastid,

mitochondria and cytoplasmic rRNA references. The alignment rates for Q20 reads from each individual library mapped to the individual *C. macra* transcriptome generated by de novo assembling each individual library was 83.88%, 78.74%, 63.22% and 84.22% respectively.

Contig length	Filtered contig	Total contig	SOAP contig
200 bp-500 bp	164,572	620004	304203
501 bp-1,000 bp	76,245	184,362	139,350
1,001 bp-2,000 bp	39,933	72,646	54,516
2,001 bp-3,000 bp	11,223	17,131	11,885
3,001 bp-4,000 bp	3,655	5,161	3,393
4,001 bp-5,000 bp	1,207	1,562	1,259
>5,001 bp	860	1,003	983
In total	297,695	901,869	5155,89
N50	1,009	674	783
GC%	47.01%	47.12%	45.69%

Table S6.4 Statistics of assembled contigs in the *C. macra* transcriptome. Filtered contig (the *C. macra* transcriptome) and total contig (the raw *C. macra* transcriptome) were from Trinity assembler. SOAP contigs were from SOAPdenovo-Trans assembler (v1.03).

S6.2.4 Assembly completeness

The raw *C. macra* transcriptome (901,869 contigs), the *C. macra* transcriptome (297,695 contigs) and the filtered-out transcriptome were assessed in order to investigate the completeness of the transcriptome and whether useful information was lost in the filtering process (Table S6.5). The *C. macra* transcriptome matched 428 of the 458 core proteins (93.43%) that are present in a wide range of eukaryotic species with scores above the threshold in CEGMA. This transcriptome also matched 213 of the 248 highly conserved proteins (85.89%) with full alignments and 238 of the 248 highly conserved proteins (95.97%) including partially aligned, with an average of 3.24 orthologues per CEGs. The raw transcriptome and the filtered-out transcriptome matched 432 of the 458 (94.32%) and 184 of the 458 (40.17%) core proteins, and 217 of the 248 (87.50%) and 60 of the 248 (24.19%) highly conserved proteins with complete length respectively. The transcriptome generated by SOAPdenovo-Trans was also examined, with 338 of the 458 (73.80%) core proteins and 126 of the 248

(50.81%) highly conserved proteins matched. Thus, the transcriptome generated by the Trinity assembler was of better quality than SOAPdenovo-Trans in the dataset.

	458 core proteins	248 highly core proteins	
	Match	Complete match	Match
Trinity Raw transcriptome	94.32%	87.50%	96.37%
Trinity *C. macra* transcriptome	93.43%	85.98%	95.97%
Trinity filtered-out transcriptome	40.17%	24.19%	45.56%
SOAP Raw transcriptome (merged)	73.80%	50.81%	78.23%
Trinity Raw transcriptome macra1	93.88%	88.31%	95.16%
Trinity Raw transcriptome macra2	93.88%	83.47%	95.16%
Trinity Raw transcriptome macra3	91.27%	74.60%	92.74%
Trinity Raw transcriptome macra4	95.41%	89.92%	96.77%

Table S6.5 The completeness of the *C. macra* transcriptome. Core proteins were derived from CEGMA database. Complete match was defined as matching more than 70% alignment length with core proteins. Match was defined as including complete match and partial match (<70%). Merged libraries, the combination of four libraries in the assembly of a *C. macra* transcriptome by the Trinity.

The completeness of the individual *C. macra* transcriptome assembled with each single library was also assessed. The four individual *C. macra* transcriptome matched 438 (93.88%), 430 (93.88%), 418 (91.27%) and 437 (95.41%) of the 458 conserved core proteins present in most eukaryotic species, and also matched 219 (88.31%), 207 (83.47%), 185 (74.60%) and 223 (89.92%) of the 248 highly conserved core proteins with full alignments respectively, which were slightly higher than that of the *C. macra* transcriptome assembled with the merged libraries.

There were 297 nucleotide sequences and 341 protein sequences of *Chionochloa* species in the NCBI GeneBank, among which only 16 nucleotide sequences and 178 protein sequences were from the *C. macra*. Our transcriptome matched all 297 nucleotide sequence and 311 protein sequences from *Chionochloa* species. It also matched 175 protein sequences from *C. macra* in GeneBank with at least 73% similarity.

As there was no transcriptome or genome available for any close relatives of *Chionochloa* in the *Danthonioideae*, we chose one of the closest species in the *Poaceae* family, *Eragrostis tef*, the transcriptome of which was available, to assess the completeness of the *C. macra* transcriptome. Five model plants (*Oryza sativa Japonica, Seteria italica, Sorghum bicolor, Brachypodium distachyon, and Zea mays*) were also included in the analysis. The predicted ORFs from the *C. macra* transcriptome matched 16.46%, 22.12%, 23.03%, 26.60%, 30.33% and 37.25%

respectively of the transcriptome in the six species (*Sorghum bicolor*, *Oryza sativa Japonica*, *Zea mays*, *Seteria italica*, *Brachypodium distachyon* and *Eragrostis tef*) with a protein similarity of no less than 40.0% (Figure S6.1). *Eragrostis tef* had the largest number of proteins (13,075 transcripts of proteins with similarities no less than 50%) showing high levesl of similarity with those of the *C. macra* transcriptome.

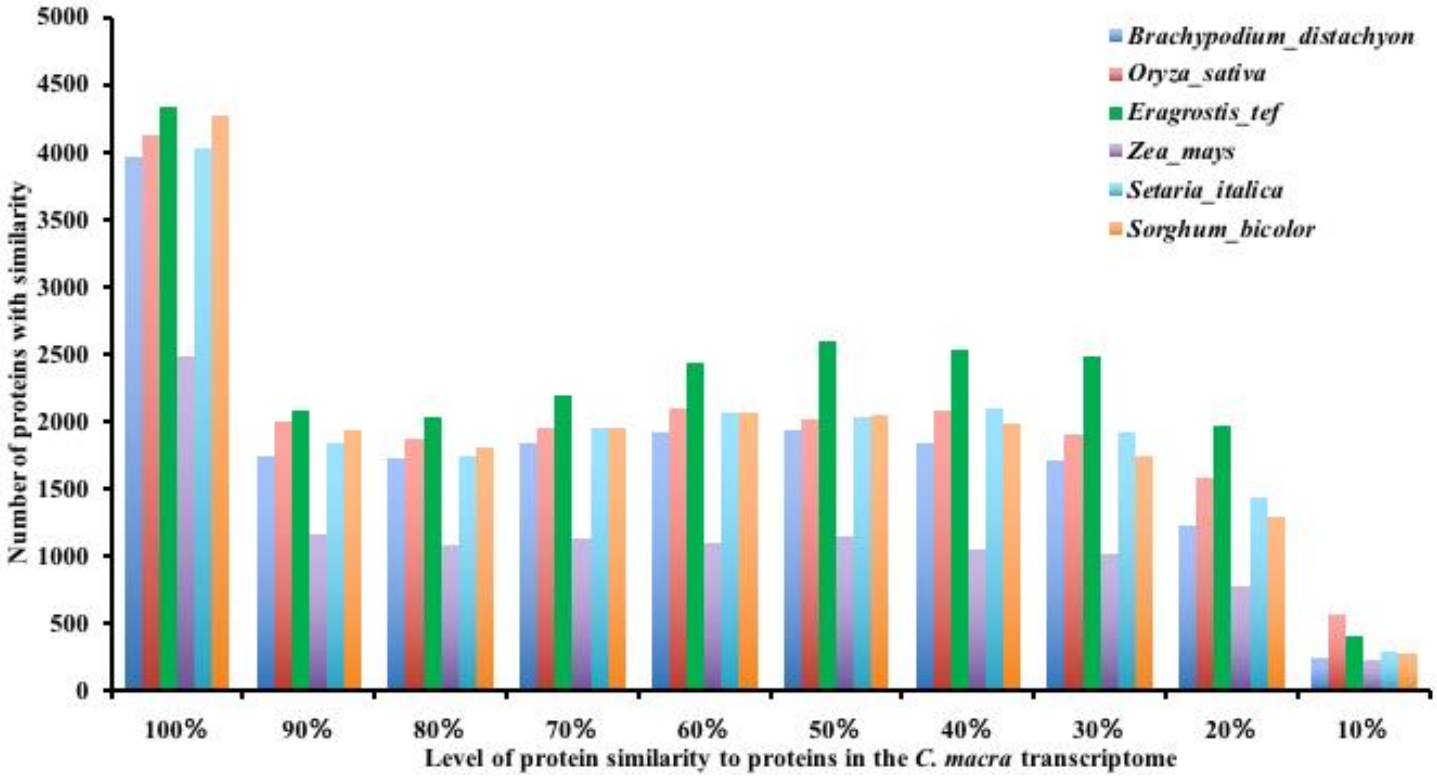


Figure S6.1 Transcriptome similarities between the *C. macra* and six plant species. *Eragrostis tef* showed the highest and *Zea may* showed the lowest similarities to the *C. macra* transcriptome.

S6.2.5 Transcriptome annotation

Annotation was performed on the *C. macra* transcriptome (the 297,695 contigs). A total of 157,491 (52.90%) contigs gave at least one Blastx/Blastn hit against the nr, UniProt, TrEMBL and nt database (e-value<10^{-5}). Of the 297,695 contigs, 122,557 gave domain matches by Interproscan and 175,138 contigs were without an Interproscan hit. The top five annotated domains from Interproscan with the largest number of contigs hits were non-cytoplasmic domains, transmembrane domains, TMhelix, cytoplasmic domains, and signal peptides (Figure S6.2). The top five Pfam domain hits with e-values less than 10^{-3} were Leucine rich repeats (PF00560), Leucine rich repeats (PF12799), and the Pentatricopeptide repeat family (PPR_1,PF01535; PPR_2, PF13041; PPR_3 PF13812) (Figure S6.4). There were 140,204 (47.10%) contigs without any Blast hit against either of the mentioned databases. Species that had the largest number of the Blastx hits with the *C. macra* transcriptome in Blast2GO were *Oryza sativa* (18.78%), *Zea mays* (15.17%), *Setaria italica* (7.21%), *Sorghum bicolor* (6.25%) and *Brachypodium distachyon* (6.23%) (Figure S6.3).

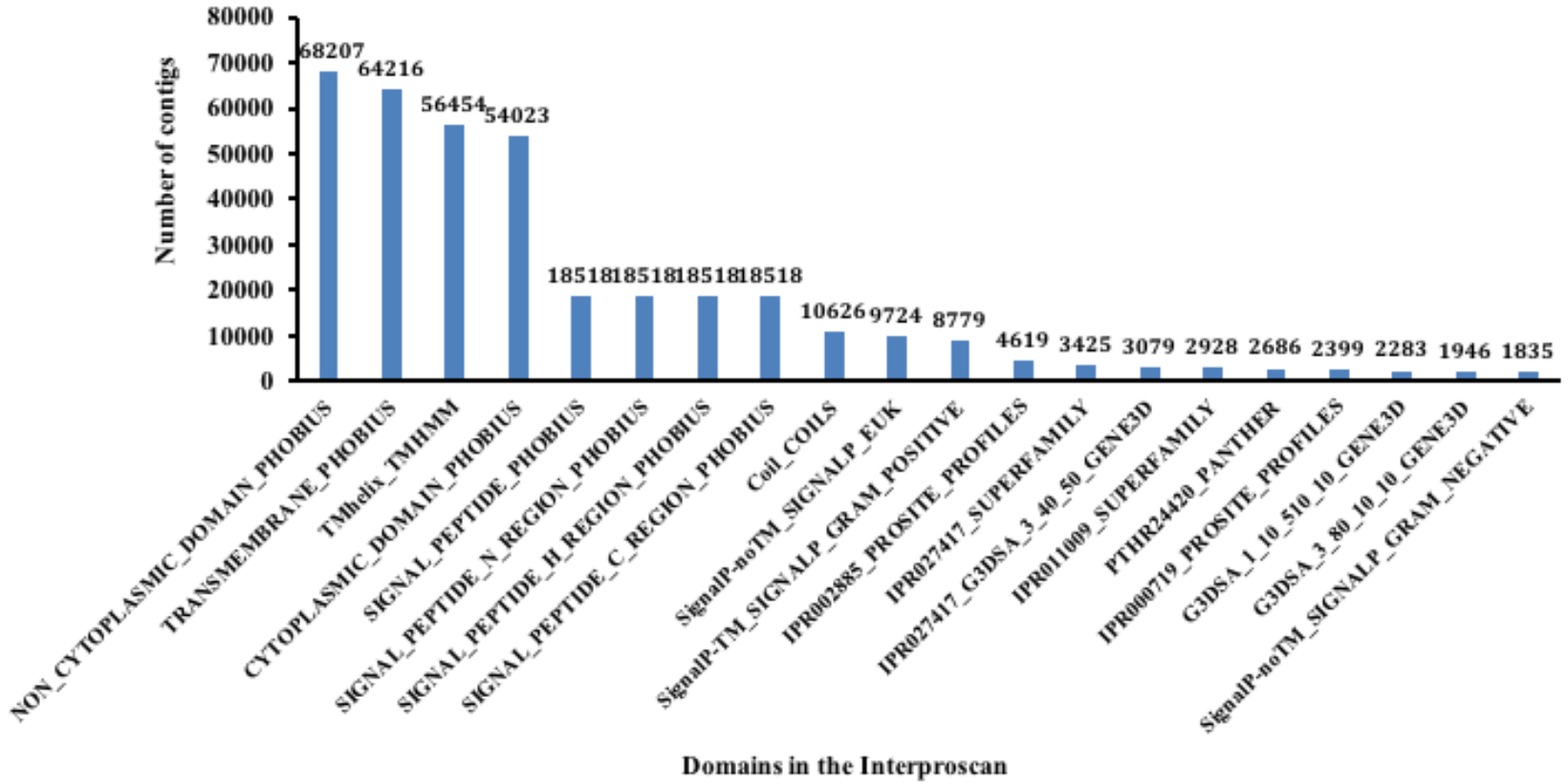


Figure S6.2 Protein domains in the *C. macra* transcriptome summarized by Interproscan. Non-cytoplasmic domain, transmembrane, TMhelix and cytoplasmic domain had the highest number of contigs in the *C. macra* transcriptome.

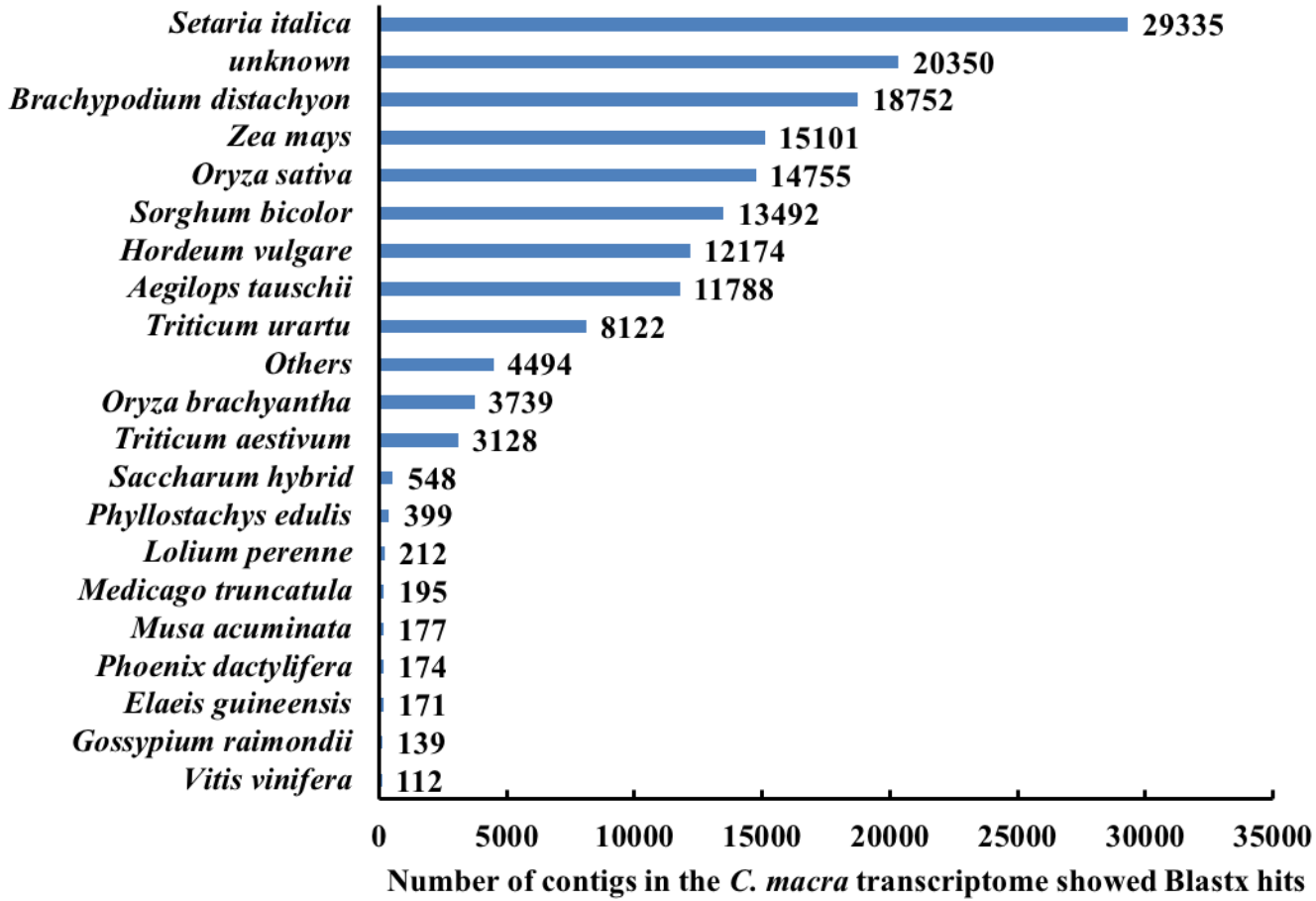


Figure S6.3 Species distributions of Blastx hits against the *C. macra* transcriptome summarized by Blast2GO. Unknown, no Blastx hit between these contigs from the *C. macra* transcriptome and proteins in any other plant species.

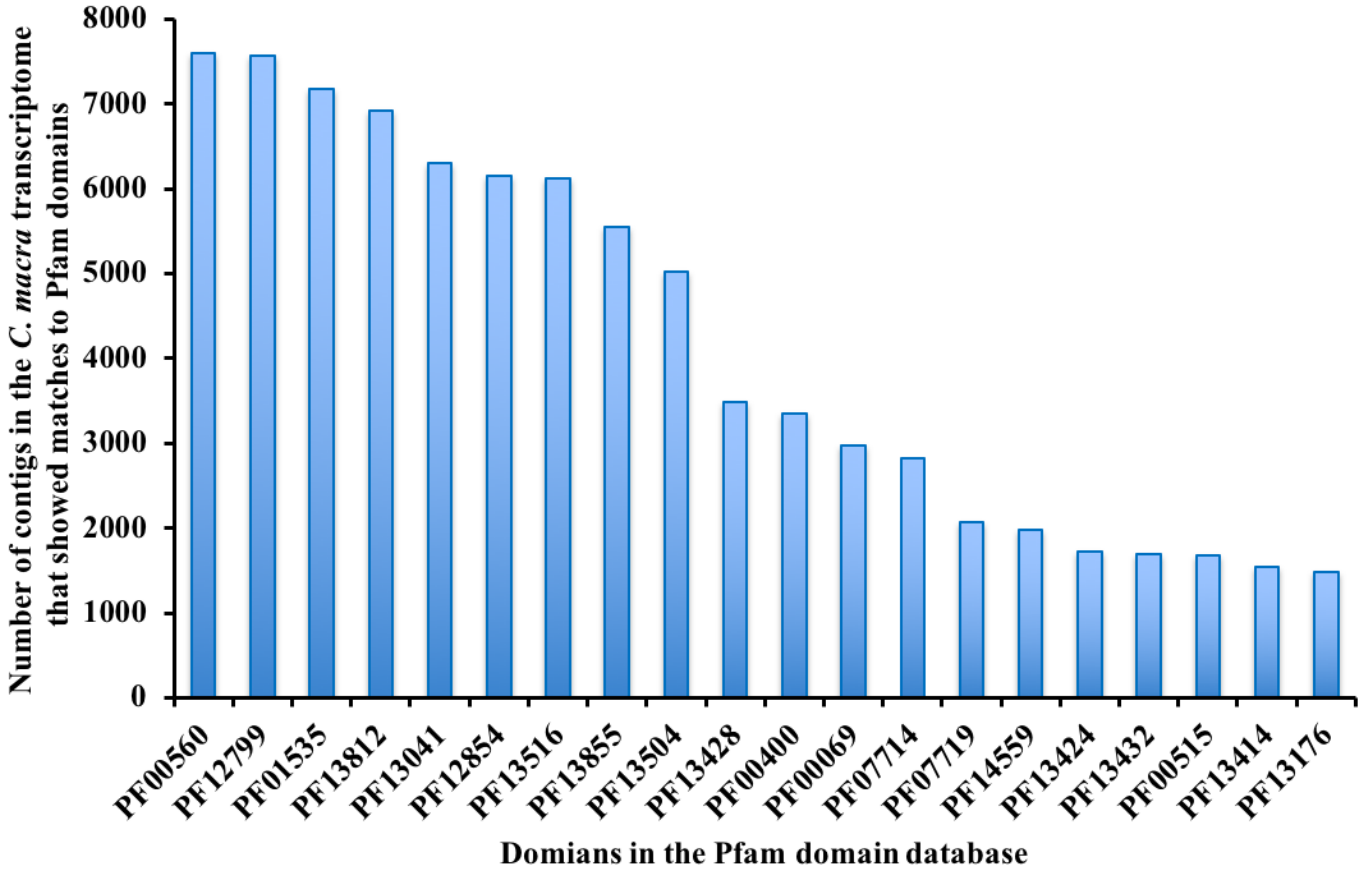


Figure S6.4 Domains matches in the *C. macra* transcriptome summarized by Pfam domain search.

Blast2GO assigned GO annotations to 109,850 (36.9%) contigs in the *C. macra* transcriptome. The annotated contigs were assigned to three main GO catalogs: Molecular function (GO: 0003674), Cell component (GO: 0005575) and Biological process (GO: 0008150) (Figure S6.5). The annotated contigs were assigned to 52 GO terms. In the Biological process catalog, most contigs were assigned to metabolic process (GO: 0008152) (73,543 contigs), cellular process (GO: 0009987) (67,434 contigs) and single-organism process (GO: 0044699) (52,864 contigs). In the Molecular function GO catalog, the majority of the contigs were assigned to binding (GO: 0005488) (63,558 contigs), catalytic activity (GO: 0003824) (55,435 contigs), transporter activity (GO: 0005215) (6,146 contigs), nucleic acid binding transcription factor activity (GO: 0001071) (2,690 contigs) and molecular transducer activity (GO: 0060089) (2,687 contigs). In the Cell component catalog, the top four GO terms that included most contigs were cell (GO: 0005623) (73,424 contigs), organelle (GO: 0043226) (62185 contigs), membrane (GO: 0016020) (26,254 contigs) and macromolecular (GO: 0032991) (18,827 contigs).

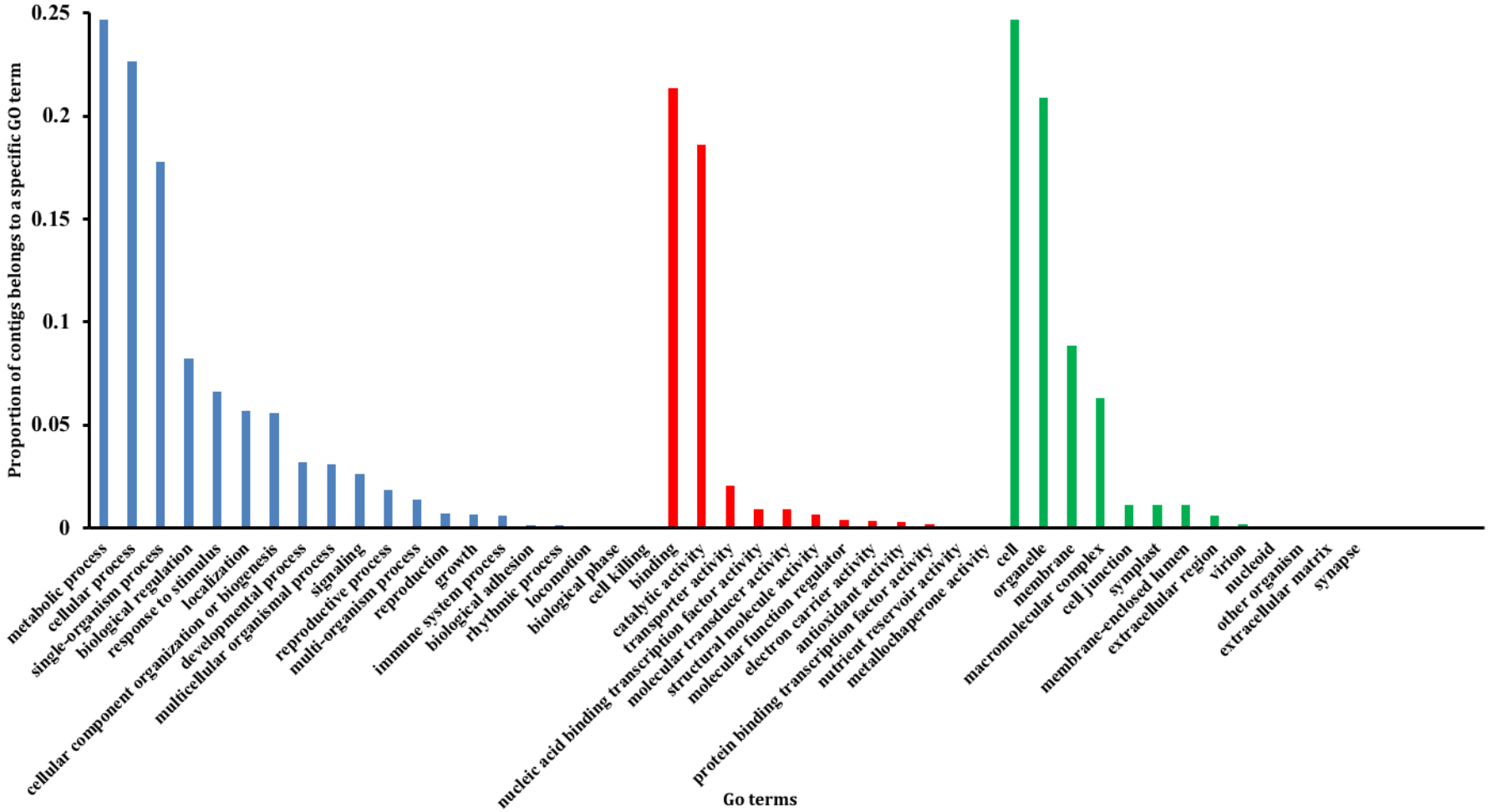


Figure S6.5 Gene ontology classifications of the *C. macra* transcriptome. Biological process (blue), Cellular component (red) and Molecular function (green).

There were 36,477 contigs mapped to 1,130 enzymes from 143 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Figure S6.6). Most enzyme-related contigs were grouped into the following four enzyme classes: hydrolases (39.00%), transferases (38.93%), oxidoreductases (12.07%) and ligase (3.66%). The top five KEGG pathways with largest number of contigs were purine metabolism (7,286 contigs, 58 enzymes), thiamine metabolism (5,856 contigs, 10 enzymes), biosynthesis of antibiotics (2,043 contigs, 163 enzymes), starch and sucrose metabolism (2,043 contigs, 47 enzymes) and aminobenzoate degradation (1,645 contigs, eight enzymes).

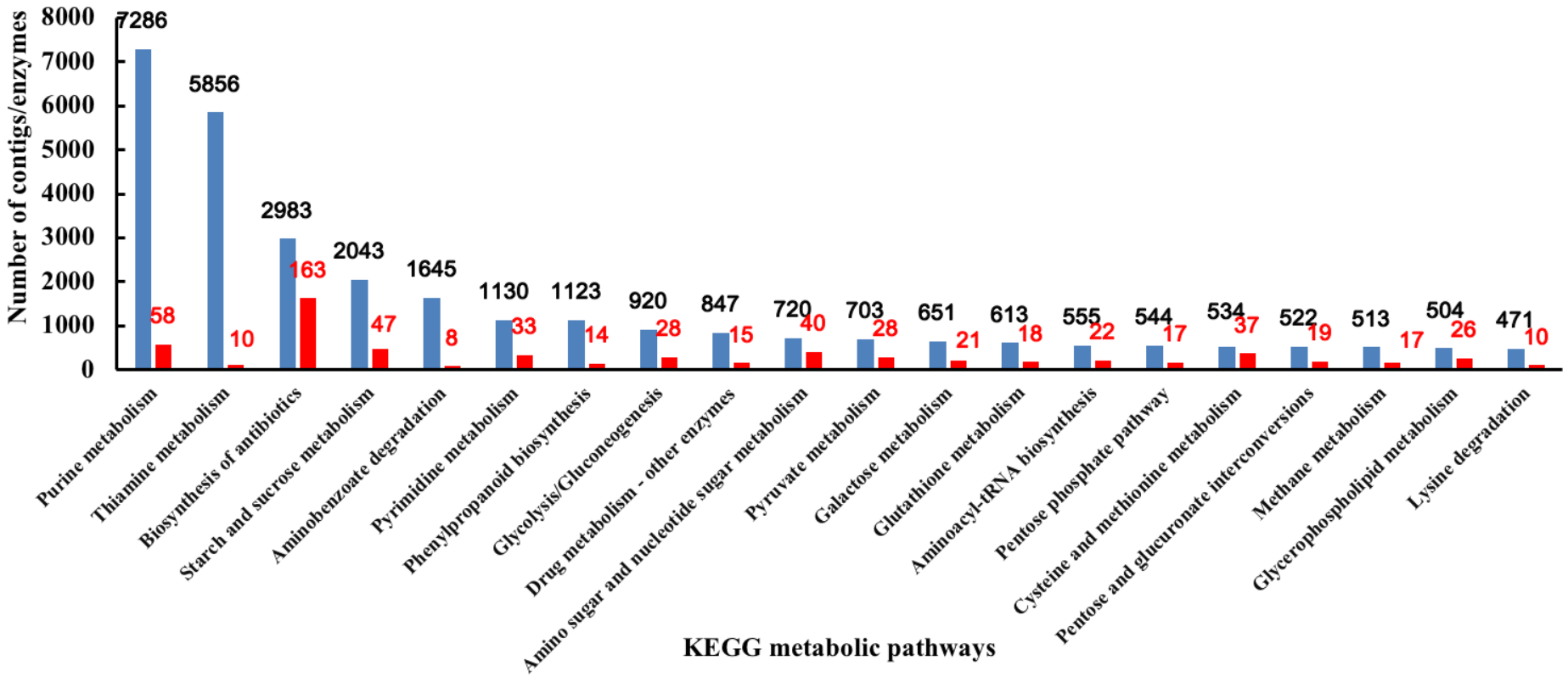


Figure S6.6 Top 20 KEGG pathways mapping to the *C. macra* transcriptome. Y axis, colour in blue, number of contigs in each KEGG pathway; colour in red, number of related enzymes (expressed in the *C. macra* transcriptome) in each KEGG pathway.

Contigs were also mapped to the 25 functional categories in COG (Clusters of Orthologous Groups) and KOG (Eukaryotic Orthologous Groups) database to identify ortholog and paralog proteins (Figure S6.7). The top five COG functional categories with the largest number of contigs were R (General function prediction only) (18.72%), L (Replication, recombination and repair) (9.70%), K (Transcription) 9.28%), T (Signal transduction mechanisms) (8.43%) and O (Posttranslational modification, protein turnover, chaperones) (7.79%). Whereas the top five KOG catalogs with the largest number of contigs were T (Signal transduction mechanisms) (15.64%), R (General function prediction only) (12.00%), O (Posttranslational modification, protein turnover, chaperones) (9.87%), K (Transcription) (6.88%), U (Intracellular trafficking, secretion, and vesicular transport) (6.38%) and S (Function unknown) (5.82%).

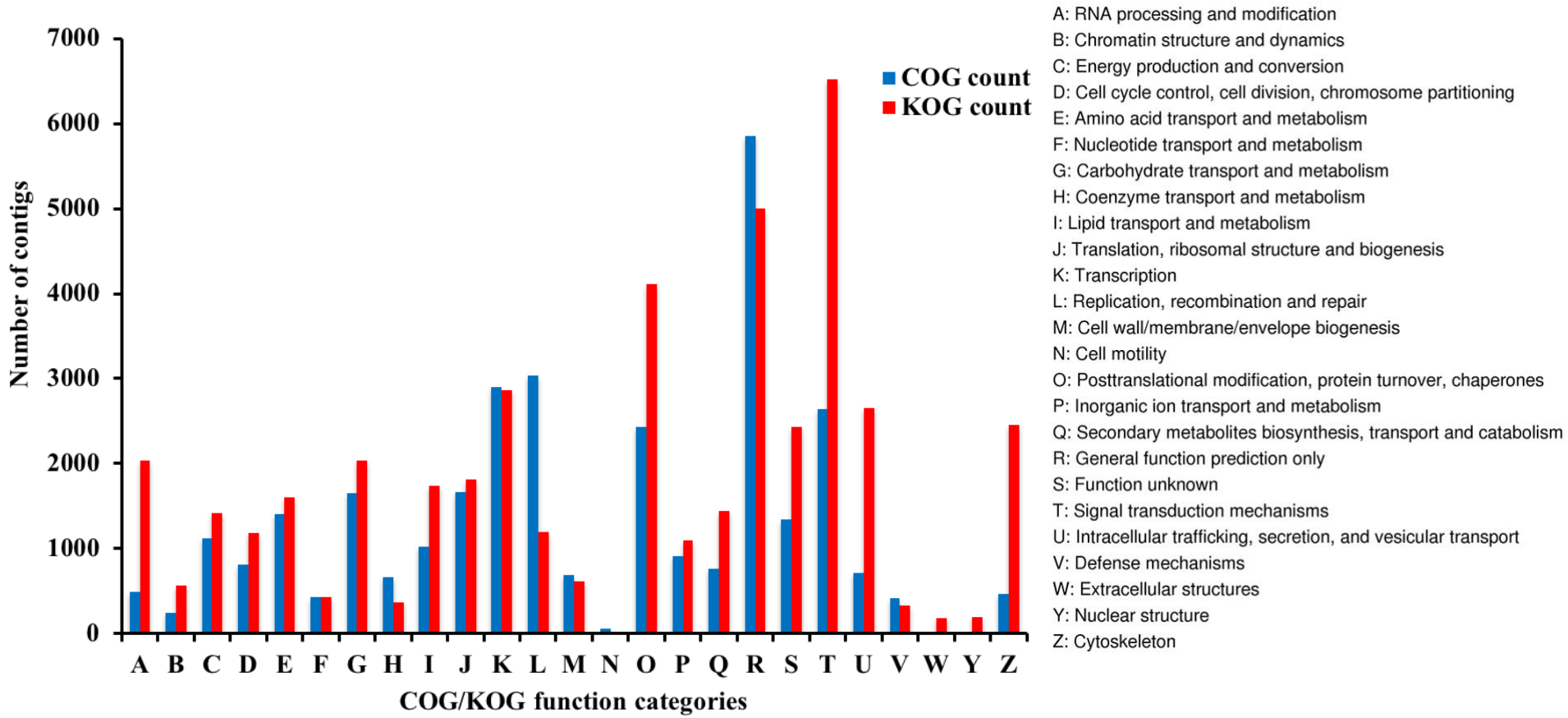


Figure S6.7 COG/KOG classifications of the *C. macra* transcriptome into the 25 functional categories. COG, Clusters of Orthologous Groups; KOG, Eukaryotic Orthologous Groups.

S6.2.6 Unannotated transcriptome analysis

The repeat elements were assessed by the RepeatMasker with the Repbase database (Table S6.6). Different retroelements (SINE, LINE and LTR elements) and DNA transposons were detected in the *C. macra* transcriptome and the unannotated transcriptome (contigs without any Blast hit). The LTR elements and the LINEs represented the majority of retroelements in the *C. macra* transcriptome, with Gypsy/DIRS1 to be the most abundant LTR elements and L1/CIN4 to be the most abundant LINEs. The DNA transposons were also highly represented in the *C. macra* transcriptome, with hobo-Activator, Tc1-IS630-Pogo and Tourist/Harbinger the most represented. The total interspersed repeats took 4.30% of the entire length of the *C. macra* transcriptome. Simple repeats, low complexity repeat, and small RNA related repeats were also abundant, with length proportions of 1.07%, 0.20% and 0.08% respectively.

There were obvious decreases in the numbers of repeats reported in the unannotated *C. macra* transcriptome compared with those in the *C. macra* transcriptome, with only 5,822 retroelements, 6,306 DNA transposons, 17,608 simple repeats and 3,687 low complexity repeat reported in the unannotated *C. macra* transcriptome. The total interspersed repeats reduced to 1.93% of the entire length of the unannotated dataset.

The repeat elements in the filter-out transcriptome (the 604,174 contigs) were also assessed. The number of repeats and the percentage of repeat occupied length in most repeats types of the filtered-out transcriptome was significantly higher than those of the *C. macra* transcriptome whereas the number of repeat related small RNA and other DNA transposon types were higher in the *C. macra* transcriptome than the filtered-out transcriptome (Table 6.6).

Repeat types	Filtered-out Transcriptome			*C. macra* Transcriptome			Unannotated Transcriptome		
	Number	Length	Percentage	Number	Length	Percentage	Number	Length	Percentage
	604174	276842238bp	100%	297694	210455460bp	100%	153070	69426723bp	100%
Retroelements	57473	12929831bp	4.67%	22957	5884489bp	2.80%	5822	587274bp	0.85%
SINEs:	845	72779bp	0.03%	350	32301bp	0.02%	217	20613bp	0.03%
Penelope	3	186bp	0.00%	2	202bp	0.00%	2	202bp	0.00%
LINEs:	24703	5990454bp	2.16%	8223	2189895bp	1.04%	1504	120552bp	0.17%
RTE/Bov-B	606	51254bp	0.02%	557	54647bp	0.03%	304	20385bp	0.03%
L1/CIN4	24077	5938007bp	2.14%	7658	2134500bp	1.01%	1196	99714bp	0.14%
LTR elements:	31925	6866598bp	2.48%	14384	3662293bp	1.74%	4101	446109bp	0.64%
Ty1/Copia	15692	3398266bp	1.23%	6310	1362749bp	0.65%	1796	196644bp	0.28%
Gypsy/DIRS1	14646	3283452bp	1.19%	7434	2218966bp	1.05%	1912	204798bp	0.29%
DNA transposons	29140	4141013bp	1.50%	16010	2671665bp	1.27%	6306	680470bp	0.98%
hobo-Activator	5624	793186bp	0.29%	3251	580214bp	0.28%	1223	117599bp	0.17%
Tc1-IS630-Pogo	3290	391476bp	0.14%	1612	213409bp	0.10%	968	120590bp	0.17%
Tourist/Harbinger	4184	605867bp	0.22%	2164	373331bp	0.18%	833	98963bp	0.14%
Other	22	1295bp	0.00%	26	1621bp	0.00%	10	645bp	0.00%
Unclassified:	4634	601042bp	0.22%	3094	483799bp	0.23%	815	68761bp	0.10%
Total interspersed repeats:		17671886bp	6.38%		9039953bp	4.30%		1336505bp	1.93%
Small RNA:	1035	103244bp	0.04%	1089	165586bp	0.08%	539	67193bp	0.10%
Satellites:	159	14768bp	0.01%	71	7199bp	0.00%	33	3597bp	0.01%
Simple repeats:	62806	2673474bp	0.97%	51821	2255985bp	1.07%	17608	754220bp	1.09%
Low complexity:	11089	554168bp	0.20%	8448	427639bp	0.20%	3687	187002bp	0.27%

Table S6.6 Repeats information in the *C. macra* transcriptome mapped by RepeatMasker. LINE, Long Interspersed Nuclear Elements. SINE, Short Interspersed Elements. Other, include repeat types of Mirage, P-element, Transib.

Sequences without Blastx hit against nr database took 47.1% (153,070 contigs) of the *C. macra* transcriptome. There were 12,796 of the 153,070 contigs showed at least one Blast hit against nt database. The 153,070 contigs were assessed for redundancy by the CD-HIT-EST at a similarity threshold of 90.0% and were clustered into 145,379 non-redundant clusters.

S6.2.7 Comparative analysis with model plants

The Venn diagram summarized by OrthoVenn indicated the distribution of shared gene families among the six plant species studied. Proteins in the six species formed 25,338 orthologous clusters (one cluster contains at least two species) and 2,616 single-copy gene clusters.

Species	Protein	Cluster	Singleton
Chionochloa macra	72197	18699	44321
Eragrostis tef	42052	16823	14570
Oryza sativa	35679	20737	10463
Setaria italica	35471	21177	8964
Brachypodium distachyon	26552	19040	3785
Sorghum bicolor	34496	21413	6265

Table S6.7 Summary of the six plant species studied by OrthoVenn.

There were 72,197 protein sequences identified by the Transdecoder from the *C. macra* transcriptome (Table S6.7). These protein sequences were clustered into 18,699 functional clusters with 44,321 singletons (Table S6.7). Of the 18,699 clusters, 8,504 clusters were shared by all the six plant species and 3,116 clusters were only belonged to *C. macra* (Figure S6.8). Other clusters of *C. macra* were shared by at least one plant species with *C. macra*.

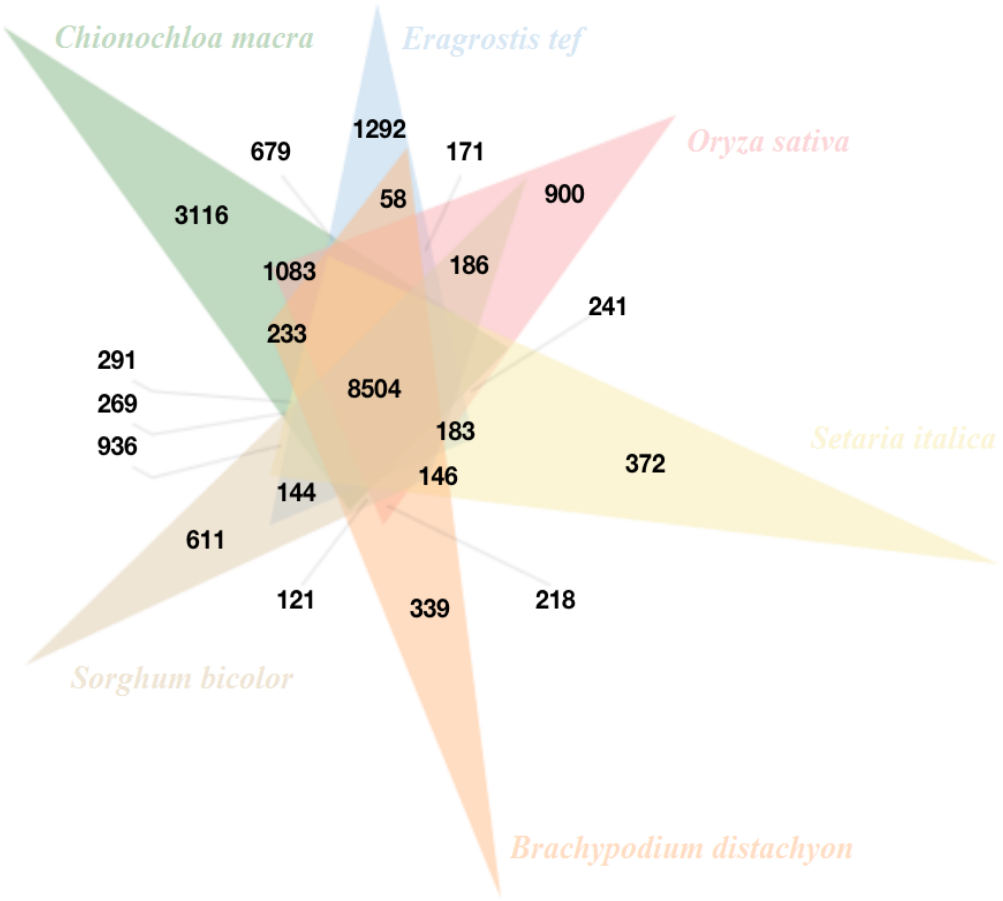


Figure S6.8 Distribution of shared gene families between the *C. macra* and the other five plant species by the Venn diagram. Number in the figure indicated the number of gene family clusters shared by species.

A total of 10,570 protein sequences identified from the *C. macra* transcriptome were included the 8,504 shared clusters, with 4,242 protein sequences included in 3275 clusters returning no GO annotation. The *C. macra* specific clusters consisted of 11,422 protein sequences with 2,335 sequences in 664 clusters given GO annotations. There was no GO annotation for the 2,452 of the 3,116 clusters containing 9,087 protein sequences from the *C. macra* transcriptome.

In the six species shared cluster, Gene Ontology annotation indicated GO terms of biological process (GO: 000850), metabolic process (GO: 0008152), cellular metabolic process (GO: 0044237), cellular process (GO: 0009987) and nitrogen compound metabolic process (GO: 0006807) contained the largest number of protein sequences of the *C. macra* transcriptome with 1,971, 1,718, 1,466, 1,388 and 963 sequences respectively in the Biological process GO catalog (Figure S6.8). In the

Molecular function GO catalog, GO terms of ion binding (GO: 0043167), transferase activity (GO: 0016740), nucleotide binding (GO: 0000166), nucleic acid binding (GO: 0003676) and binding (GO: 0005488) contained the largest number of sequences from the *C. macra* transcriptome with 1,709, 974, 971, 906 and 832 sequences respectively (Figure S6.9). In the Cellular component GO catalog, the top five GO terms including the largest number of sequences from the *C. macra* transcriptome were cell part (GO: 0044464), cellular component (GO: 0005575), intracellular (GO: 0005622), membrane (GO: 0016020), with 1,336, 1,292, 1,274, 1,106 and 809 sequences respectively.

GO classifications in the *C. macra* species-specific cluster were similar to the shared cluster but with fewer sequences (Figure S6.10). GO terms of DNA integration (GO: 0015074), transposition (GO: 0032196), cotyledon vascular tissue pattern formation (GO: 0010588), nodulation (GO: 0009877) and negative regulation of gene expression, epigenetic (GO: 0045814) were enriched in the Biological process catalog (Hypergeometric test, $p<0.05$). RNA-directed DNA polymerase activity (GO: 0003964), structural constituent of cell wall (GO: 0005199), RNA-DNA hybrid ribonuclease activity (GO: 0004523), protein binding (GO: 0005515) and endonuclease activity (GO: 0004519) were enriched in the Molecular function catalog ($p<0.05$). The enriched GO terms in the Cell component catalog were nuclear envelope ($p<0.05$).

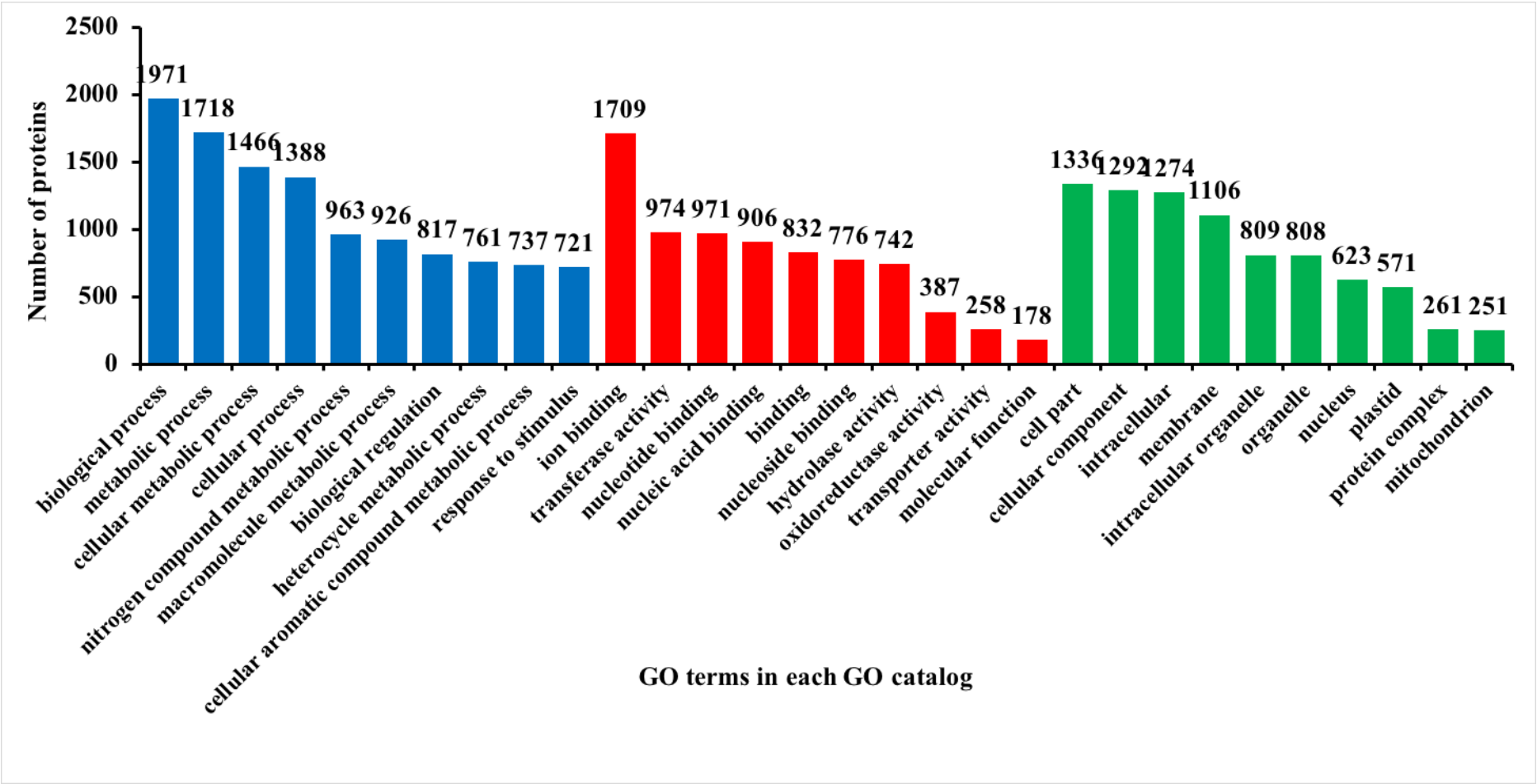


Figure S6.9 GO classification of the six species shared clusters in the Venn diagram. Blue, Biological process GO catalog; Red, Molecular function GO catalog; Green, Cell component GO catalog.

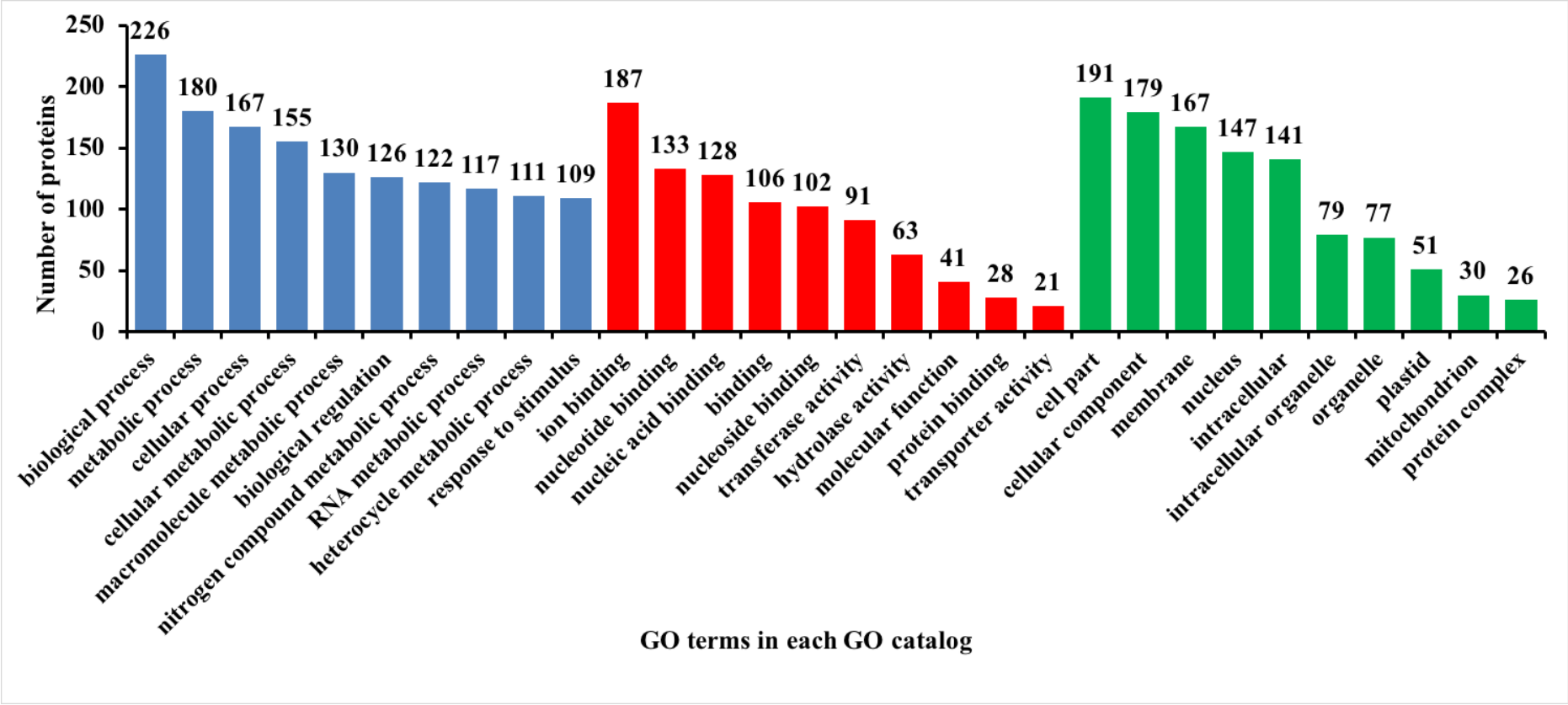


Figure S6.10 GO classification of the *Chionochloa macra* specifics clusters in the Venn diagram. Blue, Biological process GO catalog; Red, Molecular function GO catalog; Green, Cell component GO catalog.

S6.2.8 Highly expressed genes in the *C. macra* transcriptome

The threshold of minimum FPKM value 0.5 filtered out a total of 604,125 contigs in the raw *C. macra* transcriptome (Table S6.8). There were 175,077 contigs (58.8% of the *C. macra* transcriptome) with FPKM values ranging from 0.5 to 1.0, 117,965 contigs (39.6% of the *C. macra* transcriptome) with FPKM values ranging from 1.0 to 10, 3,868 contigs (1.3% of the *C. macra* transcriptome) with FPKM values between 10.0 and 20.0, and 2,358 contigs (0.8% of the *C. macra* transcriptome) with FPKM values higher than 20.0 (Table S6.8).

FPKM range	Number of contig
0	96443
0-0.5	507682
0.5-1.0	175077
1.0-10.0	117965
10.0-20.0	3868
>20.0	2358

Table S6.8 FPKM range of contigs in the raw *C. macra* transcriptome.

The *C. macra* transcriptome (297,695 contigs) had a quartile FPKM value of 1.92 and approximate 10 times of this value (FPKM = 20) was considered to be comparatively highly expressed in this transcriptome. There were 2,358 contigs with FPKM >20. Of the 2,358 contigs, 1,677 gave GO mappings by Blast2GO and 343 contigs gave no Blastx hits.

The top 20 highly expressed contigs are listed in Table S6.9. The contig, c352768_g1_i1, which included the gene encoding the cytochrome p450 monooxygenase (EC: 1.14.14.1) and presented in many pathways including aminobenzoate degradation, fatty acid degradation, the metabolisms of caffeine, tryptophan, retinol, xenobiotics, linoleic acid and arachidonic acid, had the highest FPKM value (34,332.6) in the *C. macra* transcriptome (Table S6.9). The contig, c346811_g1_i2, with a second highest FPKM value in the transcriptome, edcoding the purple acid phosphatase 2, which catalyzed the dephosphorylation of a-xylosidase

and b-glucosidase, and regulated their activity in the plant cell wall (Kaida *et al*. 2010). The contig with the third highest FPKM value, c336809_g1_i1, encoding the maturase k, a plastid product that spliced the introns.

Gene Ontology enrichment analysis revealed that GO terms of translation, ribonucleoprotein complex biogenesis, ribosome biogenesis, cellular component biogenesis and oxidative phosphorylation in the Biological process catalog and GO terms of structural constituent of ribosome, structural molecule activity, NADH dehydrogenase activity, NADH dehydrogenase (ubiquinone) activity and NADH dehydrogenase (quinone) activity in the Molecular function catalog were significantly enriched in the highly expressed transcripts.

Contig	FPKM	Blastx hit description	Possible function
c352768_g1_i1	34332.6	t02955probable cytochrome p450 monooxygenase	Involved in biosynthetic and detoxification pathways (Persans *et al.* 2001).
c346811_g1_i2	12791.8	Purple acid phosphatase 2	Metalloenzyme that hydrolyses phosphate esters, controls the activity of a-xylosidase and b-glucosidase (Kaida *et al.* 2010).
c336809_g1_i1	10882.4	Maturase k	An intron maturase that splices introns (Mohr *et al.* 1993).
c352768_g3_i1	7490.2	Senescence-associated protein	Function unknown.
c352768_g1_i2	6776.7	Senescence-associated protein	Function unknown.
c364052_g5_i1	6767.1	YCF68 protein	A chloroplast protein. Function unknown.
c346811_g1_i1	5380.0	Domain-containing receptor-like kinase	Transmembrane protein. Involved in intracellular signaling (Greeff *et al.* 2012).
c267094_g1_i1	4876.5	DNA repair rada-like protein	Involved in DNA repair.
c140565_g1_i1	4816.3	NA	Function unknown.
c316189_g1_i2	4135.5	Mitochondrial protein	Mitochondrial protein
c299264_g1_i1	3430.4	Hypothetical protein B456_003G156400	Function unknown.
c162532_g1_i1	3161.7	Cell wall-associated hydrolase	Involved in cell wall polysaccharide degradation (Minic and Jouanin 2006).
c296344_g1_i1	2888.7	Chloroplast orf91	Function unknown.
c364052_g1_i1	2596.3	Photosystem ii cp47 chlorophyll apoprotein	Transmembrane protein. Involved in photosystem II (Mullet *et al.* 1990).
c550343_g1_i1	2290.4	*Phragmites australis* complete genome	Plastid protein.
c354389_g9_i1	2155.3	ATP synthase cf1 beta subunit	Hosted catalytic sites in ADP binding (Walker *et al.* 1982).
c162532_g2_i1	2099.4	Hypothetical protein (mitochondrion)	Mitochondrial protein
c342361_g1_i1	2003.5	Photosystem ii cp43 chlorophyll partial	Transmembrane protein. Involved in photosystem II (Mullet *et al.* 1990).
c267093_g1_i1	1944.8	ATP synthase subunit alpha	Regulatory role in ADP binding (Walker *et al.* 1982).
c366494_g1_i1	1875.4	Photosystem i p700 apoprotein a1	Involved in the photosystem. Electron donor (Kruip *et al.* 1993).

Table S6.9 Highly expressed genes in the *C. macra* transcriptome and their estimated functions.

S6.2.9 Featured KEGG pathways in the *C. macra* transcriptome

S6.2.9.1 Highly expressed KEGG pathways

There were 548 of the 2,358 contigs (FPKM>20) coding 154 enzymes and mapped to 94 KEGG pathways. Of the 660 contigs (FPKM>50), 331 encoded enzymes and mapped to 53 pathways. There were 109 contigs with FPKM value no less than 100, 76 of which mapped to 17 KEGG pathways. The top four KEGG pathways that contained the largest number of enzymes with FPKM no less than 50 were: biosynthesis of antibiotics, purine metabolism, starch and sucrose metabolism, and glycolysis/gluconeogenesis. Other highly expressed metabolic pathways were thiamine metabolism, pyrimidine metabolism, and oxidative phosphorylation and carbon fixation in photosynthetic organisms.

In the top 20 highly expressed contigs, c354389_g9_i1, c267093_g1_i1, c323117_g1_i1, c340097_g2_i1, c354389_g9_i2 and c358494_g13_i3 encoded enzymes in the pathways, purine metabolism and thiamine metabolism (Table S6.10). Contigs, c366494_g1_i1, c364052_g3_i3 and c364052_g3_i1 c364052_g3_i2, encoded enzymes in the pathways, purine metabolism and pyrimidine metabolism. Contigs, c348075_g4_i1, c364643_g2_i1, c309404_g2_i1, c364052_g4_i1, c358494_g2_i1, c364052_g6_i1, c364052_g2_i2 and encoded enzymes in the pathways, oxidative phosphorylation. Contigs, c358028_g2_i1 and c339202_g1_i1, encoded enzymes in the pathways, biosynthesis of antibiotics, glycolysis/gluconeogenesis and carbon fixation in photosynthetic organisms.

GO enrichment analysis by BINGO indicated that GO terms of translation, ribonucleoprotein complex biogenesis, ribosome biogenesis, cellular component biogenesis and oxidative phosphorylation were significantly enriched in the Biological process GO catalog (p-value<e^{-5}). Structural constituent of ribosome, structural molecule activity, NADH dehydrogenase activity, NADH dehydrogenase

(ubiquinone) activity and NADH dehydrogenase (quinone) activity were also significantly enriched in the Molecular function GO catalog (p-value<e^{-5}).

Contig	FPKM	Enzyme (EC)	Pathway
c354389_g9_i1	2155.3	3.6.1; 3.6.1.3; 3.6.1.15	Purine metabolism and thiamine metabolism
c267093_g1_i1	1944.8	3.6.1; 3.6.1.3; 3.6.1.15	
c323117_g1_i1	1442.1	3.6.1; 3.6.1.3; 3.6.1.15	
c340097_g2_i1	663.8	3.6.1; 3.6.1.3; 3.6.1.15	
c354389_g9_i2	586.0	3.6.1; 3.6.1.3; 3.6.1.15	
c358494_g13_i3	519.8	3.6.1; 3.6.1.3; 3.6.1.15	
c366494_g1_i1	1875.4	2.7.7.6	Purine metabolism and pyrimidine metabolism
c364052_g3_i3	1166.5	2.7.7.6	
c364052_g3_i1	1023.2	2.7.7.6	
c364052_g3_i2	553.7	2.7.7.6	
c348075_g4_i1	1274	1.6.99.5; 1.6.99.3; 1.6.5.3	Oxidative phosphorylation
c364643_g2_i1	1117.8	1.9.3.1	
c309404_g2_i1	1051.2	1.9.3.1	
c364052_g4_i1	770.9	1.6.99.5; 1.6.99.3; 1.6.5.3	
c358494_g2_i1	580.3	1.6.99.5; 1.6.99.3; 1.6.5.3	
c364052_g6_i1	553.7	2.7.7.6	
c358028_g2_i1	322.9	3.6.1; 3.6.1.3; 3.6.1.15	Biosynthesis of antibiotics, glycolysis/gluconeogenesis and carbon fixation in photosynthetic organisms
c339202_g1_i1	256.3	4.1.1.39	
c364052_g2_i2	253.6	1.6.5.3	Oxidative phosphorylation
c352947_g1_i3	245.6	1.6.5.3	

Table S6.10 Enzymes and related KEGG pathways in the top 20 high expressed contigs from the *C. macra* transcriptome.

S6.2.9.2 SSRs, SNPs, miRNAs and transcription factors

A total of 56,648 SSRs were identified by MISA in 45,860 contigs (15.41%) from the *C. macra* transcriptome and 887 SSRs were presented in the compound formation where the maximal number of bases interrupting two SSRs no more than 10. There were 8,387 contigs containing more than one SSR. The number of SSRs for the unit size 1, 2, 3, 4, 5 and 6 were 14,513, 6,707, 9,650, 16,718, 5,851 and 3,209 respectively (Table S6.11). A/T repeat was the majority (13,436 SSRs) in the one nucleotide repeats and AG/CT repeat was the most abundant (4,636 SSRs) in the de-nucleotide repeats. In the tri-nucleotide repeats, CCG/CGG repeat, AGG/CCT repeat and AGC/CTG repeat had the greatest number with 3,207, 2,018 and 1,364 SSRs respectively. AAAG/CTTT repeat, AAAT/ATTT repeat and ATCC/ATGG repeat were the most dominant in the tetra-nucleotide repeats, with 1,615 SSRs, 1,347 SSRs and 1,292 SSRs respectively. In the penta-nucleotide repeats, AAAAG/CTTTT repeat, AAAAT/ATTTT repeat and AGAGG/CCTCT repeat were the most dominant with 611 SSRs, 354 SSRs and 295 SSRs respectively. In the hexa-nucleotide repeats, AAAAAG/CTTTTT repeat, AGGCGG/CCGCCT repeat and CCGGCG/CCGGCG repeat were the most dominant with 170 SSRs, 154 SSRs and 136 SSRs respectively.

SSR size	Total number	SSR type	Number
1	14513	A/T	13436
		C/G	1077
2	6707	AG/CT	4636
		AC/GT	682
		AT/AT	1045
3	9650	CCG/CGG	3207
		AGG/CCT	2018
		AGC/CTG	1364
4	16718	AAAG/CTTT	1615
		AAAT/ATTT	1347
		ATCC/ATGG	1292
5	5851	AAAAG/CTTTT	611
		AAAAT/ATTTT	354
		AGAGG/CCTCT	295
6	3209	AAAAAG/CTTTTT	170
		AGGCGG/CCGCCT	154
		CCGGCG/CCGGCG	136

Table S6.11 SSRs in the *C. macra* transcriptome. The top three SSRs in each SSR type were shown.

A total of 709,367 single nucleotide polymorphisms (SNP), 38,738 of insertion (INS) and 29,640 deletion (DEL) were identified in 69,235 contigs by the SAMtool in the transcriptome assembled from the merged library. There were 127,491 missense variants, which took 48.3% of the variants effects by the three functional classes (missense, nonsense and silent). A number of 2,159 variants were nonsense which took 0.8%, and a number of 134,034 variants were silent, which took 50.8%. The missense/silent ratio is 0.95. There were 30.0% of the SNPs in the exon region and 41.4% of the SNPs in the intergenic region (Figure S6.11). G/A, A/G, C/T and T/C variants were the most abundant changes in the *C. macra* transcriptome (Figure S6.12).

Type

Type (alphabetical order)	Count	Percent
3_prime_UTR_variant	136,452	14.819%
5_prime_UTR_premature_start_codon_gain_variant	16,259	1.766%
5_prime_UTR_variant	105,230	11.428%
chromosome_number_variation	4,270	0.464%
disruptive_inframe_deletion	845	0.092%
disruptive_inframe_insertion	1,339	0.145%
disruptive_inframe_insertion+splice_region_variant	10	0.001%
feature_elongation	1,493	0.162%
frameshift_variant	5,606	0.609%
frameshift_variant+splice_region_variant	19	0.002%
frameshift_variant+start_lost	31	0.003%
frameshift_variant+stop_gained	201	0.022%
frameshift_variant+stop_gained+splice_region_variant	2	0%
frameshift_variant+stop_lost	76	0.008%
inframe_deletion	1,569	0.17%
inframe_insertion	2,612	0.284%
initiator_codon_variant	36	0.004%
intergenic_region	380,875	41.363%
missense_variant	126,627	13.752%
missense_variant+splice_region_variant	10	0.001%
splice_region_variant	72	0.008%
splice_region_variant+synonymous_variant	2	0%
start_lost	261	0.028%
start_lost+disruptive_inframe_deletion	1	0%
start_lost+disruptive_inframe_insertion	3	0%
start_lost+inframe_deletion	6	0.001%
start_lost+inframe_insertion	7	0.001%
stop_gained	2,159	0.234%
stop_gained+disruptive_inframe_deletion	3	0%
stop_gained+disruptive_inframe_insertion	38	0.004%
stop_gained+disruptive_inframe_insertion+splice_region_variant	1	0%
stop_gained+inframe_insertion	80	0.009%
stop_lost	557	0.06%
stop_lost+disruptive_inframe_deletion	1	0%
stop_lost+inframe_deletion	17	0.002%
stop_lost+inframe_insertion	2	0%
stop_retained_variant	130	0.014%
synonymous_variant	133,902	14.542%

Region

Type (alphabetical order)	Count	Percent
EXON	275,837	29.956%
INTERGENIC	380,875	41.363%
NONE	5,763	0.626%
SPLICE_SITE_REGION	74	0.008%
TRANSCRIPT	314	0.034%
UTR_3_PRIME	136,452	14.819%
UTR_5_PRIME	121,489	13.194%

Figure S6.11 SNPs found in the *C. macra* transcriptome. Effects to amino acid changes were estimated by snpEff. Different SNP types were shown.

	Nucleotide in the sequence			
	A	C	G	T
A		33607	**104658**	30689
C	37540		36573	**111653**
G	**112756**	37034		37213
T	30492	**103450**	33702	

Figure S6.12 Summary of base changes in the *C. macra* transcriptome. Numbers in red, the most frequent types of base changes, G/A, A/G, C/T and T/C.

The *C. macra* transcriptome was searched against the plant transcription factor (TF) database and 826 contigs in the transcriptome matched transcription factors in the database with protein similarity no less than 70%. There were 19 contigs that have FPKM value larger than 20 and 5 of them have FPKM larger than 50 (Table S6.12). The ethylene-responsive transcription factors gave the highest FPKM value (100.0) in

this search. Several TFs such as auxin response factor and squamosa promoter-binding-like protein involved in plant development are highly expressed.

Contig	FPKM	Blastx description	PTFD description
c364283_g3_i1	99.97	Ethylene-responsive transcription factor 1-like	Ap2 domain containing protein
c364283_g3_i2	84.2	Ethylene-responsive transcription factor 1-like	Ap2 domain containing protein
c365631_g6_i1	82.89	Ketol-acid chloroplastic-like	Myb family protein
c367940_g1_i1	74.1	Squamosa promoter-binding-like protein 15	Sbp family protein
c363836_g1_i1	61.78	Sucrose-phosphate synthase	Hd-zip family protein
c366008_g1_i5	48.29	Auxin response factor 4	Auxin response
c354437_g2_i3	37.91	Knotted-like transcription factor family protein	Homeobox protein knotted
c368966_g1_i1	36.99	Auxin response factor expressed	Auxin response
c363264_g1_i2	33.77	Auxin response factor 7-like	Arf family protein
c359400_g2_i1	32.84	Salt tolerance-like protein	B-box type zinc finger family protein
c358818_g2_i4	30.31	Probable histone h2a variant 3	Histone h2a 2
c360037_g1_i3	30.1	Auxin response factor 24-like	Auxin response
c362014_g1_i1	28.21	Zinc finger ccch domain-containing protein 13-like isoform x1	Zinc finger c-x8-c-x5-c-x3-h type family
c355694_g3_i2	27.68	Homeobox protein knotted-1-like 3 isoform x1	Tale family protein
c369856_g2_i2	27.43	Callose synthase 10	Myb related family protein
c362795_g2_i4	26.39	Nuclear factor yb2	Nf-yb family protein

c336811_g1_i2	25.98	Mixed-linked glucan synthase 8	Bzip family protein
c360416_g2_i1	25.47	Telomere repeat-binding factor 1-like isoform x2	Myb related family protein
c368966_g1_i2	24.35	Auxin response factor expressed	Auxin response

Table S6.12 High expressed transcription factors in the *C. macra* transcriptome. PTFD, plant transcriptional factor database.

There were 54 contigs showed Blast hits to the sequences deposited in miRNA database. Two contigs, c470538_g1_i1 and c357593_g3_i1, had FPKM >20 and four contigs, c351945_g4_i3, c351945_g4_i4, c351945_g4_i2 and c351945_g4_i5, had FPKM between 10 and 20 (Table S6.13). The complimentary miRNA sequences were searched from the corresponding transcriptome sequences to predict the potential targets. The contig c470538_g1_i1 matched c364052_g4_i2 and c364052_g4_i1, which encoded dehydrogenase (EC: 1.6.99.3) in oxidative phosphorylation. The contigs c357593_g3_i1 matched c151653_g1_i1 and c151653_g1_i2, and c351945_g4_i3, c351945_g4_i4 and c351945_g4_i5 matched c64524_g2_i1 and c64524_g1_i1. There was no annotation available for these contigs.

Contig	miRNA
c470538_g1_i1	MIR5079
c357593_g3_i1	MIR5082
c351945_g4_i3	MIR1119
c351945_g4_i4	MIR1119
c351945_g4_i2	MIR1119
c351945_g4_i5	MIR1119

Table S6.13 Contigs (FPKM>10) in the *C. macra* transcriptome that matched miRNA in the miRNA database.

S6.2.10 Alternative splicing events

There were 45,579 of the 205,924 (22.1%) Trinity genes or gene fragments in the *C. macra* transcriptome containing at least two isoforms (Figure S6.13). BLAT showed that 57,653 Trinity isoforms belonging to 24,836 Trinity genes in the *C. macra* transcriptome were undergoing BLAT alternative splicing (AS) events. Trinity gene c362878_g1 contained the largest number of AS isoforms (15), followed by c364323_g1 (14) and c360026_g2 (14) (Table S6.14).

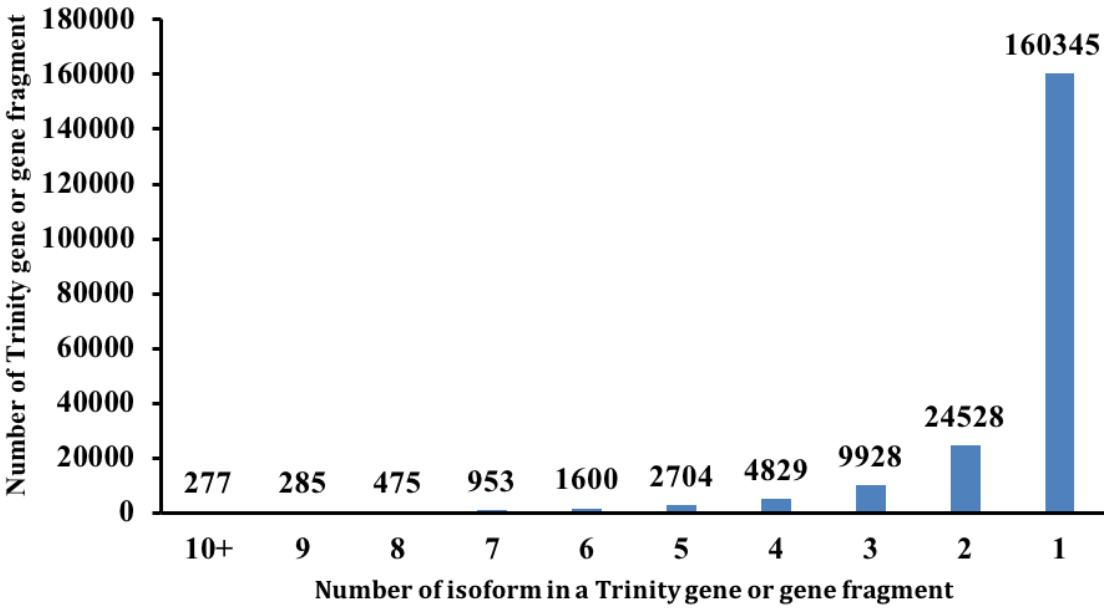


Figure S6.13 Summary of isoforms of Trinity genes in the *C. macra* transcriptome. Most Trinity gene or gene fragment contained only one isoform.

Numbers of gene isoform	Trinity gene	Blastx description	Numbers of AS isoform	Trinity gene	Blastx description
19	c341478_g1	NA	15	c362878_g1	NA
18	c356275_g1	Transcription factor myb1r1-like	14	c364323_g1	Microtubule-associated protein 70-4-like
16	c362878_g1	NA	14	c360026_g2	Splicing factor 3b subunit 2
16	c354745_g1	NA	14	c341478_g1	NA
15	c363150_g1	NA	13	c368208_g1	Udp-glucose pyrophosphorylase 3 isoform
15	c358716_g2	Cysteine proteinase	13	c359605_g1	B3 domain-containing protein
15	c340411_g1	Predicted protein	13	c359505_g3	Chaperonin-like protein
14	c367172_g2	Regulatory particle non-atpase	13	c356275_g1	Transcription factor myb1r1-like
14	c366142_g1	Hypothetical protein SORBIDRAFT_05g005960	12	c366142_g1	Hypothetical protein SORBIDRAFT_05g005960
14	c364323_g1	Microtubule-associated protein	12	c365771_g2	D-lactate dehydrogenase

Table S6.14 Top 10 Trinity genes in the *C. macra* transcriptome contained the largest number of isoforms and AS isoforms (BLAT). NA, contig with no Blastx annotation.

Kissplice was also applied to identify AS events and compare with the AS events reported by Trinity. Defined in the Kissplice, each Kissplice alternative splicing (AS) events contained two Kissplice AS contigs. There were 141,986 Kissplice alternative splicing (AS) events (contained 283,972 Kissplice AS contigs) reported by Kissplice (Figure S6.14). Among the Kissplice AS contigs, 246,231 Kissplice AS contigs (in 126,422 Kissplice AS events) mapped to a number of 45,883 *C. macra* contigs (15.41% of the *C. macra* transcriptome), which belonged to 29,552 Trinity genes or gene fragments. There were half of the Kissplice AS contigs in 6,613 Kissplice AS events mapped to 5,456 *C. macra* contigs (in 4,938 Trinity genes), with the other half returned no hit to the *C. macra* transcriptome. 37,741 of the 283,972 Kissplice AS contigs gave no Blastx hit to the *C. macra* transcriptome, 31,128 of which formed 15,564 AS event pairs, indicating these events were not detected by Trinity. There were 119,809 pairs of Kissplice AS contigs showing Blastx hits to 40,427 filtered Trinity contigs from 27,380 Trinity genes, indicating 119,809 AS events in the 27,380 Trinity genes. There were 1,027,892 SNPs, 922 inexact tandem repeats, 61,407 short indels (<3 nt) reported by Kissplice.

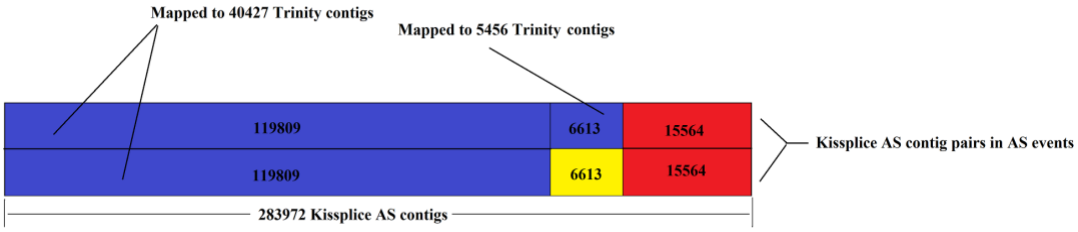


Figure S6.14 Kissplice AS contig annotations to the *C. macra* transcriptome. Blue color, Kissplice contigs mapped to the *C. macra* transcriptome. Yellow and red color, Kissplice contigs unmapped to the *C. macra* transcriptome.

A total of 82,594 Trinity contigs from 33,956 (16.49%) Trinity genes contained AS events as reported by Trinity, BLAT and Kissplice, with 14,112 Trinity genes that underwent AS events reported by both Trinity-BLAT and Kissplice. An additional 15,564 AS events contained 31,128 Kissplice AS contigs which did not show blast hit to the Trinity *C. macra* transcriptome were also reported. The top four interproscan domains in these *C. macra* AS contigs were transmembrane (26,650), non-

cytoplasmic domain (25,215), TMhelix (23,196) and cytoplasmic domain (20,819). Searching against Pfam-A database with e-values no less than e^{-3} suggested that leucine rich repeats (PF12799), pentatricopeptide repeat (PF01535), pentatricopeptide repeat_3 (PF13812), leucine rich repeat (PF00560) and PPR repeat family (PF13041) were most frequent protein domains in the *C. macra* AS contigs (Figure S6.15).

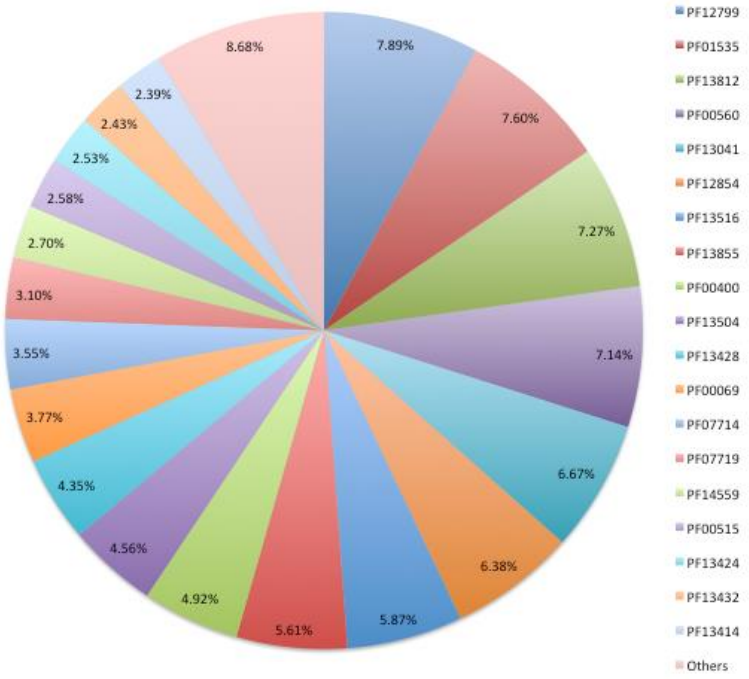


Figure S6.15 Frequencies of Pfam domain in the *C. macra* AS contigs. Top 20 Pfam domains with the largest numbers of contigs were present. 8.68% of the total contigs are included in the other Pfam domians.

There were 977 of the 1,130 enzymes in 140 KEGG pathways containing alternative splicing isoforms. The top five KEGG pathways containing the largest number of AS contigs were purine metabolism, thiamine metabolism, biosynthesis of antibiotics, aminobenzoate degradation and starch and sucrose metabolism (Figure S6.16). There were 17 enzymes in terpenoid backbone biosynthesis, nine enzymes in steroid biosynthesis and three enzymes in sesquiterpenoid and triterpenoid biosynthesis underwent alternative splicing events.

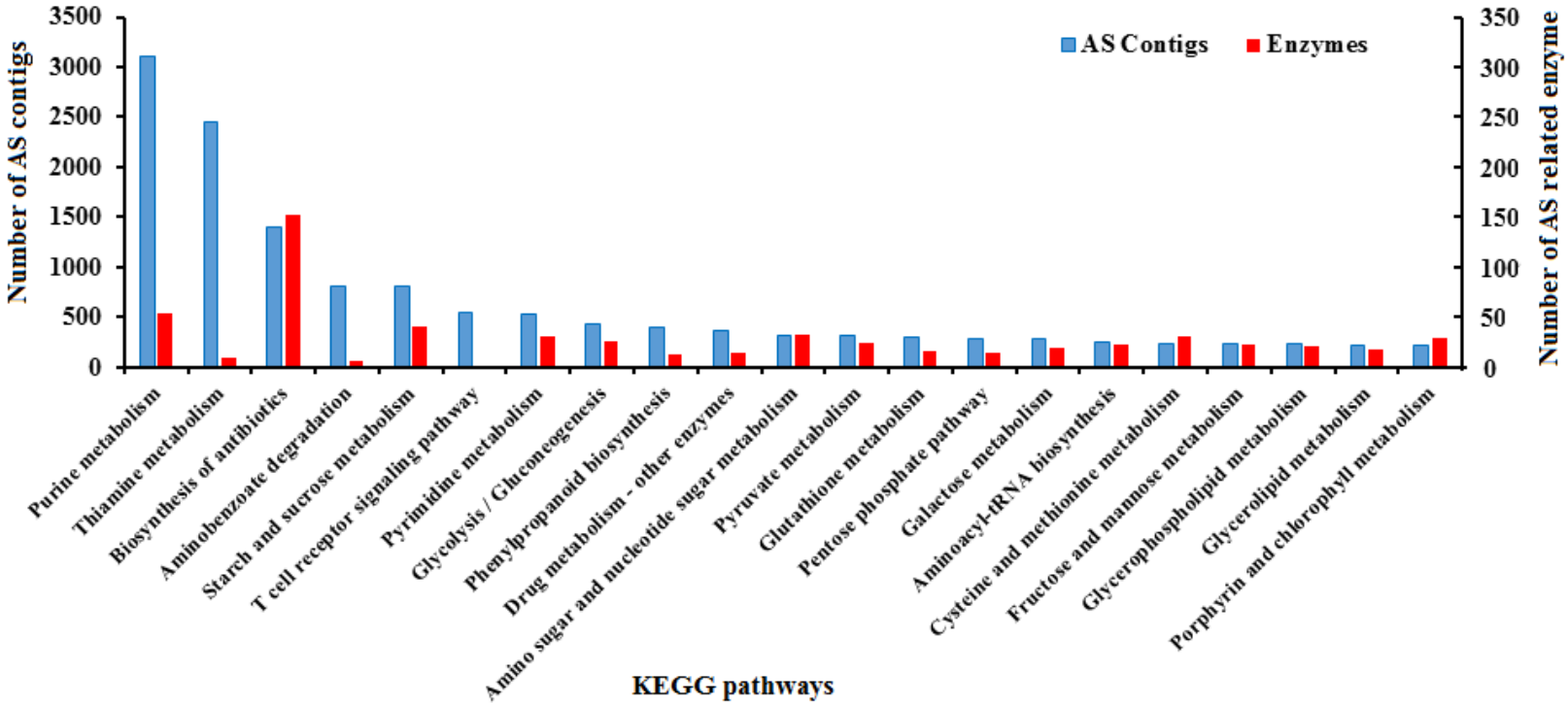
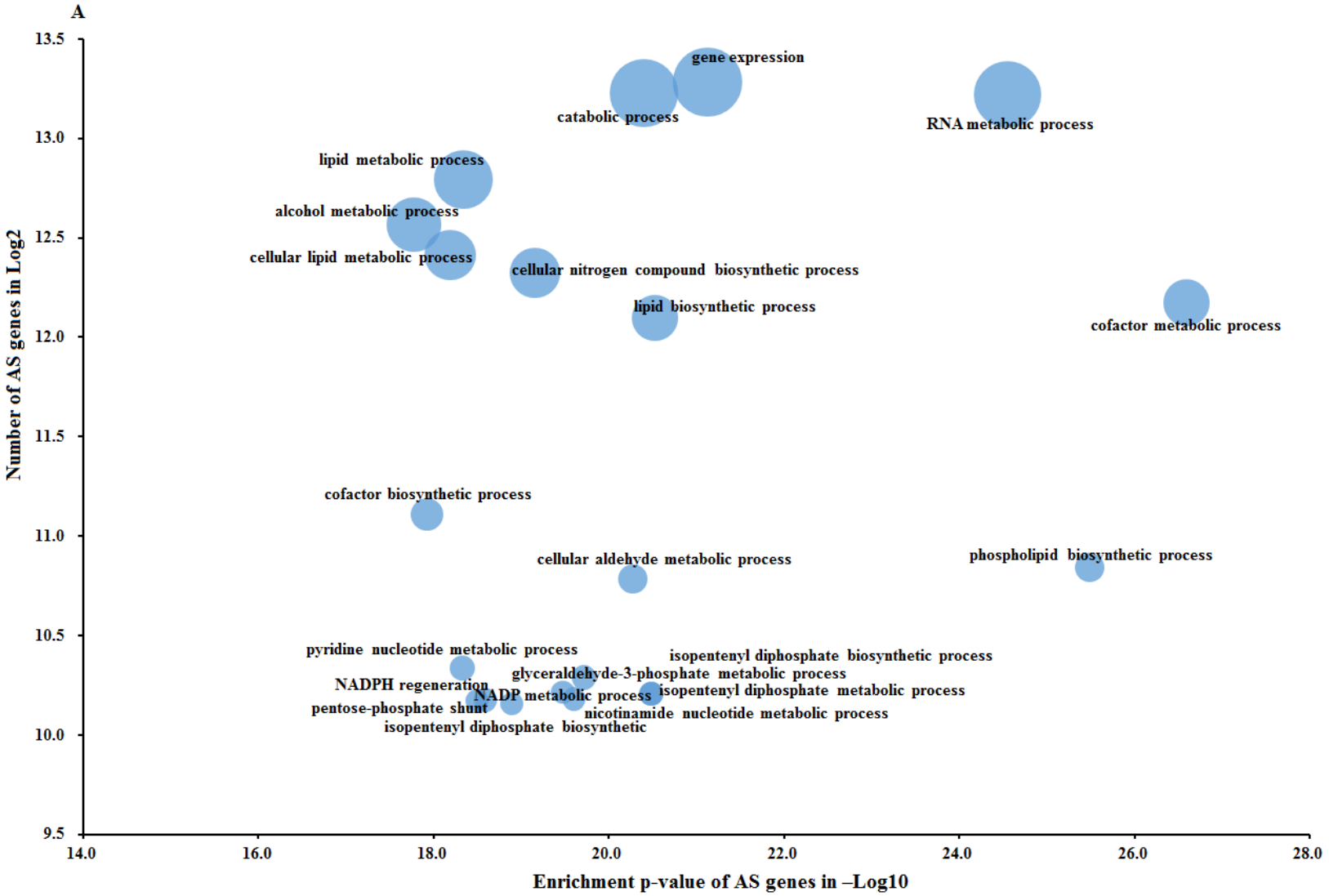
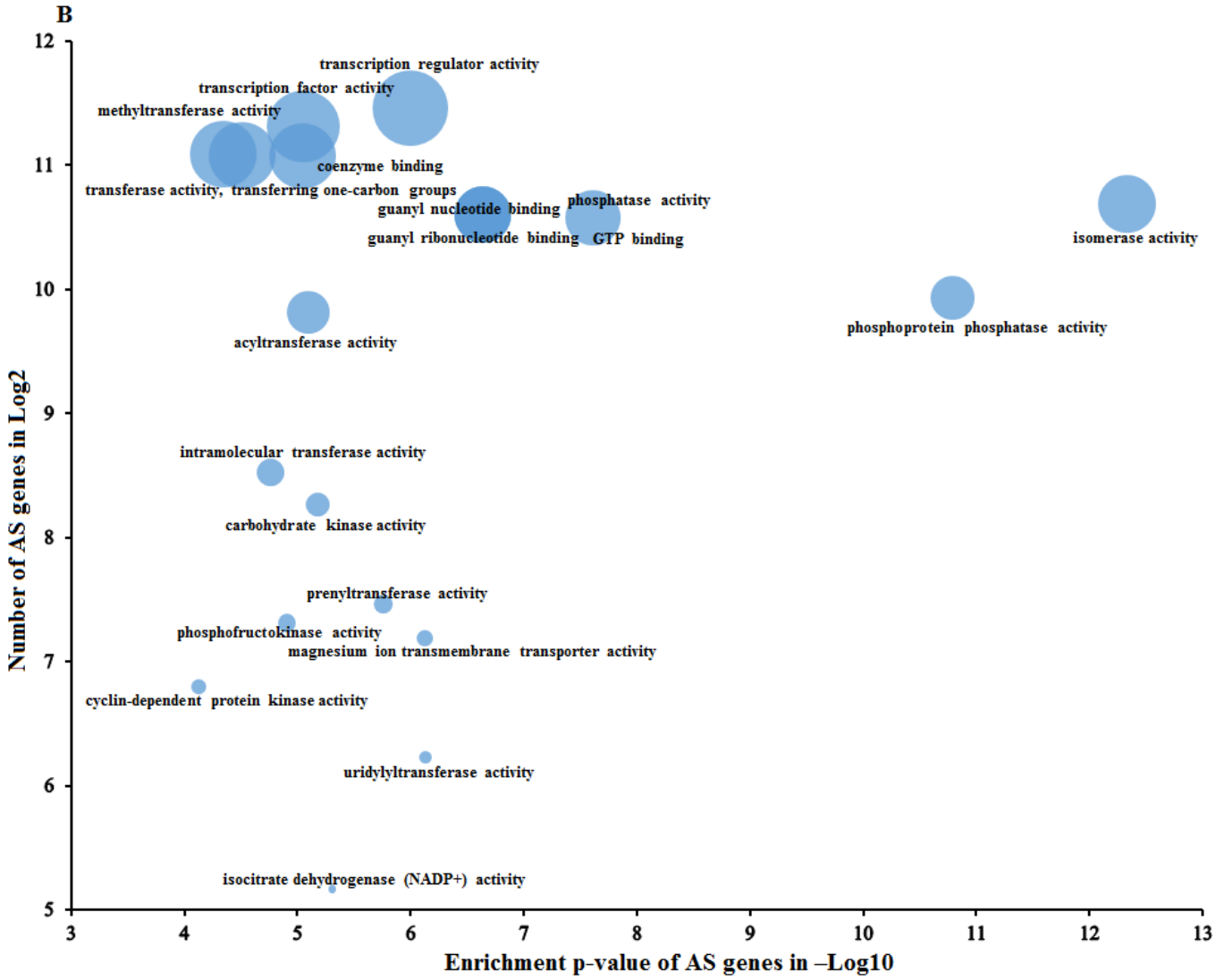


Figure S6.16 KEGG pathways in the *C. macra* transcriptome underwent AS events. Purine metabolism and biosynthesis of antibiotics contain the largest numbers of AS contigs and AS enzyme respectively.

GO terms of the *C. macra* AS contigs were assigned by Blast2GO. A number of 247,319 GO terms were extracted from the *C. macra* AS contigs and 4,869 unique GO terms were identified. The BINGO plugin in the Cytoscape was used to enrich GO terms from the AS genes in the *C. macra* transcriptome (Figure S6.17). There were 888 GO terms enriched into the Biological process catalog (GO: 0008150), 222 GO terms in the Molecular function catalog (GO: 0003074) and 129 GO terms in the Cellular component catalog (GO: 0005575). GO terms including cofactor metabolic process, phospholipid biosynthetic process, RNA metabolic process, gene expression and isopentenyl diphosphate biosynthetic process were enriched in the Biological process from AS genes with p-value threshold e^{-5}. In the Molecular function, GO terms including isomerase activity, phosphoprotein phosphatase activity, phosphatase activity, guanyl ribonucleotide binding and guanyl nucleotide binding were enriched (p-value < e^{-5}). In the Cell component, cytoplasmic part, chloroplast, chloroplast part, plastid part and organelle part were enriched (p-value < e^{-5}).


A
13.5
13.0
12.5
12.0
11.5
11.0
10.5
10.0
9.5
Number of AS genes in Log2
14.0
16.0
18.0
20.0
22.0
24.0
26.0
28.0
Enrichment p-value of AS genes in –Log10
gene expression
catabolic process
RNA metabolic process
lipid metabolic process
alcohol metabolic process
cellular lipid metabolic process
cellular nitrogen compound biosynthetic process
lipid biosynthetic process
cofactor metabolic process
cofactor biosynthetic process
cellular aldehyde metabolic process
phospholipid biosynthetic process
pyridine nucleotide metabolic process
isopentenyl diphosphate biosynthetic process
glyceraldehyde-3-phosphate metabolic process
NADPH regeneration
isopentenyl diphosphate metabolic process
NADP metabolic process
pentose-phosphate shunt
nicotinamide nucleotide metabolic process
isopentenyl diphosphate biosynthetic


B
transcription regulator activity
transcription factor activity
methyltransferase activity
coenzyme binding
transferase activity, transferring one-carbon groups
guanyl nucleotide binding
phosphatase activity
guanyl ribonucleotide binding GTP binding
isomerase activity
acyltransferase activity
phosphoprotein phosphatase activity
intramolecular transferase activity
carbohydrate kinase activity
prenyltransferase activity
phosphofructokinase activity
magnesium ion transmembrane transporter activity
cyclin-dependent protein kinase activity
uridylyltransferase activity
isocitrate dehydrogenase (NADP+) activity
Number of AS genes in Log2
Enrichment p-value of AS genes in –Log10
12
11
10
9
8
7
6
5
3
4
5
6
7
8
9
10
11
12
13

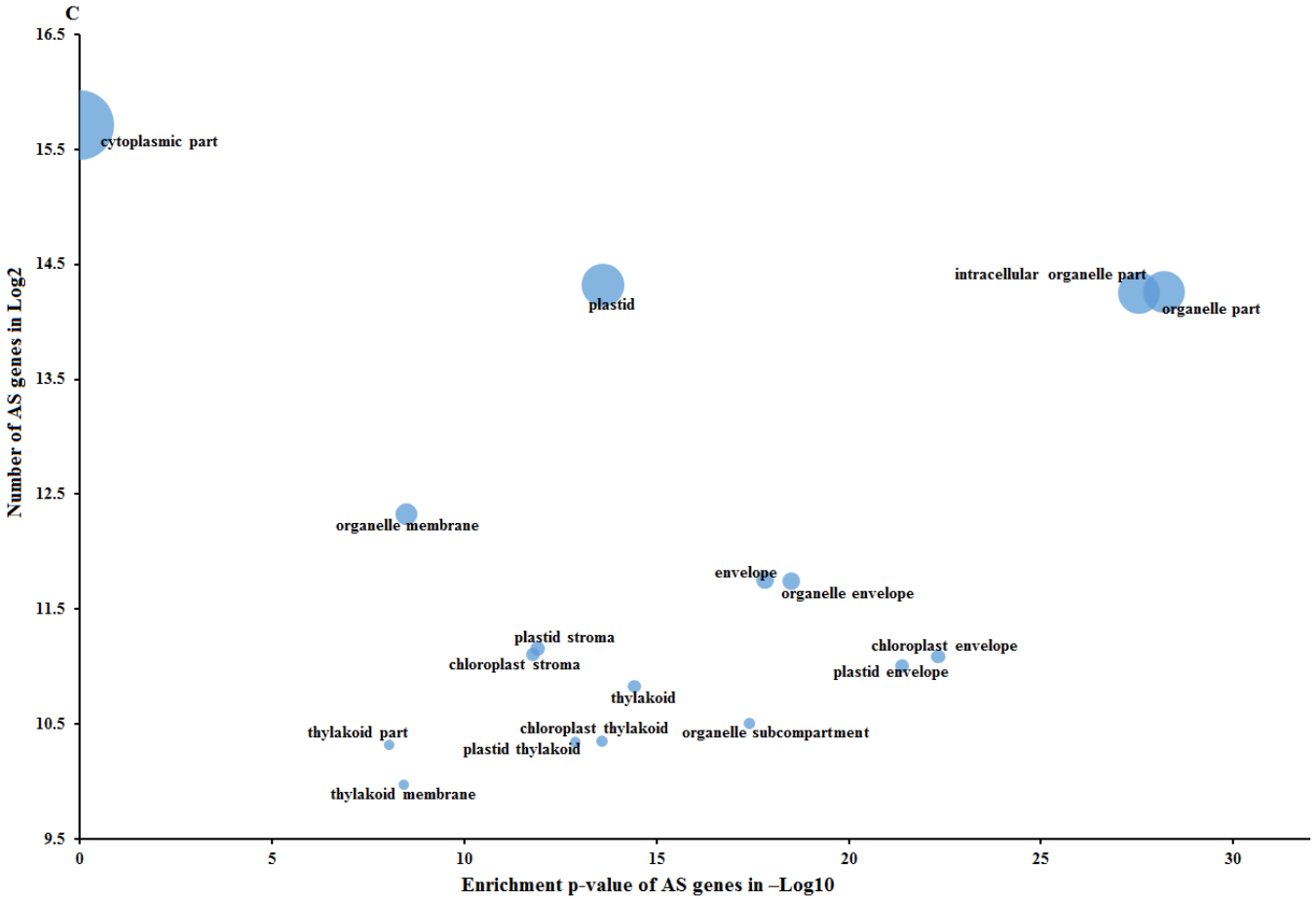


Figure S6.17 Top 20 of most significant enriched GO terms of AS genes in the three GO catalogs, Biological process (A), Molecular function (B) and Cell component (C). The size of each enriched GO term (color in blue) presented its percentage of the total number of AS genes associated GO terms.