Tracing Molecular Patterns of Adaptation in Arctic Brassicaceae

- Evolutionary repeatability and adaptations to extreme abiotic stress

Siri Birkeland

Dissertation for the degree of Philosophiae Doctor

Natural History Museum
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

2020

Main supervisor: Michael D. Nowak

Co-supervisors: A. Lovisa S. Gustafsson, Anne K. Brysting, Tanja Slotte

Summary

Introduction. Extreme environments are natural laboratories for studying how different organisms adapt to similar selection pressures. This thesis explores how three Arctic plant species independently adapted to some of the coldest biomes on Earth, and how they developed similar adaptations to extremes in light and temperature. It addresses fundamental questions in plant evolutionary biology, such as: "To what extent does adaptation follow the same genetic trajectories in different species?", and "What is the genetic basis for plant adaptation to extreme environments?". The thesis has two main objectives that are addressed through three papers (**Papers I-III**): 1) estimate the degree of adaptive molecular convergence in the three Arctic Brassicaceae *Cardamine bellidifolia*, *Cochlearia groenlandica*, and *Draba nivalis*, and 2) identify molecular adaptations to the Arctic environment in the same three species.

Approach. The first two papers address the degree of evolutionary repeatability in how *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* adapted to the Arctic environment at the genetic level (objective 1). In **Paper I**, we estimated molecular convergence at the level of codons, genes, and functional pathways, by comparing genome-wide patterns of positive selection and identifying convergent substitutions in the three species. In **Paper II**, we conducted a time series experiment to examine the transcriptional responses of the Arctic Brassicaceae to low temperatures, and to identify potential convergent expression patterns in cold response.

In all three papers we identified putative molecular adaptations to extremes in light and temperature (objective 2). In **Paper I**, we identified candidate genes for adaptation to the Arctic environment by searching for positively selected genes associated with abiotic stresses common in the Arctic. In **Paper II**, we explored the molecular basis of cold tolerance in Arctic Brassicaceae, and described how their cold-induced transcriptomes differ from that of the temperate model species, *Arabidopsis thaliana*. In **Paper III**, we assembled the genome of *D. nivalis* and explored the genomic characteristics of Arctic plant adaptation, by conducting comparative analyses of chromosomal evolution and functional genomics with other species in the Brassicaceae.

Main findings and discussion. The findings in **Papers I-II** suggest that the three Arctic Brassicaceae have adapted to the Arctic environment through independent genetic trajectories (objective 1). In **Paper I**, we found that positive selection has been acting on different genes, but similar functional pathways in the three species. The positively selected gene sets showed convergent functional profiles associated with abiotic stresses common in the Arctic. However, we found little evidence for convergent substitutions at the same sites, or for positive selection acting on the same genes in the three species. In **Paper II**, we found that the cold-response of *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* is highly species-specific. Most differentially expressed genes (DEGs) were unique for each species, and the Arctic Brassicaceae shared more DEGs with *A. thaliana* than with each other. This suggests that the cold response in Arctic Brassicaceae mainly evolved independently, but with some components likely conserved across the family. The low levels of molecular convergence could be explained by the many evolutionary trajectories leading to better performance under temperature and light stress in plants, and/or less repeatable patterns of adaptation in highly polygenic traits such as cold tolerance.

Papers I-III present some of the first molecular evidence for plant adaptations to the Arctic environment (objective 2). In **Paper I**, we found multiple candidate genes for Arctic adaptation associated with cold stress, freezing stress, oxidative stress and light stress in all species. Adaptations associated with the plasma membrane seemed to be particularly important in all species, possibly due to the membrane's crucial role in freezing tolerance. In **Paper II**, we found that the Arctic species' cold response followed trends similar *A. thaliana*, but a few DEGs and characteristics were specific for the Arctic species alone. In **Paper III**, we presented a 302 Mb genome assembly of *D. nivalis* that is highly contiguous with 91.6 % assembled into eight "chromosomes" (the base chromosome number of the species). We found that the *D. nivalis* genome contains expanded suites of genes associated with common Arctic stresses, and the expansion of these gene families appear partly driven by the activity of transposable elements.

To summarize, one of the main findings in this dissertation is that plant adaptation to the extreme abiotic stresses of the Arctic can happen through multiple genetic trajectories. The results provide both a framework and tools (in form of the *D. nivalis* genome assembly) for more in depth studies on Arctic adaptation in Brassicaceae and other flowering plants.

Acknowledgements

I have really loved my time as a PhD student at the **Natural History Museum**, and I am a bit sad that it has come to an end. The museum and botanical garden is a great working place, and I especially enjoyed all the small museum perks like being able to search the Zoological collections for traces of Nansen's Fram expedition, browse the botanical garden for plants to the BIO4250 course, and helping to raise a huge model of an ichthyosaur up to the ceiling at the Zoological museum.

During this period, I have met many inspiring people who have made my time as a doctoral student extra enjoyable. I am grateful that many people have supported and guided me during the course of this PhD.

To my supervisors and project leader:

Mike, thank you for being passionate about evolutionary biology, for lending your razor-sharp mind to scientific discussions, and for giving advice on career and the science world. You introduced me to many great things, like big conferences in evolution, bioinformatics, Texan bbq and karaoke (what a mix!). **Lovisa**, thank you for always having my back, for making the most amazing wine and cheese celebration when I submitted my first paper, and for being the best conference buddy ever. We shared some awesome (and not so awesome – Jamaica in NYC?) moments all around the world :) **Anne**, I sometimes feel like you are the rock that holds everything together. Thank you for always being there and for giving advice on all levels. Your endless compassion for other people is really admirable. You rock! **Tanja**, I really enjoyed the time I spent in your lab. Thank you for setting up meetings with people working on similar topics, for guiding me, and for showing me all the great lunch places on the Stockholm University campus. It meant a lot. **Christian**, I hope I can one day write as elegantly as you. Thank you so much for taking the time to teach me some tricks, and for giving me the opportunity to do this really cool PhD in Arctic plant evolution.

To the SpArc & Speciation Clock teams:

Abel, (with the kindest soul), thank you for always being so supportive, for being my partner in crime in surprise birthday celebrations, and for being great company at the Triple AAA course in Ascona. I can't wait to see where you will be in a few years. **Desalegn**, thank you for spreading your good moods and always being so chill. I hope I get to see some of your Elon Musk-like research ideas in real life :) **Filip**, thank you for always being so enthusiastic and full of initiative. It has been really great to have you in the group. **Lise**, I am so glad you joined our team, and that we get to hang out for a few more months. I also want to thank the rest of the SpArc and Speciation Clock teams for all help and support during our meetings.

To the PET-group:
Hugo, thank you for being such a uniting group leader and always keeping an eye out for everyone in the group. It is very much appreciated. **Brecht**, it was really fun sharing an office with you! Thank you for taking the time to give good advice on small and big problems, and for always keeping cookies in your drawer. I also want to thank **Anneleen**, **Irene**, **Marcella**, **Marcele**, **Eliot** and the other **PET-members** for creating such a positive and supportive environment.

To the other PhD-students:
José, thank you for making this PhD period so fun, for trips to Schous, for our many discussions about science and the science world, and for being there for me at the end when I was starting to get really tired (chocolate croissants ftw!). **Nanna <3**, I am so glad we started at the same time! We shared so many good times, and I hope I soon can visit you in Denmark :) **Silje**, **Trude**, and **Hanna**: Thank you for great company at "shut up at write", NHM parties, café trips, and conferences (the cathedral party in Montpellier, Hanna?). I also want to thank the members of the student activity committee, **Sonja, Franzi**, **Heidrun**, and **Lene** for fun collaborations, **Tiril** for taking the time to stop by my office to chat about everything (I am glad we can continue our chats online), **Vincent** for bringing vanilla from Madagascar and making the most amazing chicken and apricot jam I ever tasted (like a true French:), **Anca** for being such a lovely office mate (I will always think about you when I hear "despacito"), and **Peter** for all the chats about how amazingly slow our PhDs were going (it was actually a great comfort!). To **Margret**, **María**, **Feli**, **Solveig**, **Eva**, **Adrian** and **Lasse**: Thank you for bringing new life into the Botanical museum and making everything fun and new again. **Marie Brandrud**: it has been really cool to keep in touch while you did your PhD in Vienna. **Angelica Cuevas**: our Blindern lunches and discussions (read: complaining) on PhD life always cheered me up. - And **Peter Hoitinga**: Thank you for coming up with a cool idea just when I needed a distraction.

To my PhD committee and Elisabeth A.:
Torsten, **Lutz**, **Brita**, and **Einar**: Thank you for supporting me during all my semester reports, and for coming with valuable suggestions to how I could improve my project. **Elisabeth**, thank you for great help with all the practical things during the PhD.

To my friends and family in Oslo <3
I have been very fortunate to be close to my family and friends from Oslo during this PhD. I want to thank **Maria**, **Helene**, **Ingrid**, **Christine, Tim**, **Luca**, **Monica**, **Idunn**, **Schmidt**, **Rødbeth, Linda,** and **Yann** for being such great friends, for good times, and for helping me think of other things than my PhD ;) I hope to be a bit more present now that I delivered my thesis! - And **Idunn**: our many great field trips in the Arctic have always been in the back of my head during this project! **Bente and Dominique**: Thank you for cosy apéritifs and dinners and for always being so

welcoming. **Marianne**, **PC**, **Nikolai**, **Emilie and Ventus**, thank you for always being there and for lightening up the life of a tired PhD student. I hope to see you a bit more from now on ;) **Mamma og Pappa**, thank you for always supporting and encouraging me, for making big family dinners, and for taking the time to discuss all of my strange PhD problems like they were your own. It meant a lot. And last but not least, the chillest and best guy I know, **Nils!** Thank you for your unconditional support, for making me laugh even when things are hard, and for always reminding me what is truly important in life. I'm looking forward to many more chill mornings with Alpha Boys School radio now that I handed in my thesis <3

- Siri

From the serpentine *Draba nivalis* expedition to Sweden.

a. "*(...) if variations useful to any organic being do occur, assuredly individuals thus characterised will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance they will tend to produce offspring similarly characterised. This principle of preservation, I have called, for the sake of brevity, Natural Selection.*"

b. "*(...) in nearly the same way as two men have sometimes hit on the very same invention, so natural selection (...) has sometimes modified in very nearly the same manner two parts in two organic beings, which we owe but little of their structure in common to inheritance from the same ancestor*"

Charles Darwin on natural selection (Chapter IV), and evolutionary repeatability (VI), On the origin of species, published 1859

Contents

Abbreviations

ABA	Abscisic acid
BP	biological process domain
BUSCO	Benchmarking Universal Single-Copy Orthologs
CAT1	catalase 1
CAT2	catalase 2
CBF pathway	C-repeat binding factor pathway
CBF4	C-repeat binding factor 4
CC	cellular component domain
CCP	comparative chromosome painting
COR15B	cold regulated 15B
COR27	cold regulated 27
COR28	cold regulated 28
CRY2	cryptochrome 2
CSDP1	cold shock domain protein 1
ddRAD	double-digest restriction-associated DNA
DEG	differentially expressed gene
DNA	deoxyribonucleic acid
DREB	dehydration responsive element binding
DSD	dispersed duplication
EMB2742	embryo defective 2742
ER	endoplasmic reticulum
GO	gene ontology
HY5	hypocotyl 5
IAA16	indoleacetic acid-induced protein 16
LEA4-5	late embryogenesis abundant 4-5
LINEs	long interspersed nuclear elements
LTR-RTs	long terminal repeat retrotransposons
MAPK	mitogen activated protein kinase
MAPKKK14	mitogen activated protein kinase kinase kinase 14
MF	molecular function domain
OGOX1	oligogalacturonide oxidase 1
PAC	pale cress
PCR2	plant cadmium resistance 2
PD	proximal duplication
PLC1	phosphoinositide-specific phospholipase C 1
PSG	positively selected gene
RAP2.2	related to AP2 2
RAP2.10	related to AP2 10
RAV1	related to ABI3/VP1 1

RNA	ribonucleic acid
ROS	reactive oxygen species
SFR2	sensitive to freezing 2
SINEs	short interspersed nuclear elements
TD	tandem duplication
TIRs	terminal inverted repeat transposons
TRD	transposed duplication
WGD	whole-genome duplication

List of papers

This thesis includes three papers, whereof one is published (Paper I), one is in manuscript form (Paper II), and one is accepted (Paper III).

I. **Birkeland S**, Gustafsson ALS, Brysting AK, Brochmann C, & Nowak MD (2020) *Multiple genetic trajectories to extreme abiotic stress adaptation in Arctic Brassicaceae*. Molecular Biology and Evolution 37(7): 2052–2068. doi:10.1093/molbev/msaa068.

II. **Birkeland S**, Slotte T, Brochmann C, Brysting AK, Gustafsson ALS, & Nowak MD. *What can the cold-induced transcriptomes of Arctic Brassicaceae tell us about the evolution of cold tolerance?* (Manuscript)

III. Nowak MD, **Birkeland S**, Mandáková T, Choudhury RR, Guo X, Gustafsson ALS, Gizaw A, Schrøder-Nielsen A, Fracassetti M, Brysting AK, Rieseberg L, Slotte T, Parisod C, Lysak MA, Brochmann C. *The genome of Draba nivalis shows signatures of adaptation to the extreme environmental stresses of the Arctic*. Molecular Ecology Resources; *accepted*.

1. Introduction

A central objective of evolutionary biology is to understand how organisms evolve to become better suited to live and reproduce in an environment (i.e. the process of adaptation). In the book often viewed as the foundation of the field, "On the origin of species" (1859), Charles Darwin was the first to thoroughly describe how natural selection can lead to adaptation and even speciation (citation a, page 9). Many decades later, the synthesis between Darwin's theory and Mendelian genetics enabled us to ask how adaptation happens at the genetic level. Many fundamental questions about the genetic basis of adaptation are still unanswered. Such questions include whether adaptation usually involves new mutations or standing genetic variation (Barrett and Schluter 2008), if adaptation occurs through a few genes of large effect or many genes of small effect (Orr 2005), and whether adaptation follows predictable genetic trajectories (Hendry 2013). In this thesis I aim to trace molecular patterns of adaptation in Arctic plant species, in order to explore some of these questions, and to learn more about the genetic basis of plant adaptation to extreme environments.

Extreme environments provide powerful opportunities to study the genetic basis of adaptation for several reasons. First, extremophilic organisms often develop conspicuous traits that are easy to target and compare among species. In plants, such examples include heavy metal tolerance enabling plants to live in metalliferous soils (e.g. Preite et al. 2019), photosynthetic pathways that minimize photorespiration in hot and dry environments (e.g. Heyduk et al. 2019), and tolerance to high salinity and hypoxia in intertidal zones (e.g. Xu et al. 2017). Second, strong selection pressures imposed by extreme environments tend to leave particularly clear signatures on the genome, which makes it easier for us to deduce the genomic landscape of adaptation. These strong selection pressures can even lead to the independent evolution of similar morphological and/or physiological traits in different lineages (i.e. convergent evolution; Sackton and Clark 2019). In this way, extreme environments can function as natural laboratories for studying how different organisms adapt to the same selection pressures, and even bear a slight resemblance to "replaying life's tape", as once famously coined by paleontologist Stephen Jay Gould (1989).

Stephen Jay Gould imagined that if we were to rewind and replay "the tape of life", we would end up with a completely different outcome due to the contingency and unrepeatability of the evolutionary process (1989). However, most evolutionary biologists recognize that evolution has a tendency to repeat itself, and even Darwin

touched upon the subject in “On the origin of species” (citation b, page 9). In plants, carnivory has evolved independently at least six times in five different angiosperm orders (Ellison and Gotelli 2009), and photosynthetic pathways that minimize photorespiration, like C_4 and CAM photosynthesis, are estimated to have evolved an astonishingly sixty times each (Heyduk et al. 2019). Convergent evolution thus seems to be an ubiquitous pattern in nature. However, convergent traits do not necessarily evolve as a result of similar genetic changes, and the prevalence of molecular convergence remains a heavily debated topic in evolutionary biology today (Losos 2011; Conte et al. 2012; Hendry 2013; Bolnick et al. 2018; Sackton and Clark 2019).

In this thesis, I ask if adaptation has followed the same genetic trajectories when three plant species independently evolved to cope with the extreme temperature and light conditions of the Arctic region. For this, I leveraged recent advances in using comparative genomics to test for convergent molecular evolution across genomes. A substantial part of this thesis also focuses on tracing the exact adaptations enabling plants to live in the Arctic, and throughout three different studies (together with colleagues), I have searched for putative molecular adaptations to the Arctic environment. In this work, I use the word “adaptation” interchangeably about three different evolutionary aspects (Dobzhansky 1970; Tigerstedt 1994; Oxford Dictionary of Biology 2004; Futuyma 2005): 1) the process whereby an organism evolves to become better suited to its environment, 2) a characteristic that enhances the survival and reproduction of the organisms that bears it (an adaptation), and 3) the state of being adapted - the degree to which an organism is suited to live and reproduce in an environment (adaptiveness).

1.1 Where do Arctic plants come from?

To understand how plants have adapted to Arctic environments, it may be useful to know something about their origin. The Arctic is often defined as the areas north of the Arctic Circle (i.e. the northernmost boundary of the midnight sun and polar night), or north of the mean July 10 °C isotherm (roughly equivalent to the northernmost treeline). Today, the region includes some of the coldest biomes in the world, and holds less than 1% of the world's vascular plant diversity (~2218 species; CAFF 2013). It can therefore be hard to imagine that the Arctic once was covered by extensive forests (Hoffmann et al. 2010). According to Moran et al. (2006), the Arctic likely transitioned from a warm “greenhouse world” to the colder “icehouse world”

that we now know, during the late Paleocene and early Eocene epochs (the Paleocene-Eocene transition ~56 Mya).

The characteristic tundra ecosystems of today's Arctic likely emerged only ~3.2 Mya, at the same time of the ice sheet expansion in Greenland (Matthews 1979; Matthews and Ovenden 1990; Hoffmann et al. 2010). It is believed that parts of today's Arctic flora likely are derived from alpine plants in Asia and North America that migrated northwards during the climate cooling (Savile 1972). However, fossils also indicate that many genera or species currently found in the Arctic were already present long before the emergence of Arctic tundras (Hoffmann et al. 2010), and parts of the flora may have evolved from shrubby or herbaceous elements of the prehistoric forests (Abbott and Brochmann 2003).

The Quaternary period (~2.6 Mya to present) was characterized by repeated glaciations and deglaciations that had profound consequences for distributions of plants and animals in the Arctic (Birks 2008). During the periods of glaciation, parts of the Arctic flora likely migrated to unglaciated areas in central Europe and Eurasia, Siberia, Beringia, and North America, and then re-colonized the Arctic during the interglacials (Abbott and Brochmann 2003; Birks 2008). Long-distance plant colonization is indeed frequent in the Arctic, and Arctic plants are known to have great dispersal abilities (Alsos et al. 2007). A few Arctic plant species may also have survived the glaciations *in situ* (Schneeweiss and Schönswetter 2011; i.e. in ice-free refugias within the Arctic; Westergaard et al. 2011). There are also Arctic plant species of very recent origin, like for instance allopolyploid species that originated within the period of the last major ice ages (Brochmann and Brysting 2008).

The history of today's Arctic flora is a reminder that adaptation to the Arctic did not start from a clean slate. According to Savile (1972), there are three likely scenarios to Arctic plant adaptation: 1) Arctic plants may have been pre-adapted to cold temperatures and short growing seasons through evolution in Alpine environments, 2) Arctic plants may have already possessed broad plasticity that enabled them to cope with extreme temperatures and light conditions, or 3) Arctic plants may have developed new adaptations to cope with the new environment.

1.2 Extreme abiotic stresses imposed by the Arctic environment

Contrary to what many believe, the Arctic summer is probably more challenging for plant life than the Arctic winter. This is because plants must maintain all normal life functions at unusually low temperatures during summer, while they are dormant and covered with protective layers of snow and withered leaves during the winter period (retaining wilted leaves is a strategy for coping with low temperatures; Körner 2003; Körner and Alsos 2009). Arctic plants are able to cope with this extremely short and cold growing season, which in summer can hold up to 24 hours of daylight. Furthermore, they must endure the possibility of freezing temperatures in the warmest summer months. This is a crucial difference from temperate areas where freezing stress is limited to the beginning and end of the growing season. In addition, there are a number of other abiotic stresses that accompany the extremes in light and temperature found in the Arctic. Here, I classify major Arctic stresses into five categories that I will focus on throughout the thesis:

Physiological cold stress: Low temperatures can affect almost all aspects of plant physiology, and temperature is one of the most important factors determining plant distributions world-wide (Billings and Mooney 1968; Wiens and Donoghue 2004). In order to discuss the effects of low temperatures on plants, it is practical to differentiate between chilling stress (low temperatures above zero) and freezing stress (temperatures below zero; discussed in detail below). Chilling stress impairs photosynthesis, changes the physical properties of proteins, causes cell membranes to rigidify, and retards plant growth (Heino and Palva 2004; Shi et al. 2018). It may also thermodynamically reduce the kinetics of metabolic reactions (Ruelland et al. 2009), and lead to reduced fertility through meiotic failure (Bomblies et al. 2015).

Ice and snow stress: In the Arctic, ice and snow formation can occur throughout the year, and lead to mechanical stress on exposed plant tissue. Due to permafrost, all soil 20-100 cm below the surface is permanently frozen, which limits root growth to the top layers (Billings and Mooney 1968). Ice formation may lead to cell death if crystallization initiates inside the cells, but some plants can tolerate freezing of the apoplast (the space between the cells; Körner 2003). Freezing of the apoplast will draw liquid water out of the cells, leading to osmotic stress and cellular dehydration,

as well as increased concentrations of salts and toxic solutes inside the cells (Steponkus 1984; Wisniewski and Fuller 1999; Körner 2003; Wisniewski et al. 2004).

Protein misfolding and membrane stress: This category is related to both chilling stress and freezing stress, but is separated as its own category because of its putative importance in Arctic adaptation. Cell membrane systems are the primary site of freezing injury in plants (Steponkus 1984), which is mainly caused by membrane lesions during freezing induced dehydration (Thomashow 1999). The rigidification of membrane systems during chilling will also disturb all membrane processes, such as the opening of ion channels (Ruelland et al. 2009). In addition, low temperature will alter normal biogenesis and stability of proteins (Guy et al. 1998), and even lead to secondary structures in RNA with important implications for gene and protein expression (Ruelland et al. 2009).

Light stress: Due to the Earth's axillary tilt, the Arctic light environment is characterized by having large fluctuations in the amount of sunlight between summer and winter (Salisbury 1985). The number of days with midnight sun varies from six months at the North Pole, to just one day at the Arctic Circle. The Arctic's placement on the top of our spherical Earth also leads to the sun always being low on the horizon. This influences spectral composition, and the Arctic light environment is enriched in e.g. blue and far-red light (Nilsen 1985). These characteristics are especially important for plants, which often use photoperiod and spectral composition as cues for development, reproduction, and the onset of dormancy (Chory et al. 1996; Jackson 2009; Piskurewicz et al. 2009).

Oxidative stress: The accumulation of reactive oxygen species (ROS), is a common stress-response in plants and accompanies many of the above-mentioned abiotic stresses (Kilian et al. 2007). Arctic summers are in addition characterized by the combination of low temperatures and up to 24 hours of daylight, and the combination of high light intensity and low temperature is thought to lead to considerable oxidative stress (Heino and Palva 2004; Lütz 2010). We can thus assume that oxidative stress is ubiquitous in Arctic environments.

In addition to these categories of abiotic stresses, there is a range of other challenges important for Arctic plant life that will not be addressed here. These include, for instance, low nitrogen supply in Arctic soils, strong winds, the very short growing season, and scarcity of pollinators, some of which are discussed in detail in other works (e.g. Billings and Mooney 1968; Savile 1972; Wullschleger et al. 2015).

1.3 What do we know about morphological and physiological adaptations in Arctic plants?

Arctic plants possess suites of similar morphological and physiological adaptations to extremes in light and temperature (Billings and Mooney 1968; Chapin 1983; Lütz 2010). Morphologically, most Arctic plants are herbaceous with a caespitose or low growth form that take advantage of favorable climatic conditions close to the ground (Bliss 1962). Their leaves are often tough and leathery, sometimes even hairy (Hultén 1968), providing protection from desiccation, abrasive winds and mechanical stress from ice and snow. Furthermore, almost all Arctic plants are perennials with large underground storage systems (Billings and Mooney 1968), which is an advantage in an unpredictable environment where summers occasionally may be too unfavorable to even reproduce.

Physiologically, Arctic plants have optimal photosynthetic rates at lower temperatures than other plants (Chapin 1983), and are unusually resistant to cold and freezing stress. It has, for instance, been reported that some Arctic plants may tolerate being frozen stiff in the middle of flowering (Bliss 1962), and that they are extraordinarily resistant to ice encasement (Bjerke et al. 2018). Little is known about the exact mechanisms behind this cold/freezing tolerance, but from other cold tolerant plants, we know that common strategies include synthesis of cryoprotectants (i.e. compounds/solutes that in various ways hinder ice nucleation, or alleviate the effects of ice formation by protecting plant tissues against freezing damage), changes to the lipid composition of the membrane, and activation of active oxygen scavenging systems (Uemura and Steponkus 1999; Ruelland et al. 2009). In temperate environments, plants prepare for periods of cold through a process called cold acclimation, signifying an increase in freezing tolerance in response to low non-freezing temperatures (Thomashow 2010). This usually involves a complete reorganization of the transcriptome, and is partly regulated by the C-repeat binding factor (CBF) pathway (Kreps et al. 2002; Kilian et al. 2007; Thomashow 2010). We

still do not know if Arctic plants are hard-wired to always be prepared for freezing temperatures, or if they acclimate in similar ways as temperate plants.

When it comes to adaptations to the Arctic light environment, Arctic plants seem to be physiologically adapted to a 24 hour photoperiod (Mooney and Billings 1961). There are also reports of Arctic plants requiring a longer critical day length for induction of flowering (Wehrmeister and Bonde 1977; Heide 2005), while others have found that the main factor controlling flowering at high latitudes is temperature and not light (Bliss 1962). Nilsen (1985) suggested that changes in the far-red to red light ratio might be especially important for the onset of dormancy when there is still midnight sun in late summer. Several Arctic plants are also known to exhibit heliotropism, i.e. the turning or bending of plant parts directly to and with the sun (Kevan 1972). This could be a way to increase metabolic rates in flowers and inflorescences in an environment that has a very low heat budget (Kevan 1972).

Although several studies have explored morphological and physiological adaptations in Arctic plants (e.g. Mooney and Billings 1961; Chapin 1983; Lütz 2010; Jónsdóttir 2011), the molecular basis of these adaptations has rarely been investigated (Wullschleger et al. 2015). By tracing molecular adaptation in plant lineages independently adapted to the Arctic, we can learn more about the general traits and genes that are important for plant survival at high latitudes.

1.4 A system for studying Arctic plant adaptation: Arctic Brassicaceae

The focus of this thesis is three plant species from the mustard family (Brassicaceae), all of which have their main distribution above the Arctic Circle (Figure 1-2). The species, *Cardamine bellidifolia*, *Draba nivalis*, and *Cochlearia groenlandica*, are diploid and represent three main Brassicaceae clades (denoted clade A, B and C respectively; Huang et al. 2016; Guo et al. 2017). These clades likely diverged ~30 Mya, and the three species in question are thought to have colonized the Arctic independently of each other (see e.g. Carlsen et al. 2009; Jordon-Thaden et al. 2010; Koch 2012). With 133 species, the Brassicaceae is one of the most well-represented families in the Arctic (CAFF 2013), and Arctic Brassicaceae is therefore a representative study system for Arctic plant adaptation. Furthermore, many molecular resources exist for Brassicaceae (TAIR 2020), and the family has been thoroughly studied in terms of cold tolerance (Kilian et al. 2007; e.g. Thomashow

2010; Park et al. 2015). The three focal species have distributions that cover the entire Arctic (even the coldest bioclimatic subzone called Arctic polar desert; Figure 2; Elvebakk 1999; Elven et al. 2011), and they are therefore suitable candidates for studying adaptation to low sun angle and limited heat budget.

Figure 1. Pictures of the focal species. From left to right: *Cardamine bellidifolia* (Photo: I. Pospelov, iNaturalist, CC BY-NC 4.0), *Cochlearia groenlandica* (Photo: M. Goff, iNaturalist, CC BY-NC-SA 4.0), and *Draba nivalis* (Photo: Z. Harris, iNaturalist, CC BY-NC 4.0).

1.5 Main objectives and research questions of this thesis

This thesis has two main objectives (and two main research questions; see Figure 3). **The first objective** is to estimate the degree of adaptive molecular convergence in the three Arctic species, *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*. This objective is addressed in Papers I-II. In Paper I we focus on identifying signatures of convergent evolution across codons, genes, and pathways in the three Arctic species. By disentangling the impacts of these different levels of molecular convergence, we may gain a more comprehensive perspective on the genomic landscape of adaptation. As the focus of Paper I is mainly limited to convergent changes in protein coding regions, we also wanted to investigate convergent changes in expression patterns. In Paper II, we therefore aim to explore gene expression responses to cold in the same three species, and to identify potential convergent gene expression responses.

The second objective of this thesis is to identify putative molecular adaptations to a life in the Arctic in the same three species. This objective is addressed in Papers I-III. In Paper I, we aim to identify sets of candidate genes for Arctic adaptation, by testing for positive selection and identifying convergent substitutions in orthologous gene alignments. This paper represents the first

molecular screening for putative molecular adaptations in Arctic plants. In Paper II, we delve deeper into the molecular basis of cold and freezing tolerance in *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*. Here, we aim to characterize their cold induced transcriptomes, and describe how they differ from the temperate model species *Arabidopsis thaliana*, which is probably one of the best studied plant species in terms of cold tolerance. In Paper III, we aim to assemble and characterize the first genome of an Arctic plant. This is a case study of *D. nivalis*, in which we explore how Arctic adaptation has left its marks across the genome.

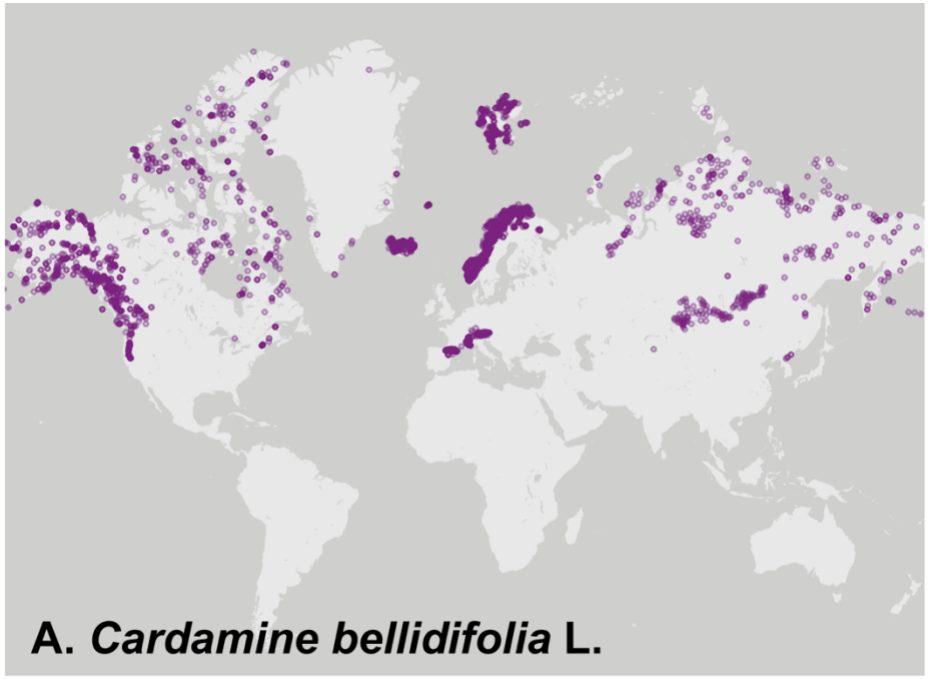
There have so far been few studies investigating genome-wide patterns of convergence between plant genera (but see e.g. Yeaman et al. 2016; Xu et al. 2017), and even fewer (if any) that have estimated molecular convergence at multiple levels of genetic similarity (i.e. codons, genes and pathways). An assessment of convergence across codons, genes, and pathways within a plant system may aid in elucidating evolutionary and ecological factors that influence the causes of convergent evolution, and thus also contribute to derive generally applicable rules governing the genetics of adaptation. Furthermore, by studying putative molecular adaptations and convergence in these three Arctic Brassicaceae, we may learn more about how plants adapt to novel and extreme environments.

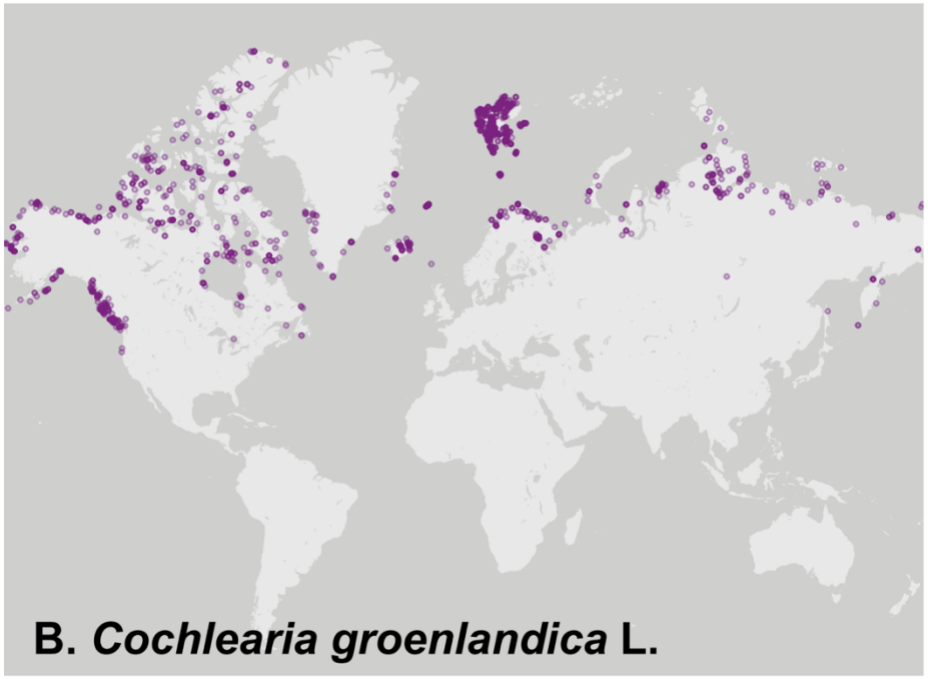
The specific aims and questions of each paper

Paper I aims to 1) identify sets of candidate genes for Arctic adaptation in *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*, 2) search for evidence of convergent evolution across codons, genes and pathways, and 3) test for a signature of convergent evolution that exceeds neutral expectations between Arctic species pairs.

Paper II aims to 1) characterize the cold induced transcriptomes of *C. bellidifolia*, *C. groenlandica* and *D. nivalis*, 2) describe how their cold response differ from that of the model species *A. thaliana*, and 3) identify potential convergent expression patterns in the three Arctic species.

Paper III aims to 1) assemble the genome of *D. nivalis* (the first genome of an Arctic plant), and 2) shed light on the genomic characteristics of a plant adapted to the extreme abiotic stresses of the Arctic.





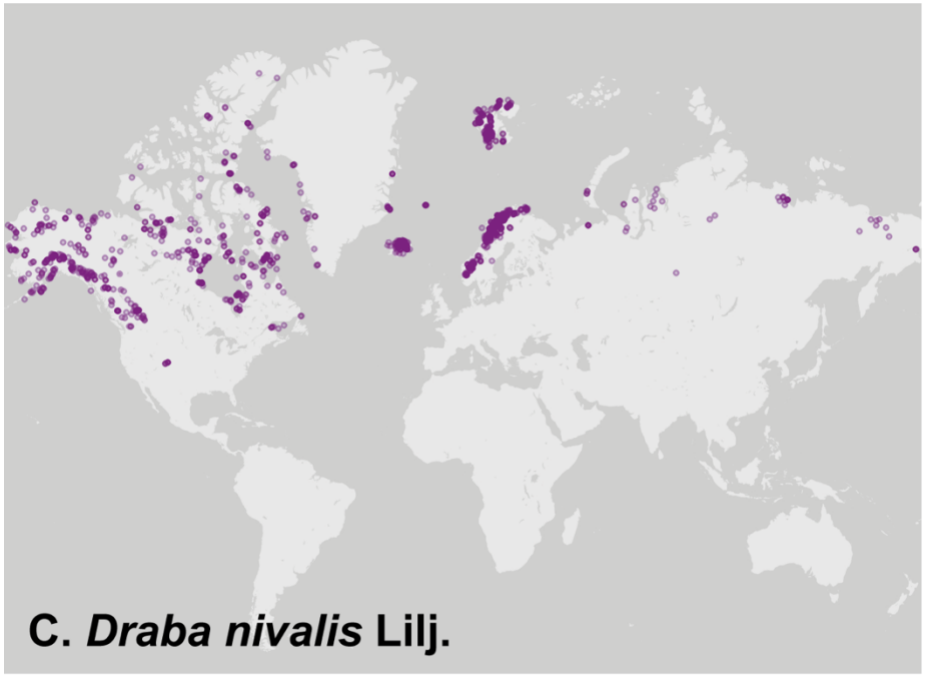


Figure 2. Approximate distribution of A) *C. bellidifolia*, B) *C. groenlandica* and C) *D. nivalis* based on occurrences from www.gbif.org.

GBIF OCCURRENCE DOWNLOAD: A) https://doi.org/10.15468/dl.km2jrr, B) https://doi.org/10.15468/dl.smsnzr, C) https://doi.org/10.15468/dl.auzsqk

The table below shows distribution across the Arctic vegetation zones as defined by Elvebakk (1999). Data from Elven et al. (2011). The zones greatly reflect the North-South temperature gradient.

	C. bellidifolia	*Co. groenlandica*	*D. nivalis*
Arctic Polar Desert	Frequent	Frequent	Scattered
Northern Arctic Tundra	Frequent	Frequent	Frequent
Middle Arctic Tundra	Frequent	Frequent	Frequent
Southern Arctic Tundra	Frequent	Frequent	Frequent
Arctic Shrub Tundra	Frequent	Frequent	Frequent
Bordering boreal or alpine areas	Frequent	Scattered	Frequent

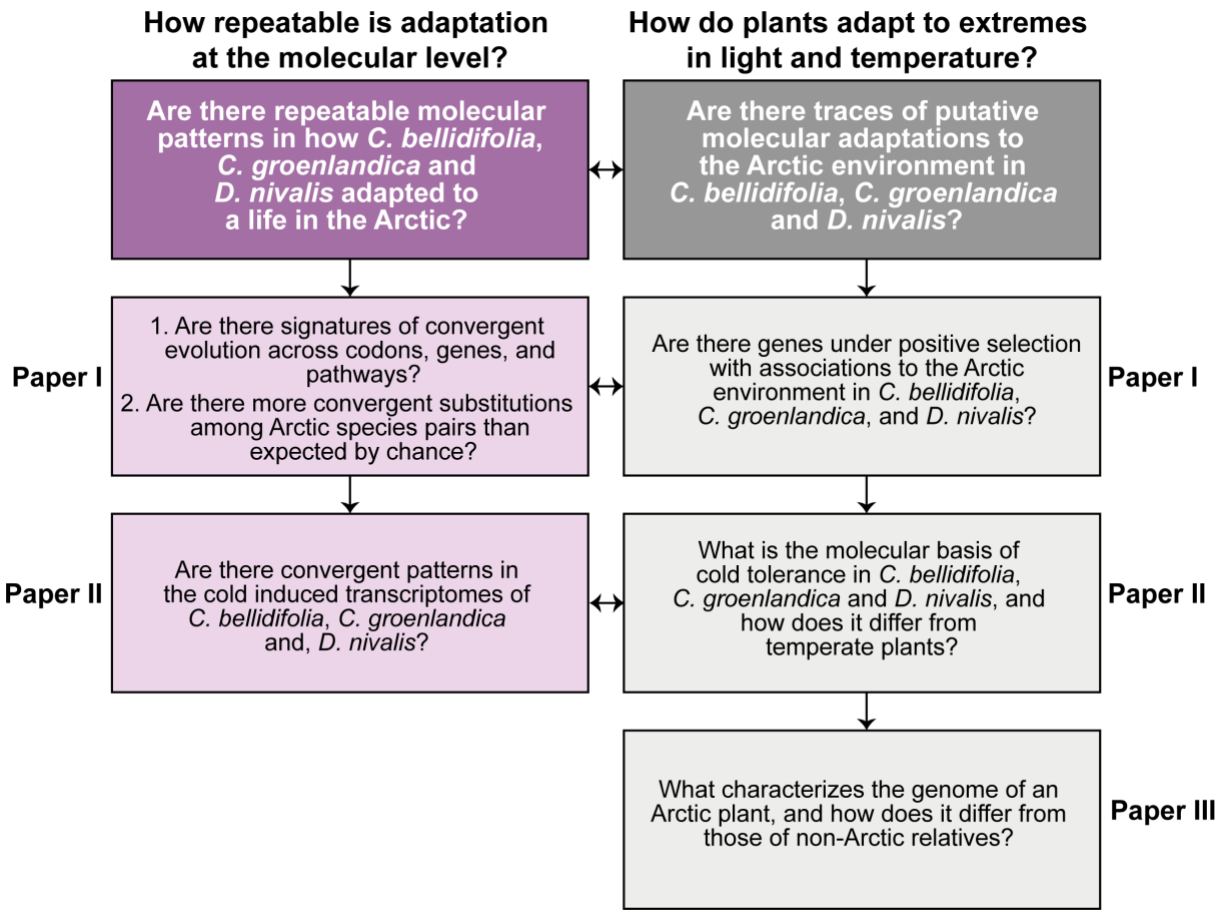


Figure 3. Flow chart showing main research questions in Papers I-III.

2. Material and Methods

2.1 Data acquisition

2.1.1 Plant material

The Arctic plant material for this thesis stems from American and Norwegian accessions of *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*. Different accessions were used in each paper, except for *D. nivalis* 008-7, which is included in all papers (see the supplementary material of Papers I-II for details on all accessions). Paper I includes one American and one Norwegian individual for each of the three species, Paper II includes four replicates derived from one American individual per species, and Paper III is the *D. nivalis* genome assembly of the American 008-7 accession. To construct a genetic map used for scaffolding the *D. nivalis* genome assembly, we also

generated an F_2-mapping population of 480 individuals by self-pollinating an F_1-hybrid obtained from a cross between an American (008-7) and a Norwegian plant (see Paper III for details).

2.1.2 Plant cultivation

All Arctic plant material was derived from plants grown in the phytotron located at the University of Oslo, Norway. Plants for paper I and III were cultivated in a growth chamber according to conditions specified in Brochmann et al. (1992). Plants for Paper II were grown in a natural daylight room (22 °C/18 °C day/night), for eight weeks, until they were subjected to a cold shock treatment (described below).

2.1.3 Cold shock treatment

To characterize and compare the cold induced transcriptomes of *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*, the plants of Paper II were given a 24-hour cold shock prior to RNA extractions (Figure 4). The treatment simulated a sudden drop in temperature during an Arctic summer day. At 13:00 (minimizing correlation with circadian change), all plants were moved from the 22 °C daylight room to a 2 °C growth chamber. Leaf tissue was sampled at four different time points: just before the plants were transferred (0h, i.e. the control point), and then after 3h, 6h, and 24h with 2 °C. Four replicates were sampled per species and per time point (i.e. four pots with seeds derived from the same parent).

2.1.4 RNA/DNA extractions

The Arctic sequence data for Paper I-II is entirely based on RNA extracted from leaves, while Paper III also includes RNA extracted from flowers, flower buds, and roots (used for annotating genes in the *D. nivalis* genome assembly). All RNA extractions were performed with the Ambion RNAqueous Kit (Thermo Fisher Scientific, Waltham, USA), following the manufacturer's protocol for fresh plant tissue. In addition, DNA for the *D. nivalis* genome assembly and F_1-mapping population (Paper III) was extracted from young leaf tissue using the Qiagen Plant Mini Kit (Qiagen, Hilden, Germany).

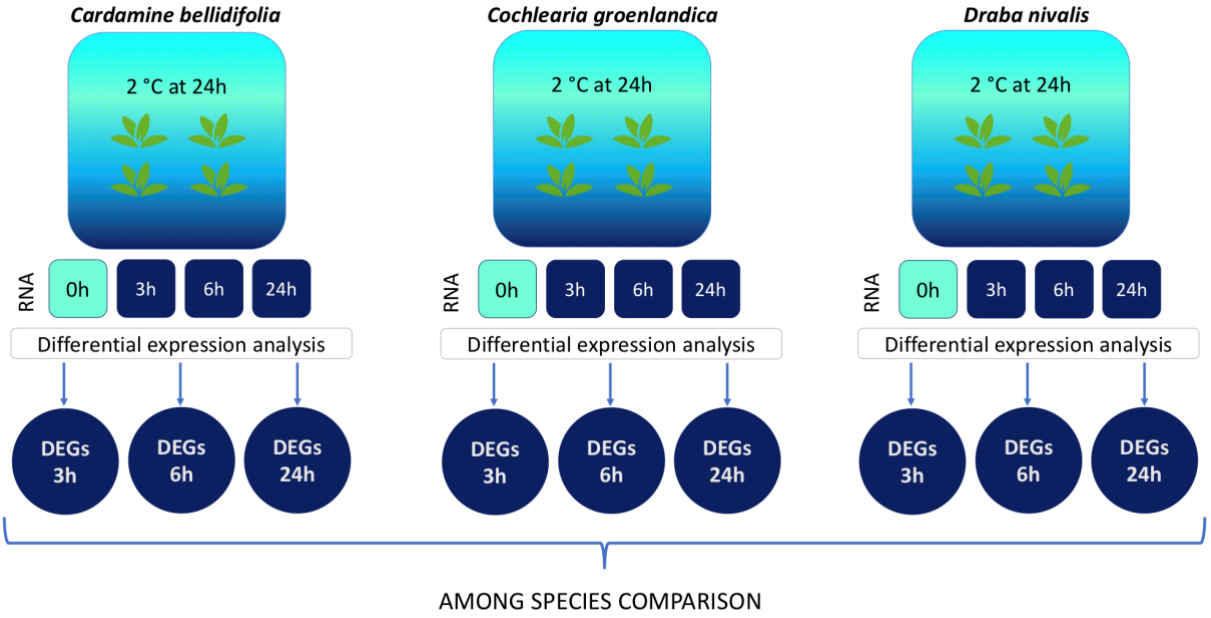


Figure 4. Experimental setup for Paper II. The three Arctic species were given a 24 hour cold shock of 2 °C, and leaf tissue was sampled for RNA extractions at four different time points: 0h (before treatment), and after 3h, 6h, and 24h with 2 °C. Differentially Expressed Genes (DEGs) were identified at 3h, 6h, and 24h, and compared among species.

2.1.5 RNA sequencing

The Norwegian Sequencing Centre (www.sequencing.no) performed the library preparations and RNA sequencing for Paper I-III. In Paper I and III, RNA was sequenced on an Illumina HiSeq 2000 (Illumina, San Diego, USA), producing paired-end reads of 125 bp with an insert size of approximately 300 bp. In Paper II, RNA was sequenced on an Illumina HiSeq 3000 (Illumina, San Diego, USA), producing paired-end reads of 150 bp with an insert size of approximately 350 bp.

2.1.6 DNA sequencing

For the genome assembly of *D. nivalis* (Paper III), several different sequencing approaches were used. The Norwegian Sequencing Centre prepared the libraries and performed short-read DNA sequencing on an Illumina HiSeq 2500 (Illumina, San Diego, USA), producing paired-end reads of 250 bp with an insert size of approximately 550 bp. We also produced single DNA-molecule long-reads with an Oxford Nanopore MinION (Oxford Nanopore Technologies, Oxford, United Kingdom). For this approach, four libraries (prepared from 8 µg of genomic DNA) were sequenced on a single MinION flow cell for ~48 hours. In addition, two flash

frozen *D. nivalis* plants were sent to Dovetail Genomics for Chicago proximity ligation (Dovetail Genomics, Scotts Valley, USA) and subsequent 150 bp paired-end sequencing on an Illumina HiSeq 2500 platform (Illumina, San Diego, USA). Finally, the F_2 plants used for the genetic map were sequenced using a double-digest restriction-associated DNA approach (ddRAD; see Paper III for details).

2.1.7 Acquisition of open source data for analyses of molecular evolution/comparative analyses

In order to perform analyses of molecular evolution (Papers I and III), and to differentiate putative Arctic patterns of adaptation (Papers I-III), we acquired genome-wide coding sequences from non-Arctic relatives within the Brassicaceae family. This included twelve additional species for Paper I, one additional species for Paper II (i.e. the model species, *A. thaliana*), and nine additional species for Paper III. The open source data for Papers II-III included only published genomes, while the data for Paper I also included raw RNA sequences from NCBI (https://www.ncbi.nlm.nih.gov/genbank/), and one already assembled transcriptome.

2.1.8 Transcriptome assemblies

Papers I-III include a total of fifteen *de novo* transcriptome assemblies: eleven for Paper I (both Arctic and non-Arctic species), three for Paper II (only Arctic species), and one for Paper III (only *D. nivalis*). Note that the transcriptomes of *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* were assembled once for Paper I, and again for Paper II to ensure that all cold induced genes were present in the reference transcriptomes used for estimating differentially expressed genes. We also assembled the *D. nivalis* transcriptome for Paper III separately, as we wanted to include genes exclusively expressed in flowers, flower buds, and roots. All fifteen transcriptomes were assembled using Trinity (Grabherr et al. 2011; Haas et al. 2013). Trinity was run with the integrated Trimmomatic option (Bolger et al. 2014), strand-specificity, and a minimum assembled contig length of 300 bp.

2.1.9 Transcriptome filtering and annotation

For Paper I, transcript abundance in each assembly was estimated with kallisto (Bray et al. 2016), and only the highest expressed isoform was retained for each Trinity gene (see Paper I for details). Chimeric transcripts (possibly resulting from

misassemblies) were filtered out following the procedure of Yang and Smith (2013), and coding regions were predicted with TransDecoder (Haas et al. 2013) using default settings. For Paper II, we used EnTAP (Eukaryotic Non-Model Transcriptome Annotation Pipeline; Hart et al. 2020) to filter and annotate the transcriptomes in two separate rounds: First to apply the EnTAP filtering option on raw transcriptomes to reduce inflated transcript estimates, and a second time to annotate the highest expressed isoform and filter out contaminants (used to annotate differentially expressed genes; DEGs). The number of complete and fragmented BUSCOs (Benchmarking Universal Single-Copy Orthologs; Simão et al. 2015) were assessed for all transcriptome assemblies in Paper I-III.

2.1.10 The genome assembly of *D. nivalis*

For the genome assembly of *D. nivalis*, we used an iterative scaffolding strategy that combined data from short DNA-reads, single DNA-molecule long-reads, proximity ligation data, and a genetic map (see Paper III for a more detailed description of our approach). A first draft assembly was performed with DISCOVAR *de novo* (https://software.broadinstitute.org/software/discovar/blog/) based on the short-read Illumina data. In order to increase contiguity, the first draft assembly (scaffold N50 = 0.03 Mb) was sent to Dovetail Genomics for scaffolding with the HiRise software based on Chicago *in vitro* proximity ligation data (see paragraph on RNA/DNA sequencing). After receiving the improved assembly from Dovetail Genomics (scaffold N50 = 2.92 Mb), we used the Oxford Nanopore MinION long reads to further scaffold the assembly using SSPACE-LongRead (Boetzer and Pirovano 2014), and LINKS (Warren et al. 2015). Finally, we also scaffolded the assembly with a genetic map using the software Chromonomer (http://catchenlab.life.illinois.edu/chromonomer/). The genetic map was based on ddRAD data from the F_2 mapping population, and was constructed using R/qtl (Broman et al. 2003; Arends et al. 2010), and ASMap (Taylor and Butler 2017; see Paper III for a more detailed description.).

2.1.11 Annotation of transposable elements in the *D. nivalis* genome assembly

Long terminal repeat retrotransposons (LTR-RTs) were annotated following Choudhury et al. (2017), and classified into *Gypsy* and *Copia* superfamilies using

LTRDIGEST (Steinbiss et al. 2009). Terminal inverted repeat transposons (TIR) were identified using GenomeTools tirvish (Gremme et al. 2013), and Helitron sequences were identified with HelitronScanner (Xiong et al. 2014). The resulting copies of transposable elements were then classified into families by clustering sequences with at least 80 % similarity using CD-HIT-EST (Li and Godzik 2006; but see Paper III for details on further classification). We annotated transposable elements along the genome using all copies of LTR-RTs, TIR transposons and helitrons from *D. nivalis* together with non-LTR-RTs (i.e. LINEs and SINEs from Viridiplantae in Repbase as a reference; www.girinst.org/repbase/). This reference was used in RepeatMasker (http://www.repeatmasker.org) with RM-BLAST as search engine and divergence set to 20 %. For more details on our transposable element annotation, see Paper III.

2.1.12 Gene annotation in the *D. nivalis* genome assembly

For gene prediction, we applied the Maker pipeline (Holt and Yandell 2011), using Augustus (Stanke et al. 2008) and SNAP (Korf 2004) as *ab initio* gene predictors. The assembled transcriptome was used as transcript evidence, while protein sequences from *Arabis nemorensis*, *Arabis alpina*, *Arabidopsis lyrata*, *Arabidopsis thaliana*, and *Eutrema salsugineum* were used as homology-based evidence. The *D. nivalis* repeat library was used to mask the repetitive elements from annotation. The resulting set of genes was annotated with gene ontology (GO) terms through Blast2GO (Conesa et al. 2005) using the UniProt database for Viridiplantae, and also Pfam domains (El-Gebali et al. 2019) using InterProScan (Jones et al. 2014). Finally, we estimated the number of complete and fragmented Benchmarking Universal Single-Copy Orthologs in the *D. nivalis* genome assembly/gene set using BUSCO (Simão et al. 2015).

2.2. Orthogroup inference and alignment generation

2.2.1 Orthogroup inference

All three papers include comparative analyses between Arctic and non-Arctic species that are based on identifying groups of genes descended from a single gene in the last common ancestor of the species being compared (i.e. orthogroups, Emms and Kelly 2015). In Paper II, we used this approach solely to compare differentially expressed genes among species based on orthogroup identity (i.e. among the three Arctic

species, and the temperate *A. thaliana*; see below; Table 1). In Paper I and III, orthogroups were used as a basis for making gene alignments for analyses of convergent substitutions, and analyses of positive selection (Table 1). The program OrthoFinder (Emms and Kelly 2015; Emms and Kelly 2019) was used to identify orthogroups in all papers.

2.2.2 Alignment generation

For Paper I and III, we generated gene alignments for analyses of positive selection and identification of convergent substitutions, based on the orthogroups delineated above (Table 1). Due to the low number of single-copy orthogroups, we divided the orthogroups with multiple gene copies/transcripts per species into subsets based on smallest genetic distance (see Paper I and III for details). The resulting orthogroup subsets contained only one gene copy per species/Arctic accession, and were realigned with PRANK (Löytynoja and Goldman 2005) through the program GUIDANCE2 (Sela et al. 2015) with ten bootstraps. Analyses of positive selection and identification of convergent substitutions can be sensitive to sequence and alignment errors, and GUIDANCE2 allows easy identification and filtering of unreliable aligned columns and sequences. Furthermore, studies have shown that a phylogeny aware aligner such as PRANK in combination with alignment filtering through GUIDANCE2 improves positive selection inference on simulated data (Jordan and Goldman 2012; Privman et al. 2012). Note that the final filtered alignments contained all species in question.

2.2.3 Species trees for analyses of molecular evolution

We generated phylogenetic species trees for the analyses of positive selection, identification of convergent substitutions, and gene family evolution in Paper I and III. The species trees for the positive selection analyses in Paper I were generated using RAxML on codon alignments of ~2000-4000 single copy genes (Stamatakis 2014; see Paper I for details). All other species trees were generated using the software OrthoFinder, which was recently updated to infer rooted species trees (Emms and Kelly 2019).

Table 1. Species included for orthogroup inference in Paper I-III.

	Positive selection	Convergent substitutions	Differential expression	Gene family evolution
I	*Arabidopsis thaliana* *Barbarea vulgaris* ***Cardamine bellidifolia*** **US/NO** *Cardamine hirsuta* *Cardamine hupingshanensis* *Nasturtium officinale* *Arabidopsis thaliana* *Arabis alpina* *Arabis verna* *Aubrieta canescens* *Draba hispanica* *Draba nemorosa* ***Draba nivalis*** **US/NO** *Arabidopsis thaliana* *Arabis alpina* *Arabis verna* ***Cochlearia groenlandica*** **US/NO** *Cochlearia pyrenaica* *Pugionium cornutum*	*Arabidopsis thaliana* *Arabis alpina* *Arabis verna* *Aubrieta canescens* *Barbarea vulgaris* ***Cardamine bellidifolia*** **US/NO** *Cardamine hirsuta* *Cardamine hupingshanensis* ***Cochlearia groenlandica*** **US/NO** *Cochlearia pyrenaica* *Draba hispanica* *Draba nemorosa* ***Draba nivalis*** **US/NO** *Nasturtium officinale* *Pugionium cornutum*		
II			*Arabidopsis thaliana* ***Cardamine bellidifolia*** **US** ***Cochlearia groenlandica*** **US** ***Draba nivalis*** **US**	
III	*Arabidopsis lyrata* *Arabidopsis thaliana* *Arabis alpina* *Brassica oleracea* *Capsella rubella* ***Draba nivalis*** **US** *Raphanus raphanistrum* *Thellungiella parvula*			*Aethionema arabicum* *Arabidopsis lyrata* *Arabidopsis thaliana* *Arabis alpina* *Brassica oleracea* *Capsella rubella* ***Draba nivalis*** **US** *Raphanus raphanistrum* *Thellungiella parvula* *Thlaspi arvense*

2.3 Comparative analyses between *C. bellidifolia*, *C. groenlandica* and *D. nivalis*

2.3.1 Identifying and comparing positively selected genes

In Paper I, we performed tests of positive selection to identify candidate genes for Arctic adaptation (i.e. genes under positive selection associated with the five categories of abiotic stresses described in the introduction), and to estimate the number of orthologous genes under positive selection in the three Arctic species (i.e. molecular convergence at the level of genes). These tests were run separately for each species to minimize the effect of missing data (i.e. cladewise; see Paper I for more details; Table 1).

We used the branch-site test (model A; Zhang et al. 2005) implemented in codeml (a part of the PAML software suite; Yang 1997), to test for positive selection on individual codons along the lineage leading to each Arctic species. The test is performed on the species trees generated with RAxML (see above). In the branch-site test, a model allowing positive selection on the foreground lineage (e.g. one of the Arctic lineages) is compared to a null-model that does not allow for such positive selection using a likelihood-ratio test (Zhang et al. 2005). Codeml was run 4-6 times per model with different initial parameter values, and the run with the highest likelihood score was used in subsequent analyses (Wong et al. 2004). The likelihood-ratio tests were performed with the R package extRemes (Gilleland and Katz 2016), using a significance threshold of 0.05 and one degree of freedom. We also ran the branch-site test for a set of close non-Arctic relatives (*Cardamine hirsuta*, *Cochlearia pyrenaica*, and *Draba nemorosa*), in order to filter out genes that were always under positive selection and to get a better understanding of which genes were tied to Arctic adaptation. See Paper I and III for more details on codeml runs.

Sets of positively selected genes (PSGs) were compared between species in order to identify orthologous genes under selection in more than one Arctic species. Genes were considered as being under "convergent selection" if the same *A. thaliana* ortholog was found to be under positive selection in two or more species. We also tested the significance of these PSG intersections with the supertest function in SuperExactTest (Wang et al. 2015) and visualized the PSG overlaps with UpSetR (Conway et al. 2017). All PSGs were annotated with DAVID (Huang et al. 2009) based on the *A. thaliana* ortholog.

2.3.2 Identification of convergent substitutions and testing for excess convergence

When identifying convergent substitutions over entire genomes, it has proven important to account for the random accumulation of convergent substitutions with increasing divergence (i.e. the expected background convergence; Thomas and Hahn 2015; Xu et al. 2017). We therefore used the program Grand-Convergence (Qian et al. 2018; Qian and de Koning 2018) to calculate posterior expected numbers of convergent and divergent substitutions between a set of branch-pairs in a phylogeny of all sampled species (see Table 1 for species list). In Grand-Convergence, a convergent substitution is defined as a change resulting in the same amino acid at the same site in two lineages, and a divergent substitution is defined as a change resulting in different amino acids at the same site in two lineages (Castoe et al. 2009).

To test if Arctic species pairs had more convergent substitutions than expected from their divergence level (i.e. "excess convergence"), we randomly chose 50 branch pairs and summed all pairwise convergent substitutions and all pairwise divergent substitutions over all gene alignments. The branch-pairs were randomly chosen, but included the Arctic species pairs and the non-Arctic reference species pairs from the positive selection tests (see Paper I for details). Using the branch totals from the 50 branch pairs, we performed a simple linear regression to investigate whether our data fitted a linear model as expected from other studies (Castoe et al. 2009; Thomas and Hahn 2015).

In addition to testing for excess convergence, we also counted the number of convergent genes in Arctic species pairs. This was taken as a measure of molecular convergence at the level of codons. Genes with at least one convergent site with posterior probability >0.80 were considered as convergent. The resulting sets of convergent genes were compared to identify genes overlapping between more than one Arctic species pair.

2.3.3 Identifying and comparing differentially expressed genes in response to cold

In Paper II, we identified differentially expressed genes (DEGs) in response to cold, and compared the resulting DEGs among the Arctic species, as well as among the Arctic species and *A. thaliana*. First, we mapped all reads to the reference transcriptomes using the alignment free mapper Salmon with a GC content bias

correction (Patro et al. 2017). We then used DESeq2 (Love et al. 2014) to identify DEGs at 3h, 6h, and 24h with cold treatment, using a design formula controlling for the effect of the pot number (design = ~ pot number + time). This means that we tested for the effect of time with 2 °C treatment, while controlling for the individual effects of the sampled pots. A generalized linear model was fitted to each gene, and a Wald test (Love et al. 2014) was applied to contrast the 3h, 6h, and 24h model coefficients to the 0h model coefficients. A gene was considered as being differentially expressed if the transcript level exhibited ≥ twofold change in response to 3h, 6h, and 24h with cold (log2 fold change threshold = 1). A Benjamini-Hochberg adjusted p-value (Benjamini and Hochberg 1995) was used to evaluate the significance of each differentially expressed gene (False Discovery Rate cutoff of alpha = 0.05). We also generated heatmaps of the top 30 differentially expressed genes with the lowest false discovery rate using the pheatmap package in R (Kolde 2019) using regularized log function (rld) on original count data. See Paper II for additional details.

2.4 Genome evolution in *D. nivalis*

2.4.1 Comparative chromosome painting and analyses of synteny

In Paper III, we examined chromosome evolution in *D. nivalis*, by performing comparative chromosome painting (CCP) and pairwise synteny analyses with other species in the Brassicaceae. The chromosome painting (i.e. fluorescence *in situ* hybridisation of chromosome-specific DNA probes) was used to identify genomic blocks of the ancestral crucifer karyotype in *D. nivalis* (Schranz et al. 2006; Lysak et al. 2016) after the procedure of Mandáková and Lysak (2016a; 2016b). Syntenic relationships to *Arabis alpina* and *Arabidopsis lyrata* were inferred by aligning the *D. nivalis* genome to the two other species using NUCmer (Kurtz et al. 2004), and the results were synthesized with those of the CCP to infer the structure of the *D. nivalis* chromosomes. See Paper III for a more extensive description of these approaches.

Table 2. Duplication categories used in analyses of gene evolution

Abbrev.	Duplication category	Definition (Qiao et al. 2019), see Paper III
WGD	whole-genome duplication	Genes that reside in relatively large collinear chromosomal regions
TD	tandem duplication	Gene pairs that are located next to each other on the same chromosome. Possibly resulting from unequal crossing over.
PD	proximal duplication	Gene copies that are near each other but separated by several genes (10 or fewer genes). Might occur through localized transposon activities or from ancient tandem duplicates interrupted by other genes.
TRD	transposed duplication	Gene pair consisting of an ancestral and a novel locus. Presumed to arise through distantly transposed duplications by DNA-based or RNA-based mechanisms
DSD	dispersed duplication	Two gene copies that are neither neighboring nor collinear. Occurs through unpredictable and random patterns by mechanisms that remain unclear.

2.4.2 Gene and gene family evolution

In order to explore the specialization in the *D. nivalis* gene set, we first compared the abundance of protein family (Pfam) annotations to nine other Brassicaceae genomes (see Table 1 for species list). Genes were annotated with Pfam domains using InterProScan (Jones et al. 2014) as described for *D. nivalis* above. The number of Pfam domains were quantified for each species, and domains with a Z-score above 1.96 or below -1.96 in *D. nivalis* were considered as significantly enriched or significantly contracted, respectively.

To test for significant contractions and expansion of gene families in the genome of *D. nivalis*, we used CAFE (Han et al. 2013) to analyze gene family sizes using the orthogroups inferred for the ten species being compared (Table 1), as well as the species tree generated by Orthofinder (but see details on tree generation in Paper III). Gene duplications within the *D. nivalis* genome were classified into five categories using the Dup_GenFinder pipeline (Table 2; Wang et al. 2012; Qiao et al. 2019). This approach is based on an all-versus-all BlastP of the *D. nivalis* gene set to itself, and a BlastP of the *D. nivalis* gene set to the closely related *Arabis alpina*, in order to identify homologous gene pairs. The Dup_GenFinder pipeline then builds on MCScanX (Wang et al. 2012) to identify patterns of synteny and collinearity in the

duplicated genes, both within the *D. nivalis* genome and between the *D. nivalis* and *A. alpina*, to make the duplication classification.

2.4.3 Additional positive selection analyses using the *D. nivalis* genome

In Paper III, we also performed genome-wide tests of positive selection specifically to identify candidate genes for Arctic adaptation in the *D. nivalis* genome. These tests included a different set of non-Arctic species than in Paper I, and were only based on coding sequences from genome assemblies (Table 1). The branch-site test was run on gene alignments like described above.

2.5 Gene Ontology enrichment analyses

All three papers contain functional analyses of gene sets based on Gene Ontology (GO) enrichment analyses. In Paper I, we used GO-enrichment tests to functionally characterize sets of PSGs and convergent genes, and to assess if positive selection has acted on similar functional pathways in the three Arctic species. In Paper II, they were used to functionally characterize sets of upregulated and downregulated genes in response to cold in the three Arctic species and *A. thaliana*. Finally, in Paper III, we used GO-enrichment tests to functionally characterize PSGs and sets of expanded and contracted gene families in the *D. nivalis* genome.

The GO-enrichment analyses were largely performed in the same way for Paper I-III. A Fisher's exact test was used in combination with the "classic," "elim," and "weight" algorithms implemented in topGO of Bioconductor (Gentleman et al. 2004; Alexa et al. 2006), to test for overrepresented GO-terms within the three domains: Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). The three algorithms differ in that the "classic" algorithm processes each term independently without taking the GO-hierarchy into account, the "elim" algorithm traverses the GO-graph bottom-up, discarding genes that have already been mapped to significant child terms, and the "weight" algorithm is weighing genes annotated to GO-terms based on the scores from neighboring GO-terms (Alexa et al. 2006). Based on simulated data, the "elim" and "weight" algorithms tend to be more conservative than the "classic" algorithm (Alexa et al. 2006).

In Paper I, sets of PSGs and convergent genes were annotated based on homology to *A. thaliana* using annotations from org.At.tair.db (Carlson 2018), and the *A. thaliana* gene set was used as the background gene universe in all analyses. For

this paper we also performed gene set enrichment analyses for the three non-Arctic reference species that were included in the positive selection analyses and analyses of convergent substitutions. In Paper II, the EnTAP gene annotations of each transcriptome were used as a background gene universe for each species (but see Paper II for details on *A. thaliana*). In Paper III, we used the *D. nivalis* annotated gene set as a background for all analyses.

The significance level was set to $p < 0.05$ in all analyses, and we did not correct for multiple testing following the recommendations of the creators of the topGO package (Alexa and Rahnenfuhrer 2018). Enriched GO-terms that might be tied to Arctic adaptation were assessed based on literature. Finally, in Paper I, we also generated a heatmap based on the proportions of PSGs associated with the five categories of Arctic stresses (see Paper I for more details).

3. Main findings and discussion

3.1 An assessment of molecular convergence in Arctic Brassicaceae

In Paper I-II we examined whether adaptation to the Arctic environment shared repeatable molecular patterns in *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*. In our first study, we found little evidence for independently fixed mutations at the same sites, or for positive selection acting on the same genes in the three species (Figure 5; Paper I). These results were based on ~12-13,000 gene alignments for the positive selection analyses, and ~8000 gene alignments for the identification of convergent substitutions. Neither did we find evidence that there were more convergent substitutions between the Arctic species than what would be expected by chance (i.e. the Arctic branch pairs fitted a linear relationship between divergent and convergent substitutions; Figure 5B). Although there were few overlaps among the positively selected gene sets identified in each Arctic species, they showed convergent functional profiles associated with the extreme abiotic stresses characteristic of the Arctic (discussed in 3.3). Based on these findings, we concluded that the three species likely had adapted to the Arctic environment by modifying different components in similar stress response pathways, implying that adaptation had followed independent genetic trajectories in each species.

In Paper II, we compared the cold-induced transcriptomes of *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*, and found that their cold-responses were highly species-specific. Not only were most differentially expressed genes unique for each species, but the number of differentially expressed genes shared by the three Arctic species and the temperate *A. thaliana* were higher than the number shared by the three Arctic species alone (Figure 6). Many of the shared genes between *A. thaliana* and the Arctic species included components of the CBF pathway. This is one of the main regulatory pathways leading to cold acclimation and increased freezing tolerance in plants, and parts of this pathway is also conserved in other plant lineages (Jaglo et al. 2001; Shi et al. 2018). This suggests that the cold response in Arctic Brassicaceae mainly evolved independently, but with some components likely conserved across the family. These findings are thus in line with the main conclusion of Paper I.

3.2 Causes of low levels of molecular convergence in Arctic Brassicaceae

The low levels of molecular convergence in Arctic Brassicaceae were somewhat surprising, as there is a growing body of literature showing that adaptation may involve a high degree of repeatability also at the level of substitutions and genes (Conte et al. 2012; Stern 2013). The last years have even seen several reports of genome-wide convergence, i.e. hundreds of convergent substitutions spanning entire genomes (see e.g. Parker et al. 2013; Xu et al. 2017). However, some of these studies have been called into question due to methodological issues (see e.g. Thomas and Hahn 2015; Zou and Zhang 2015), and in parallel with the publication of Paper I, there have been several genome-wide comparisons describing low levels of molecular convergence, despite high levels of phenotypic convergence (e.g. Davies et al. 2018; Corbett-Detig et al. 2020). The debate concerning the prevalence of molecular convergence thus seems to continue. To gain a better understanding of when molecular convergence is common, we need to deduce the underlying causes influencing levels of evolutionary repeatability. In Paper I, we propose several hypotheses for the observed levels of molecular convergence in Arctic Brassicaceae. By synthesizing the findings in Paper I and Paper II, we might gain new insights into whether these hypotheses hold water.

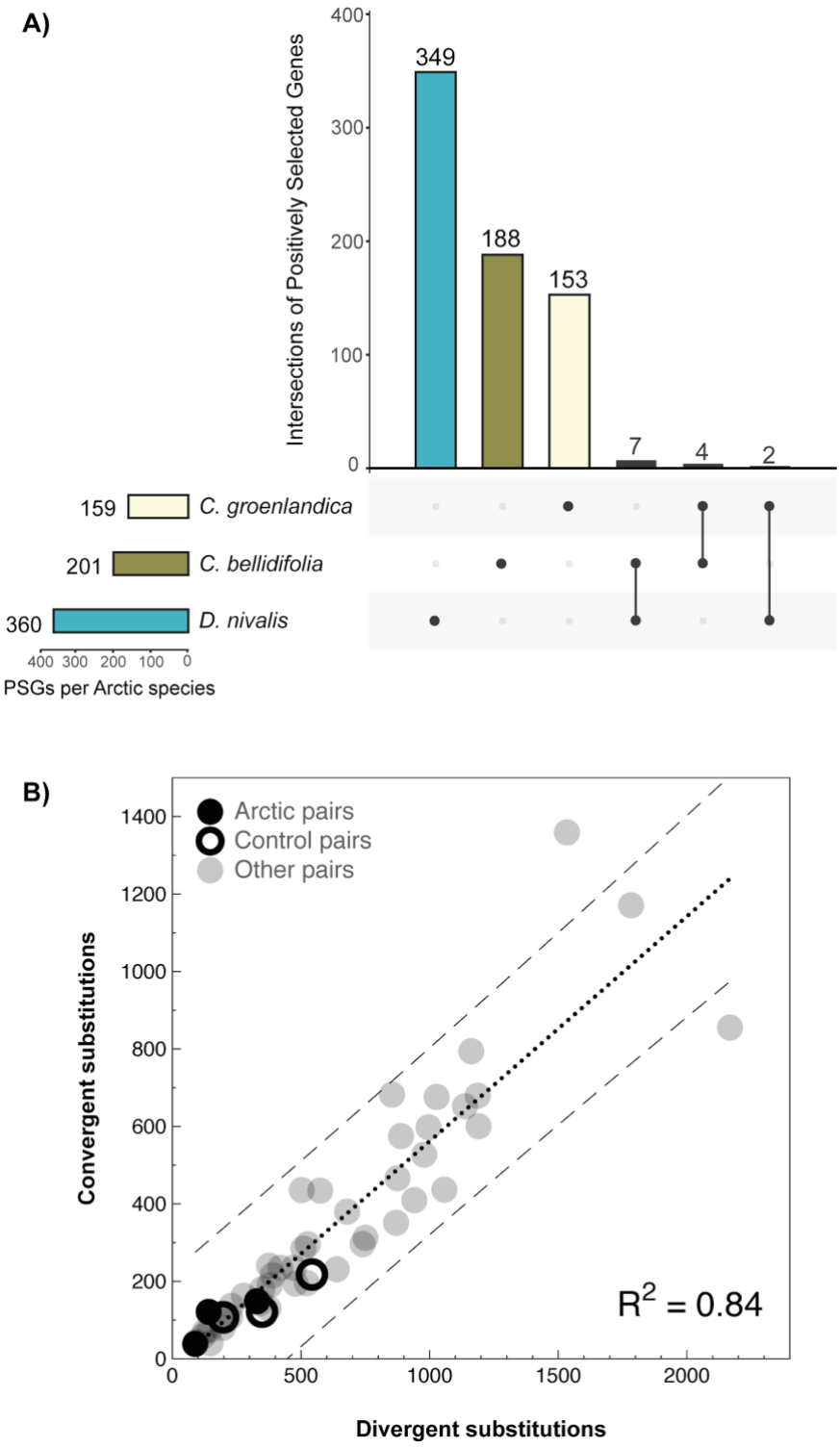


Figure 5. A) UpSet plot of orthologous PSGs identified in Arctic species (Paper I). The plot in the lower left corner shows the total number of PSGs, and the main plot shows the number of unique PSGs, followed by PSG intersections/overlaps between species (connected dots). **B) Posterior expected numbers of convergent and divergent substitutions between 50 branch pairs from the full species phylogeny (Paper I).** Numbers are sums over all alignments. Dotted line, the regression line; dashed line, 95% prediction intervals; filled black circles, the three Arctic branch pair comparisons; open circles, the branch pairs of the non-Arctic reference species; filled gray circles, remaining branch pairs.

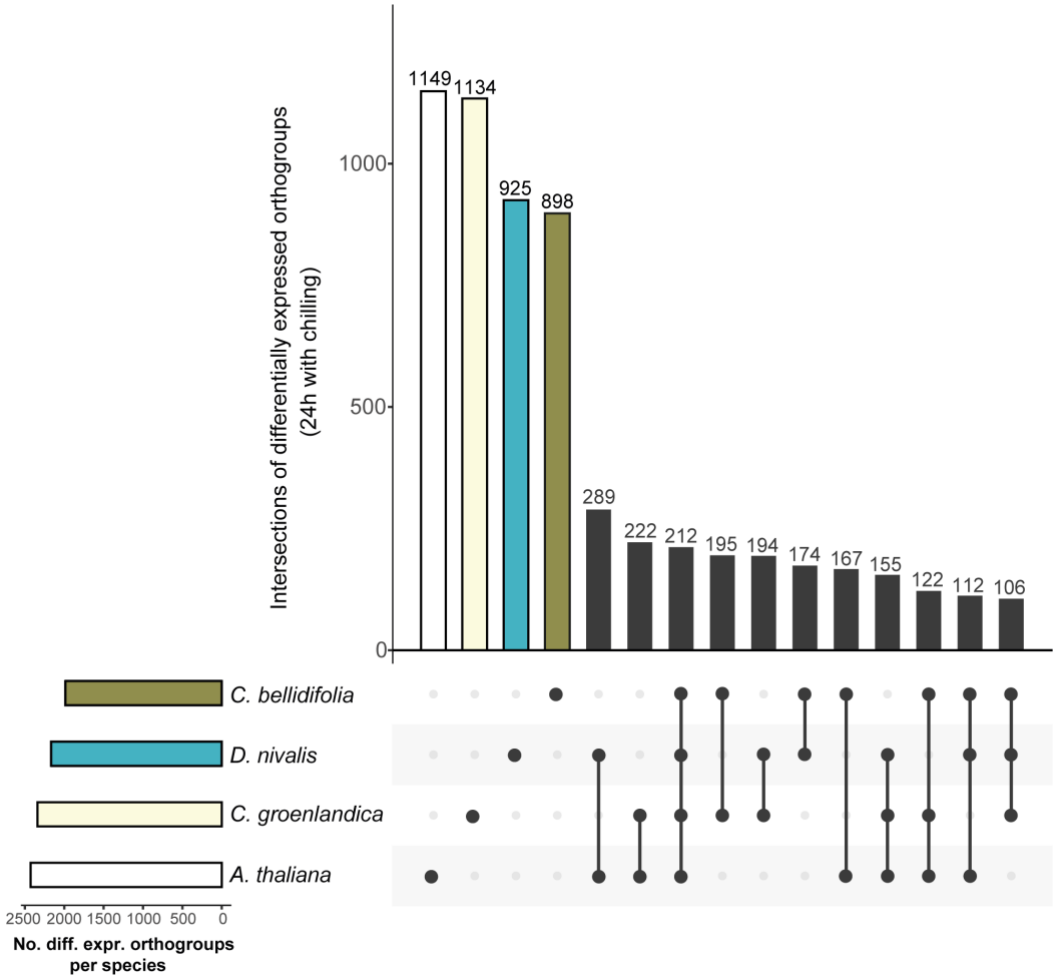


Figure 6. UpSet plot of differentially expressed orthogroups (Paper II). The plot in the lower left corner shows total numbers of differentially expressed orthogroups, and the main plot shows the number of unique differentially expressed orthogroups, followed by orthogroups intersections/overlaps between species (connected dots).

3.2.1 Vast opportunity space for adaptation

In Paper I, one of our main hypotheses for the low levels of molecular convergence in Arctic Brassicaceae was tied to the vast opportunity space for adaptation. Molecular convergence in Arctic plants could be uncommon solely due to the number of possible evolutionary trajectories leading to better performance under temperature and light stress. This is because light and temperature affect nearly all aspects of plant physiology and biochemistry (Salisbury 1985; Heino and Palva 2004), and many of the traits in question are also highly polygenic and include multiple molecular pathways. Such traits may have great genetic redundancy, and should also show less repeatable patterns of adaptation than monogenic and oligogenic traits (Yeaman et al. 2016). This was supported by our findings in Paper II, where the cold

response of Arctic Brassicaceae seemed to be highly species-specific, while involving thousands of genes. Yet, there are also studies showing high levels of molecular convergence in polygenic traits. Yeaman et al. (2016) studied convergent local adaptation to climate in two distantly related conifers, and found that local adaptation was surprisingly convergent at the genomic level. They reported finding a suite of 47 genes containing variants associated with spatial variation in temperature or cold hardiness in both species, despite the highly polygenic nature of the traits in question (Yeaman et al. 2016). However, their approach is very different from ours as they used a population genetic approach based on genotype-phenotype and genotype-environment association studies, and our results might not be directly comparable.

Intriguingly, a study published in parallel with Paper I, presents a very similar hypothesis for low levels of molecular convergence in a lizard system (Corbett-Detig et al. 2020). Anolis lizards have independently converged on suites of adaptive behavioral and morphological traits tied to specific habitats on Caribbean islands (ecomorphs), and represent one of the best-studied examples of phenotypic convergence (e.g. Schoener 1970; Sanger et al. 2012; Mahler et al. 2013). Corbett-Detig et al. (2020) suggested that the lack of genome-wide convergence in Anolis lizards might be tied to the vast number of molecular mechanisms through which a lineage can evolve the characteristic phenotypes associated with a given ecomorph. They also similarly point to the traits in question being quantitative (Corbett-Detig et al. 2020). Our similar set-ups and findings thus support the idea that phenotypic convergence is not mirrored at molecular level if the basis of the traits in question are highly polygenic, and when the opportunity space for adaptation is high.

3.2.2 The multifaceted nature of stress-response pathways in plants

Another possibility is that the general nature of stress-response pathways in plants might influence convergence levels in Arctic Brassicaceae. Arctic plants must cope with a wide range of abiotic stresses, and research on plant stress responses have shown considerable overlap in gene expression patterns between various stress stimuli (Kilian et al. 2007). This might indicate that there are substantial constraints of these pathways and/or that adaptation to the Arctic environment might happen through modification of different elements in a variety of general stress response pathways. In our study of cold response (Paper II), we find support for both

possibilities. Many of the genes showing putative conserved expression patterns were responsive to other abiotic stresses (e.g. salt and dehydration stress), and it is also known that there is a certain shared basis between response to cold stress and other abiotic stresses (Kreps et al. 2002; Bartels and Souer 2004; Kilian et al. 2007). On the other hand, the cold response of Arctic Brassicaceae was also highly species-specific, which supports that adaptation to low temperature might involve different elements in multiple molecular pathways. One could, for instance, imagine that evolution of cold tolerance might take different genetic trajectories if the ancestors were preadapted to different stresses like drought or salt stress. Very little is known about the evolutionary history of the species in question, but it is assumed that plants colonizing the Arctic might show different levels of preadaptation (Savile 1972).

3.3.3 Divergence-time, species-specific habitat preferences, and a conservative methodology

Our study compares plant species representing three clades that diverged ~30 Mya, and the age of the common ancestor is following relatively high. The degree of relatedness of the compared species is only briefly touched upon in Paper I, but is one of the most debated factors believed to influence levels of molecular convergence (Schluter et al. 2004; Arendt and Reznick 2008). Haldane (1932) suggested that related species might evolve in parallel, as they will tend to vary in similar directions and be subject to similar selective influences. It has also been shown that gene reuse decreases with increasing age of the common ancestor of compared taxa (Conte et al. 2012; Bohutínská et al.). Accordingly, the levels of molecular convergence in Arctic Brassicaceae might be low because they represent distantly related lineages. However, as pointed out in Paper I, there are also several comparative studies of plant genera with even higher divergence times showing surprisingly high levels of molecular convergence (e.g. ~140 Mya in Yeaman et al. 2016, and ~100 Mya in Xu et al. 2017 based on APG IV 2016).

Finally, in Paper I we also discuss how our findings could be caused by species-specific habitat preferences, and the conservative methodology used to study molecular convergence. Both of these possibilities should be less relevant for the results of Paper II, as the three species are exposed to relatively similar temperature spectres across the Arctic (see Figure 2; Elven et al. 2011), and because we used an entirely different approach in the two papers. Based on these findings, it is therefore the three first hypotheses (vast opportunity space for adaptation, the multifaceted

nature of stress response pathways, and high divergence levels), that are the most likely candidates for low convergence levels in Arctic plants.

3.3 Possible molecular adaptations to extreme abiotic stress

This thesis presents some of the first molecular evidence for putative plant adaptation to the Arctic environment. By testing for positive selection, identifying convergent substitutions, exploring gene expression in response to cold, and investigating larger patterns of genome evolution, we find that the extreme abiotic stresses of the Arctic environment have shaped *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* on various genetic levels. These putative molecular adaptations to the Arctic environment can be attributed to general abiotic stress responses, but also 1) physiological cold stress, 2) ice and snow stress, 3) protein misfolding and membrane stress, 4) light stress, and 5) oxidative stress.

3.3.1 Adaptations to general abiotic stress

In Paper I, we found that genes associated with general abiotic stress responses, such as "positive regulation of abscisic acid-activated pathway" (BP), were overrepresented in the Arctic convergent gene sets. Abscisic acid (ABA) is a plant hormone that plays important roles in stress signalling under abiotic stress, and has important roles in the regulation of cold response in plants (Ruelland et al. 2009; de Zelicourt et al. 2016). Consistent with this, we found that genes associated with abscisic acid response were overrepresented in the Arctic cold-induced transcriptomes (Paper II), and among the expanded gene families in *D. nivalis* (Paper III). Other general stress responses overrepresented among Arctic convergent genes (Paper I), included "calcium channel activity" (MF), and "regulation of stomatal movement" (BP). Most abiotic stress responses are initiated with an increase in cytoplasmic calcium (Knight and Knight 2012), and stomatal regulation has an important role in plant responses to environmental changes (plants will close their stomata in response to drought or cold stress; Wilkinson et al. 2001). Genes associated with stomatal movement were also overrepresented among the PSGs of *D. nivalis* (Paper I and III). We can thus hypothesize that the abiotic stresses associated with extreme light and temperature conditions have influenced the evolution of general stress response pathways in Arctic Brassicaceae.

3.3.2 Adaptations to physiological cold stress

From our three papers, it is obvious that low temperatures have posed strong selection pressures on *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*. In all species, we found sets of candidate genes for cold adaptation that were under positive selection, or contained convergent substitutions in two or more Arctic species (Paper I, but also Paper III). In *C. groenlandica* and *D. nivalis*, genes annotated with "response to cold" were even overrepresented among the PSGs (Paper I). Many of the candidate genes for cold adaptation were confirmed to be upregulated in response to cold in Paper II, and several had important functions in cold acclimation and freezing tolerance. This included genes such as *LEA4-5* (protects enzyme activities from freeze-thaw damage; PSG in *C. groenlandica*), *RAV1* (transcription factor that may be involved in growth regulation; PSG in *C. groenlandica*), *COR15B* (protects chloroplast membranes during freezing; PSG in *D. nivalis*), and *CSDP1* (blocks secondary structure of mRNA which in turn facilitates translation at low temperatures; PSGs in *D. nivalis*; all gene information from TAIR 2020). An interesting candidate gene for cold adaptation in all species was *EMB2742*. This gene contained convergent substitutions in all Arctic species, and was also upregulated in response to cold. Its exact function remains to be discovered, but the gene is reported to be involved in embryo development ending in seed dormancy in *A. thaliana* (TAIR 2020).

Paper II revealed that the cold induced transcriptomes of *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* followed many similar patterns as in the temperate *A. thaliana* (i.e. involving a constant increase in DEGs during a time course of 24h, and including similar numbers of DEGs). This indicates that Arctic plants are not always prepared for low temperatures, but need a cold acclimation period to increase freezing tolerance in similarity with temperate relatives. Several genes and biological processes were also shared with *A. thaliana* (see Paper II), but some were unique to the three Arctic species alone. The Arctic DEG sets shared several transcription factors with potential important roles in cold acclimation, such as *CBF4*, *RAP2.10* and *RAP2.2*. We also found traces of a putative Arctic MAPK (mitogen activated protein kinase) cascade, which included a putative convergent gene from Paper I (*MAPKKK14*). MAPK cascades are known to be important in regulating the CBF-pathway (Teige et al. 2004; Shi et al. 2018). Intriguingly, we also found that the PSG sets of *C. bellidifolia* and *D. nivalis* had an overrepresentation of genes associated

with “MAPK cascade” in Paper I. Finally, the Arctic upregulated DEGs were overrepresented by genes associated with “spermidine biosynthetic process”. Based on studies in other plants, spermidine might have a role in maintaining photosynthesis rates at low temperatures (He et al. 2002), and Arctic species are also known to exhibit optimal photosynthetic rates at lower temperatures than other plants (Chapin 1983).

In Paper III, we found that expanded gene families in the *D. nivalis* genome are enriched for functions associated with meiosis, and specifically with the assembly of the synaptonemal complex. Four PSGs were associated with Meiosis I, and two of these were homologs of the *A. thaliana ZYP1A*. This is one of three synaptonemal complex transverse filament proteins whose function is disrupted by temperature stress (Bomblies et al. 2015). One could therefore hypothesize that meiosis in *D. nivalis* has been tuned to the low temperatures in the Arctic. There was also an overrepresentation of PSGs associated with vernalization response in *D. nivalis* (i.e. the acceleration of flowering by a long period of low temperature; Gendall et al. 2001). As temperatures in the Arctic are low year round, one could expect that the vernalization response in *D. nivalis* has evolved to fit the current conditions.

3.3.3 Adaptations to ice and snow stress

Arctic plants are always at risk of experiencing ice and snow stress, and this was also reflected in our findings in Paper I-III. Consistent with ice formation being accompanied by cellular dehydration and increased concentrations of salts and toxic solutes (cf. introduction), we observed that genes annotated with “response to salt stress”, “response to water deprivation”, and “response to toxic substance” (all BP) were overrepresented among the cold-induced genes in *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* (and in *A. thaliana*; Paper II). We also found that there was a large number of PSGs associated with salt stress and osmotic stress in all Arctic species (Paper I), and that gene families associated with desiccation resistance are expanded in *D. nivalis* (Paper III). Stress response pathways partially homologous with those of freezing (Bartels and Souer 2004) thus seem to be under evolution in Arctic Brassicaceae.

In Paper II, we saw that genes associated with possible cryoprotectants, like flavonoids, sucrose, and proline, were upregulated in response to cold in Arctic species. Accumulation of such compounds/solutes are known to contribute to

freezing tolerance in plants (Ruelland et al. 2009). There was also a significant up-regulation of many Late Embryogenesis Abundant (LEA) genes, (e.g. the above mentioned *COR15B* and *LEA4-5*), which are known to be involved in protection from freezing damage (Hundertmark and Hincha 2008; Cuevas-Velazquez et al. 2014). Other candidate genes for adaptation to ice and snow stress included *SFR2*, which is involved in protecting the chloroplast envelope from freezing damage (a PSG in *C. groenlandica*; Fourrier et al. 2008). These findings suggest that Arctic Brassicaceae are well adapted to ice and snow stress, which is consistent with a recent study demonstrating high tolerance to ice encasement in other Arctic plants (Bjerke et al. 2018).

3.3.4 Adaptations to protein misfolding and membrane stress

In line with cell membrane systems being the primary site of freezing injury in plants (Steponkus 1984), we found that genes associated with the plasma membrane were overrepresented among the PSGs in all Arctic species (Paper I). The plasma membrane is also known to have important functions under freezing stress, such as allowing water to flow out of the cell in presence of ice, while also preventing the spread of ice crystals within the cytosol (Uemura and Steponkus 1999). Consistent with these findings, we observed many cold induced DEGs that are involved in protecting cell membranes from freezing injury in Paper II (e.g. the above mentioned *COR15*). These results highlight the important role of cell membranes in adaptation to the Arctic environment.

Low temperatures affect the structure of proteins, and lead to accumulation of misfolded proteins in the endoplasmic reticulum (ER; Zhu 2016). We found that three of the thirteen orthologous genes under positive selection in more than one Arctic species were tied to protein folding/traffic and ER morphology. Genes annotated with protein misfolding were also overrepresented among the PSGs in *D. nivalis* (Paper I). In addition, all Arctic species had many heat-shock related proteins under positive selection (Paper I). Heat-shock proteins are involved in protein folding and membrane stabilization under both normal and stressful conditions (Wang et al. 2004). Challenges associated with protein structure under low temperatures thus seem to have imposed strong selection pressures on Arctic Brassicaceae.

3.3.5 Adaptations to light stress

All Arctic species seem to possess some putative molecular adaptations to the Arctic light environment. In Paper I, we found many PSGs and convergent genes annotated with response to far-red or blue light in Arctic species (e.g. *CRY2*, a blue light receptor involved in cotyledon expansion and flowering time; TAIR 2020). Two genes associated with light-controlled growth and development were also under positive selection in two different Arctic species (*IAA16* and *PAC*; Rinaldi et al. 2012; TAIR 2020). However, most putative molecular adaptations to the Arctic light environment were observed in *D. nivalis*, which had especially many PSGs associated with photoperiodism (Paper I and III).

Although Paper II focused on cold response, we observed several upregulated DEGs that were associated with both circadian regulation, and regulation of freezing tolerance in all species (i.e. *COR27* and *COR28*; Li et al. 2016). These DEGs were shared with *A. thaliana*, and fit well with the CBF pathway being gated by the circadian clock (Fowler et al. 2005; Dong et al. 2011). Some of the cold induced DEGs shared only by the Arctic species were also related to circadian rhythm (e.g. *REVEILLE 2*, *PLC1* and *HY5*). These might be putative adaptations for balancing freezing tolerance and plant growth under Arctic light conditions (but see Discussion in Paper II).

The evolution of the above mentioned genes might be tied to changes in day length, and/or the enrichment of blue and far-red light with increasing latitude (Nilsen 1985; cf. Salisbury 1985). Such cues are important for development, reproduction, growth, and the onset of dormancy in plants (Chory et al. 1996; Jackson 2009; Piskurewicz et al. 2009).

3.3.6 Adaptations to oxidative stress

In papers I-III, we observed several pieces of evidence for oxidative stress being especially challenging for Arctic plants. This is consistent with 1) high irradiance in combination with low temperatures leading to excessive production of ROS (Inzé and Van Montagu 1995), and 2) accumulation of ROS accompanying many other abiotic stresses common in the Arctic (cf. introduction). In Paper I, we found that genes annotated with “response to oxidative stress” (BP) were overrepresented among the PSGs of *D. nivalis* and *C. bellidifolia*, and also among Arctic convergent genes. These candidate genes for adaptation to oxidative stress included e.g. *PCR2* and *OGOX1*

(both under positive selection in *C. bellidifolia* and *C. groenlandica*), and the two catalase genes *CAT1* and *CAT2* (both PSGs in *D. nivalis*, but the latter also contained a convergent substitution in *D. nivalis* and *C. groenlandica*).

In Paper II, we found that genes associated with "response to oxidative stress" and "response to karrikin" were overrepresented among the cold induced DEGs in all Arctic species. This pattern was also found in *A. thaliana*, but karrikin responsive genes were especially common among the DEGs found in Arctic species only. The role of karrikin in abiotic stress is just being discovered, but it may improve plant vigour through regulation of redox homeostasis under stressful conditions (Shah et al. 2020).

In Paper III, we also observed that the *D. nivalis* gene set was enriched for Pfam domains associated with "oxidation-reduction process" (BP), and the expanded gene families for various terms related to redox-activity. Traces of oxidative stress is, followingly, apparent on many genetic levels in Arctic Brassicaceae.

3.3.7 A functional syndrome of Arctic adaptation

Our findings show that the *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* possess putative molecular adaptation to cold stress, freezing stress, oxidative stress, and the unique Arctic light regime, although the exact substitutions and genes vary from species to species. The three species thus seem to have evolved similar ecological preferences and phenotypes through different genetic trajectories. The identified suites of molecular adaptations could represent a functional syndrome of Arctic adaptation in Brassicaceae and perhaps flowering plants in general, but this needs experimental verification. The candidate genes and traits highlighted in this thesis might serve as a good starting point for future studies.

3.4 The genome assembly of *D. nivalis* - Additional findings

Paper III presents the genome assembly of *D. nivalis*, which is the first complete assembly of a plant adapted to the high Arctic. We created a 302 Mb assembly that is highly contiguous with 91.6 % assembled into eight chromosomes (the base chromosome number of the species). The structure of the *D. nivalis* genome seem to be very similar to that of the closely related *Arabis alpina* (Willing et al. 2015), and also consistent with genome structures determined for other species in the Arabideae tribe (including three *Draba* species; Mandáková et al. 2020). However, we also

identified several rearrangements and extensive centromere repositioning relative to the ancestral crucifer karyotype.

As much as one third (94.8 Mb) of the genome was annotated to be direct remnants of repetitive elements, dominated by LTR-RTs (60.1 Mb), TIRs (20.1) and Helitrons (13.4 Mb). TIR-CACTA elements and Helitrons seem to be particularly abundant (Hu et al. 2019), and several *Copia* and *Gypsy* LTR-RTs showed evidence of recent transposition bursts within the genome. Intriguingly, gene-duplications in *D. nivali*s were dominated by transposed (TRDs) and dispersed (DSDs) duplicates, both of which are product of transposition events possibly mediated by transposable elements (Qiao et al. 2019).

We annotated 33,557 gene models in the *D. nivalis* assembly, of which 74 % were functionally annotated based on entries in the UniProt knowledge base, and 70 % were annotated with InterPro domains. Comparison of Pfam domains with nine other species in the Brassicaceae family, revealed that 226 Pfam domains were significantly enriched in *D. nivalis*, and 32 significantly depleted. Among the significantly enriched Pfam domains, were many associated with oxidation-reduction (as discussed in section 3.3 above). We also found that 198 gene families were significantly expanded, and that 31 gene families were significantly retracted in the *D. nivalis* genome. Many of these expanded gene families had functions possibly associated with abiotic stresses characteristic of the Arctic, like e.g. stress signaling, desiccation resistance, and redox activity (cf. section 3.3). Relevant for adaptation to the Arctic was also that LTR-RTs, TIR transposons and helitrons may have contributed to the proliferation of the significantly expanded *D. nivalis* gene families, but processes of proximal and tandem duplication (PD and TD) also seem to have been important.

The results of the positive selection analyses greatly reflected those of *D. nivalis* in Paper I, and included many links to the five categories of Arctic stresses (summarized in section 3.3). However, we identified many more PSGs in Paper III than in Paper I (i.e. 1307 PSGs in Paper III compared to 360 PSGs in Paper I), which could be due to 1) the Paper III dataset only being based on genome assemblies with more complete gene sets, 2) the dataset in Paper I including several species of *Draba*, and 3) the dataset in Paper I containing two individuals of *D. nivalis* from different geographic locations. Point 2) and 3) should make the results from Paper I more

conservative/accurate, while point 1) should make the results from Paper III more complete.

The in depth examination of the *D. nivalis* genome highlights how adaptation to the Arctic might have been facilitated through different evolutionary events like, for instance, gene family expansions partly mediated by transposable element activity, as well as positive selection. The *D. nivalis* genome assembly may become an important tool for future studies of plant adaptation to the Arctic.

4. Concluding remarks and future perspectives

In Papers I-III, we used what can be considered a top-down approach, which provided us with a broad overview of putative molecular adaptations and convergence levels in Arctic Brassicaceae. This served well for a first time study of the molecular basis of Arctic plant adaptation. However, our approach limited us from examining specific traits and genes in more detail. In addition, candidate genes for Arctic adaptation were largely based on information about gene function acquired from other species, and this could be misleading if gene function has diverged in Arctic lineages. Future studies should therefore include detailed assessments of phenotypic convergence in specific traits, as well as studies of the exact functions of Arctic candidate genes.

I also note that our approach in Paper I and Paper III is largely based on patterns of positive selection, and that selection is not synonymous with adaptation (e.g. Orr 2005). As pointed out by Orr (2005) in his review on the genetic theory of adaptation, Fisher (1930) emphasized that "Adaptation is characterized by the movement of a population towards a phenotype that best fits the present environment". Our proposed molecular adaptations should therefore be treated as hypotheses for experimental verification as suggested above. However, we also expect our findings to be good indicators for which traits and genes are important for Arctic plant life, as adaptation is thought to happen mainly through natural selection (some even define adaptations simply as traits that have been shaped by natural selection; Futuyma 2005). The findings in Paper I-III were also largely consistent, making them more credible.

Fisher's point also reminds us that adaptation is a process that takes place in natural populations. Some might therefore argue that adaptation is best studied through population genomic approaches. Still, an advantage with our methodology is

that we easily could include a broad taxonomic sampling, and compare patterns of adaptation across multiple clades in the Brassicaceae family. This is especially important in studies of evolutionary repeatability where inappropriate taxonomic sampling can lead to overestimation of molecular convergence (Thomas et al. 2017). In addition, our approach enabled us to investigate broader patterns of genome evolution in Paper III. Future studies could include population genomic investigations of adaptation and convergence, to explore if there are important local differences within the Arctic.

Acknowledgements

I thank A. Lovisa S. Gustafsson, Anne K. Brysting, Tanja Slotte, Helene Lo Cascio Sætre, and Tim Dunker for valuable comments on this PhD synthesis.

References

Abbott RJ, Brochmann C. 2003. History and evolution of the arctic flora: in the footsteps of Eric Hulten. *Molecular Ecology*. 12(2):299–313.

Alexa A, Rahnenfuhrer J. 2018. topGO: enrichment analysis for Gene Ontology. http://bioconductor.org/packages/release/bioc/html/topGO.html.

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 22(13):1600–1607.

Alsos IG, Eidesen PB, Ehrich D, Skrede I, Westergaard K, Jacobsen GH, Landvik JY, Taberlet P, Brochmann C. 2007. Frequent long-distance plant colonization in the changing Arctic. *Science*. 316(5831):1606–1609.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*. 181(1):1–20.

Arends D, Prins P, Jansen RC, Broman KW. 2010. R/qtl: high-throughput multiple QTL mapping. *Bioinformatics*. 26(23):2990–2992.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology and Evolution*. 23(1):26–32.

Barrett R, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*. 23(1):38–44.

Bartels D, Souer E. 2004. Molecular responses of higher plants to dehydration. In: Hirt H, Shinozaki K, editors. Plant responses to abiotic stress. Springer. p. 9–38.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and

Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 57(1):289–300.
Billings WD, Mooney HA. 1968. The Ecology of Arctic and Alpine Plants. *Biological Reviews*. 43(4):481–529.
Birks HH. 2008. The late-quaternary history of arctic and alpine plants. *Plant Ecology & Diversity*. 1(2):135–146.
Bjerke JW, Elverland E, Jaakola L, Lund L, Bochenek Z, Kłos A, Tømmervik H. 2018. High tolerance of a high-arctic willow and graminoid to simulated ice encasement. *Boreal Environment Research*. 23:329–338.
Bliss LC. 1962. Adaptations of Arctic and Alpine Plants to Environmental Conditions. *Arctic*. 15(2).
Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*. 15:211.
Bohutínská M, Vlček J, Yair S, Laenen B, Konečná V, Fracassetti M, Slotte T, Kolář F. Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *bioRxiv*. http://dx.doi.org/10.1101/2020.03.24.005397.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–2120.
Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. 2018. (Non)Parallel Evolution. *Annual Review of Ecology, Evolution, and Systematics*. 49(1):303–330.
Bomblies K, Higgins JD, Yant L. 2015. Meiosis evolves: adaptation to external and internal environments. *New Phytologist*. 208(2):306–323.
Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 34(5):525–527.
Brochmann C, Brysting AK. 2008. The Arctic – an evolutionary freezer? *Plant Ecology & Diversity*. 1(2):181–195.
Brochmann C, Soltis PS, Soltis DE. 1992. Multiple origins of the octoploid Scandinavian endemic *Draba cacuminum*: electrophoretic and morphological evidence. Nordic Journal of Botany. 12(3):257–272.
Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 19(7):889–890.
CAFF. 2013. Arctic Biodiversity Assessment: Status and Trends in Arctic Biodiversity. Synthesis. Conservation of Arctic Flora and Fauna (CAFF), Arctic Council.
Carlsen T, Bleeker W, Hurka H, Elven R, Brochmann C. 2009. Biogeography and Phylogeny of *Cardamine* (Brassicaceae). *Annals of the Missouri Botanical Garden*. 96(2):215–236.
Carlson M. 2018. org.At.tair.db: Genome wide annotation for *Arabidopsis*. http://bioconductor.org/packages/release/data/annotation/html/org.At.tair.db.html.
Castoe TA, de Koning APJ, Kim H-M, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 106(22):8986–8991.

Chapin FS. 1983. Direct and indirect effects of temperature on arctic plants. *Polar Biology*. 2(1):47–52.
Chory J, Chatterjee M, Cook RK, Elich T, Fankhauser C, Li J, Nagpal P, Neff M, Pepper A, Poole D, et al. 1996. From seed germination to flowering, light controls plant development via the pigment phytochrome. *Proceedings of the National Academy of Sciences of the United States of America*. 93(22):12066–12071.
Choudhury RR, Neuhaus J-M, Parisod C. 2017. Resolving fine-grained dynamics of retrotransposons: comparative analysis of inferential methods and genomic resources. *Plant Journal*. 90(5):979–993.
Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 21(18):3674–3676.
Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences*. 279(1749):5039–5047.
Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 33(18):2938–2940.
Corbett-Detig RB, Russell SL, Nielsen R, Losos J. 2020. Phenotypic Convergence Is Not Mirrored at the Protein Level in a Lizard Adaptive Radiation. *Molecular Biology and Evolution*. 37(6):1604–1614.
Cuevas-Velazquez CL, Rendón-Luna DF, Covarrubias AA. 2014. Dissecting the cryoprotection mechanisms for dehydrins. *Frontiers in Plant Science*. 5:583.
Darwin C. 1859. On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life. J. Murray.
Davies KTJ, Bennett NC, Faulkes CG, Rossiter SJ. 2018. Limited Evidence for Parallel Molecular Adaptations Associated with the Subterranean Niche in Mammals: A Comparative Study of Three Superorders. *Molecular Biology and Evolution*. 35(10):2544–2559.
Dobzhansky T. 1970. Genetics of the Evolutionary Process. Columbia University Press.
Dong MA, Farré EM, Thomashow MF. 2011. Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*. 108(17):7241–7246.
El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research*. 47(D1):D427–D432.
Ellison AM, Gotelli NJ. 2009. Energetics and the evolution of carnivorous plants—Darwin's "most wonderful plants in the world." *Journal of Experimental Botany*. 60(1):19–42.
Elvebakk A. 1999. Bioclimatic delimitation and subdivision of the Arctic. In: Nordal I, Razzhivin VY, editors. The species concept in the high north: a panarctic flora initiative. The Norwegian Academy of Science and Letters. p. 81–112.

Elven R, Murray DF, Razzhivin VY, Yurtsev BA. 2011. Annotated checklist of the Panarctic Flora (PAF). Vascular plants. [accessed 17 of July 2020]. http://panarcticflora.org/.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*. 16:157.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*. 20(1):238.
Fisher R. 1930. The Genetical theory of natural selection. Clarendon Pr.
Fourrier N, Bédard J, Lopez-Juez E, Barbrook A, Bowyer J, Jarvis P, Warren G, Thorlby G. 2008. A role for *SENSITIVE TO FREEZING 2* in protecting chloroplasts against freeze-induced damage in *Arabidopsis*. *The Plant Journal*. 55(5):734–745.
Fowler SG, Cook D, Thomashow MF. 2005. Low temperature induction of *Arabidopsis CBF1*, *2*, and *3* is gated by the circadian clock. *Plant Physiology*. 137(3):961–968.
Futuyma DJ. 2005. Evolution. Sinauer Associates.
Gendall AR, Levy YY, Wilson A, Dean C. 2001. The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. Cell. 107(4):525–535.
Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*. 5(10):R80.
Gilleland E, Katz RW. 2016. extRemes 2.0: an extreme value analysis package in R. *Journal of Statistical Software*. 72(8):1–39.
Gould SJ. 1989. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton & Company.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. 29(7):644–652.
Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 10(3):645–656.
Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D, Ma T, Hu Q, Al-Shehbaz IA, et al. 2017. Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics*. 18(1):176.
Guy C, Haskell D, Li Q-B. 1998. Association of Proteins with the Stress 70 Molecular Chaperones at Low Temperature: Evidence for the Existence of Cold Labile Proteins in Spinach. *Cryobiology*. 36(4):301–314.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference

generation and analysis. *Nature Protocols*. 8(8):1494–1512.
Haldane J. 1932. The causes of evolution. Longmans, Green and co.
Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Molecular Biology and Evolution*. 30(8):1987–1997.
Hart AJ, Ginzburg S, Xu MS, Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn JL. 2020. EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Molecular Ecology Resources*. 20(2):591–604.
Heide OM. 2005. Ecotypic Variation among European Arctic and Alpine Populations of *Oxyria digyna*. *Arctic, Antarctic, and Alpine Research*. 37(2):233–238.
Heino P, Palva ET. 2004. Signal transduction in plant cold acclimation. In: Hirt H, Shinozaki K, editors. Plant Responses to Abiotic Stress. Springer. p. 151–186.
He L, Nada K, Tachibana S. 2002. Effects of Spermidine Pretreatment through the Roots on Growth and Photosynthesis of Chilled Cucumber Plants (*Cucumis sativus* L.). *Journal of the Japanese Society for Horticultural Science*. 71(4):490–498.
Hendry AP. 2013. Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity* . 111(6):456–466.
Heyduk K, Moreno-Villena JJ, Gilman IS, Christin P-A, Edwards EJ. 2019. The genetics of convergent evolution: insights from plant photosynthesis. *Nature Reviews Genetics*. 20(8):485–493.
Hoffmann MH, von Hagen KB, Hörandl E, Röser M, Tkach NV. 2010. Sources of the Arctic flora: Origins of Arctic species in *Ranunculus* and related genera. *International Journal of Plant Sciences*. 171(1):90–106.
Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 12(1).
Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, et al. 2016. Resolution of Brassicaceae Phylogeny Using Nuclear Genes Uncovers Nested Radiations and Supports Convergent Morphological Evolution. *Molecular Biology and Evolution*. 33(2):394–412.
Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 4(1):44–57.
Hu K, Xu K, Wen J, Yi B, Shen J, Ma C, Fu T, Ouyang Y, Tu J. 2019. Helitron distribution in Brassicaceae and whole Genome Helitron density as a character for distinguishing plant species. *BMC Bioinformatics*. 20(1):354.
Hultén E. 1968. Flora of Alaska and neighboring territories : a manual of the vascular plants. Stanford University Press.
Hundertmark M, Hincha DK. 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*. 9:118.
Inzé D, Van Montagu M. 1995. Oxidative stress in plants. *Current Opinion in Biotechnology*. 6(2):153–158.

Jackson SD. 2009. Plant responses to photoperiod. *New Phytologist*. 181(3):517–531.
Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF. 2001. Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiology*. 127(3):910–917.
Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 30(9):1236–1240.
Jónsdóttir IS. 2011. Diversity of plant life histories in the Arctic. *Preslia*. 83:281–300.
Jordan G, Goldman N. 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Molecular Biology and Evolution*. 29(4):1125–1139.
Jordon-Thaden I, Hase I, Al-Shehbaz I, Koch MA. 2010. Molecular phylogeny and systematics of the genus *Draba* (Brassicaceae) and identification of its most closely related genera. *Molecular Phylogenetics and Evolution*. 55(2):524–540.
Kevan PG. 1972. Heliotropism in some Arctic flowers. *Canadian Field Naturalist*. 86:41–44.
Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. 2007. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant Journal*. 50(2):347–363.
Knight MR, Knight H. 2012. Low-temperature perception leading to gene expression and cold tolerance in higher plants. *New Phytologist*. 195(4):737–751.
Koch MA. 2012. Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). *Taxon*. 61(1):76–92.
Kolde R. 2019. pheatmap: Pretty Heatmaps. https://CRAN.R-project.org/package=pheatmap.
Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics*. 5(59):1-9.
Körner C. 2003. Alpine plant life. Functional plant ecology of high mountain ecosystems. Springer.
Körner C, Alsos IG. 2009. Freezing resistance in high arctic plant species of Svalbard in mid-summer. *Bauhinia*. 21:25-32
Kreps JA, Wu Y, Chang H-S, Zhu T, Wang X, Harper JF. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiology*. 130(4):2129–2141.
Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biology*. 5(2):R12.
Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 22(13):1658–1659.
Li X, Ma D, Lu SX, Hu X, Huang R, Liang T, Xu T, Tobin EM, Liu H. 2016. Blue Light- and Low Temperature-Regulated *COR27* and *COR28* Play Roles in the

Arabidopsis Circadian Clock. Plant Cell. 28(11):2755–2769.
Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution*. 65(7):1827–1840.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15(550):1-9.
Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*. 102(30):10557–10562.
Lütz C. 2010. Cell physiology of plants growing in cold environments. *Protoplasma*. 244(1-4):53–73.
Lysak MA, Mandáková T, Eric Schranz M. 2016. Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Current Opinion in Plant Biology*. 30:108–115.
Mahler DL, Ingram T, Revell LJ, Losos JB. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*. 341(6143):292–295.
Mandáková T, Hloušková P, Koch MA, Lysak MA. 2020. Genome Evolution in Arabideae Was Marked by Frequent Centromere Repositioning. *Plant Cell*. 32(3):650–665.
Mandáková T, Lysak MA. 2016a. Chromosome Preparation for Cytogenetic Analyses in *Arabidopsis*. *Current Protocols in Plant Biology*. 1(1):43–51.
Mandáková T, Lysak MA. 2016b. Painting of *Arabidopsis* Chromosomes with Chromosome-Specific BAC Clones. *Current Protocols in Plant Biology*. 1(2):359–371.
Matthews JV. 1979. Tertiary and Quarternary environments: Historical background for an analysis of the Canadian insect fauna. *Memoirs of the Entomological Society of Canada*. 111(S108):31–86.
Matthews JV, Ovenden LE. 1990. Late Tertiary Plant Macrofossils from Localities in Arctic/Subarctic North America: A Review of the Data. *Arctic*. 43(4):364–392.
Mooney HA, Billings WD. 1961. Comparative Physiological Ecology of Arctic and Alpine Populations of *Oxyria digyna*. Ecological Monographs. 31(1):1–29.
Moran K, Backman J, Brinkhuis H, Clemens SC, Cronin T, Dickens GR, Eynaud F, Gattacceca J, Jakobsson M, Jordan RW, et al. 2006. The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature*. 441(7093):601–605.
Nilsen J. 1985. Light climate in northern areas. In: Kaurin Å, Junttila O, Nilsen J, editors. Plant production in the north: Proceedings from "Plant adaptation workshop" Tromsø, Norway. September 4-9, 1983. Norwegian University Press. p. 62–82.
Orr HA. 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*. 6(2):119–127.
Oxford Dictionary of Biology. 2004. Fifth edition. Oxford University Press.
Parker J, Tsagkogeorga G, Cotton JA, Liu Y, Provero P, Stupka E, Rossiter SJ. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*. 502(7470):228–231.

Park S, Lee C-M, Doherty CJ, Gilmour SJ, Kim Y, Thomashow MF. 2015. Regulation of the Arabidopsis CBF regulon by a complex low-temperature regulatory network. *Plant Journal*. 82(2):193–207.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*. 14(4):417–419.
Piskurewicz U, Turecková V, Lacombe E, Lopez-Molina L. 2009. Far-red light inhibits germination through DELLA-dependent stimulation of ABA synthesis and *ABI3* activity. *EMBO Journal*. 28(15):2259–2271.
Preite V, Sailer C, Syllwasschy L, Bray S, Ahmadi H, Krämer U, Yant L. 2019. Convergent evolution in Arabidopsis halleri and Arabidopsis arenosa on calamine metalliferous soils. *Philosophical Transactions of the Royal Society B*. 374(1777):20180243.
Privman E, Penn O, Pupko T. 2012. Improving the performance of positive selection inference by filtering unreliable alignment regions. *Molecular Biology and Evolution*. 29(1):1–5.
Qian C, Bryans N, Kruykov I, de Koning APJ. 2018. Visualization and analysis of statistical signatures of convergent molecular evolution. [accessed 17 of July 2020]. http://lab.jasondk.io.
Qian C, de Koning APJ. 2018. Rapid Discovery of Convergent Molecular Evolution Across Entire Phylogenies. [accessed 17 of July 2020]. http://lab.jasondk.io.
Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biology*. 20(38):1-23
Rinaldi MA, Liu J, Enders TA, Bartel B, Strader LC. 2012. A gain-of-function mutation in *IAA16* confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. *Plant Molecular Biology*. 79(4-5):359–373.
Ruelland E, Vaultier M-N, Zachowski A, Hurry V. 2009. Cold Signalling and Cold Acclimation in Plants. In: Advances in Botanical Research. Vol. 49. p. 35–150.
Sackton TB, Clark N. 2019. Convergent evolution in the genomics era: new insights and directions. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 374(1777):20190102.
Salisbury FB. 1985. Plant adaptations to the light environment. In: Kaurin Å, Junttila O, Nilsen J, editors. Plant production in the north: Proceedings from “Plant adaptation workshop” Tromsø, Norway. September 4-9, 1983. Norwegian University Press. p. 43–61.
Sanger TJ, Revell LJ, Gibson-Brown JJ, Losos JB. 2012. Repeated modification of early limb morphogenesis programmes underlies the convergence of relative limb length in Anolis lizards. *Proceedings of the Royal Society B: Biological Sciences*. 279(1729):739–748.
Savile DBO. 1972. Arctic adaptations in plants. Plant Research Institute, Ottawa Report No.: Monograph No. 6.
Schluter D, Clifford EA, Nemethy M, McKinnon JS. 2004. Parallel evolution and inheritance of quantitative traits. *American Naturalist*. 163(6):809–822.

Schneeweiss GM, Schönswetter P. 2011. A re-appraisal of nunatak survival in arctic-alpine phylogeography. *Molecular ecology*. 20(2):190–192.
Schoener TW. 1970. Size Patterns in West Indian Anolis Lizards. II. Correlations with the Sizes of Particular Sympatric Species-Displacement and Convergence. *The American Naturalist*. 104(936):155–174.
Schranz M, Lysak M, Mitchellolds T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends in Plant Science*. 11(11):535–542.
Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*. 43(W1):W7–14.
Shah FA, Wei X, Wang Q, Liu W, Wang D, Yao Y, Hu H, Chen X, Huang S, Hou J, et al. 2020. Karrikin Improves Osmotic and Salt Stress Tolerance the Regulation of the Redox Homeostasis in the Oil Plant. *Frontiers in Plant Science*. 11:216:1-14
Shi Y, Ding Y, Yang S. 2018. Molecular Regulation of CBF Signaling in Cold Acclimation. *Trends in Plant Science*. 23(7):623–637.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31(19):3210–3212.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313.
Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 24(5):637–644.
Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Research*. 37(21):7002–7013.
Steponkus PL. 1984. Role of the Plasma Membrane in Freezing Injury and Cold Acclimation. *Annual Review of Plant Physiology*. 35(1):543–584.
Stern DL. 2013. The genetic causes of convergent evolution. Nature reviews Genetics. 14(11):751–764.
TAIR. 2020. The Arabidopsis Information Resource. [accessed 17 of July 2020]. https://www.arabidopsis.org/.
Taylor J, Butler D. 2017. R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis. *Journal of Statistical Software*. 79(6).
Teige M, Scheikl E, Eulgem T, Dóczi R, Ichimura K, Shinozaki K, Dangl JL, Hirt H. 2004. The *MKK2* pathway mediates cold and salt stress signaling in *Arabidopsis*. *Molecular Cell*. 15(1):141–152.
Thomas GWC, Hahn MW. 2015. Determining the Null Model for Detecting Adaptive Convergence from Genomic Data: A Case Study using Echolocating Mammals. *Molecular Biology and Evolution*. 32(5):1232–1236.
Thomas GWC, Hahn MW, Hahn Y. 2017. The Effects of Increasing the Number of Taxa on Inferences of Molecular Convergence. *Genome Biology and Evolution*.

9(1):213–221.
Thomashow MF. 1999. Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. *Annual review of plant physiology and plant molecular biology*. 50:571–599.
Thomashow MF. 2010. Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. *Plant Physiology*. 154(2):571–577.
Tigerstedt P. 1994. Adaptation, variation and selection in marginal areas. *Euphytica*. 77(3):171-174
Uemura M, Steponkus PL. 1999. Cold Acclimation in Plants: Relationship Between the Lipid Composition and the Cryostability of the Plasma Membrane. *Journal of Plant Research*. 112(2):245–254.
Wang M, Zhao Y, Zhang B. 2015. Efficient Test and Visualization of Multi-Set Intersections. *Scientific Reports*. 5(16923):1-12
Wang W, Vinocur B, Shoseyov O, Altman A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*. 9(5):13–15.
Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*. 40(7):e49–e49.
Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, Birol I. 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience*. 4(35):1-11
Wehrmeister RR, Bonde EK. 1977. Comparative Aspects of Growth and Reproductive Biology in Arctic and Alpine Populations of *Saxifraga cernua* L. *Arctic and Alpine Research*. 9(4):401-406.
Westergaard KB, Alsos IG, Popp M, Engelskjøn T, Flatberg KI, Brochmann C. 2011. Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. *Molecular Ecology*. 20(2):376–393.
Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*. 19(12):639–644.
Wilkinson S, Clephan AL, Davies WJ. 2001. Rapid Low Temperature-Induced Stomatal Closure Occurs in Cold-Tolerant *Commelina communis* Leaves But Not in Cold-Sensitive Tobacco Leaves, via a Mechanism That Involves Apoplastic Calcium But Not Abscisic Acid. *Plant Physiology*. 126(4):1566–1578.
Willing E-M, Rawat V, Mandáková T, Maumus F, James GV, Nordström KJV, Becker C, Warthmann N, Chica C, Szarzynska B, et al. 2015. Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants*. 1(14023):1-7
Wisniewski M, Fuller M. 1999. Ice nucleation and deep supercooling in plants: new insights using infrared thermography. In: Margesin R, Schinner F, editors. Cold-Adapted Organisms. Springer. p. 105–118.
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.
Wong WSW, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*. 168(2):1041–1051.
Wullschleger SD, Breen AL, Iversen CM, Olson MS, Näsholm T, Ganeteg U, Wallenstein MD, Weston DJ. 2015. Genomics in a changing arctic: critical questions await the molecular ecologist. *Molecular Ecology*. 24(10):2301–2309.
Xiong W, He L, Lai J, Dooner HK, Du C. 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences*. 111(28):10263–10268.
Xu S, He Z, Guo Z, Zhang Z, Wyckoff GJ, Greenberg A, Wu C-I, Shi S. 2017. Genome-Wide Convergence during Evolution of Mangroves from Woody Plants. *Molecular Biology and Evolution*. 34(4):1008–1015.
Yang Y, Smith SA. 2013. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics*. 14:328.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in BioSciences*. 13(5):555–556.
Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science*. 353(6306):1431–1433.
de Zelicourt A, Colcombet J, Hirt H. 2016. The Role of MAPK Modules and ABA during Abiotic Stress Signaling. *Trends in Plant Science*. 21(8):677–685.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*. 22(12):2472–2479.
Zhu J-K. 2016. Abiotic Stress Signaling and Responses in Plants. *Cell*. 167(2):313–324.
Zou Z, Zhang J. 2015. No genome-wide protein sequence convergence for echolocation. *Molecular Biology and Evolution*. 32(5):1237–1241.

I

Multiple Genetic Trajectories to Extreme Abiotic Stress Adaptation in Arctic Brassicaceae

Siri Birkeland,*[1] A. Lovisa S. Gustafsson,[1] Anne K. Brysting,[2] Christian Brochmann,[1] and Michael D. Nowak[1]

[1]Natural History Museum, University of Oslo, Oslo, Norway

[2]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway

***Corresponding author:** E-mail: siri.birkeland@nhm.uio.no.

Associate editor: Michael Purugganan

Abstract

Extreme environments offer powerful opportunities to study how different organisms have adapted to similar selection pressures at the molecular level. Arctic plants have adapted to some of the coldest and driest biomes on Earth and typically possess suites of similar morphological and physiological adaptations to extremes in light and temperature. Here, we compare patterns of molecular evolution in three Brassicaceae species that have independently colonized the Arctic and present some of the first genetic evidence for plant adaptations to the Arctic environment. By testing for positive selection and identifying convergent substitutions in orthologous gene alignments for a total of 15 Brassicaceae species, we find that positive selection has been acting on different genes, but similar functional pathways in the three Arctic lineages. The positively selected gene sets identified in the three Arctic species showed convergent functional profiles associated with extreme abiotic stress characteristic of the Arctic. However, there was little evidence for independently fixed mutations at the same sites and for positive selection acting on the same genes. The three species appear to have evolved similar suites of adaptations by modifying different components in similar stress response pathways, implying that there could be many genetic trajectories for adaptation to the Arctic environment. By identifying candidate genes and functional pathways potentially involved in Arctic adaptation, our results provide a framework for future studies aimed at testing for the existence of a functional syndrome of Arctic adaptation in the Brassicaceae and perhaps flowering plants in general.

Key words: **molecular convergence, adaptation, positive selection, extremophiles.**

Article

Introduction

Strong selection pressures imposed by extreme environments and specialized niches can sometimes lead to independent evolution of similar morphological and/or physiological traits in different lineages (i.e., convergent evolution; Sackton and Clark 2019). Some striking examples are the evolution of antifreeze glycoproteins in Arctic cod and Antarctic notothenioid fishes (Chen et al. 1997), increased hemoglobin-oxygen affinity in high-altitude dwelling birds (Natarajan et al. 2016), and photosynthetic pathways that minimize photorespiration in hot and dry environments (Heyduk et al. 2019). Organisms adapted to extreme environments are, therefore, especially interesting for studying one of the fundamental questions in evolutionary biology: To what degree does adaptation follow predictable genetic trajectories? Convergent traits do not necessarily result from similar genetic changes, and the prevalence of molecular convergence is still heavily debated (Losos 2011; Conte et al. 2012; Hendry 2013; Bolnick et al. 2018; Sackton and Clark 2019). Increasingly, molecular convergence is viewed as a quantitative continuum rather than a presence–absence pattern (Bolnick et al. 2018). It may manifest at different levels of similarity, that is, 1) independently fixed nonsynonymous mutations at the same site, 2) positive selection acting on the same genes, or 3) positive selection acting on different genes that are part of the same functional pathway (Sackton and Clark 2019). By disentangling the impacts of these different levels of molecular convergence, we gain a more comprehensive perspective on the genomic landscape of adaptation.

In this study, we aim to identify sets of candidate genes for Arctic adaptations and search for signatures of codon, gene, and pathway convergence in plants that have adapted to live in some of the harshest terrestrial environments on Earth. The Arctic is characterized by very low year-round temperatures (average temperature of the warmest month $\leq 10\,^{\circ}C$) as well as major annual and minor diurnal fluctuations in light regime (Billings and Mooney 1968; Nilsen 1985; Elvebakk 1999). Arctic plants must therefore withstand a number of stresses directly or indirectly linked to highly fluctuating light and temperature conditions. Although there is considerable climatic variation within the Arctic, the combination of these abiotic stresses is overall more extreme relative to lower latitude environments from which most Arctic lineages are derived (see, e.g., Abbott and Brochmann [2003] on the origin of the Arctic flora). Here, we classify major Arctic stresses into five categories that we will focus on throughout the study: 1) protein misfolding and membrane stress, 2) physiological cold

stress, 3) ice and snow stress, 4) light stress, and 5) oxidative stress. Low temperatures can affect nearly all aspects of plant cell biochemistry and metabolism, including the physical properties of cell membranes and proteins (categories 1 and 2; Heino and Palva 2004). Snowfall and ice formation can occur year-round, and Arctic soils are characterized by permafrost, with all soil 20–100 cm below the surface permanently frozen (Billings and Mooney 1968). The effects of such ice and snow stress (category 3) can manifest in many ways. It may, for example, cause mechanical stress and cellular rupture, and ice formation in extracellular spaces will draw water out of cells and increase solute concentration in the cytoplasm, leading to cellular dehydration stress (Wisniewski and Fuller 1999; Körner 2003; Wisniewski et al. 2004). The Arctic light environment is enriched in far-red and blue light, and at the highest latitudes in the Arctic, the period of midnight sun lasts the entire growing season (Salisbury 1985). The actual period available for reproduction, growth, and development can also be extremely short (e.g., ~1 month in Arctic polar deserts; Jónsdóttir 2005). The Arctic *light stress* (category 4) might therefore be especially challenging for plants, as photoperiod and spectral composition are important cues for, for example, development, reproduction, and the onset of dormancy (Chory et al. 1996; Jackson 2009; Piskurewicz et al. 2009). Finally, the combined effects of high light intensity and low temperature can lead to increased *oxidative stress* (category 5; Heino and Palva 2004; Lütz 2010). This comes in addition to the increased accumulation of reactive oxygen species (ROS) within the cells due to many of the above-mentioned abiotic stresses (accumulation of ROS is a common stress response in plants; Kilian et al. 2007). These five categories of stresses will serve as a framework for discussing molecular convergence in plant adaptation to the Arctic environment.

Arctic plants exhibit a range of similar morphological and physiological adaptations to cope with these five stress categories (Chapin 1983; Lütz 2010; Jónsdóttir 2011). Many Arctic plants are small, herbaceous perennials that cling to the ground, often forming cushions or tussocks that minimize exposure to abrasive winds and low temperatures (Billings and Mooney 1968). The leaves are often tough and leathery, and/or hairy (Hultén 1968; Alsos et al. 2019), providing protection from desiccation, mechanical stress from ice and snow, and strong winds. Arctic plants have adapted their metabolism, growth, and reproduction to the low temperatures, and have for instance optimum photosynthesis rates at lower temperatures than other plants (Chapin 1983). It is also likely that the Arctic photoperiod and spectral composition (such as the enrichment of far-red light and blue light) play important roles in plant growth and timing of development (Nilsen 1985; Salisbury 1985). It has been shown that some Arctic plants require a long critical day length for induction of flowering (Wehrmeister and Bonde 1977; Heide 2005) and that changes in the red to far-red light ratio might be important for the onset of dormancy when there still is midnight sun in late summer (Nilsen 1985). We can also hypothesize that the plasma membrane plays an important role in plant adaptation to Arctic cold and freezing stress (cf., Thomashow 1999; Knight and Knight 2012). Many temperate plants will, for example, acclimate to low temperatures by changing the lipid composition of the plasma membrane to avoid membrane rigidification and rupture during cold stress (Thomashow 1999; Wisniewski and Fuller 1999; Körner 2003). Response to freezing stress involves many of the same molecular components as response to dehydration and high salinity (Ryu et al. 1995; Bartels and Souer 2004; Verslues et al. 2006), and we could therefore also expect Arctic plants to be resistant against these stresses. Detailed studies of genes involved in plant cold response pathways have been conducted in temperate plants such as *Arabidopsis thaliana* (Jaglo-Ottosen et al. 1998; Kreps et al. 2002; Kilian et al. 2007; Thomashow 2010), but little is known for Arctic plants. By comparing molecular evolution in plant lineages independently adapted to the Arctic, we can learn more about the general traits and genes that are important for plant survival at high latitudes.

The focus of our study is three plant species in the Brassicaceae that have their main distribution above the Arctic Circle. The Brassicaceae is an optimal study system with its many publicly available molecular resources (TAIR 2019), and the family comprises several lineages that have independently colonized the Arctic (Carlsen et al. 2009; Jordon-Thaden et al. 2010; Koch 2012). The three selected species are *Cardamine bellidifolia*, *Cochlearia groenlandica*, and *Draba nivalis*, which all are diploid and represent three divergent clades that shared a common ancestor ~30 Ma (Elven et al. 2011; Huang et al. 2016; Guo et al. 2017; fig. 1). Their distributions cover the entire Arctic (even the coldest bioclimatic subzone called Arctic polar desert; Elvebakk 1999; Elven et al. 2011), and they are therefore suitable candidates for studying adaptations to stresses associated with limited heat budget and low sun angle. We generated new transcriptome data for each species and conducted comparative analyses with previously published transcriptome/genome data from several temperate relatives. To identify candidate genes for Arctic adaptations, we performed genome-wide tests of positive selection and searched for positively selected genes (PSGs) associated with the five categories of Arctic stresses outlined above. To identify patterns of molecular convergence, we identified convergent amino acid substitutions and compared PSG sets in terms of orthology and functional annotations. This enabled us to estimate molecular convergence among Arctic lineages at the level of codons (the number of genes with convergent substitutions), genes (the number of orthologous genes under positive selection), and functional pathways (the degree of similarity between functional profiles of the PSGs). However, when estimating convergent substitutions, it has proven important to account for random accumulation of convergent substitutions with increasing divergence (i.e., the expected background convergence; Thomas and Hahn 2015; Xu et al. 2017). By estimating the relationship between convergent and divergent substitutions using species pairs in a phylogeny, we therefore tested whether Arctic lineage pairs exhibit an excess of convergence relative to neutral expectations (Castoe et al. 2009; Thomas and Hahn 2015). The aims of this study are to

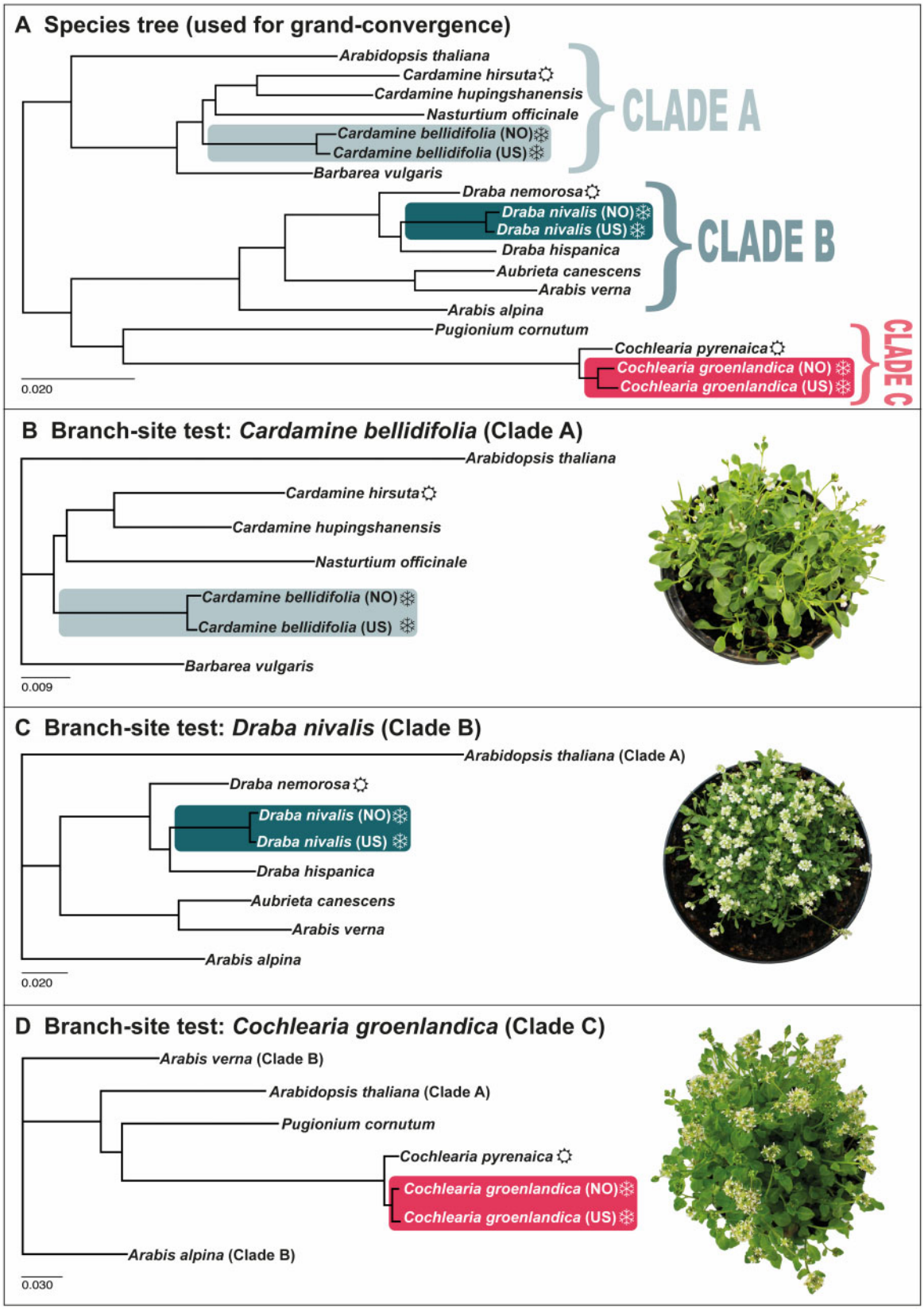


FIG. 1. Species trees used in the estimation of convergent and divergent substitutions (*A*), and in the branch-site tests of positive selection (*B*–*D*). Arctic focal species are indicated with a snow crystal, and non-Arctic comparisons are indicated with a sun. NO, Arctic genotype from Norway; US, Arctic genotype from the United States. Scale bars = substitutions/site. The topologies are in accordance with Huang et al. (2016).

Table 1. Number of PSGs Found in the Branch-Site Tests, as well as Corresponding topGO Results.

Species	Lineage Tested	Alignments Retained for the Branch-Site Test	PSGs $P < 0.05$	Enriched GO-terms[a]BP, CC, MF
Cardamine bellidifolia	Arctic	11,874	201	274 (159), 38 (25), 57 (50)
Cochlearia groenlandica	Arctic	13,306	159	198 (92), 28 (18), 34 (23)
Draba nivalis	Arctic	13,891	360	359 (256), 54 (36), 87 (73)
Cardamine hirsuta	Non-Arctic	11,874	395	268 (168), 105 (48), 72 (67)
Cochlearia pyrenaica	Non-Arctic	13,305	350	217 (168), 52 (49), 77 (69)
Draba nemorosa	Non-Arctic	13,881	380	298 (179), 22 (18), 67 (66)

[a]Numbers of significant GO-terms ($P < 0.05$) when applying the classic algorithm. Numbers of enriched GO-terms when applying the elim algorithm are given in brackets.

1) identify sets of candidate genes for Arctic adaptations in *C. bellidifolia*, *D. nivalis* and *Co. groenlandica*, 2) search for evidence of convergent molecular evolution across codons, genes, and pathways, and 3) test for a signature of convergent evolution that exceeds neutral expectations between Arctic species pairs. We show that there is little evidence for convergence at the codon and gene level, but that all positively selected and convergent gene sets were enriched for genes associated with typical Arctic stresses. Thus, our results suggest that there are many genetic trajectories to Arctic adaptation, perhaps due to the multifaceted nature of stress response pathways in plants.

Results

Final Data Set

We acquired genome-wide coding sequences from 15 species representing clades A, B, and C in the Brassicaceae (fig. 1, supplementary tables S1–S3, Supplementary Material online). These included the three selected Arctic species, for which we generated RNA sequences for this study. In addition to already published genomes, the data set consisted of de novo transcriptome assemblies yielding 18,830–34,226 transcripts per assembly and containing 67.0–89.8% universal single-copy orthologs (complete/fragmented BUSCOs; supplementary table S2, Supplementary Material online). Orthologous gene alignments were generated separately for the selection tests and for the estimation of convergent substitutions (supplementary fig. S1, Supplementary Material online) and were based on gene subsets from 8,572 (clade A), 9,941 (clade B), and 8,705 (clade C) orthogroups (~gene families) in the tests of positive selection and 6,325 orthogroups in the estimation of convergent substitutions.

Tests of Positive Selection

To detect genes that evolved under positive selection in the branches leading to the Arctic species, we ran branch-site tests separately for clades A, B, and C (fig. 1*B–D*; ~12–13,000 orthologous gene alignments/clade; supplementary fig. S1, Supplementary Material online). We identified 360 PSGs in *D. nivalis*, 201 in *C. bellidifolia*, and 159 in *Co. groenlandica* (table 1, supplementary tables S4–S6 and S7–S9, Supplementary Material online). To identify genes that might be under positive selection in both Arctic and non-Arctic lineages (e.g., immunity genes) and to evaluate whether gene functional annotations exclusively were associated with the Arctic, we performed complementary analyses in three non-Arctic reference species (fig. 1*B–D*), resulting in 380 PSGs in *D. nemorosa*, 395 in *C. hirsuta*, and 350 in *Co. pyrenaica* (table 1, supplementary tables S10–S12 and S13–S15, Supplementary Material online).

To identify candidate genes for Arctic adaptations, we manually reviewed the PSG annotations for documented associations with at least one of the five categories of Arctic stresses (i.e., gene ontology [GO] terms or publications indicating a mechanistic association with some physiological or biochemical aspect of the stress categories). We identified a total of 138 candidate genes spanning the five categories, of which 40 were found in *C. bellidifolia*, 31 in *Co. groenlandica*, and 67 in *D. nivalis* (supplementary table S16, Supplementary Material online). In *C. bellidifolia* and *D. nivalis*, there were many candidate genes associated with *light stress*, for example, blue and far-red light response (like *FAR1*-related sequences found in both PSG sets, and *ZTL* found in *D. nivalis*), whereas many candidate genes in *D. nivalis* and *Co. groenlandica* were associated with *physiological cold stress* and *ice and snow stress* (e.g., *COR15b* and *CSDP1* in *D. nivalis*, and *SFR2* and *RAV1* in *Co. groenlandica*). All Arctic PSG sets contained one or more late embryogenesis abundant (*LEA*) genes, putatively associated with *physiological cold stress* and *ice and snow stress*. Heat shock protein-related genes (e.g., heat shock chaperonin-binding or containing heat shock protein domains), which are putative candidates for adaptation to *protein misfolding and membrane stress*, were especially common in the PSG sets of *D. nivalis* and *C. bellidifolia*. Examples of candidate genes from the *oxidative stress* category were *CAT1* and *CAT2*, important for controlling ROS (both found in the *D. nivalis* PSG set). The three Arctic sets contained many MYB and MYC type transcription factors, involved in abiotic stress-activated signal transduction pathways. We also generated a heatmap showing the proportion of PSGs associated with the five categories, demonstrating that all Arctic PSG sets included considerably more genes associated with *protein misfolding and membrane stress* and *ice and snow stress* than the non-Arctic reference species (supplementary fig. S2, Supplementary Material online).

To assess whether positive selection is acting on the same genes in the three Arctic species, we compared the PSG sets with respect to orthology. We did not find any orthologous genes under positive selection in all three species. Thirteen PSGs were orthologous in two of the Arctic species (fig. 2) and four of these were identified as PSGs in at least one non-Arctic reference species (supplementary table S17, Supplementary Material online). Significance testing (the supertest; Wang et al. 2015) indicated that the PSG intersections observed

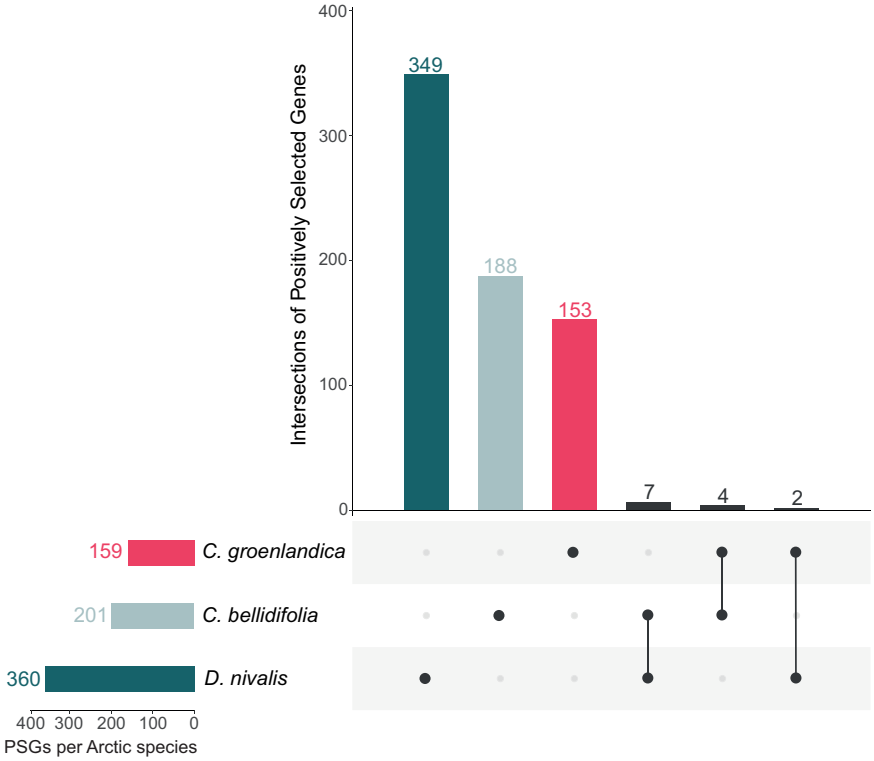


FIG. 2. UpSet plot of orthologous, PSGs identified in Arctic species. The plot in the left corner shows total number of PSGs, and the main plot shows the number of unique PSGs, followed by PSG intersections/overlaps between species (connected dots).

between any pair of Arctic species were not greater than expected by chance (supplementary table S18, Supplementary Material online). Among the 13 PSGs orthologous between Arctic species pairs, we found seven candidates possibly associated with three categories of Arctic stresses: *protein misfolding and membrane stress* (three genes involved in endoplasmic reticulum morphology or protein folding; *ERMO3*, DNAJ heat shock N-terminal domain-containing protein, and prefoldin 2), *light stress* (two genes involved in photoautotrophic growth or auxin-regulated plant growth and development; *IAA16* and *PAC*), and *oxidative stress* (*PCR2* and *OGOX1*; supplementary table S17, Supplementary Material online).

To assess whether positive selection has acted on different genes, but on similar functional pathways, we performed GO enrichment tests within the Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) domains for each of the three Arctic PSG sets (table 1, supplementary tables S19–S21 and fig. S3A–C, Supplementary Material online). We compared the sets of enriched GO-terms and used the results from the non-Arctic reference species to filter out terms unlikely to be associated with Arctic adaptation (supplementary tables S22–S24, Supplementary Material online). All Arctic PSG sets were enriched for genes associated with "response to stress" (BP, fig. 3A) and two sets for "response to abiotic stimulus" (BP, fig. 3B). The enriched "children" of these two terms (i.e., more specific terms nested under "response to stress" and "response to abiotic stimulus" in the GO hierarchy) varied somewhat among species (fig. 4), but included several potential associations with the five stress categories. All three Arctic PSG sets were significantly enriched for "response to salt stress," and the PSG sets of *D. nivalis* and *Co. groenlandica* were significantly enriched for "response to temperature stimulus," "response to cold," and "response to osmotic stress." *Draba nivalis* and *C. bellidifolia* shared an enrichment for "MAPK-cascade" and "response to oxidative stress." *Draba nivalis* stood out in being significantly enriched for a range of GO-terms associated with all five categories, including "response to light stimulus," "photoperiodism," and "response to misfolded protein." Notably, "plasma membrane" (CC domain) was significantly enriched in all Arctic PSG sets, but in none of the non-Arctic reference sets (supplementary fig. S4, Supplementary Material online), and the related BP term "endomembrane system organization" was also significantly enriched in *D. nivalis* and *C. bellidifolia*. Thus, we found evidence both for positive selection on genes associated with abiotic stress in all three Arctic species and for considerable convergence in the functional pathways in which these genes might be involved.

Convergent Amino Acid Substitutions

To estimate molecular convergence at the level of codons, we used Grand Convergence (Qian and de Koning 2018; Qian et al. 2018) to calculate posterior expected numbers of convergent and divergent substitutions for 7,765 gene alignments between pairs of Arctic and non-Arctic branches (fig. 1A). We identified 217 convergent genes between Arctic branch pairs (defined as containing at least one substitution with >0.8 posterior probability (PP) of convergence; table 2, supplementary tables S25–S27, Supplementary Material online). The

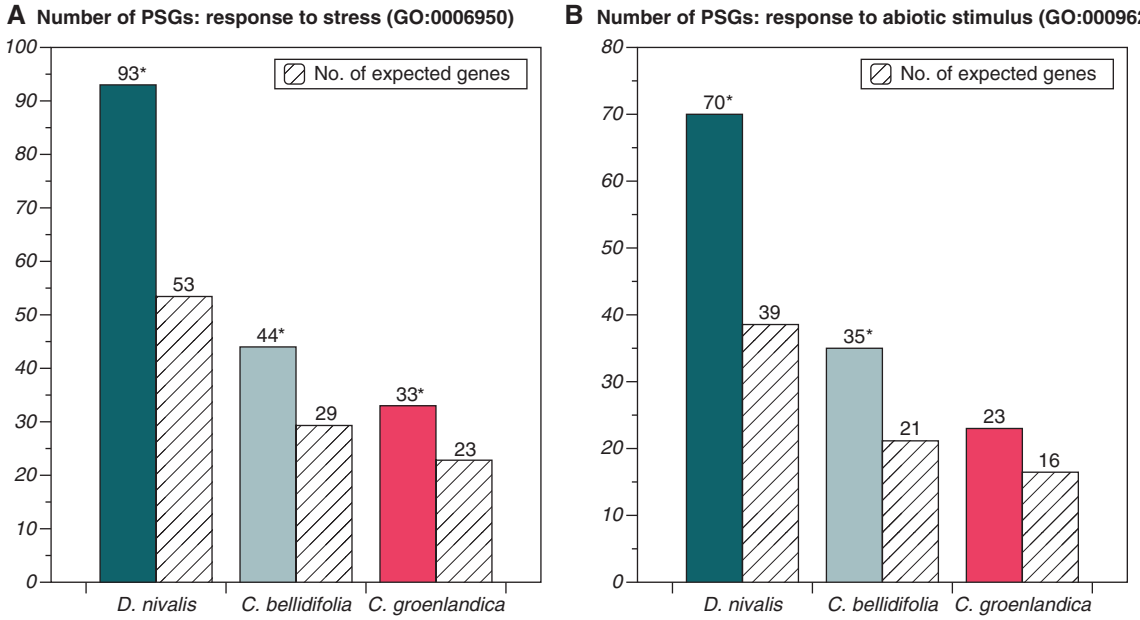


FIG. 3. Number of PSGs annotated with (*A*) "response to stress" and (*B*) "response to abiotic stimulus" in the Arctic species. Expected numbers of genes (if randomly drawn) are indicated as hatched bars. Stars (*) indicate that the GO-term was significantly enriched in the PSG set of this species when applying the classic algorithm in topGO ($P < 0.05$).

highest number of convergent genes (126) was observed between *D. nivalis* and *C. bellidifolia*, followed by *Co. groenlandica* and *C. bellidifolia* (58), and *D. nivalis* and *Co. groenlandica* (33, fig. 5, table 2). For comparison, a total of 362 convergent genes were identified between pairs of the three non-Arctic reference species (table 2, supplementary tables S28–S30, Supplementary Material online). Only a single gene was identified as convergent in all three Arctic pairwise comparisons (fig. 5): *EMB2742*, involved in, for example, embryo development ending in seed dormancy (supplementary table S31, Supplementary Material online). Three convergent genes were identified in both the *D. nivalis*–*C. bellidifolia* and the *C. bellidifolia*–*Co. groenlandica* branch pairs (*ORRM6*, *SK4*, and *MAPKKK14*), but only one of these had an apparent putative association with the five categories of stresses: *MAPKKK14*, belonging to a group of protein kinases involved in stress-induced signaling pathways, potentially involved in far-red light induced seedling deetiolation (supplementary table S31, Supplementary Material online).

To test for an excess of convergent substitutions in Arctic species pairs, we summed the probabilities for divergent and convergent substitutions over all alignments in 50 branch pairs from figure 1A, using one (arbitrarily chosen) alignment per orthogroup. These included the three Arctic branch pairs, the three non-Arctic reference branch pairs, and 44 random nonsister branch pairs from the phylogeny (supplementary table S32, Supplementary Material online). The number of convergent substitutions (y) could be predicted from the number of divergent substitutions (x) with the linear model $y = (-18.9642) + 0.5807x$ with $R^2 = 0.84$, indicating a strong linear relationship (fig. 6). Removing three outliers resulted in $y = (-32.566) + 0.59324x$ with $R^2 = 0.91$ (supplementary fig. S5, Supplementary Material online; outliers defined as falling outside 1.5σ, where σ is the standard deviation of the residual). None of the Arctic branch pairs could be categorized as outliers, and thus they did not show any sign of excess convergence.

Next, we identified convergent genes that had functional associations with the five categories of Arctic stresses and compared the functional profiles of the convergent gene sets. In addition to the genes that had convergent substitutions in more than two comparisons (*EMB2742*, *MAPKKK14*, *ORRM6*, and *SK4*; cf., above), we found candidates for Arctic adaptation in the individual convergent gene sets (supplementary table S31, Supplementary Material online). Examples of the latter were *CRY2* (a blue light receptor mediating blue-light cotyledon expansion and flowering time), the cold and light responsive *KCS1* (encoding a condensing enzyme involved in the critical fatty acid elongation in wax biosynthesis), the cold and drought stress-activated *CAT2* (also found to be under positive selection in *D. nivalis* and possibly associated with *oxidative stress*, *physiological cold stress*, and *light stress*), *LSH4* (encoding a light-dependent short hypocotyls-like protein), and several genes possibly associated with *oxidative stress*, such as *GPX7* and *GCTF9*.

GO-term enrichment tests (table 2, supplementary tables S33–S35, Supplementary Material online) showed that "CTP synthase activity" (MF) was significantly enriched in all three Arctic comparisons. This GO-term was only associated with *EMB2742* (see above), and there were several cases where only a few genes were associated with a significantly enriched GO-term (perhaps a statistical artifact for GO-terms with few annotations in the background). Several GO-terms possibly associated with Arctic stresses were significantly enriched in the convergent genes shared by two Arctic comparisons, and none of the non-Arctic comparisons (supplementary tables

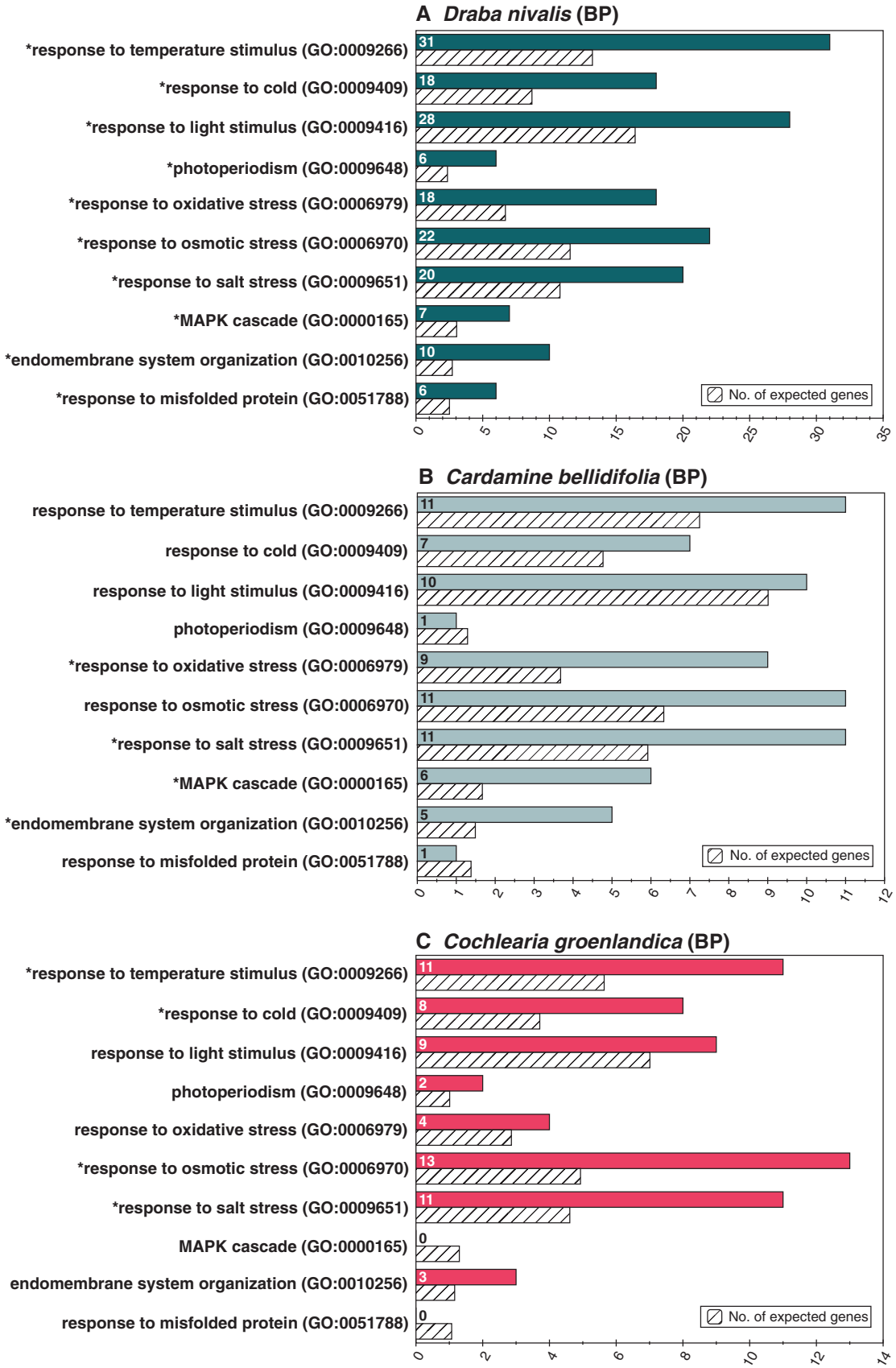


Fig. 4. Number of PSGs annotated with ten selected GO-terms in the Biological Process domain (BP). (*A*) *Draba nivalis*, (*B*) *Cardamine bellidifolia*, and (*C*) *Cochlearia groenlandica*. Expected numbers of genes (if randomly drawn) are indicated as hatched bars. An asterisk (*) indicates that the GO-term was significantly enriched in the PSG set of this species when applying the classic algorithm in topGO ($P < 0.05$).

Table 2. Number of Convergent Genes (one result per orthogroup), as well as Corresponding topGO Results.

Branches Compared [7,765 alignments]	Lineages Tested	No. of Convergent Genes (at least one site PP $\geq$ 0.80)[a]	Enriched GO-terms[b] (BP, CC, MF)
Draba nivalis, Cardamine bellidifolia	Arctic, Arctic	126	258 (161), 39 (21), 90 (59)
Draba nivalis, Cochlearia groenlandica	Arctic, Arctic	33	149 (80), 19 (17), 26 (19)
Cardamine bellidifolia, Cochlearia groenlandica	Arctic, Arctic	58	144 (81), 38 (22), 41 (31)
Draba nemorosa, Cardamine hirsuta	Non-Arctic, Non-Arctic	181	432 (253), 59 (41), 78 (64)
Draba nemorosa, Cochlearia pyrenaica	Non-Arctic, Non-Arctic	109	191 (127), 57 (37), 74 (51)
Cardamine hirsuta, Cochlearia pyrenaica	Non-Arctic, Non-Arctic	72	180 (105), 65 (36), 69 (51)

[a]Reporting only one result per orthogroup.
[b]Numbers of significant GO-terms ($P < 0.05$) when applying the classic algorithm. Numbers of enriched GO-terms when applying the elim algorithm are given in brackets.

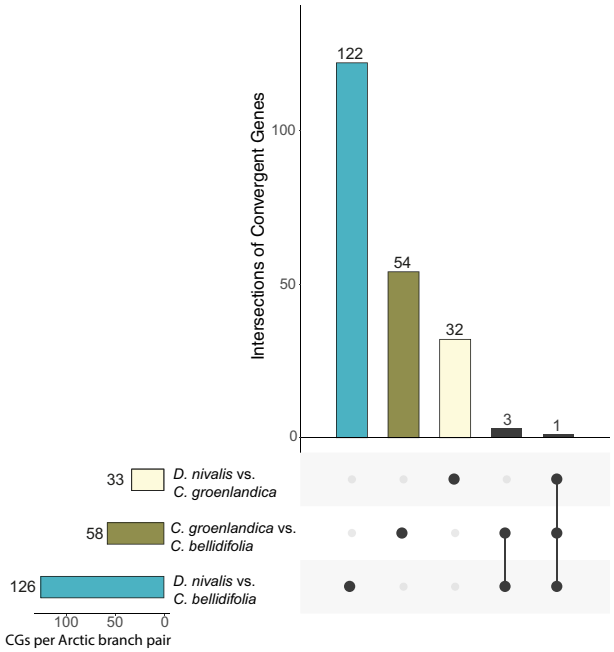


Fig. 5. UpSet plot of convergent genes (CGs) identified between Arctic branch pairs (containing at least one substitution with > 0.8 PP of convergence). The plot on the left shows total number of CGs, and the plot on the right shows the number of unique CGs, followed by CG intersections/overlaps between branch pairs (connected dots).

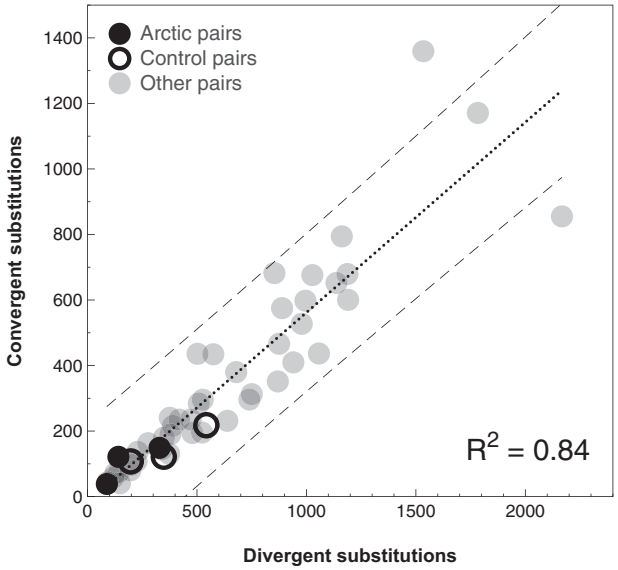


Fig. 6. Posterior expected numbers of convergent and divergent substitutions between 50 branch pairs from the full species phylogeny (fig. 1A). Numbers are sums over all alignments. Dotted line, the regression line; dashed line, 95% prediction intervals; filled black circles, the three Arctic branch pair comparisons; open circles, the branch pairs of the non-Arctic reference species; filled gray circles, remaining branch pairs.

S36–S38, Supplementary Material online). Examples are terms associated with general abiotic stress responses such as "positive regulation of abscisic acid-activated signaling pathway" (BP), "regulation of stomatal movement" (BP), and "calcium channel activity" (MF), and terms associated with protection of plant cells from oxidative stress (i.e., "glutathione peroxidase activity"; MF). We found several putative connections to the five Arctic stress categories in the individual convergent gene sets (i.e., GO-terms enriched in only one set); and interestingly, some of these terms also occurred in the functional profiles of the PSGs. Examples include "response to temperature stimulus," "cellular response to osmotic stress," "cellular response to salt stress," and GO-terms associated with light stress (e.g., "circadian regulation of calcium ion oscillation," "response to far red light," "response to blue light," "blue light photoreceptor activity," "cellular response to UV-B," "response to high light intensity," and "regulation of shade avoidance"). Thus, whereas we found limited evidence for convergence at the codon level in Arctic lineages, the convergent gene sets contained several putative candidate genes for Arctic adaptations.

Limited Overlap between PSG Sets and Convergent Gene Sets

Because genes that have evolved under positive selection and contain convergent substitutions provide strong evidence for adaptive convergence at the level of codons, we investigated intersections of PSG sets and convergent gene sets. We found that none of the genes carrying convergent substitutions in Arctic branch pairs had evolved under positive selection in both species, and only 8–12% of the convergent genes likely experienced positive selection in one of the Arctic species (supplementary tables S25–S27, Supplementary Material online). We also found that 9–16% of the convergent genes

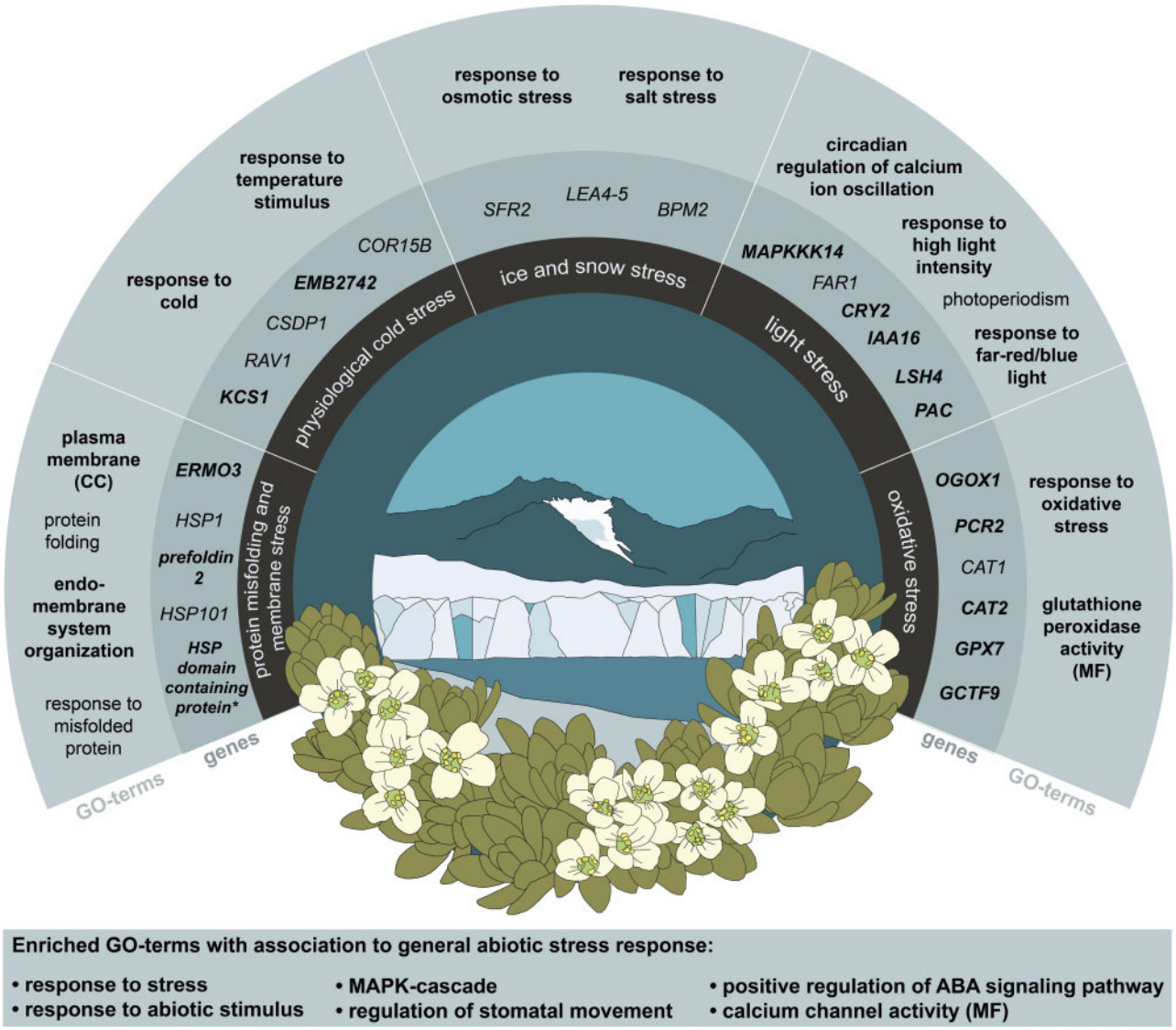


FIG. 7. Summary of putative molecular adaptations to five categories of Arctic stresses (inner black circle). The outer gray circle shows examples of enriched GO-terms (within the BP domain, unless indicated with CC or MF). The inner gray circle shows examples of candidate genes. Bold text indicates that the finding was made in more than one Arctic lineage. The results are based on PSG sets and convergent gene sets. *DNAJ heat shock N-terminal domain-containing protein.

probably evolved under positive selection in non-Arctic species, possibly indicating that these represent rapidly evolving genes. Thus, we found limited evidence for positive selection being responsible for fixing convergent substitutions in Arctic lineages per se.

Discussion

Our findings reveal that positive selection has been acting on different genes associated with similar functional pathways in the three Arctic species and that these pathways can be associated with abiotic stress types frequently encountered in the Arctic (fig. 7). However, there was little evidence for independently fixed mutations at the same sites and/or positive selection acting on the same genes. The three species might therefore have evolved similar suites of adaptations through distinct evolutionary pathways, implying that there can be many genetic trajectories for adaptation to the Arctic environment.

Limited Evidence for Molecular Convergence in Arctic Brassicaceae: Comparisons and Causes

Considering the growing body of literature showing that the genomic landscape of adaptation may involve a high degree of repeatability also at the level of substitutions and genes (Conte et al. 2012; Stern 2013), an important question arises: Why do we primarily find molecular convergence at the level of pathways in Arctic Brassicaceae? The amount of molecular convergence reported here is considerably lower than that found in other studies (Foote et al. 2015; Hu et al. 2017; Xu et al. 2017; Davies et al. 2018). Notably, we found no orthologous genes under positive selection in all three Arctic species, and only 2% of all PSGs were found to be under positive selection in Arctic species pairs. None of these genes

contained convergent substitutions, which implies that gene function could potentially have diverged rather than converged in the two lineages (Conte et al. 2012). The numbers are lower than those found in a study reporting limited evidence for parallel molecular adaptation in four lineages of subterranean mammals (Davies et al. 2018). In their study, Davies et al. (2018) found that 11% of all PSGs were under selection in more than one lineage, <1% were under selection in three lineages, and none in all four lineages. Our methodology is similar to that of Davies et al. (2018), but the results may not be directly comparable as they represent quite different organismal groups and divergence times (~90 Ma for subterranean mammals vs. ~30 Ma for Arctic Brassicaceae; Foley et al. 2016; Huang et al. 2016). In plants, a comparable study has been conducted in three species of mangroves (Xu et al. 2017). Xu et al. (2017) calculated the amount of nonneutral molecular convergence using only conserved sites and found that as many as ~400 genes contained nonneutral convergent substitutions in at least two of three mangrove species (this was interpreted as a genome-wide pattern of convergent evolution; Xu et al. 2017). In comparison, we found half as many genes with convergent substitutions in Arctic species pairs (~200 genes) without distinguishing between nonneutral and neutral substitutions. Xu et al. (2017) may have identified more evidence for convergent evolution because they used whole genome data rather than transcriptomes, which are more likely to contain the complete set of genes for a given species, but they also limited their analyses to a subset of the sites and included fewer alignments. It is also possible that specific features of the Arctic environment make molecular convergence less likely, or that our conservative study design made it difficult to detect molecular convergence.

One reason why molecular convergence may be less likely in the Arctic is the vast opportunity space for adaptation. There are numerous potential evolutionary changes that could lead to increased fitness in the Arctic, because temperature and light affect nearly all aspects of plant cell biochemistry and metabolism (Salisbury 1985; Heino and Palva 2004). This implies that adaptive molecular convergence might be uncommon in Arctic plants solely due to the high number of possible evolutionary trajectories leading to better performance under low temperatures and specific light conditions. Furthermore, plants living in the Arctic need to cope with a wide range of abiotic stresses, and research on plant stress responses has shown considerable overlap in gene expression patterns between various stress stimuli (Kilian et al. 2007). This could imply that there is considerable evolutionary constraint on these pathways and/or that adaptation to the Arctic environment might have occurred by modifying different elements in a variety of general stress response pathways. Arctic plants may thus represent an interesting example of how general pathways may lead to the evolution of different solutions to similar environmental challenges. It is also possible that the low levels of molecular convergence we document in Arctic Brassicaceae could be due to different levels of preadaptation to cold and stressful environments in non-Arctic ancestral lineages. The major clades of the Brassicaceae radiated around the Eocene–Oligocene boundary (33.9 Ma), possibly in response to a cooler and drier environment (Huang et al. 2016), and parts of the Arctic flora are likely derived from montane ancestors that migrated northwards as global temperatures dropped in the mid-Miocene and onwards (Abbott and Brochmann 2003). The Arctic also spans a wide range of growing season lengths and summer temperatures (Elvebakk 1999) and may be quite environmentally heterogeneous (see, e.g., Jónsdóttir 2005 on the heterogeneity of a High Arctic Archipelago). Although there are broad similarities in the selection pressures encountered by Arctic plants, the specific combination and strength of these pressures likely vary among different Arctic habitats. Such variation may lead to different evolutionary trade-offs and a reduced probability of molecular convergence. The three Arctic species examined here occur in the same range of bioclimatic subzones (ranging from extreme Arctic polar desert to Arctic shrub tundra; Elvebakk 1999; Elven et al. 2011), but they are often found in species-specific habitats. *Cochlearia groenlandica* grows, for example, along sea shores and on lush bird cliffs, *C. bellidifolia* on moist mineral soils and in snowbeds, and *D. nivalis* in well-drained habitats such as rocky outcrops, crevices, and rock ledges (Alsos et al. 2019). Nonetheless, our results provide evidence for considerable molecular convergence at the level of pathways, which means that the three lineages likely have experienced similar types of selection pressure. It is also possible that there may be other molecular mechanisms of Arctic adaptation that we were unable to detect. For example, convergent evolution in gene regulatory regions has been shown to be more prevalent than convergent evolution in protein-coding regions in the independent loss of flight in several lineages of paleognathous birds (Sackton et al. 2019). Future studies on molecular convergence in Arctic Brassicaceae should therefore include data on sequence evolution in regulatory regions and patterns of differential gene expression.

Conservative Estimates of Molecular Convergence

The choice of methods and study design makes our estimates of molecular convergence in Arctic Brassicaceae relatively conservative in four ways: First, the branch-site test of positive selection has proven to be relatively powerful and result in few false positives (Zhang et al. 2005; Yang and dos Reis 2011). Second, since the branch-site test can be sensitive to sequence and alignment errors (Yang and dos Reis 2011), we applied a stringent sequence and alignment filtering scheme that led to the removal of many alignments. This means that we may to some degree have underestimated the number of PSGs, as well as genes with convergent substitutions. Third, we included several species representatives from each of the three relevant clades in the Brassicaceae, because inappropriate taxonomic sampling can lead to overestimation of molecular convergence (Thomas et al. 2017). We also included two genotypes from different areas in the Arctic for each of our focal species, which narrowed our focus to evolutionary changes that were likely fixed at species level. Finally, because this study largely is based on transcriptome assemblies, the whole gene space was not recovered, and there might be

genes that are only expressed in certain tissues or that require specific environmental cues to initiate expression which are not included in our data set. This could also contribute to relatively conservative estimates of molecular convergence. Once whole-genome sequences are available for these Arctic species and temperate relatives, it is possible that more convergent genes could be identified. Nonetheless, we expect signatures of adaptation to be present in genes that are constitutively expressed in these species, because Arctic plants experience relatively low temperatures and extreme light regimes year-round, and this was confirmed by the strong signal of Arctic adaptation in our analyses. It is therefore likely that our results provide a good, but conservative estimate of the amount of molecular convergence in protein-coding regions in these three lineages of Arctic Brassicaceae. Furthermore, the relationship between convergent and divergent substitutions should be relatively unaffected by the stringent filtering scheme, and we found no indication of excess convergence, consistent with our low estimates of molecular convergence.

A Functional Syndrome of Arctic Adaptation: A Hypothesis for Experimental Verification

In this study, we have presented some of the first genetic evidence for putative plant adaptations to the Arctic environment. We found signatures of positive selection and convergent substitutions in genes associated with general abiotic stress responses, such as the "MAPK-cascade," "calcium channel activity," "positive regulation of abscisic acid-activated pathway," and "regulation of stomatal movement" (figs. 4 and 7). Most abiotic stress responses are initiated with an increase in cytoplasmic calcium, and numerous abiotic stresses also activate MAPK-cascades (Knight and Knight 2001). The plant hormone abscisic acid similarly plays important roles in stress signaling (de Zelicourt et al. 2016), and abscisic acid-level increase in response to both high light and during cold acclimation (Heino and Palva 2004; Tuteja 2007; Galvez-Valdivieso et al. 2009). Finally, stomatal regulation plays an important role in plant responses to environmental changes, and plants will, for example, close their stomata in response to drought or cold stress (Wilkinson et al. 2001). It is likely that extreme abiotic stress directly and/or indirectly linked to the dramatic light and temperature conditions of the Arctic may have influenced the evolution of such general stress response pathways.

Our results also revealed more specific pathways and genes related to the five categories of Arctic stresses. Relevant to *protein misfolding and membrane stress* (category 1), we found strong support for the plasma membrane playing a vital role in adaptation to the Arctic environment (see fig. 7 and supplementary fig. S4, Supplementary Material online). In cold-adapted organisms, the plasma membrane will often act as a semipermeable barrier that allows water to flow out of the cell in the presence of ice, while also preventing the spread of ice crystals within the cell (Uemura and Steponkus 1999). The Arctic species also contained many heat shock-related proteins (Hsps/chaperones) under positive selection, which are involved in membrane stabilization and protein folding under normal and stressful cellular conditions (Wang et al. 2004). Three genes under positive selection in more than one Arctic species were involved in protein folding and/or traffic, as well as endoplasmic reticulum morphology. These results highlight the important role that protein misfolding and the accumulation of misfolded proteins in the endoplasmic reticulum can play in cellular responses to the abiotic stresses, particularly cold stress, imposed by Arctic environments (Zhu 2016).

Of relevance to *physiological cold stress* (category 2), we found that many cold-regulated genes are under positive selection and carry convergent substitutions in at least two Arctic species. Convergent substitutions were identified in *EMB2742* (Embryo defective 2742/CTP synthase family protein) in all Arctic branch pairs. This gene has been found to be upregulated during cold acclimation in *A. thaliana* (Oono et al. 2006). In addition, all three Arctic species had one or more *LEA* genes under positive selection. These genes may be associated with *physiological cold stress*, as most *LEA* genes in *A. thaliana* contain abscisic acid response elements and/or low-temperature response elements in their promoters (Hundertmark and Hincha 2008). One of the *LEA* genes under positive selection in *D. nivalis* is the well-known cold-responsive *COR15b* (Cold Regulated 15b), which functions in stabilizing chloroplast membranes, and whose overexpression yields increased freezing tolerance in *A. thaliana* (Navarro-Retamal et al. 2016). Our identification of *COR15b* as a candidate gene for Arctic adaptation is consistent with a previous study showing that *Draba* contains a lineage-specific *COR15* duplication and that the *COR15b* duplicate appears to be evolving under positive selection (Zhou et al. 2009). Other important cold-regulated genes include *CSDP1* (Cold Shock Domain Protein 1; *D. nivalis*) and *RAV1* (related to ABI2/VP1; *Co. groenlandica*). Cold shock domains are conserved nucleic acid-binding domains found in a variety of organisms, and overexpression of *CSDP1* in *A. thaliana* has been shown to be associated with delayed seed germination in stressful conditions (Sasaki et al. 2013). The low-temperature induced transcription factor *RAV1* is involved in the regulation of plant growth under cold stress in *A. thaliana* and is known to exhibit circadian regulation (Chinnusamy et al. 2007; TAIR 2019).

Arctic plants are always at risk of experiencing *ice and snow stress* (category 3), and we found one candidate gene associated with "response to freezing" in *Co. groenlandica* (*SFR2*, sensitive to freezing 2) that plays a role in the protection of the chloroplast envelope against freezing injuries (Fourrier et al. 2008). *Ice and snow stress* can also lead to loss of cellular water and increase solute concentration in the cytoplasm (Wisniewski and Fuller 1999; Körner 2003). In this way, the stress response pathways associated with freezing tolerance are partially homologous with those of dehydration and salt tolerance (Bartels and Souer 2004), and thus the >50 candidate genes functionally associated with salt stress and osmotic stress may also be connected to this stress category. Some of the previously mentioned LEA proteins are also known to be involved in protection from freezing damage

(Hundertmark and Hincha 2008; Cuevas-Velazquez et al. 2014). That Arctic Brassicaceae may be well adapted to ice stress is consistent with a recent study demonstrating high tolerance to ice encasement in other Arctic plants (Bjerke et al. 2018).

Our results also highlight several putative molecular adaptations to the Arctic light regime (category 4; *light stress*). Many genes annotated with far-red or blue light response were found to be under positive selection or containing convergent substitutions in the Arctic species. Specifically, we identified convergent substitutions in Arctic branch pairs in the genes *MAPKKK14* (involved in far-red light induced change to photomorphogenesis in seedlings; Khanna et al. 2006) and *CRY2* (a blue light receptor involved in cotyledon expansion and flowering time; TAIR 2019). We also found that *IAA16* and *PAC* (both involved in light-controlled growth and development; Rinaldi et al. 2012; TAIR 2019) are under positive selection in two different Arctic species. The evolution of these genes may be connected to changes in day length and/or the enrichment of far-red and blue light with increasing latitude (cf., Nilsen 1985; Salisbury 1985). These genes can be interesting starting points for future experiments aimed at understanding plant adaptations to polar light climates.

Consistent with the fact that high irradiance in combination with low temperatures leads to excessive production of ROS (*oxidative stress*, category 5; Inzé and Montagu 1995; Heino and Palva 2004), we found that genes associated with oxidative stress were enriched in the PSG sets of *D. nivalis* and *C. bellidifolia*. Genes associated with "glutathione peroxidase activity" were also overrepresented among convergent genes in Arctic branch pairs, and these enzymes are involved in both cellular redox homeostasis and H_2O_2 detoxification. Their activity can be induced by a range of environmental stresses, including, for example, cold stress, drought, and UV-B light (Bela et al. 2015). Some interesting candidate genes for adaptation to oxidative stress in Arctic environments include *PCR2* and *OGOX1* (both under positive selection in *C. bellidifolia* and *Co. groenlandica*) and two catalase genes under positive selection in *D. nivalis* (*CAT1* and *CAT2*). *CAT1* and *CAT2* are involved in controlling cellular redox homeostasis (Du et al. 2008). *CAT2* was also found to contain convergent substitutions in *Co. groenlandica* and *D. nivalis*. Intriguingly, *CAT2* is both cold and drought stress activated and involved in preventing oxidative stress under photorespiration, while also being conditioned by photoperiod (Queval et al. 2007; TAIR 2019).

To summarize, we found putative molecular adaptations to cold stress, freezing stress, oxidative stress, and the unique light regime of polar environments in all three Arctic species. We also found ample evidence that Arctic Brassicaceae possess adaptations associated with the plasma membrane and protein folding, a finding which is highly relevant for coping with typical Arctic stresses. Our results thus suggest that Arctic plants may have evolved similar ecological preferences and phenotypes, but through different genetic trajectories. This potential "functional syndrome" of Arctic adaptation will need experimental verification, and the hypothesized molecular adaptations suggested here can serve as a starting point for future studies.

Materials and Methods

Taxon Sampling, Transcriptome Assemblies, and Final Data Set

We acquired genome-wide coding sequences from 15 species belonging to clades A, B, and C in the Brassicaceae, including the three Arctic focal species *C. bellidifolia*, *Co. groenlandica*, and *D. nivalis* (fig. 1A). RNA sequences for the three Arctic species were generated de novo for this study (Supplementary Material online, GenBank SRA ID: SRP240404 and BioProject ID: PRJNA599963), whereas data for the other 12 species were derived from published genomes or transcriptomes (supplementary table S3, Supplementary Material online). Because we were mainly interested in substitutions fixed at the species level, we included two genotypes from each focal species, representing two distant Arctic regions (NO/US, supplementary table S1, Supplementary Material online). Seeds were harvested from field-collected plants and germinated in a growth chamber at the University of Oslo (cultivation conditions as specified by Brochmann et al. [1992]).

We downloaded RNA sequences from eight of the non-Arctic species (supplementary tables S2 and S3, Supplementary Material online). Short-read data were selected and downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genbank/) based on the following criteria: 1) The species should belong to the same genus or clade as one of the Arctic focal species in the phylogenies of Huang et al. (2016) and Guo et al. (2017), 2) the data should have been generated on an Illumina machine, and 3) the files should contain more than 11 M fragments in total. The raw RNA-seq reads were quality-checked with FastQC v.0.11.4 (Andrews 2010) and assembled with Trinity v.2.4.0 (Grabherr et al. 2011) using the integrated Trimmomatic option and a minimum assembled contig length of 300 bp. The transcript abundance in each assembly was estimated with kallisto v.0.43.0 (Bray et al. 2016), and only the highest expressed isoform for each "trinity gene" was retained. We also acquired an early version of the now published transcriptome of the non-Arctic species *Co. pyrenaica* (for assembly details, see Lopez et al. 2017), which was filtered based on the longest isoform as we did not have access to the raw reads at the time. The following assembly processing was performed on all transcriptome assemblies: Chimeric transcripts, likely resulting from misassemblies, were filtered out according to the procedure of Yang and Smith (2013), and coding regions within the isoform-filtered transcripts were predicted with TransDecoder v.3.0.0 (Haas et al. 2013) using default settings. The final transcriptome quality was evaluated based on several metrics (see Supplementary Material online), including the number of complete and fragmented BUSCOs (BUSCO v.3.0.1, Simão et al. 2015). The transcriptomic data were then combined with coding sequences derived from the already published genomes of three additional non-Arctic species (*A. thaliana*,

Arabis alpina, and *C. hirsuta*; supplementary table S3, Supplementary Material online). Sequences containing missing data or ambiguity characters were excluded, and the sequence sets of *Arabis alpina* and *C. hirsuta* were run through TransDecoder to ensure that they, for example, did not contain stop codons. The filtered Arctic transcriptome assemblies are available at https://doi.org/10.5061/dryad.v41ns1rs0.

Orthogroup Inference and Alignment Generation

Alignments were generated separately for the positive selection tests and for the calculation of posterior expected numbers of convergent and divergent substitutions. For the positive selection tests, we ran OrthoFinder v.1.1.8 (Emms and Kelly 2015) separately for each of the three clades/groups illustrated in figure 1*B–D*. This was done in order to identify sets of genes descended from a single gene in the last common ancestor of the species (i.e., orthogroups; Emms and Kelly 2015). These orthogroups were used as a basis for subsequent alignment construction. Due to the low number of single-copy orthogroups (not uncommon in plants), orthogroups that contained multiple gene copies/transcripts per species were divided into subsets based on smallest genetic distance. All orthogroups were aligned based on protein sequence using MAFFT v.7.300 (Katoh et al. 2005), genetic distances between the gene copies/transcripts were calculated as Kimura protein distances (Kimura 1983), and one gene copy/transcript per species or accession was extracted based on the smallest genetic distance to the transcript of the Alaskan accession of either *C. bellidifolia*, *D. nivalis* or *Co. groenlandica*, depending on the clade in question. If there were more than one transcript from the Arctic accessions, a new subset was made for each such transcript. The resulting orthogroup subsets contained only one transcript per species/accession and were realigned with PRANK (Löytynoja and Goldman 2005) using GUIDANCE v2.02 (Sela et al. 2015) with ten bootstraps. GUIDANCE allows assessment of alignment confidence and subsequent removal of unreliable aligned columns and sequences. Previous studies have shown that a phylogeny aware aligner such as PRANK in combination with alignment filtering with GUIDANCE improves positive selection inference on simulated data (Jordan and Goldman 2012; Privman et al. 2012). All alignments with a sequence score <0.6, as well as all columns scoring <0.80, were removed from the data set. The three final alignment sets had an average length of 1,544–1,586 bp and contained on average 17.0–20.7% missing data, but sites containing alignment gaps were deleted within codeml (cleandata = 1).

For estimation of posterior expected numbers of convergent and divergent substitutions, we ran OrthoFinder v.2.3.1 (Emms and Kelly 2019) for the total data set (fig. 1*A*). The resulting number of single-copy orthogroups was extremely small, and we therefore extracted subsets from multi-copy orthogroups as described above for the positive selection tests. For simplicity, we extracted subsets based on only one of the Arctic focal species, that is, we extracted one transcript per species/accession based on the smallest genetic distance to the transcripts of *D. nivalis* (arbitrarily chosen). The alignments were then generated as explained above (final average alignment length was 1,524 bp with an average of 25.8% missing data). Note that all alignments for both the positive selection tests and the calculation of posterior expected numbers of convergent and divergent substitutions contained all relevant species (cf., fig. 1).

The Branch-Site Test of Positive Selection

The branch-site test (model A; Zhang et al. 2005) implemented in codeml (PAML v.4.9d, Yang 1997) was used to test for positive selection on individual codons along the lineage leading to each of the Arctic species. In this test, an alternative model allowing positive selection on the foreground lineage (e.g., one of the Arctic lineages) is contrasted to a null-model that does not allow such positive selection using a likelihood-ratio test (Zhang et al. 2005). To limit the influence of missing data (a challenge when using transcriptomic data), we ran the tests separately for clades A, B, and C (Huang et al. 2016; fig. 1*B–D*). However, the model species *A. thaliana* was included in all runs, and due to the low availability of high-quality transcriptomic data from clade C, we also included two species from clade B in the *Co. groenlandica* data set (fig. 1*D*). Codeml was run from four to six times per model with different initial parameter values, and the run with the highest likelihood score was used in subsequent analyses (Wong et al. 2004). The likelihood-ratio tests were performed in R v.3.4.4 (R Core Team 2019) using extRemes (Gilleland and Katz 2016). The significance level was set to 0.05 with 1 degree of freedom. Phylogenetic trees (representing the three clades) used in the analyses were constructed using RAxML based on codon alignments of ~2,000–4,000 single-copy genes (Stamatakis 2014; see fig. 1*B–D*, Supplementary Material online). We also ran the branch-site test for each topology in figure 1*B–D* with a close non-Arctic relative (*C. hirsuta*, *Co. pyrenaica*, and *D. nemorosa*) as the foreground branch. This was done in order to filter out genes that always seem to be under positive selection and to gain a better understanding of which genes that actually could be tied to Arctic adaptation. The resulting sets of PSGs were compared between species in order to identify genes with convergent selection patterns. Genes were considered as being under convergent selection if the same *A. thaliana* ortholog was found to be under positive selection in two or more species. The significance of the gene intersections was then assessed with the supertest function in SuperExactTest v.1.0.7 (Wang et al. 2015) and visualized with UpSetR 1.3.3 (Conway et al. 2017).

Posterior Expected Numbers of Convergent and Divergent Substitutions

We used Grand Convergence (Qian and de Koning 2018; Qian et al. 2018) to calculate the posterior expected numbers of convergent and divergent substitutions across all branch pairs in a super phylogeny generated with OrthoFinder v.2.3.1 (Emms and Kelly 2019; fig. 1*A*). This was done to identify sets of convergent genes (i.e., genes containing at least one convergent substitution in Arctic species pairs) and to examine deviations from the expected relationship between convergent and divergent substitutions. We also estimated the

number of convergent genes in the non-Arctic reference species. In Grand Convergence, a convergent substitution is defined as changes resulting in the same amino acid at the same site in two lineages, and divergent substitutions as changes resulting in different amino acids at the same site in two lineages (silent substitutions are not considered; Castoe et al. 2009). The program was run with branch length estimations from OrthoFinder, the Le Gascuel amino acid substitution model (Le and Gascuel 2008), and all gaps and ambiguity characters excluded. Genes containing at least one convergent site with PP >0.80 were considered as convergent, but since a gene/transcript potentially could be included in more than one alignment following our orthogroup subsetting procedure (see above), we only included unique TAIR IDs (i.e., *A. thaliana* orthologs) in our list of convergent genes. The resulting sets of convergent genes were compared with each other in the same way as the PSGs in order to identify genes that were shared between more than one Arctic branch pair.

To investigate the relationship between divergent and convergent substitutions in our data set, we chose one alignment per orthogroup and summed all pairwise convergent substitutions and all pairwise divergent substitutions over all alignments. Using the branch totals from 50 branch pairs, we performed a simple linear regression to investigate whether our data fitted a linear model as expected from other studies (Castoe et al. 2009; Thomas and Hahn 2015). The 50 branch pairs were mostly arbitrarily chosen, except that all pairwise combinations between the Arctic species and between the non-Arctic reference species were included (see supplementary table S32, Supplementary Material online, for a complete list of the 50 branch pairs).

Gene Annotations and GO Enrichment Tests

The final sets of PSGs and convergent genes were annotated with DAVID v.6.8 (Huang et al. 2009) based on the *A. thaliana* ortholog. We also tested for overrepresented GO-terms within the three domains: BP, CC and MF using a Fisher's exact test in combination with the "classic," "elim," and "weight" algorithms implemented in topGO v.2.32 of Bioconductor (Gentleman et al. 2004; Alexa et al. 2006). The "classic" algorithm processes each term independently without taking the GO-hierarchy into account, the "elim" algorithm traverses the GO-graph bottom-up, discarding genes that have already been mapped to significant child terms, and the "weight" algorithm decides which GO-terms best describe the gene based on a weighting scheme that takes the enrichment scores of the neighboring GO-terms into account (Alexa et al. 2006). Simulations have shown that the "elim" and "weight" algorithms tend to be more conservative than the "classic" algorithm, and the creators of topGO recommend using "elim" for identifying important areas in the graph due to its simplicity (Alexa et al. 2006). The gene sets were annotated based on homology to *A. thaliana* using annotations from org.At.tair.db v.3.6 (Carlson 2018), and *A. thaliana* was also used as the background gene universe in all gene set enrichment analyses. Although it could be considered less than ideal to apply the *A. thaliana* background in this type of analysis, it is likely to be more consistent than creating new backgrounds based on the annotation of de novo transcriptome assemblies for each species, because transcriptome assemblies represent a less complete sampling of gene space and are prone to varying degrees of sampling bias. We did not correct for multiple testing as we wanted to avoid being overly conservative, and because such corrections cannot directly be applied to the results of the "elim" and "weight" algorithms due to non-independent testing (the topGO manual; Alexa and Rahnenfuhrer 2018). Enriched GO-terms that might be tied to Arctic adaptation were assessed based on literature, and the results from the non-Arctic species were used to filter out more general GO-terms that were enriched in all species (e.g., GO-terms related to immunity). Finally, we also generated a heatmap of the proportions of PSGs associated with the five categories of Arctic stresses, by using the number of genes annotated with a selection of these GO-terms. The annotations were identical to those used in the topGO analyses, and the heatmap was generated with gplots package v.3.0.1.1 in R (Warnes et al. 2019). (For a flow chart of the analysis pipeline, see supplementary fig. S1, Supplementary Material online.)

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Patrick P. Edger and J. Chris Pires for making the *Cochlearia pyrenaica* transcriptome available to us before publication. We also thank Tanja Slotte, the Slotte Group, Filip Kolář, Magdalena Bohutínská, Christian Rellstab, and the SpArc and SpeciationClock project teams for valuable input and comments. This article was greatly improved by the feedback of two anonymous reviewers. Computational analyses were performed on the Abel Cluster, owned by the University of Oslo and Uninett/Sigma2, and operated by the Department for Research Computing at USIT, the University of Oslo IT-department (http://www.hpc.uio.no/). The study was funded by the Research Council of Norway through the *SpArc* project (RCN 240223) awarded to C.B.

References

Abbott RJ, Brochmann C. 2003. History and evolution of the Arctic flora: in the footsteps of Eric Hultén. *Mol Ecol.* 12(2):299–313.

Alexa A, Rahnenfuhrer J. 2018. topGO: enrichment analysis for Gene Ontology. R package version 2.34.0. Available from: http://bioconductor.org/packages/release/bioc/html/topGO.html.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22(13):1600–1607.

Alsos I, Arnesen G, Elven R, Sandbakk BE. 2019. The flora of Svalbard. Tromsø (Norway): The Arctic University of Norway. [cited 2019 Sep 27]. Available from: https://svalbardflora.no/

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bartels D, Souer E. 2004. Molecular responses of higher plants to dehydration. In: Hirt H, Shinozaki K, editors. Plant responses to abiotic stress. Berlin (Germany): Springer. p. 9–38.

Bela K, Horváth E, Gallé Á, Szabados L, Tari I, Csiszár J. 2015. Plant glutathione peroxidases: emerging role of the antioxidant enzymes in plant development and stress responses. *J Plant Physiol*. 176:192–201.

Billings WD, Mooney HA. 1968. The ecology of Arctic and alpine plants. *Biol Rev*. 43(4):481–529.

Bjerke JW, Elverland E, Jaakola L, Lund L, Bochenek Z, Kłos A, Tømmervik H. 2018. High tolerance of a high-arctic willow and graminoid to simulated ice encasement. *Boreal Environ Res*. 23:329–338.

Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. 2018. (Non)parallel evolution. *Annu Rev Ecol Evol Syst*. 49(1):303–330.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 34(5):525–527.

Brochmann C, Soltis PS, Soltis DE. 1992. Multiple origins of the octoploid Scandinavian endemic *Draba cacuminum*: electrophoretic and morphological evidence. *Nord J Bot*. 12(3):257–272.

Carlsen T, Bleeker W, Hurka H, Elven R, Brochmann C. 2009. Biogeography and phylogeny of *Cardamine* (Brassicaceae). *Ann Mo Bot Gard*. 96(2):215–236.

Carlson M. 2018. org.At.tair.db: Genome wide annotation for *Arabidopsis*. R package version 3.7.0. Available from: http://bioconductor.org/packages/release/data/annotation/html/org.At.tair.db.html.

Castoe TA, de Koning APJ, Kim H-M, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci U S A*. 106(22):8986–8991.

Chapin F. 1983. Direct and indirect effect of temperature on Arctic plants. *Polar Biol*. 2(1):47–52.

Chen L, DeVries AL, Cheng C-HC. 1997. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc Natl Acad Sci U S A*. 94(8):3817–3822.

Chinnusamy V, Zhu J, Zhu JK. 2007. Cold stress regulation of gene expression in plants. *Trends Plant Sci*. 12(10):444–451.

Chory J, Chatterjee M, Cook RK, Elich T, Fankhauser C, Li J, Nagpal P, Neff M, Pepper A, Poole D, et al. 1996. From seed germination to flowering, light controls plant development via the pigment phytochrome. *Proc Natl Acad Sci U S A*. 93(22):12066–12071.

Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proc Biol Sci*. 279(1749):5039–5047.

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33(18):2938–2940.

Cuevas-Velazquez CL, Rendón-Luna DF, Covarrubias AA. 2014. Dissecting the cryoprotection mechanisms for dehydrins. *Front Plant Sci*. 5:1–6.

Davies KTJ, Bennett NC, Faulkes CG, Rossiter SJ. 2018. Limited evidence for parallel molecular adaptations associated with the subterranean niche in mammals: a comparative study of three superorders. *Mol Biol Evol*. 35(10):2544–2559.

Du YY, Wang PC, Chen J, Song CP. 2008. Comprehensive functional analysis of the catalase gene family in *Arabidopsis thaliana*. *J Integr Plant Biol*. 50(10):1318–1326.

Elvebakk A. 1999. Bioclimatic delimitation and subdivision of the Arctic. In: Nordal I, Razzhivin VY, editors. The species concept in the high north: a panarctic flora initiative. Oslo (Norway): The Norwegian Academy of Science and Letters. p. 81–112.

Elven R, Murray DF, Razzhivin VY, Yurtsev BA. 2011. Annotated checklist of the Panarctic Flora (PAF). Vascular plants. Fairbanks (USA): University of Alaska. [cited 2019 Sep 27]. Available from: http://panarcticflora.org/

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 16(1):157.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20(1):1–14.

Foley NM, Springer MS, Teeling EC. 2016. Mammal madness: is the mammal tree of life not yet resolved? *Philos Trans R Soc Lond B Biol Sci*. 371(1699): pii: 20150140.

Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S, van Elk, CE, Hunter ME, Joshi V, et al. 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet*. 47(3):272–275.

Fourrier N, Bédard J, Lopez-juez E, Barbrook A, Bowyer J, Jarvis P, Warren G, Thorlby G. 2008. A role for SENSITIVE TO FREEZING2 in protecting chloroplasts against freeze-induced damage in *Arabidopsis*. *Plant J*. 55(5):734–745.

Galvez-Valdivieso G, Fryer MJ, Lawson T, Slattery K, Truman W, Smirnoff N, Asami T, Davies WJ, Jones AM, Baker NR, et al. 2009. The high light response in *Arabidopsis* involves ABA signaling between vascular and bundle sheath cells. *Plant Cell* 21(7):2143–2162.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 5(10):R80.

Gilleland E, Katz RW. 2016. extRemes 2.0: an extreme value analysis package in *R*. *J Stat Softw*. 72(8):1–39.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 29(7):644–652.

Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D, Ma T, Hu Q, Al-Shehbaz IA, et al. 2017. Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* 18(1):176.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Philip D, Bowden J, Couger MB, Eccles D, Li B, Macmanes MD, et al. 2013. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc*. 8(8):1494–1443.

Heide OM. 2005. Ecotypic variation among European arctic and alpine populations of *Oxyria digyna*. *Arct Antarct Alp Res*. 37(2):233–238.

Heino P, Palva ET. 2004. Signal transduction in plant cold acclimation. In: Hirt H, Shinozaki K, editors. Plant responses to abiotic stress. Berlin (Germany): Springer. p. 151–186.

Hendry AP. 2013. Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity* 111(6):456–466.

Heyduk K, Moreno-Villena JJ, Gilman IS, Christin P-A, Edwards EJ. 2019. The genetics of convergent evolution: insights from plant photosynthesis. *Nat Rev Genet*. 20(8):485–493.

Hu Y, Wu Q, Ma S, Ma T, Shan L, Wang X, Nie Y, Ning Z, Yan L, Xiu Y, et al. 2017. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc Natl Acad Sci U S A*. 114(5):1081–1086.

Huang CH, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz IA, Edger PP, et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol Biol Evol*. 33(2):394–412.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 4(1):44–57.

Hultén E. 1968. Flora of Alaska and neighboring territories: a manual of the vascular plants. Stanford (CA): Stanford University Press.

Hundertmark M, Hincha DK. 2008. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9(1):118–122.

Inzé D, Montagu MV. 1995. Oxidative stress in plants. *Curr Opin Biotechnol*. 6(2):153–158.

Jackson SD. 2009. Plant responses to photoperiod. *New Phytol*. 181(3):517–531.

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF. 1998. *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. *Science* 280(5360):104–107.

Jónsdóttir IS. 2005. Terrestrial ecosystems on Svalbard: heterogeneity, complexity and fragility from an Arctic island perspective. *Proc R Ir Acad*. 105B(3):155–165.

Jónsdóttir IS. 2011. Diversity of plant life histories in the Arctic. *Preslia* 83:281–300.
Jordan G, Goldman N. 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol Biol Evol*. 29(4):1125–1139.
Jordon-Thaden I, Hase I, Al-Shehbaz I, Koch MA. 2010. Molecular phylogeny and systematics of the genus *Draba* (Brassicaceae) and identification of its most closely related genera. *Mol Phylogenet Evol*. 55(2):524–540.
Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33(2):511–518.
Khanna R, Shen Y, Toledo-ortiz G, Kikis EA, Johannesson H, Hwang Y, Quail PH. 2006. Functional profiling reveals that only a small number of phytochrome-regulated early-response genes in *Arabidopsis* are necessary for optimal deetiolation. *Plant Cell* 18(9):2157–2171.
Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. 2007. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J*. 50(2):347–363.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
Knight H, Knight MR. 2001. Abiotic stress signalling pathways: specificity and cross-talk. *Trends Plant Sci*. 6(6):262–267.
Knight MR, Knight H. 2012. Low-temperature perception leading to gene expression and cold tolerance in higher plants. *New Phytol*. 195(4):737–751.
Koch MA. 2012. Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). *Taxon* 61(1):76–92.
Körner C. 2003. Climatic stress. In: Körner C, editor. Alpine plant life. Functional plant ecology of high mountain ecosystems. 2nd ed. Berlin (Germany): Springer. p. 101–119.
Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF, Mesa T, Row M, Diego S, California JAK, et al. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiol*. 130(4):2129–2141.
Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. 25(7):1307–1320.
Lopez L, Wolf EM, Pires JC, Edger PP, Koch MA. 2017. Molecular resources from transcriptomes in the Brassicaceae family. *Front Plant Sci*. 8:1–15.
Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65(7):1827–1840.
Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. 102(30):10557–10562.
Lütz C. 2010. Cell physiology of plants growing in cold environments. *Protoplasma* 244(1–4):53–73.
Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. 2016. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* 354(6310):336–340.
Navarro-Retamal C, Bremer A, Alzate-Morales J, Caballero J, Hincha DK, González W, Thalhammer A. 2016. Molecular dynamics simulations and CD spectroscopy reveal hydration-induced unfolding of the intrinsically disordered LEA proteins COR15A and COR15B from *Arabidopsis thaliana*. *Phys Chem Chem Phys*. 18(37):25806–25816.
Nilsen J. 1985. Light climate in northern areas. In: Kaurin Å, Junttila O, Nilsen J, editors. Plant production in the north: Proceedings from "Plant Adaptation Workshop"; 1983 Sep 4–9; Tromsø, Norway. 1st ed. Tromsø (Norway): Norwegian University Press. p. 62–82.
Oono Y, Seki M, Satou M, Iida K, Akiyama K, Sakurai T, Fujita M, Yamaguchi-Shinozaki K, Shinozaki K. 2006. Monitoring expression profiles of *Arabidopsis* genes during cold acclimation and deacclimation using DNA microarrays. *Funct Integr Genomics*. 6(3):212–234.
Piskurewicz U, Turečková V, Lacombe E, Lopez-Molina L. 2009. Far-red light inhibits germination through DELLA-dependent stimulation of ABA synthesis and ABI3 activity. *EMBO J*. 28(15):2259–2271.
Privman E, Penn O, Pupko T. 2012. Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol*. 29(1):1–5.
Qian C, Bryans N, Kruykov I, de Koning APJ. 2018. Visualization and analysis of statistical signatures of convergent evolution. Alberta (Canada): University of Calgary. [cited 2019 Sep 27]. Available from: https://github.com/dekoning-lab/grand-conv.
Qian C, de Koning A. 2018. Rapid discovery of convergent molecular evolution across entire phylogenies. Alberta (Canada): University of Calgary. [cited 2019 Sep 27]. Available from: https://github.com/dekoning-lab/grand-conv.
Queval G, Issakidis-Bourguet E, Hoeberichts FA, Vandorpe M, Gakière B, Vanacker H, Miginiac-Maslow M, Van Breusegem F, Noctor G. 2007. Conditional oxidative stress responses in the *Arabidopsis* photorespiratory mutant *cat2* demonstrate that redox state is a key modulator of daylength-dependent gene expression, and define photoperiod as a crucial factor in the regulation of H_2O_2-induced cell death. *Plant J*. 52(4):640–657.
R Core Team. 2019. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/.
Rinaldi MA, Liu J, Enders TA, Bartel B, Strader LC. 2012. A gain-of-function mutation in *IAA16* confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. *Plant Mol Biol*. 79(4–5):359–373.
Ryu SB, Costa A, Xin Z, Li PH. 1995. Induction of cold hardiness by salt stress involves synthesis of cold- and abscisic acid-responsive proteins in potato (*Solanum commersonii* Dun). *Plant Cell Physiol*. 36(7):1245–1251.
Sackton TB, Clark N. 2019. Convergent evolution in the genomics era: new insights and directions. *Philos Trans R Soc Lond B Biol Sci*. 374(1777):20190102.
Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* 364(6435):74–78.
Salisbury FB. 1985. Plant adaptations to the light environment. In: Kaurin Å, Junttila O, Nilsen J, editors. Plant production in the north: Proceedings from "Plant Adaptation Workshop"; 1983 Sep 4–9; Tromsø, Norway. 1st ed. Tromsø (Norway): Norwegian University Press. p. 43–61.
Sasaki K, Kim M-H, Imai R. 2013. *Arabidopsis* COLD SHOCK DOMAIN PROTEIN 2 is a negative regulator of cold acclimation. *New Phytol*. 198(1):95–102.
Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res*. 43(W1):W7–W14.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet*. 14(11):751–764.
[TAIR] The Arabidopsis Information Resource. 2019. Fremont: Phoenix Bioinformatics Corporation. [cited 2019 Oct 23]. Available from: www.arabidopsis.org.
Thomas GWC, Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Mol Biol Evol*. 32(5):1232–1236.
Thomas GWC, Hahn MW, Hahn Y. 2017. The effects of increasing the number of taxa on inferences of molecular convergence. *Genome Biol Evol*. 9(1):213–221.

Thomashow MF. 1999. Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu Rev Plant Physiol Plant Mol Biol*. 50(1):571–599.

Thomashow MF. 2010. Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. *Plant Physiol*. 154(2):571–577.

Tuteja N. 2007. Abscisic acid and abiotic stress signaling. *Plant Signal Behav*. 2(3):135–138.

Uemura M, Steponkus PL. 1999. Cold acclimation in plants: relationship between the lipid composition and the cryostability of the plasma membrane. *J Plant Res*. 112(2):245–254.

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu JK. 2006. Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J*. 45(4):523–539.

Wang M, Zhao Y, Zhang B. 2015. Efficient test and visualization of multi-set intersections. *Sci Rep*. 5:16923.

Wang W, Vinocur B, Shoseyov O, Altman A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci*. 9(5):13–15.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, et al. 2019. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.1. Available from: https://cran.r-project.org/web/packages/gplots/index.html.

Wehrmeister RR, Bonde EK. 1977. Comparative aspects of growth and reproductive biology in Arctic and Alpine populations of *Saxifraga cernua* L. *Arct Alp Res*. 9(4):401–406.

Wilkinson S, Clephan AL, Davies WJ. 2001. Rapid low temperature-induced stomatal closure occurs in cold-tolerant *Commelina communis* leaves but not in cold-sensitive Tobacco leaves, via a mechanism that involves apoplastic calcium but not abscisic acid. *Plant Physiol*. 126(4):1566–1578.

Wisniewski M, Fuller M. 1999. Ice nucleation and deep supercooling in plants: new insights using infrared thermography. In: Margesin R, Schinner F, editors. Cold-adapted organisms. Ecology, physiology, enzymology and molecular biology. Berlin (Germany): Springer. p. 105–118.

Wisniewski M, Fuller M, Palta J, Carter J, Arora R, Fuller M, Palta J, Carter J, Arora R. 2004. Ice nucleation, propagation, and deep supercooling in woody plants. *J Crop Improv*. 10(1–2):5–16.

Wong WSW, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168(2):1041–1051.

Xu S, He Z, Guo Z, Zhang Z, Wyckoff GJ, Greenberg A, Wu CI, Shi S, Zhang J. 2017. Genome-wide convergence during evolution of mangroves from woody plants. *Mol Biol Evol*. 34(4):1008–1015.

Yang Y, Smith SA. 2013. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics* 14(1):328–311.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13(5):555–556.

Yang Z, dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol*. 28(3):1217–1228.

de Zelicourt A, Colcombet J, Hirt H. 2016. The role of MAPK modules and ABA during abiotic stress signaling. *Trends Plant Sci*. 21(8):677–685.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22(12):2472–2479.

Zhou D, Zhou J, Meng L, Wang Q, Xie H, Guan Y, Ma Z, Zhong Y, Chen F, Liu J. 2009. Duplication and adaptive evolution of the *COR15* genes within the highly cold-tolerant *Draba* lineage (Brassicaceae). *Gene* 441(1–2):36–44.

Zhu J. 2016. Abiotic stress signaling and responses in plants. *Cell* 167(2):313–324.

Supplementary material, Paper I

Supplementary materials and methods

Taxon sampling and RNA sequencing of focal species

We acquired genome-wide coding sequences from 15 species within clade A, B and C in the Brassicaceae family, including the three Arctic focal species *Cardamine bellidifolia*, *Cochlearia groenlandica* and *Draba nivalis* (fig. 1A). RNA sequences for the three Arctic species were generated *de novo* for this study (GenBank SRA ID: SRP240404 and BioProject ID: PRJNA599963), while data for the other 12 species were derived from published genomes or transcriptomes (supplementary table S3). Because we were mainly interested in substitutions fixed at the species level, we included two genotypes from each focal species, representing two distant Arctic regions (NO/US, supplementary table S1). Seeds were harvested from field-collected plants and germinated in a growth chamber at the University of Oslo (cultivation conditions as specified in Brochmann et al. 1992). Total RNA from each individual was extracted from ~50 mg of leaf tissue using the Ambion RNAqueous Kit (Thermo Fisher Scientific, Waltham, USA) following the manufacturer's protocol for fresh plant tissue. The Broad Range RNA Kit was used to quantify RNA extractions using a Qubit v.2.0 fluorometer (Life Technologies, Carlsbad, USA), and RNA extract quality was assessed using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, Waltham, USA) and an Experion Automated Electrophoresis System Station (Bio-Rad Laboratories, Hercules, USA). The RNA extractions were sent to the Norwegian Sequencing Centre (www.sequencing.uio.no) for library preparations (TruSeq protocol for stranded mRNA, Illumina, San Diego, USA), and RNA sequencing (RNA-seq). The samples were indexed, pooled together and run on a single lane (10 samples/lane) of an Illumina HiSeq2000 (Illumina, San Diego, USA) producing paired-end reads of 125 bp with an insert size of approximately 300 bp.

Transcriptome assemblies and final data set

We downloaded RNA sequences from eight of the non-Arctic species (supplementary tables S2 and S3). Short-read data was selected and downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genbank/) based on the following criteria: 1) The species should belong to the same genus or clade as one of the Arctic focal species in the phylogenies of Huang et al. (2016) and Guo et al. (2017), 2) The data should have been generated on an Illumina machine, and 3) The files should contain more than 11 M fragments in total. The raw RNA-seq reads were quality-checked with FastQC v.0.11.4 (Andrews 2010) and assembled with Trinity v.2.4.0 (Grabherr et al. 2013) using the integrated Trimmomatic option and a minimum assembled contig length of 300 bp. The transcript abundance in each assembly was estimated with kallisto v.0.43.0 (Bray et al. 2016) using the “align_and_estimate_abundance.pl” script as well as the “abundance_estimates_to_matrix.pl” script included in the Trinity

v.2.4.0 software suite. The resulting transcript abundance estimate was used for filtering the assemblies with the "filter_low_expr_transcripts.pl" (also included in the Trinity v.2.4.0 software package), so that only the highest expressed isoform for each "trinity gene" was retained. We also acquired an early version of the now published transcriptome of the non-Arctic species, *Cochlearia pyrenaica* (for assembly details see Lopez et al. 2017), which was filtered based on the longest isoform as we did not have access to the raw reads at the time. The following assembly processing was performed on all transcriptome assemblies: Chimeric transcripts, likely resulting from misassemblies, were filtered out according to the procedure of Yang and Smith (2013), and coding regions within the isoform filtered transcripts were predicted with TransDecoder v.3.0.0 (Haas et al. 2013) using default settings. The final transcriptome quality was evaluated based on the E90N50 transcript contig length, the read representation calculated with Bowtie 2 (Bowtie 2 v.2.2.9, Langmead and Salzberg 2012), the number of coding regions identified with TransDecoder v.3.0.0 (Haas et al. 2013), and the number of complete and fragmented BUSCOs (BUSCO v.3.0.1, Simão et al. 2015). The transcriptomic data was then combined with coding sequences derived from the already published genomes of three additional non-Arctic species (*Arabidopsis thaliana*, *Arabis alpina* and *Cardamine hirsuta*; supplementary table S3). Sequences containing missing data or ambiguity characters were excluded, and the sequence sets of *A. alpina* and *C. hirsuta* were run through TransDecoder to ensure that they e.g. did not contain stop codons. The filtered Arctic transcriptome assemblies are available at https://doi.org/10.5061/dryad.v41ns1rs0.

Orthogroup inference and alignment generation

Alignments were generated separately for the positive selection tests and for the calculation of posterior expected numbers of convergent and divergent substitutions. For the positive selection tests, we ran OrthoFinder v.1.1.8 (Emms and Kelly 2015) separately for each of the three clades/groups illustrated in figure 1B-D. This was done in order to identify sets of genes descended from a single gene in the last common ancestor of the species (i.e. orthogroups; Emms and Kelly 2015). These orthogroups were then used as a basis for subsequent alignment construction. Due to the low number of single-copy orthogroups (not uncommon in plants), orthogroups that contained multiple gene copies/transcripts per species were divided into subsets based on smallest genetic distance. First, all orthogroups were aligned based on protein sequence using MAFFT v.7.300 with the --auto option (Katoh et al. 2005). The genetic distances between the gene copies/transcripts were then calculated as Kimura protein distances (Kimura 1983) using the distmat algorithm embedded in the EMBOSS package (v. 6.5.7, Rice et al. 2000). Next, one gene copy/transcript per species or accession was extracted based on the smallest genetic distance to the transcript of the Alaskan accession of either *C. bellidifolia*, *D. nivalis* or *C. groenlandica,* depending on the clade in question. If there were more than one transcript from the Arctic accessions, a new subset was made for each such

transcript. The resulting orthogroup subsets contained only one transcript per species/accession and were realigned with PRANK (Löytynoja and Goldman 2005) using GUIDANCE v2.02 (Sela et al. 2015) with ten bootstraps. GUIDANCE allows assessment of alignment confidence and subsequent removal of unreliable aligned columns and sequences. Previous studies have shown that a phylogeny aware aligner such as PRANK in combination with alignment filtering with GUIDANCE improves positive selection inference on simulated data (Jordan and Goldman 2012; Privman et al. 2012). All alignments with a sequence score < 0.6, as well as all columns scoring < 0.80, were removed from the data set. The three final alignment sets had an average length of 1544-1586 bp and contained on average 17.0-20.7% missing data, but sites containing alignment gaps were deleted within codeml (cleandata = 1).

For estimation of posterior expected numbers of convergent and divergent substitutions, we ran OrthoFinder v.2.3.1 (Emms and Kelly 2019) for the total data set (fig. 1A). The resulting number of single-copy orthogroups was extremely small, and we therefore extracted subsets from multiple copy orthogroups as described above for the positive selection tests. For simplicity, we extracted subsets based on only one of the Arctic focal species, i.e. we extracted one transcript per species/accession based on the smallest genetic distance to the transcripts of *D. nivalis* (US, arbitrarily chosen). The alignments were then generated as explained above (final average alignment length was 1524 bp with an average of 25.8% missing data). Note that all alignments for both the positive selection tests and the calculation of posterior expected numbers of convergent and divergent substitutions contained all relevant species (cf. fig. 1).

The branch-site test of positive selection

The branch-site model A (Zhang et al. 2005) implemented in codeml (PAML package v.4.9d, Yang 1997) was used to test for positive selection on individual codons along the lineage leading to each of the Arctic species. In the branch-site test, an alternative model (model = 2, NSsites = 2) allowing positive selection on the foreground lineage (e.g. one of the Arctic lineages) is contrasted to a null-model that does not allow such positive selection (i.e. by adjusting fix_omega and omega to 1) using a likelihood-ratio test (LRT; see below for details; Zhang et al. 2005). To limit the influence of missing data (a challenge embedded in the use of transcriptomic data), we ran the tests separately for clade A, B and C, (Huang et al. 2016; fig. 1B-D). However, the model species *Arabidopsis thaliana* was included in all runs, and due to the low availability of high-quality transcriptomic data from Clade C, we also included two species from Clade B in the *C. groenlandica* data set (fig. 1D). The trees for the branch-site tests were made with RAxML v.8.0.26 (Stamatakis 2014; see fig. 1B-D) based on codon alignments of ~2000-4000 single-copy orthogroups created with pal2nal v.14 (Suyama et al. 2006). The alignment was partitioned by codon position, and the best fit GTRGAMMAI model was applied for each partition based on the results of MrModeltest v.2.3 (Nylander 2004). Codeml was run from

four to six times per model with different initial parameter values, and the run with the highest likelihood score was used in subsequent analyzes (see e.g. Wong et al. 2004). The LRTs were performed in R v.3.4.4 (R Core Team 2019) with the package extRemes (Gilleland and Katz 2016). The significance level was set to 0.05 with one degree of freedom. We also ran the branch-site test for each topology in figure 1B-D with a close non-Arctic relative (*C. hirsuta*, *C. pyrenaica* and *D. nemorosa*) as the foreground branch. This was done in order to filter out genes that always seem to be under positive selection and also to gain a better understanding of which genes that actually could be tied to Arctic adaptation. The resulting sets of positively selected genes were compared between species in order to identify genes with convergent selection patterns. Genes were considered as being under convergent selection if the same *A. thaliana* ortholog was found to be under positive selection in two or more species. The significance of the gene intersections was then assessed with the supertest function in SuperExactTest v.1.0.7 (Wang et al. 2015) and visualized with the UpSetR package 1.3.3 (Conway et al. 2017).

Posterior expected numbers of convergent and divergent substitutions

We used the program Grand-Convergence (Qian et al. 2018; Qian and de Koning 2018), to calculate the posterior expected numbers of convergent and divergent substitutions across all branch pairs in a super phylogeny generated with OrthoFinder v.2.3.1 (Emms and Kelly 2019; fig. 1A). This was done partly to identify sets of convergent genes (i.e. genes containing at least one convergent substitution in Arctic species pairs), and partly to examine deviations from the expected relationship between convergent and divergent substitutions. We also estimated the number of convergent genes in the non-Arctic reference species. In Grand Convergence, a convergent substitution is defined as changes resulting in the same amino acid at the same site in two lineages, and divergent substitutions as changes resulting in different amino acids at the same site in two lineages (silent substitutions are not considered; Castoe et al. 2009). The program was run with branch length estimations from OrthoFinder, the LG amino acid substitution model (Le and Gascuel 2008), and with all gaps and ambiguity characters excluded. Genes containing at least one convergent site with PP > 0.80 were considered as convergent, but since a gene/transcript potentially could be included in more than one alignment following our orthogroup subsetting procedure (see above), we only included unique TAIR IDs (i.e. *A. thaliana* orthologues) in our list of convergent genes. The resulting sets of convergent genes were compared to each other in the same way as the PSGs. This was done in order to identify genes that were shared between more than one Arctic branch pair.

To investigate the relationship between divergent and convergent substitutions in our data set, we chose one alignment per orthogroup and summed all pairwise convergent substitutions and all pairwise divergent substitutions over all alignments. Using the branch totals from fifty branch pairs, we performed a simple linear

regression to investigate if our data fitted a linear model as expected from other studies (Castoe et al. 2009; Thomas and Hahn 2015). The fifty branch-pairs were mostly arbitrarily chosen, except that all pairwise combinations between the Arctic species and between the non-Arctic reference species were included (see supplementary table S32 for a complete list of the fifty branch pairs).

Gene Annotation, Gene Ontology Enrichment tests

The final sets of PSGs and convergent genes were annotated with DAVID v.6.8 (Huang et al. 2009) based on the *A. thaliana* ortholog. We also tested for overrepresented Gene Ontology (GO) terms within the three domains Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) using a Fisher's exact test in combination with the "classic", "elim" and "weight" algorithms implemented in topGO v.2.32 of Bioconductor (Gentleman et al. 2004; Alexa et al. 2006). The "classic" algorithm processes each term independently without taking the GO-hierarchy into account, the "elim" algorithm traverses the GO-graph bottom-up, discarding genes that have already been mapped to significant child terms, and the "weight" algorithm decides which GO-terms best describe the gene based on a weighting scheme that takes the enrichment scores of the neighboring GO-terms into account (Alexa et al. 2006). Simulations have shown that the "elim" and "weight" algorithms tend to be more conservative than the "classic" algorithm, and the creators of topGO recommend using "elim" for identifying important areas in the graph due to its simplicity (Alexa et al. 2006). The gene sets were annotated based on homology to *A. thaliana* using annotations from org.At.tair.db v.3.6 (Carlson 2018), and *A. thaliana* was also used as the background gene universe in all gene set enrichment analyses. Although it could be considered less than ideal to apply the *A. thaliana* background in this type of analysis, it is likely to be more consistent than creating new backgrounds based on the annotation of *de novo* transcriptome assemblies for each species, because transcriptome assemblies represent a less complete sampling of gene space and are prone to varying degrees of sampling bias. We did not correct for multiple testing as we wanted to avoid being overly conservative, and because such corrections cannot directly be applied to the results of the "elim" and "weight" algorithm due to non-independent testing (the topGO manual; Alexa and Rahnenfuhrer 2018). Enriched GO-terms that might be tied to Arctic adaptation were assessed based on literature, and the results from the non-Arctic species were used to filter out more general GO-terms that were enriched in all species (e.g. GO-terms related to immunity). Finally, we also generated a heatmap of the proportions of PSGs associated with the five categories of Arctic stresses, by using the number of genes annotated with a selection of these GO-terms. The annotations were identical to those used in the topGO analyses, and the heatmap was generated with gplots package v.3.0.1.1 in R (Warnes et al. 2019). For a flow chart of the analysis pipeline see supplementary fig. S1.

Supplementary figures and tables

Supplementary Figures

Figure S1. The pipeline chart,
Figure S2. Heatmap showing proportions of PSGs associated with the five categories of Arctic stresses
Figure S3A-C. Upsets from Arctic topGO results of PSGs: A) results from the classic algorithm, B) results from the elim algorithm and C) results from the weight algorithm
Figure S4. Bar chart showing the number of significant PSGs annotated with "plasma membrane"
Figure S5. Scatter plot of divergent and convergent substitutions: Regression without outliers

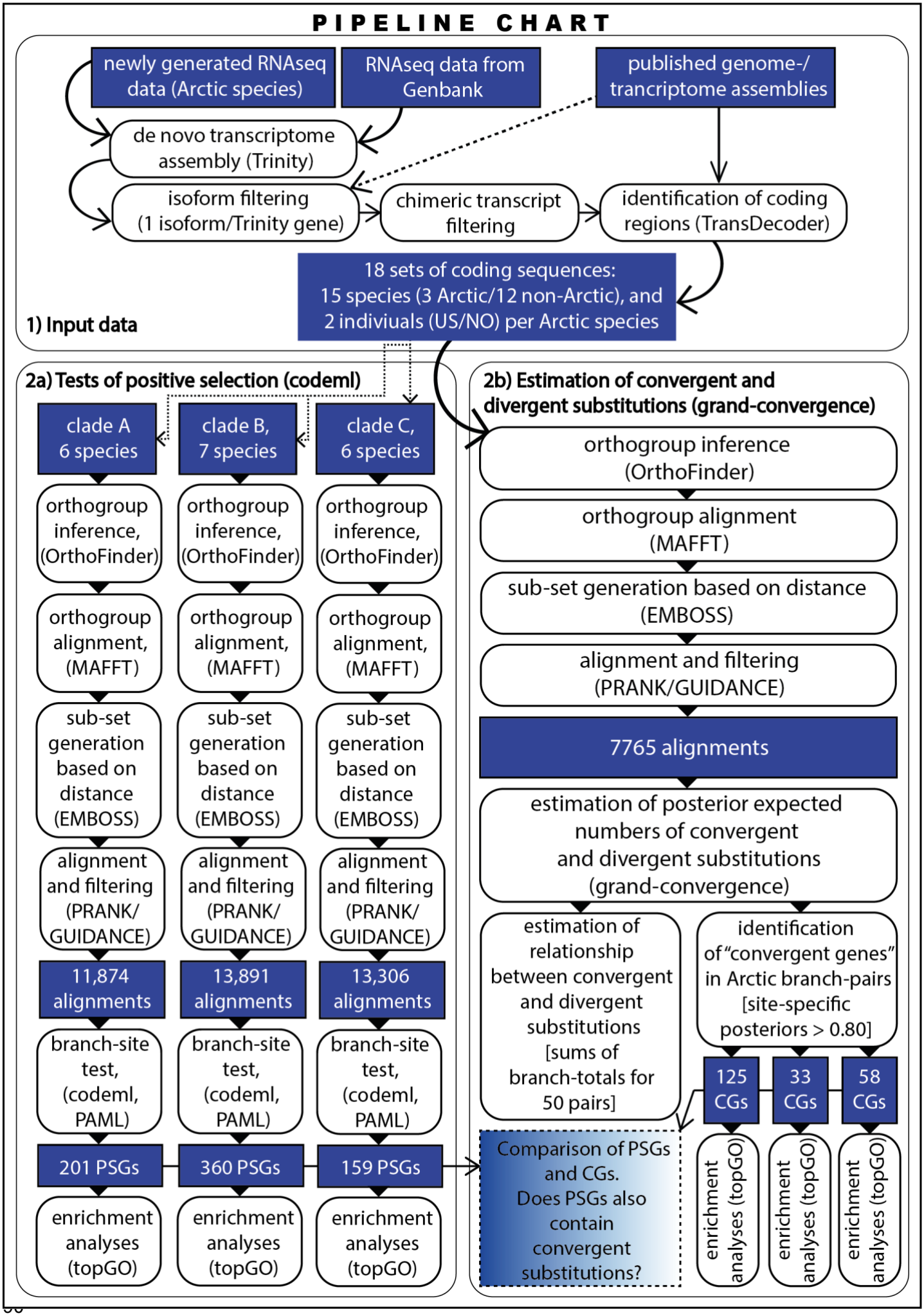
Supplementary Tables

Table S1. Sampling information for Arctic focal species
Table S2. Transcriptome assembly statistics
Table S3. Species representation
Table S4-S6. PAML tables for Arctic species. *Separate excel file.*
Table S4: *Cardamine bellidifolia*
Table S5: *Draba nivalis*
Table S6: *Cochlearia groenlandica*.
Table S7-S9. PSG annotations from DAVID for Arctic species. *Separate excel file.*
Table S7: *Cardamine bellidifolia*
Table S8: *Draba nivalis*
Table S9: *Cochlearia groenlandica*.
Table S10-S12. PAML tables for non-Arctic reference species. *Separate excel file.*
Table S10: *Cardamine hirsuta*
Table S11: *Draba nemorosa*
Table S12: *Cochlearia pyrenaica*.
Table S13-S15. PSG annotations from DAVID for non-Arctic reference species. *Separate excel file.*
Table S13: *Cardamine hirsuta*
Table S14: *Draba nemorosa*
Table S15: *Cochlearia pyrenaica*.
Table S16. Arctic candidate genes (PSGs). *Separate excel file.*
Table S17. Genes under positive selection in more than one Arctic species
Table S18. Results from the supertest. Arctic PSG intersections
Table S19-S21. topGO tables for Arctic PSGs. *Separate excel file.*
Table S19A-C: Biological Process.
Table S20A-C: Cellular Component.

Table S21A-C: Molecular Function.
Table S22-S24. topGO tables for non-Arctic PSGs. *Separate excel file.*
Table S22A-C: Biological Process.
Table S23A-C: Cellular Component.
Table S24A-C: Molecular Function.
Table S25-S30. DAVID annotations of convergent gene sets. *Separate excel file.*
Table S25: *Draba nivalis - Cardamine bellidifolia*
Table S26: *Draba nivalis - Cochlearia groenlandica*
Table S27: *Cardamine bellidifolia - Cochlearia groenlandica*
Table S28: *Draba nemorosa - Cardamine hirsuta*
Table S29: *Draba nemorosa - Cochlearia pyrenaica*
Table S30: *Cochlearia pyrenaica - Cardamine hirsuta*.
Table S31. Arctic candidate genes (Convergent genes). *Separate excel file.*
Table S32: Branch pair information for Figure 6 (Grand Convergence)
Table S33-35. topGO results for Arctic convergent gene sets. *Separate excel file.*
Table S33A-C: Biological Process.
Table S34A-C: Cellular Component.
Table S35A-C: Molecular Function.
Table S36-38. topGO results for non-Arctic convergent gene sets. *Separate excel file.*
Table S36A-C: Biological Process.
Table S37A-C: Cellular Component.
Table S38A-C: Molecular Function.

For excel files, see: https://doi.org/10.5061/dryad.7pvmcvdqx

or: https://academic.oup.com/mbe/article/37/7/2052/5804990


PIPELINE CHART
newly generated RNAseq data (Arctic species)
RNAseq data from Genbank
published genome-/ trancriptome assemblies
de novo transcriptome assembly (Trinity)
isoform filtering (1 isoform/Trinity gene)
chimeric transcript filtering
identification of coding regions (TransDecoder)
18 sets of coding sequences: 15 species (3 Arctic/12 non-Arctic), and 2 indiviuals (US/NO) per Arctic species
1) Input data
2a) Tests of positive selection (codeml)
clade A 6 species
clade B, 7 species
clade C, 6 species
orthogroup inference, (OrthoFinder)
orthogroup alignment, (MAFFT)
sub-set generation based on distance (EMBOSS)
alignment and filtering (PRANK/ GUIDANCE)
11,874 alignments
13,891 alignments
13,306 alignments
branch-site test, (codeml, PAML)
201 PSGs
360 PSGs
159 PSGs
enrichment analyses (topGO)
2b) Estimation of convergent and divergent substitutions (grand-convergence)
orthogroup inference (OrthoFinder)
orthogroup alignment (MAFFT)
sub-set generation based on distance (EMBOSS)
alignment and filtering (PRANK/GUIDANCE)
7765 alignments
estimation of posterior expected numbers of convergent and divergent substitutions (grand-convergence)
estimation of relationship between convergent and divergent substitutions [sums of branch-totals for 50 pairs]
identification of "convergent genes" in Arctic branch-pairs [site-specific posteriors > 0.80]
125 CGs
33 CGs
58 CGs
enrichment analyses (topGO)
Comparison of PSGs and CGs. Does PSGs also contain convergent substitutions?

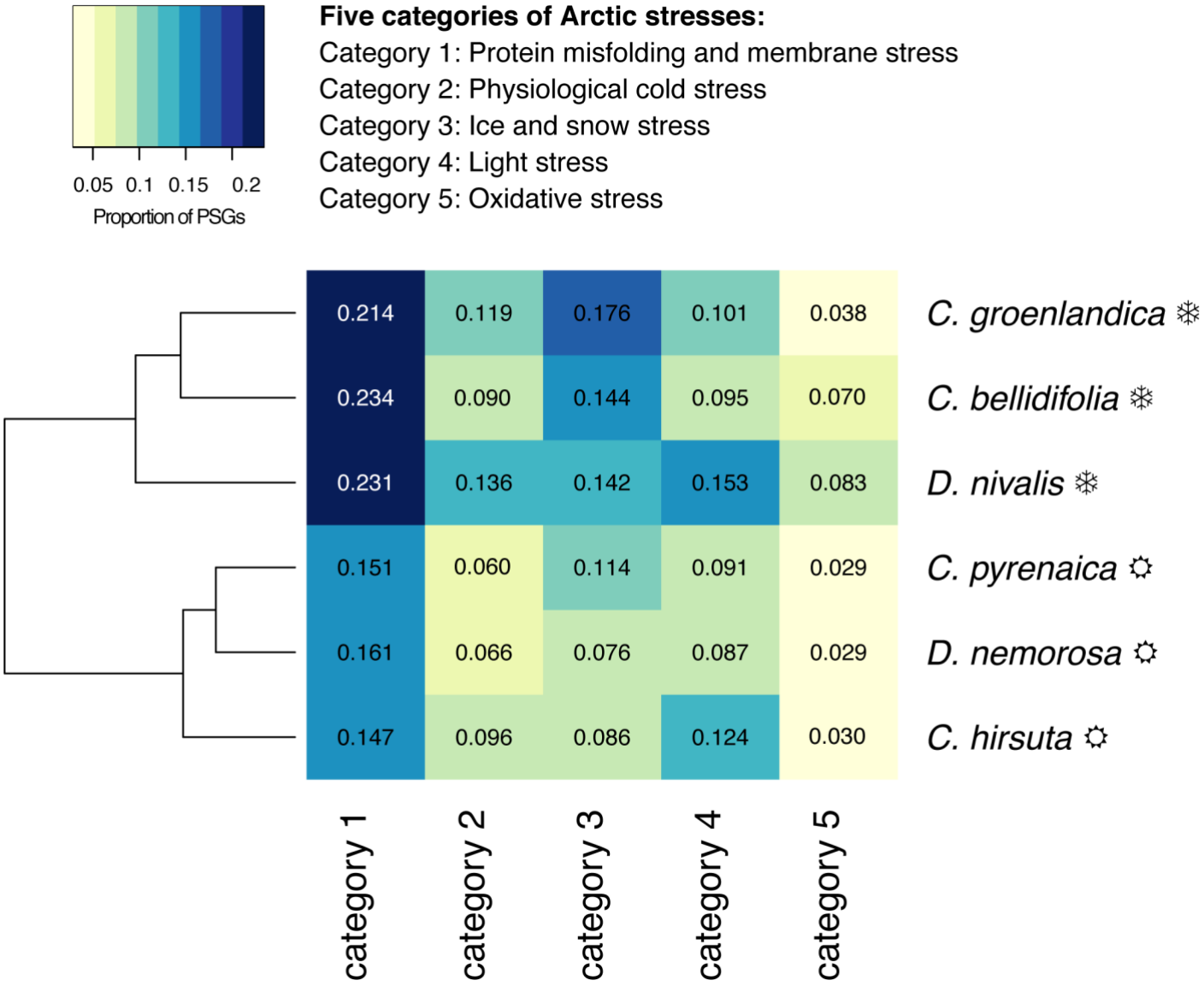


FIG S2. Heatmap showing the proportion of PSGs associated with the five categories of Arctic stresses. Arctic species are indicated with ❄ and non-Arctic species are indicated with ☼. The colors should be compared within columns. The heatmap is based on the number of genes annotated with the following selection of GO-terms (within the Biological Process domain, unless indicated with CC = Cellular Component or MF = Molecular Function): ***Category 1)*** *plasma membrane (CC), endomembrane system organization; protein folding; response to misfolded protein,* ***Category 2)*** *response to temperature; response to cold,* ***Category 3)*** *response to freezing; response to water deprivation; response to osmotic stress; response to salt stress,* ***Category 4)*** *response to light stimulus; photoperiodism; entrainment of circadian clock by photoperiod; response to blue light; response to red or far red light; response to far red light; response to light intensity,* ***Category 5)*** *response to oxidative stress; response to reactive oxygen species; glutathione biosynthetic process; glutathione peroxidase activity (MF).*

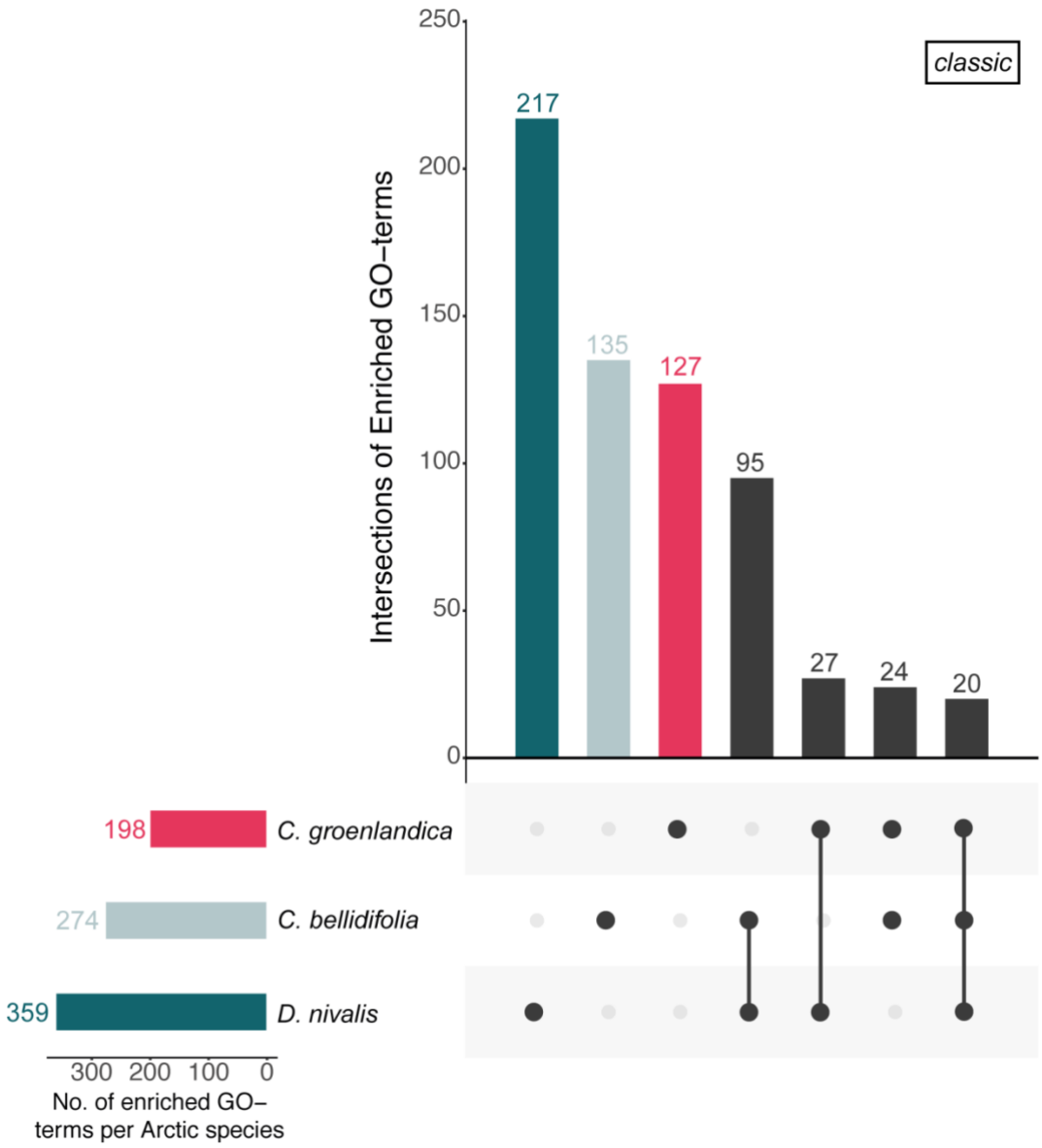


FIG S3A. UpSet plot of significantly enriched GO-terms within the Biological Process domain using the CLASSIC algorithm. The plot in the left corner shows total numbers of significantly enriched categories per species, while the main plot shows the number of enriched GO-terms that are unique per species, followed by the numbers of enriched GO-terms that are overlapping between species. Dark green = *Draba nivalis*, grey = *Cardamine bellidifolia*, red = *Cochlearia groenlandica*, black = intersections between species.

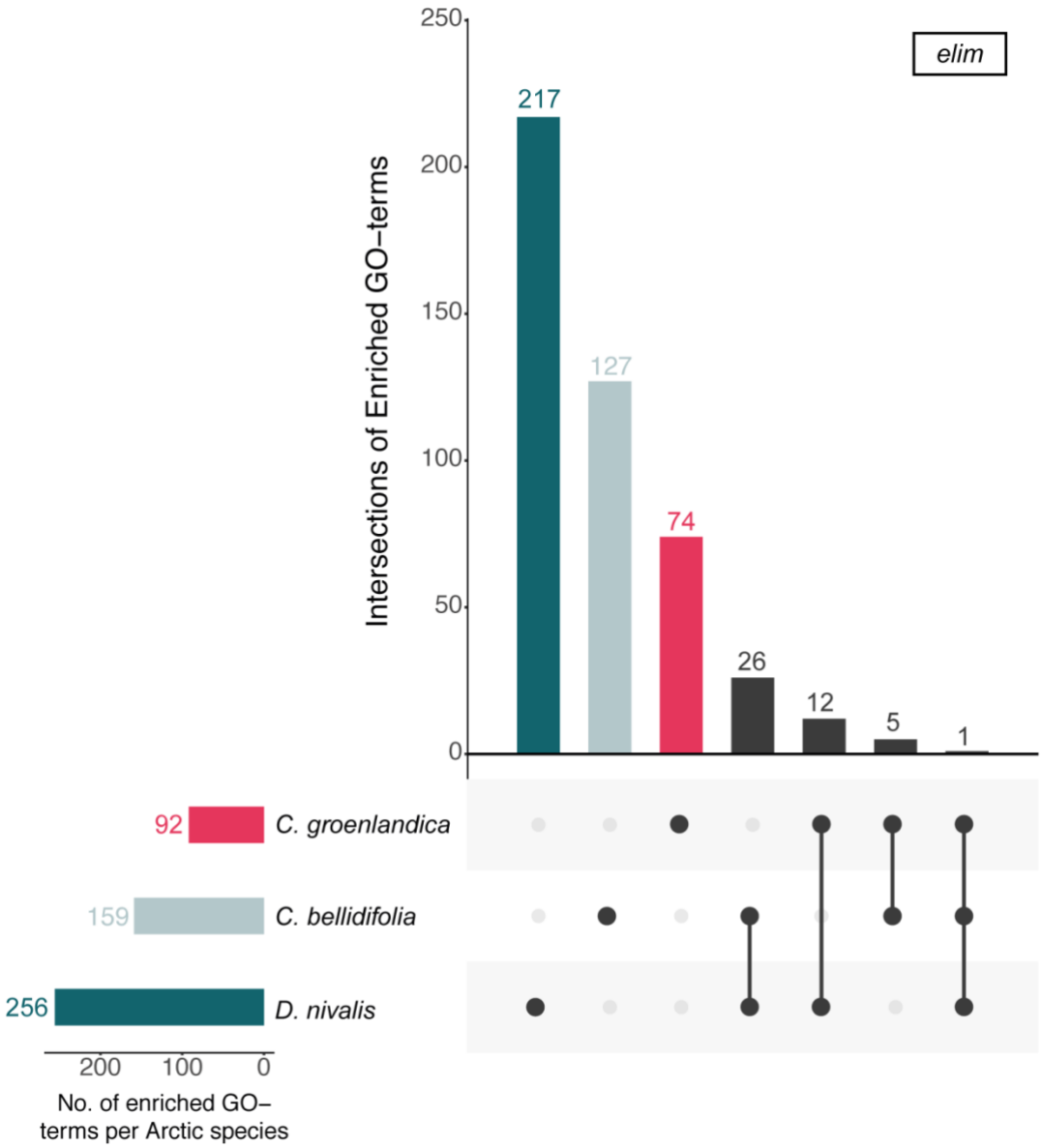


FIG S3B. UpSet plot of significantly enriched GO-terms within the Biological Process domain using the ELIM algorithm. The plot in the left corner shows total numbers of significantly enriched categories per species, while the main plot shows the number of enriched GO-terms that are unique per species, followed by the numbers of enriched GO-terms that are overlapping between species. Dark green = *Draba nivalis*, grey = *Cardamine bellidifolia*, red = *Cochlearia groenlandica*, black = intersections between species.

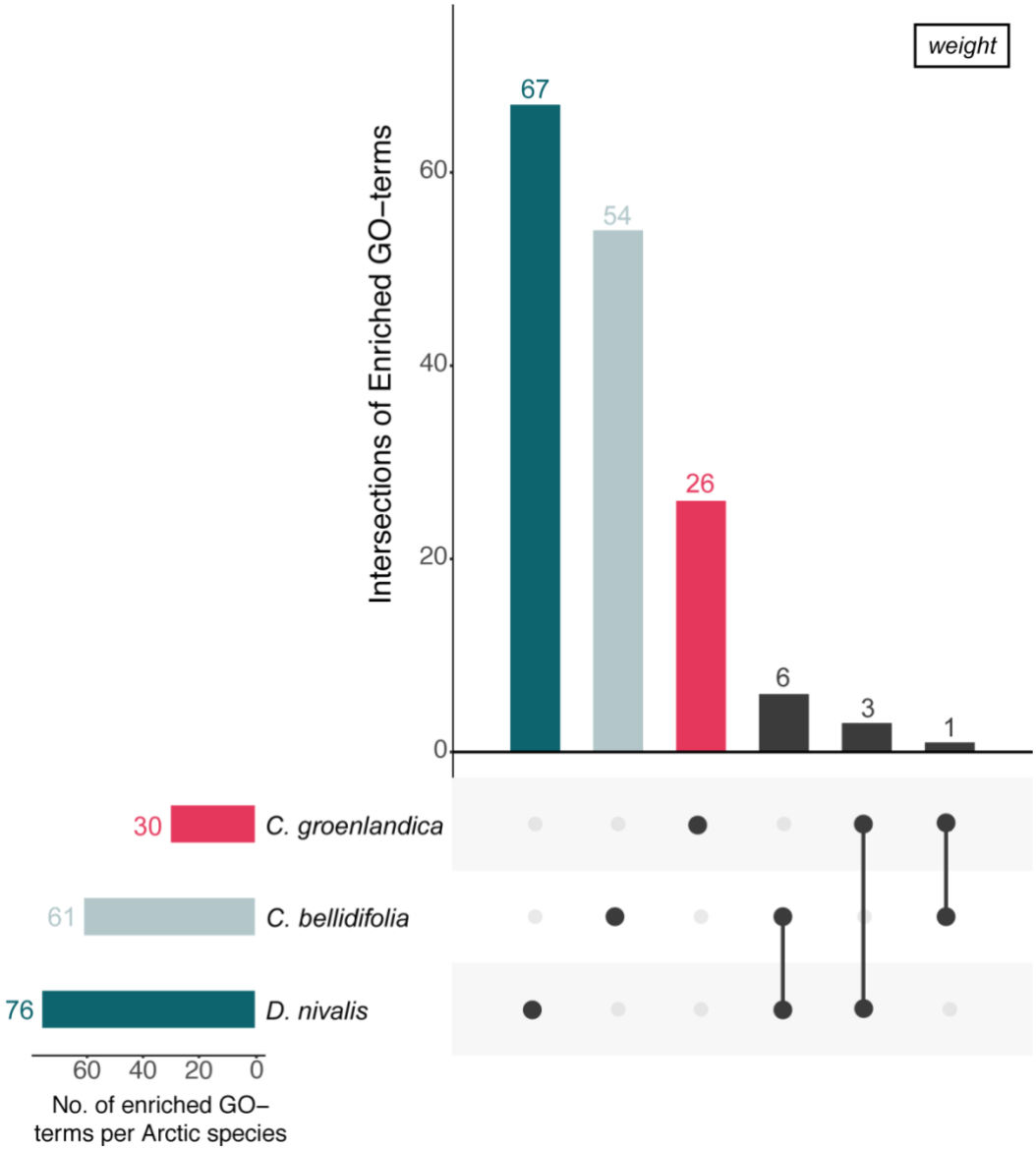


FIG S3C. UpSet plot of significantly enriched GO-terms within the Biological Process domain using the WEIGHT algorithm. The plot in the left corner shows total numbers of significantly enriched categories per species, while the main plot shows the number of enriched GO-terms that are unique per species, followed by the numbers of enriched GO-terms that are overlapping between species. Dark green = *Draba nivalis*, grey = *Cardamine bellidifolia*, red = *Cochlearia groenlandica*, black = intersections between species.

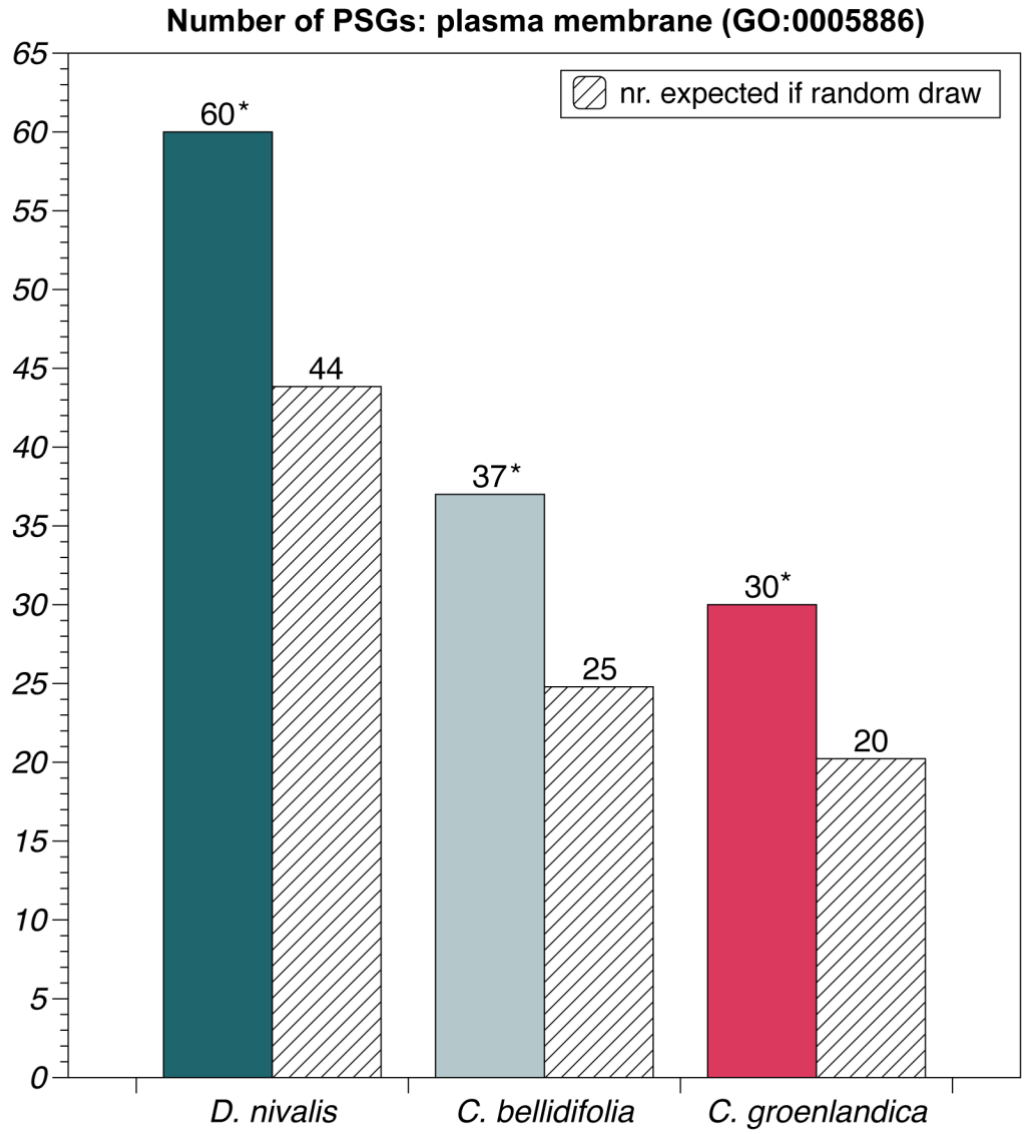


FIG S4. Bar chart showing the number of PSGs that showed significant enrichment of the GO-term "plasma membrane" (CC) in Arctic lineages. Expected numbers of genes (if randomly drawn) are indicated as hatched bars. Dark green = *Draba nivalis*, grey = *Cardamine bellidifolia*, red = *Cochlearia groenlandica*. Stars (*) indicate that the GO-term was significant when applying the classic algorithm in topGO ($p < 0.05$).

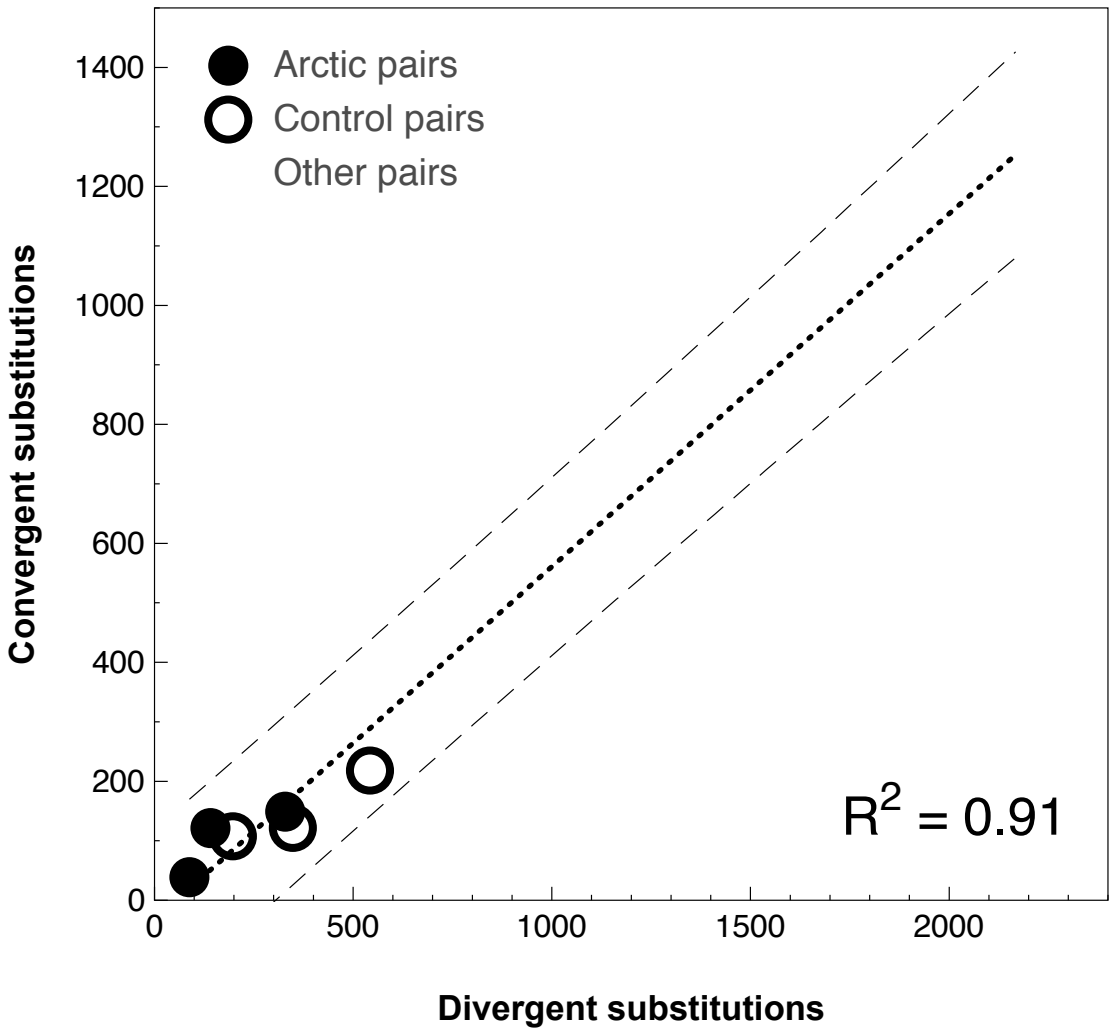


FIG S5. Posterior expected numbers of convergent and divergent substitutions between 50 branch-pairs from the full species phylogeny (fig. 1A), but with outliers removed. Numbers are sums over all alignments. The correlation $R^2 = 0.91$ indicates a strong linear relationship similar to what has been found in other studies (Castoe et al. 2009; Thomas and Hahn 2015). Dotted line = the regression line, dashed line = 95 % prediction intervals, filled black circles = the three Arctic branch-pair comparisons, open circles = the branch-pairs of the non-Arctic reference species, filled grey circles = the remaining branch-pairs.

Table S1. Sampling information for Arctic focal species

Species	Population ID	Region (Country)	Locality	Coordinates (DMS Latitude, DMS Longitude)
Cardamine bellidifolia	LG09-A-109-01	The border between Alaska (US) and Yukon (CA)	Goldensides Trail, Tombstone Territorial Park	N64°31'43.74" W138°11'57.42"
Cardamine bellidifolia	LG09-S-32-03	Svalbard (NO)	Sarkofagen, Longyearbyen	N78°9'59.9" E15°32'0.1"
Cochlearia groenlandica	LG09-A-51-08	Alaska (US)	Woolley Lagoon, Seward Peninsula	N64°51'18.7" W166°24'6.7"
Cochlearia groenlandica	LG09-S-04-10	Svalbard (NO)	Nybyen, Longyearbyen	N78°12' 20.7" E15°35'53.4"
Draba nivalis	008-7	Alaska (US)	Waterfall Creek W	N63°2'42' W147°12'3.6"
Draba nivalis	045-5	Oppland (NO)	Verkensetri, Grimsdalen, Rondane	N61°49'58.8" E9°30'0"

Table S2. Transcriptome assemblies – statistics

Species (source)	Tissue	No. raw sequences	No. of transcripts in raw assembly	No. of "genes" in raw assembly[a]	No. of transcripts in final assembly	% BUSCOs present in isoform filtered cds (Complete, Fragmented)[b]
Arabis verna	leaves	32,986,341[c]	88,177	26,146	21,810	80.4% [C: 75.3%, F: 5.1%]
Aubrieta canescens	leaves	57,955,397[c]	48,802	21,357	19,561	88.8% [C: 85.4%, F: 3.4%]
Barbarea vulgaris	leaves	11,415,972[c]	36,210	21,593	18,830	67% [C: 55.0%, F: 12.0%]
Cardamine bellidifolia (CA/US)	leaves	15,369,844	63,548	39,028	27,218	75.4% [C: 64.1%, F: 11.3%]
Cardamine bellidifolia (NO)	leaves	21,630,446	62,918	37,190	25,501	67.9% [C: 50.9%, F: 17.0%]
Cardamine hupingshanensis	roots, leaves	50,049,009[c]	111,011	31,848	23,577	80.9% [C: 73.9%, F: 7.0%]
Cochlearia groenlandica (NO)	leaves	19,795,668	60,310	36,411	26,728	77.9% [C: 66.4%, F: 11.5%]
Cochlearia groenlandica (US)	leaves	16,115,578	56,202	37,224	26,995	75.3% [C: 62.5%, F: 12.8%]

Cochlearia pyrenaica	Lopez et al. (2017)	Lopez et al. (2017)	62,716	37,632	22,633	88.1%[d] [C: 78.0%, F: 10.1%]
Draba hispanica	leaves	38,616,893[c]	51,417	22,185	19,707	88.1% [C: 84.7%, F: 3.4%]
Draba nemorosa	leaves	39,942,436[c]	51,618	21,517	19,590	89.8% [C: 86.7%, F: 3.1%]
Draba nivalis (NO)	leaves	19,683,645	64,623	41,194	28,679	78.8% [C: 66.1%, F: 12.7%]
Draba nivalis (US)	leaves	25,762,936	59,357	38,840	28,120	83.2% [C: 74.3%, F: 8.9%]
Nasturtium officinale	leaves, stem	28,128,352[c]	116,301	43,886	34,226	83.9% [C: 78.3%, F: 5.6%]
Pugionium cornutum	leaves	27,387,868[c]	61,981	25,388	20,980	71.8% [C: 59.1%, F: 12.7%]

[a]Corresponding to Trinity genes (or transcript clusters), [b]The presented BUSCO values are for coding sequences made from isoform filtered assemblies. As we only retained the highest expressed isoform per Trinity gene (the longest isoform has a higher chance of containing misassemblies), there was a decrease in complete BUSCOs and an increase in fragmented BUSCOs, [c]based on multiple reads or a subset of available reads (see Table S1), [d]Only the longest isoform was retained for *C. pyrenaica*

Table S3. Brassicaceae species representation

Species	Clade in Brassicaceae[a]	Ploidy	Distribution	Assembly/SRA ID (ref.)
Cardamine bellidifolia	A	2x	Circumpolar-alpine[b]	RNAseq performed for this study. BioProject ID: PRJNA599963, Short-read archive ID: SRP240404. Assemblies available at https://doi.org/10.5061/dryad.v41ns1rs0
Cardamine hirsuta	A	2x	World-wide, but native to Europe/Asia	carhr38.cds.fa/carhr38.aa.fa *downloaded from* http://chi.mpipz.mpg.de
Cardamine hupingshanensis	A	2x? 2n = 24?	South-East Asia, restricted to China (Bai et al. 2008; Zhou et al. 2018)	BioProject ID: PRJNA360038, Short-read archive ID: SRP097726 (Zhou et al. 2018)
Nasturtium officinale	A	4x	World-wide, but native to Europe and Asia aquatic	BioProject ID: PRJNA284126, Short-read archive ID: SRP058520 (Voutsina et al. 2016)
Barbarea vulgaris	A	2x	European - Asian (W)[b]	BioProject ID: PRJNA261637, Short-read archive ID: SRP047356 (Wei et al. 2013)
Cochlearia groenlandica	C	2x	Circumpolar[b]	RNAseq performed for this study. BioProject ID: PRJNA599963, Short-read archive ID: SRP240404. Assemblies available at https://doi.org/10.5061/dryad.v41ns1rs0
Cochlearia pyrenaica	C	2x	European Arctic-montane element, but absent from the Boreal zonobiome (Online atlas of the British and Irish flora)	The full transcriptome assembly of *C. pyrenaica* was made available from Lopez et al. (2017) pre-publication

Pugionium cornotum	C	2x? 2n=22	Central Asia (desert)	BioProject ID: PRJNA357188, Short-read archive ID: SRP095911 (Wang et al. 2017)
Draba nivalis	B	2x	Circumpolar-alpine[b]	RNAseq performed for this study. BioProject ID: PRJNA599963, Short-read archive ID: SRP240404. Assemblies available at *https://doi.org/10.5061/dryad.v41ns1rs0*
Draba hispanica	B	2x	Native to Spain (mountainous regions 1500 above sea level)	BioProject ID: PRJEB9424, Short-read archive ID: ERP010528 (Heidel et al. 2016)
Draba nemorosa	B	2x	Circumboreal[b]	BioProject ID: PRJEB9424, Short-read archive ID: ERP010528 (Heidel et al. 2016)
Arabis alpina	B	2x	Amphi-Atlantic - European - Asian (W) & tropical mountains[b]	V5 *downloaded from http://www.arabis-alpina.org* (BioProject ID: PRJNA241291)
Arabis verna	B	4x?	Mediterranean region/West Asia	BioProject ID: PRJEB9424, Short-read archive ID: ERP010528 (Heidel et al. 2016)
Aubrieta canescens	B	2x		BioProject ID: PRJEB9424, Short-read archive ID: ERP010528 (Heidel et al. 2016)
Arabidopsis thaliana	A	2x	World-wide, native to Europe, Asia and Africa	TAIR10_cds_20110103_representative_gene_model_updated.cds TAIR10_pep_20110103_representative_gene_model_updated.pep *downloaded from https://www.arabidopsis.org*

[a]Classified according to Huang et al. (2016) and Guo et al. (2017), [b]Elven et al. (2011)

Table S17. Genes found to be under positive selection in more than one Arctic species (indicated by [+] if found to be under positive selection in an Arctic species and [-] if not). If the gene also was found to be under positive selection in a non-Arctic species, the species name is given. Tair ID = Gene ID of shared *Arabidopsis thaliana* orthologue, Gene description = Gene information associated with the *A. thaliana* orthologue from The Arabidopsis Information Resource [TAIR] 2019, GO Biological Process/GO Cellular Component/GO Molecular function = Gene Ontology terms associated with the *A. thaliana* orthologue within each of the three domains (The Arabidopsis Information Resource [TAIR] 2019), Putative Arctic association = Possible connection to the five categories of Arctic stresses.

C. bellidifolia	*C. groenlandica*	*D. nivalis*	Non-Arctic species?	TAIR ID	Gene description	GO Biological Process	GO Cellular Component	GO Molecular Function	Putative Arctic association
+	-	+		AT1G54030	**ENDOPLASMIC RETICULUM MORPHOLOGY 3 (ERMO3)** encodes a vacuolar protein. Mutation causes organizational defects in the endoplasmic reticulum and aberrant protein trafficking in the plant secretory pathway. The mRNA is cell-to-cell mobile.	*involved in:* endoplasmic reticulum organization, intracellular protein transport, protein exit from endoplasmic reticulum	*located in:* ER body, cell, chloroplast, endoplasmic reticulum, intracellular, plant-type vacuole membrane, plasmodesma, vacuole	*has:* carboxylic ester hydrolase activity, hydrolase activity, acting on ester bonds, lipase activity	Protein misfolding and membrane stress
+	-	+	*C. hirsuta*	AT2G34310	**hypothetical protein**;(source:Araport11)	*involved in:* biological_process	*located in:* integral component of membrane, nucleus	*has:* GTP binding	

+	-	+		AT2G48120	The **pale cress (pac)** mutant affects chloroplast and leaf development; mutants are ABA-deficient and accumulate lower levels of carotenoids and chlorophyll compared to wild type. PAC binds 23srRNA and appears to be required for 50s ribosome assembly. Three alternative transcripts of this gene exist. PAC is essential for photoautotrophic growth and associates with psbK-psbI, ndhF, ndhD, and 23S ribosomal RNA in vivo (PMID:28805278)	*involved in:* RNA metabolic process, carotenoid biosynthetic process, cellular response to light stimulus, chloroplast mRNA processing, chloroplast organization, leaf morphogenesis, ribosome biogenesis	*located in:* amyloplast, chloroplast, chloroplast nucleoid, chloroplast stroma, chromoplast, etioplast, proplastid	*functions in:* RNA binding, protein self-association, rRNA binding	Light stress
+	-	+	*C. hirsuta*	AT3G06340	**DNAJ heat shock N-terminal domain-containing protein**;(source:Araport11)	*involved in:* protein folding	*located in:* cytoplasm		Protein misfolding and membrane stress
+	-	+		AT4G10320	**tRNA synthetase class I (I, L, M and V) family protein**;(source:Araport11)	*involved in:* isoleucyl-tRNA aminoacylation, response to cadmium ion, tRNA aminoacylation for protein translation	*located in:* cytoplasm, cytosol	*functions in:* ATP binding *has:* aminoacyl-tRNA ligase activity, isoleucine-tRNA ligase activity	
+	-	+		AT4G39190	**nucleolar-like protein**;(source:Araport11)	*involved in:* biological_process	*located in:* integral component of membrane	*has:* molecular_function	

+	-	+		AT5G04540	**MYOTUBULARIN 2 (MTM2)** Myotubularin-like phosphatases II superfamily;(source:Araport11)		*located in:* nucleus		
-	+	+		AT3G04730	**early auxin-induced (IAA16)**	*involved in:* auxin-activated signaling pathway, regulation of transcription, DNA-templated, response to auxin	*located in:* nucleus	*has:* DNA-binding transcription factor activity, protein binding	Light stress[a]
-	+	+		AT3G12340	**FKBP-like peptidyl-prolyl cis-trans isomerase family protein**;(source:Araport11)		*located in:* chloroplast thylakoid lumen, nucleus	*functions in:* FK506 binding *has:* peptidyl-prolyl cis-trans isomerase activity	
+	+	-	*D. nemorosa*	AT1G14870	**PLANT CADMIUM RESISTANCE 2 (PCR2)** encodes a membrane protein involved in zinc transport and detoxification.	*involved in:* response to oxidative stress	*located in:* cytoplasm, integral component of membrane, plasma membrane	*has:* molecular_function	Oxidative stress
+	+	-		AT3G22480	**prefoldin 2**;(source:Araport11)	*involved in:* protein folding	*located in:* cytoplasm, prefoldin complex	*has:* protein folding chaperone, unfolded protein binding	Protein misfolding and membrane stress

+	+	-		AT4G20830	**OLIGOGALACTURONIDE OXIDASE 1 (OGOX1)** encodes an oligogalacturonide oxidase that inactivates the elicitor-active oligogalacturonides (OGs). It is involved in plant immunity. Overexpressing plants are more resistant to *Botrytis cinerea*.	*involved in:* defense response to fungus, oxidation-reduction process, response to oxidative stress	*located in:* apoplast, chloroplast, cytosol, mitochondrion, plant-type cell wall, plasma membrane, plasmodesma, vacuole	*has:* FAD binding, electron transfer activity, oxidoreductase activity, acting on the CH-OH group of donors, oxygen as acceptor	Oxidative stress
+	+	-	*C. pyrenaica*	AT4G24380	**dihydrofolate reductase**;(source:Araport11)		*located in:* chloroplast, cytoplasm, mitochondrion		

[a]Rinaldi et al. (2012)

NOTE: We performed the supertest with three different types of backgrounds. In A) we extracted the number of shared *A. thaliana* orthologues that were present and tested in each comparison. For instance was 10,509 similar TairIDs (*A. thaliana* gene IDs) present and testet in both *C. groenlandica* and *D. nivalis*. This included duplicates, because the same TairID sometimes is included in different alignments. In B) we did the same as in A), but here we excluded duplicates (so only unique Tair Ids present and tested in each comparison). In C) we just used the smallest number of alignments tested (which was 11,874 for *C. bellidifolia*).

Table S18A. Results from the supertest with background = number of shared unique TAIR IDs present and tested in all data sets – duplicates included

Intersection	Background	Observed overlap	Expected overlap	Fold Enrichment	*p-value*
C. groenlandica, D. nivalis	10509	2	5.42	0.37	0.97
C. bellidifolia, D. nivalis	8859	7	8.04	0.87	0.70
C. bellidifolia, C. groenlandica	8314	4	3.81	1.05	0.53
C. groenlandica, C. bellidifolia, D. nivalis	7651	0	0.19	0.00	1.00

Significance of observed intersections between the positively selected gene sets of *D. nivalis* (358 genes), *C. groenlandica* (159 genes), and *C. bellidifolia* (199 genes). The background was set to the number of alignments containing a shared *A. thaliana* orthologue in each comparison.

Table S18B. Results from the supertest with background = number of shared unique TAIR IDs present and tested in all data sets – duplicates excluded

Intersection	Background	Observed overlap	Expected overlap	Fold Enrichment	*p-value*
C. groenlandica, D. nivalis	7632	2	7.46	0.27	1.00
C. bellidifolia, D. nivalis	7122	7	10.00	0.70	0.88
C. bellidifolia, C. groenlandica	6698	4	4.72	0.85	0.70
C. groenlandica, C. bellidifolia, D. nivalis	5562	0	0.37	0.00	1.00

Significance of observed intersections between the positively selected gene sets of *D. nivalis* (358 genes), *C. groenlandica* (159 genes), and *C. bellidifolia* (199 genes). The background was set to the number of unique *A. thaliana* orthologues present and tested in all sets.

Table S18C. Results from the supertest with background = smallest number of alignments tested (11,874)

Intersection	Observed overlap	Expected overlap	Fold Enrichment	*p-value*
C. groenlandica, D. nivalis	2	4.79	0.42	0.96
C. bellidifolia, D. nivalis	7	6.00	1.17	0.39
C. bellidifolia, C. groenlandica	4	2.66	1.50	0.28
C. groenlandica, C. bellidifolia, D. nivalis	0	0.08	0.00	1.00

Significance of observed intersections between the positively selected gene sets of *D. nivalis* (358 genes), *C. groenlandica* (159 genes), and *C. bellidifolia* (199 genes). Background was set to the lowest number of alignment tested (11,874)

Table S32. Posterior expected numbers of divergent and convergent substitutions for the 50 branch pairs in figure 6. Arctic species indicated in light blue.

	Branch 1	Branch 2	Divergent	Convergent
1	*Arabidopsis thaliana*	***Cardamine bellidifolia***	1137.38	652.08
2	*Arabis verna*	*Arabis alpina*	1027.03	676.03
3	*Arabis verna*	*Arabidopsis thaliana*	1190.31	600.44
4	*Arabis verna*	*Barbarea vulgaris*	888.72	575.13
5	*Arabis verna*	***Cardamine bellidifolia***	526.48	295.72
6	*Arabis verna*	***Draba nivalis***	389.79	215.76
7	*Arabis verna*	*Nasturtium officinale*	995.53	597.78
8	*Aubrieta canescens*	*Arabidopsis thaliana*	750.30	312.61
9	*Aubrieta canescens*	*Arabis alpina*	679.22	379.94
10	*Aubrieta canescens*	*Barbarea vulgaris*	508.56	284.50
11	*Aubrieta canescens*	*Cardamine bellidifolia*	338.91	143.88
12	*Aubrieta canescens*	***Cochlearia groenlandica***	92.34	44,00
13	*Aubrieta canescens*	***Draba nivalis***	277.22	162.99
14	*Aubrieta canescens*	*Pugionium cornutum*	875.12	466.40
15	*Barbarea vulgaris*	*Arabis alpina*	1186.95	679.30
16	*Barbarea vulgaris*	*Pugionium cornutum*	1783.14	1170.93
17	*Cardamine hirsuta*	*Arabidopsis thaliana*	1534.01	1359.20
18	*Cardamine hirsuta*	*Arabis verna*	739.38	296.24
19	*Cardamine hirsuta*	*Aubrieta canescens*	523.17	195.49
20	*Cardamine hirsuta*	***Cochlearia groenlandica***	196.17	80.52
21	*Cardamine hirsuta*	*Cochlearia pyrenaica*	348.37	121.75
22	*Cardamine hirsuta*	*Draba nemorosa*	542.59	217.88
23	*Cardamine hirsuta*	***Draba nivalis***	478.67	194.15
24	*Cardamine hupingshanensis*	*Aubrieta canescens*	378.22	189.92
25	*Cardamine hupingshanensis*	*Barbarea vulgaris*	853.69	682.03
26	*Cardamine hupingshanensis*	***Cochlearia groenlandica***	144.05	74.88
27	*Cardamine hupingshanensis*	***Draba nivalis***	373.22	130.52
28	***Cochlearia groenlandica***	***Cardamine bellidifolia***	141.07	121.39
29	*Cochlearia pyrenaica*	*Arabidopsis thaliana*	472.02	235.22
30	*Cochlearia pyrenaica*	*Arabis alpina*	349.10	179.72
31	*Cochlearia pyrenaica*	*Aubrieta canescens*	201.11	96.96
32	*Cochlearia pyrenaica*	*Cardamine bellidifolia*	224.12	108.01
33	*Cochlearia pyrenaica*	*Draba nemorosa*	197.19	107.50
34	*Cochlearia pyrenaica*	***Draba nivalis***	147.99	40.58
35	*Cochlearia pyrenaica*	*Nasturtium officinale*	419.64	234.98
36	*Draba hispanica*	*Arabidopsis thaliana*	1057.28	436.72
37	*Draba hispanica*	***Cochlearia groenlandica***	127.40	71.71
38	*Draba nemorosa*	*Arabidopsis thaliana*	940.32	409.65
39	*Draba nemorosa*	*Arabis verna*	502.62	435.76
40	*Draba nemorosa*	*Barbarea vulgaris*	574.60	434.35
41	*Draba nemorosa*	***Cardamine bellidifolia***	375.64	240.56
42	*Draba nemorosa*	***Cochlearia groenlandica***	118.21	58.07

43	*Draba nemorosa*	*Pugionium cornutum*	978.57	527.21
44	***Draba nivalis***	***Cardamine bellidifolia***	328.84	148.91
45	***Draba nivalis***	***Cochlearia groenlandica***	88.18	38.59
46	*Nasturtium officinale*	***Cochlearia groenlandica***	229.193	136.84
47	*Nasturtium officinale*	***Draba nivalis***	638.625	231.29
48	*Pugionium cornutum*	*Arabis alpina*	2167.29	855.21
49	*Pugionium cornutum*	***Cardamine belldifolia***	1161.72	794.14
50	*Pugionium cornutum*	***Draba nivalis***	869.354	351.42

References

Alexa A, Rahnenfuhrer J. 2018. topGO: Enrichment Analysis for Gene Ontology. R package version 2.34.0.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 22(13):1600–1607.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bai H-F, Chen L-B, Liu K-M, Liu L-H. 2008. A new species of *Cardamine* (Brassicaceae) from Hunan, China. *Novon A J Bot Nomencl*. 18(2):135–137.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 34(5):525–527.

Brochmann C, Soltis PS, Soltis DE. 1992. Multiple origins of the octoploid Scandinavian endemic *Draba cacuminum*: electrophoretic and morphological evidence. *Nord J Bot*. 12(3):257–272.

Carlson M. 2018. org.At.tair.db: Genome wide annotation for *Arabidopsis*. R package version 3.7.0.

Castoe TA, de Koning APJ, Kim H-M, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci U S A*. 106(22):8986–8991.

Chinnusamy V, Zhu J, Zhu JK. 2007. Cold stress regulation of gene expression in plants. *Trends Plant Sci*. 12(10):444–451.

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 33(18):2938–2940.

Du YY, Wang PC, Chen J, Song CP. 2008. Comprehensive functional analysis of the catalase gene family in Arabidopsis thaliana. *J Integr Plant Biol*. 50(10):1318–1326.

Elven R, Murray DF, Razzhivin VY, Yurtsev BA. 2011. Annotated checklist of the Panarctic Flora (PAF). Vascular plants. [cited 2019 Sep 27]. Available from: http://panarcticflora.org/.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 16(1):157.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20:1–14.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor : open software development for computational biology and bioinformatics. *Genome Biol*. 5(10):R80.

Gilleland E, Katz RW. 2016. extRemes 2.0: An extreme value analysis package in R. *J Stat Softw*. 72(8).

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 29(7):644–652.

Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D, Ma T, Hu Q, Al-Shehbaz IA, et al. 2017. Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics*. 18(1):176.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Philip D, Bowden J, Couger MB, Eccles D, Li B, Macmanes MD, et al. 2013. De novo transcript sequence reconstruction from

RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc*. 8(8):1–43.
Heidel AJ, Kiefer C, Coupland G, Rose LE. 2016. Pinpointing genes underlying annual/perennial transitions with comparative genomics. *BMC Genomics*. 17(1):921.
Huang CH, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz IA, Edger PP, et al. 2016. Resolution of brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol Biol Evol*. 33(2):394–412.
Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 4(1):44–57.
Hundertmark M, Hincha DK. 2008. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*. 9(118):1–22.
Jordan G, Goldman N. 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol Biol Evol*. 29(4):1125–1139.
Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33(2):511–518.
Khanna R, Shen Y, Toledo-ortiz G, Kikis EA, Johannesson H, Hwang Y, Quail PH. 2006. Functional profiling reveals that only a small number of phytochrome-regulated early-response genes in *Arabidopsis* are necessary for optimal deetiolation. *Plant Cell*. 18:2157–2171.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359.
Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol*. 25(7):1307–1320.
Lopez L, Wolf EM, Pires JC, Edger PP, Koch MA. 2017. Molecular resources from transcriptomes in the Brassicaceae family. *Front Plant Sci*. 8:1–15.
Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. 102(30):10557–10562.
Menges M, Dóczi R, Ökrész L, Morandini P, Mizzi L, Soloviev M, Murray JAH, Bögre L. 2008. Comprehensive gene expression atlas for the *Arabidopsis* MAP kinase signalling pathways. *New Phytol*. 179:643–662.
Navarro-Retamal C, Bremer A, Alzate-Morales J, Caballero J, Hincha DK, González W, Thalhammer A. 2016. Molecular dynamics simulations and CD spectroscopy reveal hydration-induced unfolding of the intrinsically disordered LEA proteins COR15A and COR15B from Arabidopsis thaliana. *Phys Chem Chem Phys*. 18(37):25806.
Nylander JAA. 2004. MrModeltest v2. Available from: https://github.com/nylander/MrModeltest2.
Oono Y, Seki M, Satou M, Iida K, Akiyama K, Sakurai T, Fujita M, Yamaguchi-Shinozaki K, Shinozaki K. 2006. Monitoring expression profiles of *Arabidopsis* genes during cold acclimation and deacclimation using DNA microarrays. *Funct Integr Genomics*. 6:212–234.
Privman E, Penn O, Pupko T. 2012. Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol*. 29(1):1–5.
Qian C, Bryans N, Kruykov I, de Koning APJ. 2018. Visualization and analysis of statistical signatures of convergent evolution. Alberta (CA): University of Calgary. Available from: https://github.com/dekoning-lab/grand-conv.
Qian C, de Koning APJ. 2018. Rapid discovery of convergent molecular evolution across

entire phylogenies. Alberta (CA): University of Calgary. Available from: https://github.com/dekoning-lab/grand-conv.
Queval G, Issakidis-Bourguet E, Hoeberichts FA, Vandorpe M, Gakière B, Vanacker H, Miginiac-Maslow M, Van Breusegem F, Noctor G. 2007. Conditional oxidative stress responses in the *Arabidopsis* photorespiratory mutant *cat2* demonstrate that redox state is a key modulator of daylength-dependent gene expression, and define photoperiod as a crucial factor in the regulation of H_2O_2-induced cell death. *Plant J.* 52(4):640–657.
R Core Team. 2019. R: A language and environment for statistical computing.
Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet*. 16(6):2–3.
Rinaldi MA, Liu J, Enders TA, Bartel B, Strader LC. 2012. A gain-of-function mutation in IAA16 confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. *Plant Mol Biol*. 79:359–373.
Sasaki K, Kim M-H, Imai R. 2013. Arabidopsis COLD SHOCK DOMAIN PROTEIN 2 is a negative regulator of cold acclimation. *New Phytol*. 198(1):95–102.
Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res*. 43(W1):W7–W14.
Sharma R, Singh G, Bhattacharya S, Singh A. 2018. Comparative transcriptome meta-analysis of *Arabidopsis thaliana* under drought and cold stress. *PLoS One*. 13(9):1–18.
Shinozaki K, Yamaguchi-Shinozaki K, Seki M. 2003. Regulatory network of gene expression in the drought and cold stress responses. *Curr Opin Plant Biol*. 6(5):410–417.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31(19):3210–3212.
Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.
[TAIR] The Arabidopsis Information Resource. 2019. California (US): Phoenix Bioinformatics Corporation. [cited 2019 Oct 23]. Available from: www.arabidopsis.org.
Thomas GWC, Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: A case study using echolocating mammals. *Mol Biol Evol*. 32(5):1232–1236.
Todd J, Post-Beittenmiller D, Jaworski JG. 1999. *KCS1* encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis in *Arabidopsis thaliana*. *Plant J*. 17(2):119–130.
Voutsina N, Payne AC, Hancock RD, Clarkson GJ, Rothwell SD, Chapman MA, Taylor G. 2016. Characterization of the watercress (*Nasturtium officinale* R. Br.; Brassicaceae) transcriptome using RNASeq and identification of candidate genes for important phytonutrient traits linked to human health. *BMC Genomics*. 17(378):1-15.
Wang W, Vinocur B, Shoseyov O, Altman A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci*. 9(5):13–15.
Wang M, Zhao Y, Zhang B. 2015. Efficient test and visualization of multi-set intersections. *Sci Rep*. 5(16923):1–12.

Wang P, Wang F, Yang J. 2017. *De novo* assembly and analysis of the *Pugionium cornutum* (L.) Gaertn. transcriptome and identification of genes involved in the drought response. *Gene*. 626:290–297.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, et al. 2019. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.1.

Wei X, Zhang X, Shen D, Wang H, Wu Q, Lu P, Qiu Y, Song J, Zhang Y, Li X. 2013. Transcriptome analysis of *Barbarea vulgaris* infested with Diamondback Moth (*Plutella xylostella*) Larvae. *PLoS One*. 8(5).

Wong WSW, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*. 168(2):1041–1051.

Yang Y, Smith SA. 2013. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics*. 14(328):1–11.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13(5):555–556.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22(12):2472–2479.

Zhang H, Yang B, Liu W-Z, Li H, Wang L, Wang B, Deng M, Liang W, Deyholos MK, Jiang Y-Q. 2014. Identification and characterization of CBL and CIPK gene families in canola (*Brassica napus* L.). *BMC Plant Biol*. 14(8):1–24.

Zhao Z, Tan L, Dang C, Zhang H, Wu Q, An L. 2012. Deep-sequencing transcriptome analysis of chilling tolerance mechanisms of a subnival alpine plant, *Chorispora bungeana*. *BMC Plant Biol*. 12(222):1–17.

Zhou D, Zhou J, Meng L, Wang Q, Xie H, Guan Y, Ma Z, Zhong Y, Chen F, Liu J. 2009. Duplication and adaptive evolution of the COR15 genes within the highly cold-tolerant *Draba* lineage (Brassicaceae). *Gene*. 441(1–2):36–44.

Zhou Y, Tang Q, Wu M, Mou D, Liu H, Wang S, Zhang C, Ding L, Luo J. 2018. Comparative transcriptomics provides novel insights into the mechanisms of selenium tolerance in the hyperaccumulator plant *Cardamine hupingshanensis*. *Sci Rep*. 8(1):1–17.

II

What can the cold-induced transcriptomes of Arctic Brassicaceae tell us about the evolution of cold tolerance?

Siri Birkeland[1], Tanja Slotte[1,2], Christian Brochmann[1], Anne K. Brysting[3], A. Lovisa S. Gustafsson[1], Michael D. Nowak[1]

[1]Natural History Museum, University of Oslo, Oslo, Norway [2]Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden [3]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway

Corresponding author: Siri Birkeland, siri.birkeland@nhm.uio.no

Abstract

- By studying the molecular basis of cold response in plants adapted to some of the world's coldest biomes, we can gain insight into the evolution of cold tolerance - an important factor in determining plant distributions worldwide.
- We conducted a time series experiment to examine the transcriptional responses of three Arctic Brassicaceae to low temperatures. RNA was sampled before onset of treatment, and after 3h, 6h, and 24h with 2 °C. We identified sets of genes that were differentially expressed in response to cold and compared them between species, as well as to published data from the temperate *Arabidopsis thaliana*.
- We found that the cold response was highly species-specific. Among thousands of differentially expressed genes, ~200 genes were shared among the three Arctic species and *A. thaliana*, and only ~100 genes were shared among the three Arctic species alone. This pattern was also reflected in the functional comparison.
- Our results show that the cold response of Arctic plant species has mainly evolved independently, although it likely builds on a conserved basis found across Brassicaceae. The findings also confirm that highly polygenic traits, such as cold tolerance, may show less repeatable patterns of adaptation than traits involving only a few genes.

1. Introduction

Temperature is one of the most important factors determining plant distributions across the world, and only few species have been able to occupy the cold biomes found towards the poles (Billings and Mooney 1968; Wiens and Donoghue 2004). Evolutionary history seems to play an important role in determining how plant cold tolerance is distributed globally, with tropical-to-temperate transitions being key events in plant evolution (Lancaster and Humphreys 2020). However, little is known about the evolutionary changes required for the transition from temperate to the more extreme polar zones. It is still unclear if the cold response of polar plant species is distinct from that of temperate relatives, and whether it may have evolved in a convergent fashion as a result of the same extreme selection pressures. Polar environments exhibit lower year-round temperatures and shorter growing seasons (with up to 24 hours of daylight) than temperate environments, although some alpine environments share similar characteristics (Billings 1974). The Arctic is warmer than the Antarctic, but the average temperature of the warmest Arctic summer month is not more than 10 °C (Elvebakk 1999), and the growing season can be as short as one month in the coldest areas (e.g. Jónsdóttir 2005). By studying how Arctic plants cope with low temperatures, we can gain insights into how plants acquire cold tolerance, and if there are general trends in plant adaptation to extreme polar environments.

Plants face many challenges upon the transition to colder environments. Low temperatures can affect nearly all aspects of plant cell biochemistry, and influence protein properties, photosynthesis reactions, and cell membrane fluidity (Shi et al. 2018). Ice formation comes with its own set of challenges, and is in general deadly if ice forms within the cell (Körner 2003). Some plants can tolerate freezing of the apoplast (the space between the cells), but this may draw water out of the cell and lead to severe dehydration, as well as increase the level of salts and toxic solutes (Steponkus 1984; Wisniewski and Fuller 1999; Körner 2003; Wisniewski et al. 2004). In temperate environments (from where Arctic plants most likely are derived; Abbott and Brochmann 2003), plants prepare for predictable cold periods via cold acclimation, i.e. an increase in freezing tolerance in response to low non-freezing temperatures (Thomashow 2010). Exposing temperate plants to low temperatures typically results in complete reorganization of the transcriptome, ultimately leading to increased freezing tolerance (Kreps et al. 2002; Kilian et al. 2007; Thomashow 2010). The CBF transcription factors (C-repeat-Binding Factors) are among the main

"regulatory hubs" of this cold response, and have been isolated in many different plant species (Thomashow 2010; Park et al. 2015; Shi et al. 2018). The CBFs are induced shortly after exposure to cold stress, and control the expression of >100 cold-regulated (COR) genes downstream (the CBF pathway; Park et al. 2015). *Arabidopsis thaliana* exhibits three of these cold-induced CBFs (*CBF1*, *CBF2* and *CBF3*, or *DREB1b*, *DREB1c* and *DREB1a*; Jia et al. 2016), but the same genes are not always found in other species (Zhao et al. 2012). It is also becoming increasingly clear that the CBF regulon involves extensive co-regulation by other lesser known transcription factors, and that the low temperature regulatory network is highly complex (Park et al. 2015). One could envision that polar plant species are in less need of a cold acclimation period as temperatures are low year-round (e.g. their transcriptomes could be less responsive to a drop in temperature), or that their cold response is somehow more complex (e.g. involving more genes) or more effective (e.g. faster, or involving fewer genes). There are surprisingly few in-depth studies on the cold-induced transcriptomes of polar plants, but Archambault and Strömvik (2011) have studied Arctic *Oxytropis*, Wang et al. (2017) the temperate-subarctic *Eutrema* (*Thellungiella*) *salsugineum*, and Lee et al. (2013) the Antarctic *Dechampsia antarctica*. In this study, we perform a whole-transcriptome investigation of cold stress in plant species adapted to the high Arctic, and compare their response to that of temperate relatives.

The focal species of this study, *Cardamine bellidifolia*, *Cochlearia groenlandica* and *Draba nivalis*, have independent Arctic origins (Carlsen et al. 2009; Jordon-Thaden et al. 2010; Koch 2012) and represent three of the main clades of Brassicaceae (clade A, B, and C; divergence time ~30 Mya; Huang et al. 2016). They are ideal model species for studying cold response of Arctic plants for three main reasons: 1) all have their main distribution above the Arctic Circle, 2) they are found in all Arctic bioclimatic subzones (even in polar deserts; Elven et al. 2011), and 3) they belong to the plant family in which cold response probably has been most extensively studied, as it includes the model species *A. thaliana* and various economically important crop species (Kilian et al. 2007; Park et al. 2015; Shi et al. 2018). Furthermore, we recently found evidence of positive selection in genes associated with cold stress in all three species (Birkeland et al. 2020). Different genes seem to be under positive selection in the three species, suggesting that they have independently adapted to the Arctic by modifying different components of similar

stress response pathways (Birkeland et al. 2020). However, our previous study was limited to protein coding regions, and theory predicts that there could be a higher chance of convergence in their expression profiles (e.g. Stern 2013; Sackton et al. 2019). This is mainly because changes in *cis*-regulatory regions should be less affected by pleiotropy (a mutation in a *cis*-regulatory region may change only one site while leaving the other sites of expression and protein functions unchanged; Gompel and Prud'homme 2009; Stern 2013). Expression profiles are phenotypes, and similar expression patterns may have different genetic origins (e.g. epigenetic changes, or changes in transcription factor binding motifs of promoters or enhancer regions; Romero et al. 2012; Pai and Gilad 2014), and represent convergent phenotypes. To evaluate the degree of similarity in Arctic Brassicaceae expression profiles and to describe how Arctic cold-induced transcriptomes differ from those of temperate relatives, we therefore subjected these three species to a simulated summer cold shock and identified differentially expressed genes after 3h, 6h and 24h with cold treatment. We aimed to i) characterize the cold-induced transcriptomes of *C. bellidifolia*, *C. groenlandica* and *D. nivalis*, ii) describe how their cold response differ from that of *A. thaliana*, and iii) identify potential convergent expression patterns in the three Arctic species.

2. Material and methods

2.1 Plant material

For each species, we sowed seeds from a single selfed parent derived from wild Arctic populations from Alaska (Supplementary table 1). Seeds were sown in six pots per species (several seeds per pot), of which four pots were used as biological replicates at each time point (see below). Because the plants were siblings derived from selfed parents, and because selfing is assumed to be the predominant mode of reproduction of these species in the wild (Brochmann and Steen 1999), we expected the replicates to be close to genetically identical. The plants were grown in a natural daylight room in the phytotron at the University of Oslo from 22nd of March to 18th of May 2018 with day temperature at 22 °C and night temperature at 18 °C. Supplementary artificial light was given from 08:00-24:00 (400 W high-intensity discharge lamps), and moisture was at ~65 % RH.

2.2 Cold shock treatment

Eight weeks after sowing, the plants were given a 24-hour cold shock to simulate a sudden drop in temperature during a typical Arctic summer. At 13:00 p.m. (to minimize correlation with circadian change), we transferred the pots from the 22 °C daylight room to a 2 °C cooling room with artificial light from 250 W high-intensity discharge lamps (140-160 μmol m^{-2} s^{-1} measured at plant height). Leaf tissue was sampled at four time points; just before they were transferred (0h; control), and after 3h, 6h, and 24h. We sampled all six pots for RNA extractions at each time point, but only used the four best RNA extracts per time point for sequencing (i.e. in terms of RNA quality and quantity). These constituted our four biological replicates per time point per species. RNA was immediately extracted from fresh, fully expanded leaves as described below. During the 24h cold shock treatment, the plants were given continuous light to simulate Arctic midnight sun.

2.3 RNA extraction and sequencing

For extraction of total RNA, we used the Ambion RNAqueous Kit (Thermo Fisher Scientific, Waltham, USA), following the manufacturer's protocol for fresh plant tissue: ~50 mg leaf tissue was immediately ground in Lysis/Binding Solution together with 1 volume of Plant RNA Isolation Aid and consecutively extracted. The RNA quantity was measured with Broad Range RNA Kit on a Qubit v.2.0 fluorometer (Life Technologies, Carlsbad, USA); RNA quality with an Experion Automated Electrophoresis System Station (Bio-Rad Laboratories, Hercules, USA) and a Nandrop One spectrophotometer (Thermo Fisher Scientific, Waltham, USA). The Norwegian Sequencing Centre (www.sequencing.uio.no) prepared the libraries using the TruSeq protocol for stranded mRNA (Illumina, San Diego, USA) and performed the sequencing. Samples were indexed, pooled and run on three lanes (16 samples/lane) on an Illumina HiSeq 3000 (Illumina, San Diego, USA), producing paired end reads with a default insert size of 350 bp and read lengths of 150 bp. The raw reads were quality-checked with FastQC v.0.11.8 (Andrews 2010), and a single FastQC report per species was generated with MultiQC v.1.7 (Ewels et al. 2016).

2.4 Transcriptome assembly and annotation

As there were no available genome assemblies at the start of this study, we assembled a reference transcriptome *de novo* for each species using Trinity v.2.8.5 (Grabherr et

al. 2011) based on all acquired RNA samples. Trinity was run with the integrated Trimmomatic option (Bolger et al. 2014), strand-specificity, and a minimum assembled contig length of 300 bp. The transcriptomes were filtered and annotated with EnTAP (Eukaryotic Non-Model Transcriptome Annotation Pipeline; Hart et al. 2020) in two rounds: first to apply the EnTAP filtering option on the raw transcriptome (in order to reduce inflated transcript estimates), and a second time to annotate the highest expressed isoform and filter out contaminants (used for the annotation of DEGs; see below). For expression filtering, an alignment file was generated with bowtie2 (Langmead and Salzberg 2012) in combination with RSEM (Li and Dewey 2011) using default options in the "align_and_estimate_abundance.pl" script provided with the Trinity software suite. Numbers of complete and fragmented BUSCOs (Benchmarking Universal Single-Copy Orthologs) in the filtered transcriptomes were estimated with BUSCO v4.0.6 (Simão et al. 2015). The filtered transcriptomes were used as the final reference in the differential expression analyses (see below).

2.5 Differential expression analyses

The Trimmomatic filtered reads were mapped to the reference transcriptomes using the alignment free mapper Salmon with a GC content bias correction (Patro et al. 2017). Genes that were differentially expressed after 3h, 6h and 24h of cold treatment were identified with DESeq2 v.1.22.1 (Love et al. 2014), using a design formula controlling for the effect of pot number (design = ~ pot number + time). This means that we tested for the effect of time with 2 °C treatment, while controlling for the individual effects of the sampled pots. A generalized linear model was fitted to each gene and a Wald test (Love et al. 2014) applied to test if the 3h, 6h and 24h model coefficients differed significantly from zero when contrasted to the 0h model coefficient. A gene was considered as differentially expressed if the transcript level exhibited ≥ twofold change in response to the cold treatment at the different time points (log2 fold change threshold = 1). We used a Benjamini-Hochberg adjusted p-value (Benjamini and Hochberg 1995) to evaluate the significance of each differentially expressed gene (False Discovery Rate cutoff of alpha = 0.05). Heatmaps of the top 30 differentially expressed genes with the lowest false discovery rate were generated with the pheatmap package in R (Kolde 2019) using the regularized log

function (rld) on original count data. The mean count of a gene was subtracted from each observation prior to heatmap generation.

2.6 Comparison of DEG sets among Arctic species and *A. thaliana*

To enable the comparison of DEGs among the Arctic species, and among the Arctic species and *A. thaliana*, we used already published data on differentially expressed genes in wild type *A. thaliana* in response to 24h chilling treatment of 4 °C (Ws-2; see Park et al. 2015 for analysis details). Many of our analyses thus focus on the 24h DEG sets. We used two different approaches to compare the 24h DEG sets among species. First, we ran OrthoFinder v.2.3.12 (Emms and Kelly 2019) to identify sets of genes putatively descended from a single gene in the last common ancestor of all four species (orthogroups), using the assembled transcriptomes (filtered based on highest expressed isoform) and the Araport11 peptide file of *A. thaliana* downloaded from www.arabidopsis.org. This enabled us to compare orthogroup identity among DEG sets. Second, we used the BLASTP (protein-protein) search of BLAST+ v.2.9.0 (Camacho et al. 2009) to identify putative *A. thaliana* homologs in the three Arctic species, using the Araport11 peptide file as database and each of the Arctic transcriptome files as query (with e-value < 0.01 and max target seqs = 1). This second approach enabled us to more closely compare gene identity and function among species, but with the caveat that the *A. thaliana* homolog might not always be the true homolog, or represent the same function in all species. The significance of the overlaps among differentially expressed orthogroups at 24h were evaluated using the supertest function in SuperExactTest v.1.0.7 (Wang et al. 2015). We also visualized potential unique overlaps among differentially expressed orthogroups at 24h and among differentially expressed *A. thaliana* homologs at 24h, using UpSetR v.1.4.0 (Conway et al. 2017). To compare transcription factor composition, we annotated the *A. thaliana* 24h DEG set with EnTAP (as above).

2.7 Gene Set Enrichment Analyses

To further functionally characterize sets of upregulated and downregulated genes, we performed gene ontology (GO) enrichment analyses within the Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) domains for each species and time point. We used a Fisher's exact test in combination with the elim algorithm implemented in topGO 2.34.0 of Bioconductor to test for overrepresented

GO-terms in each set of significantly upregulated or downregulated genes (Gentleman et al. 2004; Alexa et al. 2006). The elim algorithm works by traversing the GO-graph bottom-up and discarding genes that already have been annotated to significant child terms, and is the recommended algorithm by the creators of topGO due to its simplicity (Alexa et al. 2006). For the enrichment analyses, we used the gene annotations of the transcriptomes as background gene sets in each test (using the GO-annotations acquired with EnTAP). For *A. thaliana*, we used the org.At.tair.db R package v.3.7.0 to annotate the 24h DEG set of Park et al. (2015), and for creating a background gene set used in the enrichment tests (Carlson 2018). A GO-term was considered significantly enriched if $p < 0.05$. We did not correct for multiple testing as the enrichment-tests were not independent. Overlaps among enriched GO-terms in similar DEG sets (i.e. upregulated and downregulated genes at similar time points) were compared among species using UpSetR as above.

2.8 Comparison with data set on positively selected genes

Because we previously have identified convergent substitutions and tested for positive selection in the three focal species (see Birkeland et al. 2020 for details), we were able to check for potential overlaps between positively selected genes/convergent genes and the 24h DEG sets. We blasted the newly assembled transcriptomes against the transcriptomes of the previous study using BLASTP with an e-value cutoff of < 0.01 and max target seqs = 1.

3. Results

3.1 Transcriptome assemblies and Differentially Expressed Genes (DEGs)

The three *de novo* assemblies contained ~22,000-24,000 (Trinity) genes and were highly complete in terms of BUSCOs (>90 % complete; Table 1). We identified a gradual increase in the number of DEGs with time at 2 ℃. About 700-1000 DEGs were identified after 3h, with varying rates of increase at 6h (+33 DEGs in *C. bellidifolia*, +283 DEGs in *C. groenlandica*, and +895 DEGs in *D. nivalis*), and with ~2500-3000 DEGs identified at 24h (Table 2; Supplementary tables S2-4). Initially, most DEGs were upregulated, but after 24h we found similar numbers of downregulated and upregulated genes in all species (Table 2). Most 3h and 6h DEGs

were also found in the 24h set, but many DEGs were also unique for each time point (Supplementary tables S5-7).

Based on orthogroup identity, most 24h DEGs seemed to be species-specific, but we found 212 differentially expressed orthogroups shared by the three Arctic species and *A. thaliana*, and 106 differentially expressed orthogroups shared by Arctic species alone (Figure 1). These unique orthogroup intersections corresponded to overlaps of 195 and 119 genes based on *A. thaliana* homologs, respectively (Supplementary Fig. 1). Note that several genes from the same species occasionally can have the same orthogroup identity. We also found that all species (i.e. all species combinations) shared significant orthogroup overlaps in the 24h DEG sets (all $p < 0.01$ based on the supertest; Supplementary table 8). Thus, our main findings were that the shared portion of the cold response was bigger than expected by chance, and that more genes were shared by the three Arctic species and *A. thaliana*, than by the Arctic species alone.

Table 1. Statistics for de novo transcriptome assemblies

Species (source)	Total no. read pairs	No. of "genes" in final assembly[b] [raw assembly]	No. of isoforms in final assembly [raw assembly]	% complete BUSCOs in final assembly
Cardamine bellidifolia	403,256,653 (16 samples[a])	21,818 [42,151]	42,646 [98,419]	93.8%
Cochlearia groenlandica	389,172,001 (16 samples[a])	22,396 [49,768]	40,639 [102,855]	93.4%
Draba nivalis	368,925,923 (16 samples[a])	23,871 [52,096]	46,282 [109,658]	92.6%

[a]Four replicates at four time points (0h, 3h, 6h, 24h), [b]Corresponding to Trinity genes (or transcript clusters)

Table 2. Number [percentage] of differentially expressed genes (DEGs) after 3h, 6h and 24h with 2 °C. (**bold** = total number of DEGs, ↑/↓ = upregulated/downregulated DEGs)

	3h ❄	6h ❄	24h ❄
Cardamine bellidifolia	**1012** 857↑, 155↓ [85% ↑, 15% ↓]	**1045** 877↑, 168↓ [84% ↑, 16% ↓]	**2520** (1301↑, 1219↓) [52% ↑, 48% ↓]
Cochlearia groenlandica	**733** 521↑, 212↓ [71% ↑, 29% ↓]	**1016** (636↑, 380↓) [63% ↑, 37%↓]	**3010** (1534↑, 1476↓) [51% ↑, 49% ↓]
Draba nivalis	**688** (505↑, 183↓) [73% ↑, 27% ↓]	**1583** (998↑, 585↓) [63% ↑, 37% ↓]	**2839** (1484↑, 1355↓) [52% ↑, 48% ↓]

3.2 Comparison of DEGs among species

3.2.1 Transcription factors in Arctic species

Transcription factors, the genes potentially orchestrating the cold response, made up 9-14% of all Arctic DEGs at all times (Table 3, Supplementary tables 2-4). AP2/ERF (Figure 2) was the most common family of transcription factors based on the 24h DEG set; known to include important candidates for cold regulation such as CBFs/DREBs and RAVs.

Among the transcription factors in the AP2/ERF family were *CBF1* (*DREB1B*) and *CBF4* (*DREB1D*), which were upregulated upon cold treatment in all species. Other upregulated AP2/ERFs included *DREB2A* (at all time points in *C. groenlandica*, and at 24h in *C. bellidifolia*), *DREB2C* (at all time points in *C. bellidifolia* and *D. nivalis*), *DREB3* (at all time points in *C. groenlandica* and *D. nivalis*), and *RAV1* (found to be upregulated at 6h and 24h in *C. bellidifolia*). Note that some of these BLAST annotations were uncertain as the results were conflicting with other modes of annotation (e.g. *DREB2C* was predicted by EggNOG to be a DREB1, and *CBF1* was annotated as *CBF2* when only blasting against *A. thaliana*).

Other common transcription factors in all three species were those containing SANT/Myb domains, MYC-type basic helix-loop-helix (bHLH) domains, basic-leucine zipper domains, and NAC domains (Figure 2). Many of these transcription factors were among the DEGs shared only by Arctic species based on the *A. thaliana* homolog (for instance *REVEILLE 2*, *RAP2.10*, *RAP2.2*, *PCL1*, and *HY5*; Supplementary table 9).

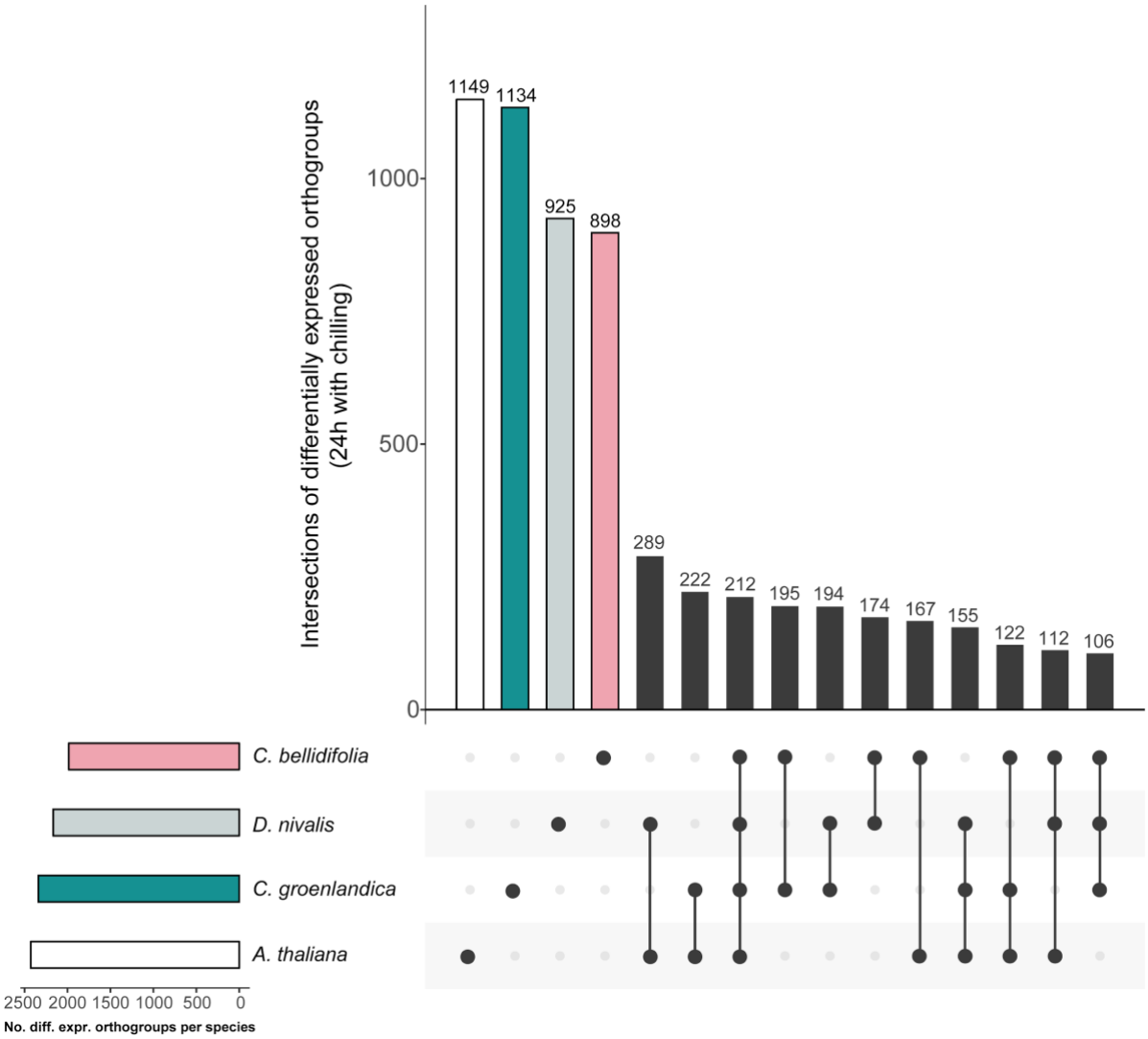


Figure 1. UpSet plot of differentially expressed orthogroups. The plot in the left corner shows total numbers of differentially expressed orthogroups, and the main plot shows the number of unique differentially expressed orthogroups, followed by orthogroups intersections/overlaps between species (connected dots).

Table 3. Number of differentially expressed genes (DEGs) annotated with "regulation of transcription" (GO:0006355) after 3h, 6h and 24h with 2 °C. The percentage of the total DEG set is given in parentheses.

	3h ❄	6h ❄	24h ❄
Cardamine bellidifolia	**123** (12%)	**139** (13%)	**266** (11%)
Cochlearia groenlandica	**106** (14%)	**143** (14%)	**344** (11%)
Draba nivalis	**92** (13%)	**180** (11%)	**260** (9%)

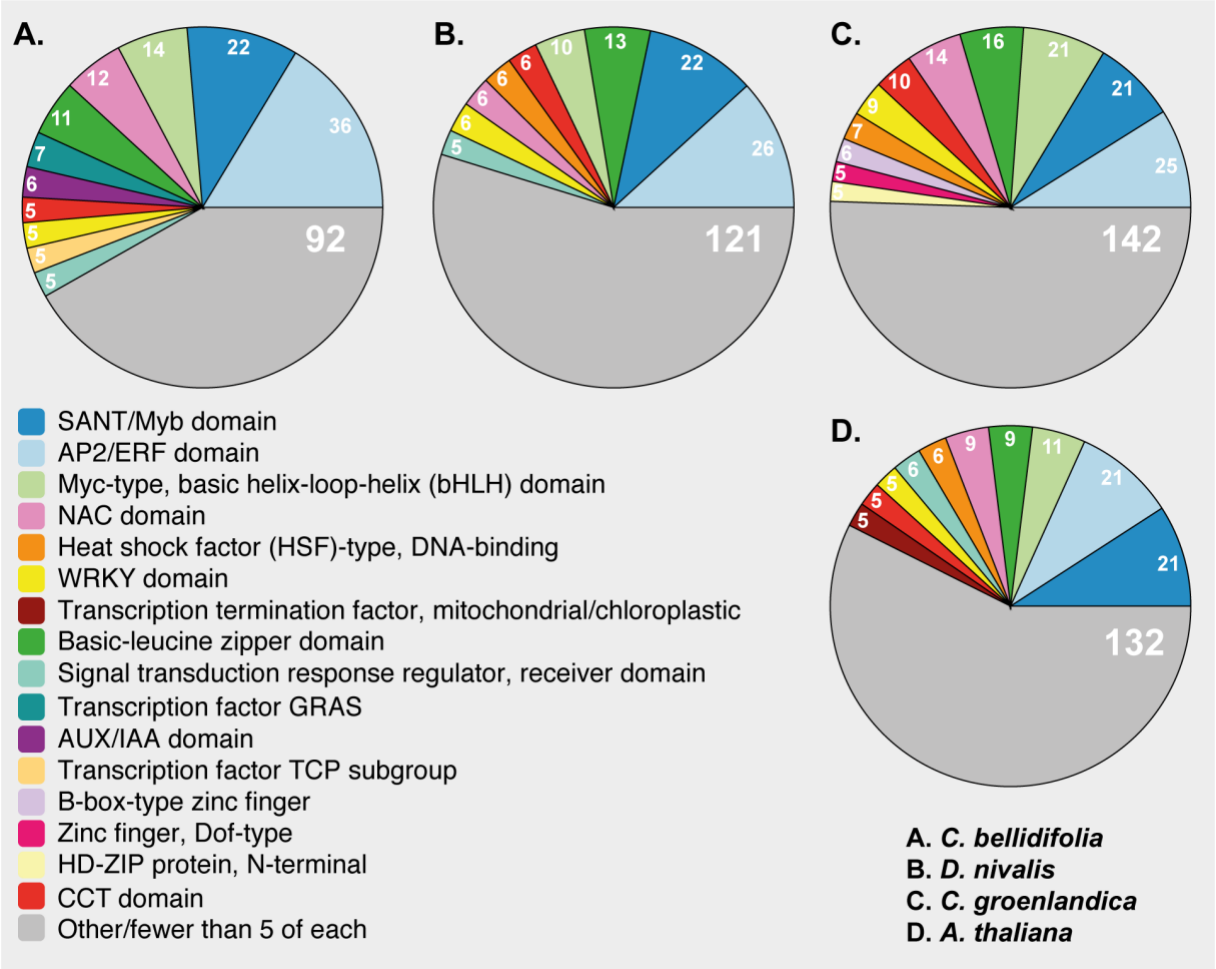


Figure 2. InterPro domains in 24h DEGs annotated with "regulation of transcription" (GO:0006355). Transcription factors that did not have InterPro domain information are not included.

3.2.2 Other shared cold regulated genes in Arctic species

The gene overlaps based on *A. thaliana* homologs were used to further investigate cold regulated genes that were common in the cold response of all Arctic species at 24h (Supplementary tables 9-13). Among the 119 uniquely shared Arctic DEGs, 109 shared similar expression patterns in all species (74 upregulated and 35 downregulated; Supplementary figure 2). The upregulated DEGs might be especially important in acquiring freezing tolerance, and included many transcription factors (15, cf. above), but also genes annotated with e.g. response to karrikin (7), circadian rhythm (6), flavonoid biosynthetic process (5), and proline transport (4; Supplementary table 9). Other upregulated DEGs shared by the Arctic species included e.g. *MAPK7*, *MAPKKK14*, *MAPKKK18*, *SUCROSE SYNTHASE 3*, *RAB18*, and a Late Embryogenesis Abundant gene (LEA; Supplementary table 9). Several of

the uniquely shared DEGs could be found in the heatmaps of the top 30 differentially expressed genes with the lowest false discovery rate (Figure 3).

3.2.3 Shared cold regulated genes among Arctic species and *A. thaliana*

Among the 195 DEGs shared by Arctic species and *A. thaliana*, 188 shared similar expression patterns in all species (Supplementary table 10). Among the 122 shared upregulated DEGs (Supplementary figure 2), were several genes annotated with e.g. response to cold (23), response to salt stress (20), response to abscisic acid (16), response to water deprivation (14), and flavonoid biosynthetic process (13; (Supplementary table 10). The shared gene list included many known cold induced genes like *LEA14*, *COR78*, *COR15B*, *TCF1*, *COR27* and *COR28* (Supplementary table 10). There were also several shared genes related to Abscisic acid and ethylene regulated pathways, like *ABF1*, *AITR5*, *ERT2* , and *ERF043* (Supplementary table 10). Several of the DEGs shared between *A. thaliana* and the three Arctic species could also be found in the heatmaps of the top 30 differentially expressed genes with the lowest false discovery rate (Figure 3).

3.3 Functional characterization of DEG sets (GO-enrichment)

Most significantly enriched GO-terms within the BP, CC, and MF domains were species-specific (Table 4, Supplementary tables 14-22). However, as many as 27 of the same GO-terms (BP and MF) were significantly enriched in the Arctic upregulated DEG sets at 24h (Supplementary table 23). Most of these (22) were also found in *A. thaliana*. Among the GO-terms shared by *A. thaliana* and the Arctic species were many BP terms like “response to cold”, and “response to heat” (Figure 4-5). Other shared BP terms included those associated with cold and freezing like “response to salt stress”, “response to water deprivation”, “response to oxidative stress”, and “response to cadmium ion” (Figure 5). We also observed a shared overrepresentation of genes associated with the hormones abscisic acid, and ethylene, and the substance karrikin, as well as genes associated with possible cryoprotectants like flavonoids and sucrose (BP, Figure 5). Among the GO-terms uniquely found in Arctic 24h upregulated DEG sets were e.g. “spermidine biosynthetic process” (BP) and “arginine decarboxylase activity” (MF). The 24h downregulated DEGs were enriched for genes associated with growth related GO-terms such as “phototropism” (BP; all species) and “auxin-activated signaling pathway” (BP; only Arctic species).

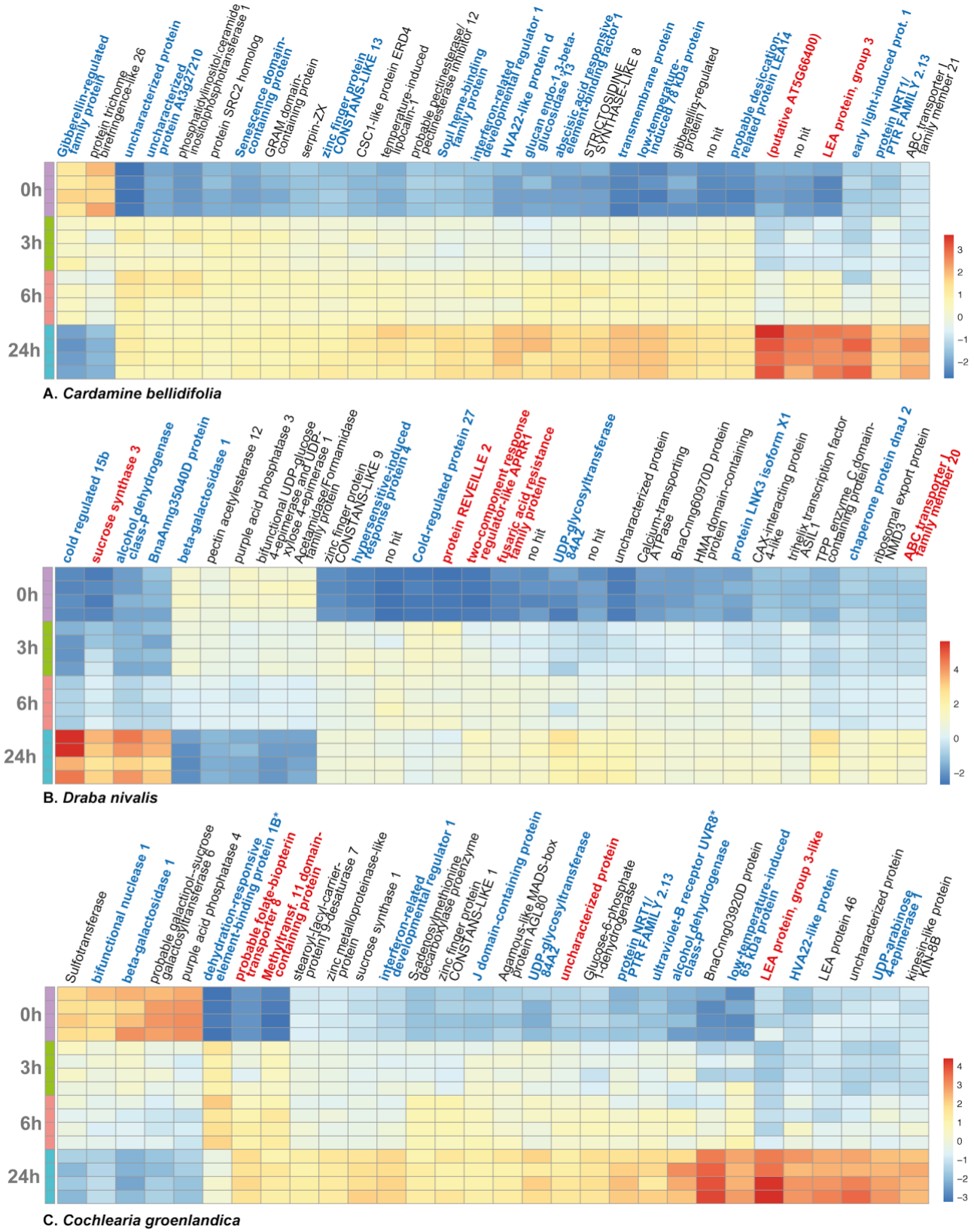


Figure 3. Heatmaps of the most significantly differentially expressed genes (DEGs): A) *Cardamine bellidifolia*, B) *Draba nivalis*, and C) *Cochlearia groenlandica*. Color scale = log2 transformed counts. Each row corresponds to a replicate, and there are four replicates at each time point (0h, 3h, 6h, 24h). Gene names in **bold and blue** = Found as 24h DEG in *Arabidopsis thaliana* and all Arctic species (based on *A. thaliana* homologs), gene names in **bold and red** = Found as 24h DEG only in Arctic species (based on *A. thaliana* homologs). *dehydration-responsive element-binding protein 1B = DREB1b/CBF1, ultraviolet-B receptor UVR8 = TCF1.

Table 4. Numbers of significantly enriched GO-terms ($p < 0.05$) in differentially expressed genes (DEGs) after 3h, 6h and 24h at 2 ℃ when applying the elim algorithm. Abbreviations: ↑ = Upregulated DEG set, ↓ = Downregulated DEG set, Biological Process = BP, Cellular Component = CC, Molecular Function = MF domains.

		C. bellidifolia	*C. groenlandica*	*D. nivalis*	Shared*
3h ❄	↑	109 BP, 12 CC, 40 MF	101 BP, 26 CC, 42 MF	102 BP, 4 CC, 38 MF	11 BP, 1 CC, 3 MF
	↓	44 BP, 7 CC, 26 MF	61 BP, 14 CC, 34 MF	101 BP, 1 CC, 44 MF	0 BP, 0 CC, 0 MF
6h ❄	↑	127 BP, 13 CC, 33 MF	108 BP, 5 CC, 37 MF	120 BP, 9 CC, 44 MF	16 BP, 1 CC, 4 MF
	↓	51 BP, 4 CC, 25 MF	73 BP, 8 CC, 37 MF	121 BP, 8 CC, 45 MF	1 BP, 0 CC, 0 MF
24h ❄	↑	156 BP, 22 CC, 62 MF	172 BP, 30 CC, 76 MF	139 BP, 24 CC, 82 MF	22 BP, 0 CC, 5 MF
	↓	138 BP, 10 CC, 78 MF	108 BP, 11 CC, 60 MF	136 BP, 25 CC, 62 MF	7 BP, 4 CC, 0 MF

*Also shared with *A. thaliana* after 24h: ↑20 BP, 0 CC, 2 MF, and ↓ 3 BP, 3 CC, 0 MF.

3.4 Note on positively selected/convergent cold-responsive genes

We identified several positively selected and convergent genes in the 24h DEG sets of the Arctic species (Table 5-6; Supplementary tables 24-25; (Birkeland et al. 2020). Upregulated positively selected genes (PSGs) included genes such as *COR15B* and *CSDP1* in *D. nivalis*, *LEA4-5* in *C. groenlandica*, and a highly upregulated transmembrane protein (putative homolog of *A. thaliana* AT1G16850) in *C. bellidifolia*. We found that a gene with convergent substitutions in all Arctic species, *EMB2742*, was upregulated in all species (and also in *A. thaliana*). Some convergent genes showed different expression patterns depending on the species. For instance, *CAT2* was downregulated in *D. nivalis* and upregulated in *C. groenlandica*. This gene has previously been found to contain convergent substitutions in *D. nivalis* and *C. groenlandica*, and to be under positive selection in *D. nivalis*. The low temperature responsive transcription factor *RAV1* has previously been found to be under positive selection in *C. groenlandica*, but was not differentially expressed in this species. We also note that *MAPKKK14*, (previously found to contain convergent substitutions in the Arctic species) was blasted to several different MAPKKKs in this study. Its status as a convergent gene is therefore uncertain.

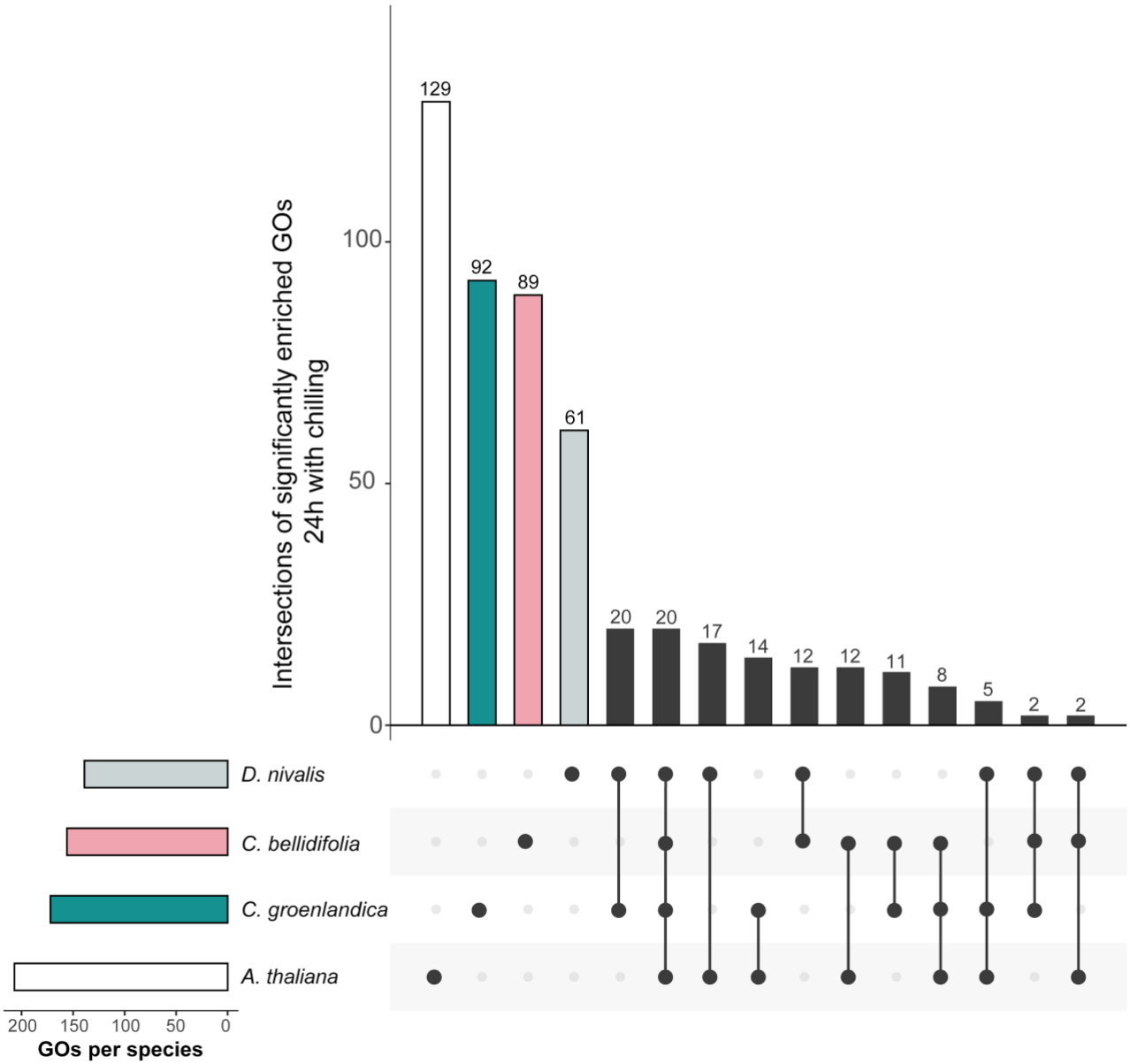


Figure 4. UpSet plot of overrepresented GO-terms in the 24h upregulated gene sets (Biological Process domain only). The plot in the left corner shows total numbers of significantly enriched GO-terms, and the main plot shows the number of unique significantly enriched GO-terms, followed by GO-term intersections/overlaps between species (connected dots).

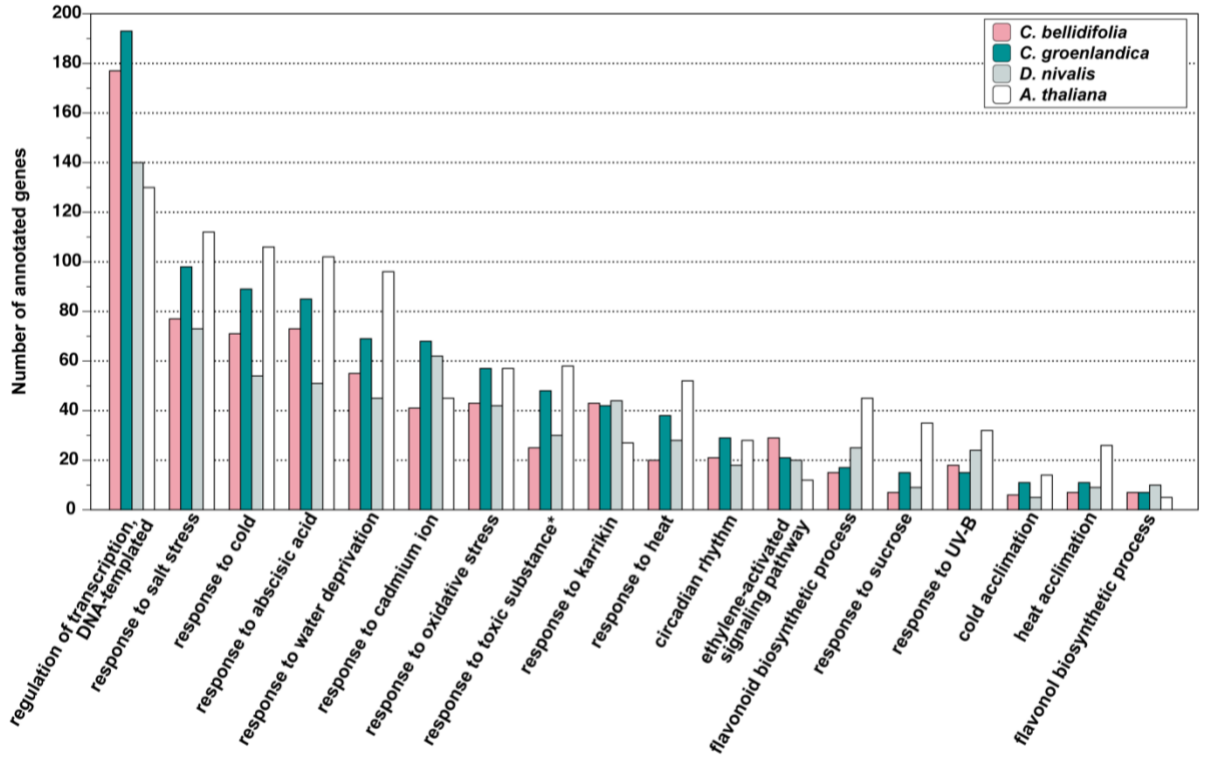


Figure 5. Barchart of genes annotated with Gene Ontology (GO) terms found to be significantly enriched in the 24h upregulated gene set of the three Arctic species. There were 22 overlapping GO-terms among the Arctic 24h upregulated DEG sets, and 20 of these overlapped with *A. thaliana*. Only terms with at least 10 annotated genes are shown. *Not significantly enriched in *A. thaliana*.

Table 5. 24h DEGs under positive selection based on Birkeland et al. (2020).
PSG = Positively selected gene

	Total number of PSGs Birkeland et al. (2020)	PSGs among all DEGs, 24h	PSGs among upregulated DEGs, 24h	PSGs among downregulated DEGs, 24h
C. bellidifolia	201	26	14	12
C. groenlandica	159	23	12	11
D. nivalis	360	46	24	22

Table 6. 24h DEGs with convergent substitutions based on Birkeland et al. (2020).
Cb = *Cardamine bellidifolia* 24h DEG set, Cg = *Cochlearia groenlandica* 24h DEG set, Dn = *Draba nivalis* 24h DEG set

	Total number of convergent genes Birkeland et al. (2020)	Convergent genes among all DEGs, 24h	Convergent genes among upregulated DEGs, 24h	Convergent genes among downregulated DEGs, 24h
C. bellidifolia*, *C. groenlandica	58	17 Cb 14 Cg	9 Cb 12 Cg	8 Cb 2 Cg
C. groenlandica*, *D. nivalis	33	8 Cg 8 Dn	5 Cg 2 Dn	3 Cg 6 Dn
D. nivalis*, *C. bellidifolia	126	19 Dn 20 Cb	13 Dn 17 Cb	6 Dn 3 Cb

4. Discussion

4.1 The cold response of Arctic Brassicaceae is highly species-specific

Our main finding was that the cold response of the three Arctic species, *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*, is highly species-specific. Not only were most DEGs unique for each species, but the number of DEGs shared by the three Arctic species and *A. thaliana* were higher than the number shared by the three Arctic species alone. This suggests that evolution of cold response in Arctic Brassicaceae followed independent genetic trajectories, but with some conserved components. These results are in concordance with our previous study of protein sequence evolution in the same three species, where we found very little overlap in positively selected genes among species, and also very few genes with convergent substitutions (Birkeland et al. 2020). Although we expected a higher degree of convergence in the cold induced expression profiles, the results agree well with those of a similar study of cold acclimation in the temperate grass subfamily Pooideae (Schubert et al. 2019). Schubert et al. (2019) found that phylogenetically diverse species of grasses showed widespread species-specific transcriptomic responses to low temperatures, but with some conserved aspects. The independent evolutionary trajectories of cold response might be tied to the polygenic nature of the trait, involving thousands of genes.

Highly polygenic traits have high genetic redundancy, and should show less repeatable patterns of adaptation compared to traits involving only a few genes (Yeaman 2015; Yeaman et al. 2016; Barghi et al. 2019). Low levels of repeatability has also been found in other polygenic traits, such as drought tolerance. Marín-de la Rosa et al. (2019) recently showed that divergent strategies underlie drought resistance in closely related Brassicaceae species. However, there are also studies demonstrating a high degree of convergence in polygenic traits, like Yeaman et al. (2016) showing convergent local adaptation to climate in two distantly related conifers, and Yang et al. (2018) showing convergent evolution of flowering time in *Capsella rubella* due to independent deletions at the same locus (flowering time is controlled by >60 genes in *A. thaliana*; Anderson et al. 2011). Although it is not known what caused evolutionary repeatability in these exact examples, adaptation may end up taking the same routes in the presence of pleiotropic constraints or restricted available standing genetic variation (as discussed in Gould and Stinchcombe 2017). Since, we find that cold tolerance mainly evolved independently in the three species, we can assume that such constraints have been of less importance. Following there seem to be many ways to evolve cold tolerance, or at least many ways to build upon cold tolerance once a basis for cold tolerance has evolved.

4.2 Conserved aspects of the Arctic Brassicaceae cold response

Another major finding in our data was that the Arctic species seem to have more in common with the temperate relative *A. thaliana*, than they have with each other. This shared aspect of the cold response may represent conserved parts of the CBF pathway, which also have been found in other plant lineages (Jaglo et al. 2001; Shi et al. 2018). Shared genes included, for instance, those that were both tied to circadian regulation and regulation of freezing tolerance (e.g. *COR27* and *COR28*; Li et al. 2016). This fits well with the CBF pathway being gated by the circadian clock (Fowler et al. 2005; Dong et al. 2011). The Arctic cold response also seemed to follow similar trends as in *A. thaliana*, with a continuous increase in the number of DEGs in response to cold temperatures (Kilian et al. 2007; Calixto et al. 2018). This indicates that Arctic plants respond to low temperatures in a similar way as temperate plants, and that they are not completely hard-wired to tolerate low temperatures.

The transcriptomic changes also triggered sets of biological processes similar to the cold response of *A. thaliana*, and that of other plant species within and outside

the Brassicaceae family (e.g. Zhao et al. 2012; Lee et al. 2013; Buti et al. 2018). First and foremost, this involved stress responses associated with low temperatures and freezing. The upregulation of genes associated with salt stress, water deprivation, and cadmium ion (a heavy metal highly toxic to plants; di Toppi and Gabbrielli 1999) could be linked to cold and freezing stress in two different ways. First, ice formation in the apoplast will draw water out of the cells and increase the concentration of salts and toxic solutes (Wisniewski and Fuller 1999; Körner 2003). This may lead to severe cell dehydration (Shi et al. 2018). Second, stress response pathways associated with cold have been shown to be partially homologous with those of dehydration and salt tolerance (Ingram and Bartels 1996; Bartels and Souer 2003; Shamustakimova et al. 2017). We also found that genes annotated with "response to oxidative stress" were overrepresented among the upregulated genes. This fits well with oxidative stress accompanying other abiotic stresses like cold, and especially high light intensity in combination with low temperatures (Heino and Palva 2004; Kilian et al. 2007; Lütz 2010).

Genes associated with the hormones abscisic acid (ABA) and ethylene were upregulated in response to cold in all species. ABA is an important hormone in plant stress signalling, and has previously been shown to increase in abundance during cold acclimation (Heino and Palva 2004; Tuteja 2007). Exogenous application of ABA will even increase freezing tolerance in *A. thaliana* and other plants (Thomashow 1999). The hormone ethylene is similarly reported to influence freezing tolerance, and regulates the CBF pathway itself (Kazan 2015). In most plant species, this entails a positive regulation of freezing tolerance (like e.g. in tomato and tobacco; Zhang and Huang 2010). However, in *A. thaliana* ethylene appears to also negatively regulate freezing tolerance depending on the growing conditions (Kazan 2015). This particular hormone's role in the Brassicaceae cold response is therefore still uncertain.

One way that plants increase their freezing tolerance is by accumulating compounds that in various ways hinder ice nucleation, or alleviate the effects of ice formation by protecting plant tissues against freezing damage (i.e. cryoprotectants; Ruelland et al. 2009). We observed that genes associated with possible cryoprotectants like flavonoids (especially flavonol) and sucrose were upregulated in response to cold in all species. Sucrose has been reported to lower the freezing point in *A. thaliana* (i.e. leading to supercooling and avoidance of freezing; Reyes-Díaz et

al. 2006), to diminish the water potential between the apoplastic space and the cell in the face of ice formation (osmotic adjustment; Ruelland et al. 2009), and even to regulate cold acclimation itself (Rekarte-Cowie et al. 2008). Similarly, accumulation of flavonoids is associated with cold acclimation and higher freezing tolerance in *A. thaliana* (Schulz et al. 2016), and flavonol might have a role in protecting cell membranes during freezing stress (Korn et al. 2008). Finally, we also observed another compound that seems to be important in acquiring cold/freezing tolerance in all species. Karrikin is a compound that has received little attention in relation to cold stress, but a recent study suggests that karrikin inhibits germination under unfavourable conditions, and also improves plant vigour in the face of abiotic stress through regulation of redox homeostasis (Shah et al. 2020). Genes responsive to karrikin have also been documented to be important in the cold response of *Chorispora bungeana* (a subnival Brassicaceae; Zhao et al. 2012). The fact that flavonoids, sucrose and karrikin seem to have important roles in the cold response of four distantly related Brassicaceae species, highlights their importance in achieving cold tolerance.

These shared trends indicate that cold response in Brassicaceae is built upon a similar scaffold, and support the claim that cold tolerance is a complex trait that is difficult to evolve (Donoghue 2008). It is highly likely that the last common ancestor of the three Arctic species had some kind of cold tolerance machinery as the major clades of the Brassicaceae are thought to have radiated in response to a colder and drier climate (~33.9 Ma in the Eocene-Oligocene transition; Huang et al. 2016). Perhaps the basis of the Arctic cold response could also have contributed to the Brassicaceae family's success in cold and dry environments. However, the highly species-specific cold responses found in *C. bellidifolia*, *C. groenlandica*, and *D. nivalis* also indicate that there is great evolutionary flexibility in cold coping strategies once there is a basis to build upon.

4.3 The Arctic cold response - is there such a thing?

The low degree of overlap in the cold response of Arctic Brassicaceae provokes the question if there actually exists anything like a true Arctic cold response. There are, however, a few shared genes and characteristics that stand out as potentially unique for the three Arctic species. For instance, among the ~100 genes shared only by the three Arctic species, there are particularly many transcription factors that potentially

could have large effects. Yeaman et al. (2016) similarly found that convergent genes in adaptation to climate in distantly related conifers were enriched for transcription factors. Although we can only speculate about their function in this study, several are related to transcription factors that have important roles in cold tolerance in other plant species. Examples included e.g. *CBF4 (mainly activated by drought in A. thaliana*; Haake et al. 2002; Supplementary table 9), *RAP2.10*, and *RAP2.2*. Considering the important role the related *CBF1-3* has in regulating the cold response, *CBF4* stands out as a potential important candidate for the regulation of an Arctic cold response. Furthermore, this gene has also been found to be cold responsive in other species (Cai et al. 2019). Two other examples, *RAP2.10*, and *RAP2.2*, are related to *DEAR1*; a transcription factor known to mediate freezing stress responses (Tsutsui et al. 2009). In addition, several transcription factors related to circadian rhythm like e.g. *REVEILLE 2*, *PCL1* and *HY5* are also among the DEGs shared by all Arctic species. One possibility is that the extreme Arctic light regime might have triggered the evolution of links between the CBF pathway and the circadian clock that are not found in temperate environments. We cannot completely rule out that certain aspects of the experimental light regime may have provoked the expression of these genes, but the interplay between the circadian clock and the CBF pathway is important for balancing freezing tolerance and plant growth (e.g. Dong et al. 2011; Shi et al. 2018).

Another interesting finding among the shared genes is the potential traces of a mitogen activated protein kinase (MAPK) cascade. Such cascades are known to be important in regulating the CBF pathway (Teige et al. 2004; Shi et al. 2018). Intriguingly, we found one MAPK (*MAPK7*), as well as two MAPKKKs (*MAPKKK14* and *MAPKKK18*) that are uniquely upregulated in the three Arctic species. In *A. thaliana*, it has been shown that the *MKK2* pathway regulates cold stress signalling (Teige et al. 2004), and that *MAPK6* is involved in releasing inhibitory effects on CBF gene expression (Kim et al. 2017). Accordingly, the shared MAPK and MAPKKKs could potentially have important roles in a putative Arctic cold response.

Functionally, “spermidine biosynthetic process” stands out as having a special status in the Arctic species, although the significant overrepresentation was caused by only a few cold induced genes. One of these genes was arginine decarboxylase gene 1 (*ADC1*), which was tied to both the shared significant enrichment of “spermidine biosynthetic process” and “arginine decarboxylase activity” in Arctic species. This

gene has previously been found to be involved in acquiring freezing tolerance in wild potato (Kou et al. 2018). It has been shown that cucumber plants pretreated with spermidine (a polyamine) show less decline in photosynthesis rates during chilling than non-treated plants (He et al. 2002). Maintaining photosynthesis rates during low temperatures should be especially important for Arctic plants as temperatures are low year-round, and accordingly, they also have optimum photosynthesis rates at lower temperatures than other plants (Chapin 1983). In addition, there was an overrepresentation of shared genes related to proline transport (an amino acid with a similar role in freezing tolerance; Ruelland et al. 2009), as well as some of the abovementioned cryoprotectants, like flavonoid biosynthesis and sucrose.

Finally, we also found other potential traces of adaptation to extreme Arctic temperatures. Some of the genes that we previously had found to be under positive selection (Birkeland et al. 2020), are upregulated in the same lineages in response to cold. This includes some of the top significantly differentially expressed genes, like e.g. *COR15B* in *D. nivalis*, a transmembrane protein in *C. bellidifolia*, and *LEA46* in *C. groenlandica*. Furthermore, *EMB2742*, which has convergent substitutions in all three species (Birkeland et al. 2020), is significantly upregulated after 24h of cold, possibly indicating that it has a unique Arctic function. As the findings in Birkeland et al. (2020) were not based on cold induced transcriptomes (just "untreated" transcriptomes), there could also be other positively selected genes among the cold induced DEGs that have not been detected. That important cold regulated genes are under positive selection, shows that the Arctic climate may have imposed strong selection pressure on the cold response of Arctic plants.

Despite a few potentially important overlaps in cold-response, it is apparent that there probably is not a single, but many different ways Arctic plants respond to low temperatures. Considering the polygenic nature of the trait, and that the three species likely descend from different temperate relatives, it would also be surprising if their cold response had converged into something uniquely Arctic. To truly understand what has made each of the three Arctic species successful in the extreme Arctic climate, a better understanding of gene functions in cold response in these species is needed. A good place to start could be to study the effects of shared transcription factors on cold and freezing tolerance.

4.4 Limitations and future perspectives

We note that there might be methodological differences between our study and the study on *A. thaliana* (Park et al. 2015) that may affect the number of DEGs considered as significant in each study. This may potentially have a small effect on the number of shared genes, both between *A. thaliana* and the Arctic species, and between the Arctic species alone. An important difference is, for instance, that our study is based on *de novo* assembled transcriptomes, which may have small inaccuracies in the delimitation of genes (e.g. isoforms of the same gene mistakenly being delineated as different genes). However, we also used a stringed filtering scheme to reduce inflated transcript numbers, and such differences should not affect the overall results.

This experiment represents one of the first snapshots of Arctic cold-induced transcriptomes, but more studies are needed to understand how Arctic plant species are able to cope with the low summer temperatures at higher latitudes. Future studies on Arctic cold tolerance could delve deeper into performance under long-term cold stress, or performance under cold stress in combination with high light (typical of Arctic environments).

Acknowledgements

We thank The Phytotron at the University of Oslo, and engineers Bjørn Langrekken, Marit Langrekken, and Ingrid Johansen for help with the technical setup for this experiment. We also thank the Speciation Clock Team, Xuyue Yang, Marian Schubert, Aelys Humphries, William Hughes, and Marco Fracassetti for valuable input and comments. Computational analyses were performed on the Saga Cluster owned by Uninett/Sigma2, as well as the Abel Cluster owned by Uninett/Sigma2 and the University of Oslo (operated by the Department for Research Computing at USIT, the University of Oslo IT-department; http://www.hpc.uio.no/). The study was funded by the Research Council of Norway through the SpArc project (RCN 240223) awarded to C.B.

References

Abbott RJ, Brochmann C. 2003. History and evolution of the arctic flora: in the footsteps of Eric Hulten. *Molecular Ecology*. 12(2):299–313.

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups

from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 22(13):1600–1607.
Anderson JT, Willis JH, Mitchell-Olds T. 2011. Evolutionary genetics of plant adaptation. *Trends in Genetics*. 27(7):258–266.
Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Archambault A, Strömvik MV. 2011. PR-10, defensin and cold dehydrin genes are among those over expressed in *Oxytropis* (Fabaceae) species adapted to the arctic. *Functional & Integrative Genomics*. 11(3):497–505.
Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, Dolezal M, Taus T, Kofler R, Schlötterer C. 2019. Genetic redundancy fuels polygenic adaptation in Drosophila. *PLOS Biology*. 17(2):e3000128.
Bartels D, Souer E. 2003. Molecular responses of higher plants to dehydration. In: Hirt H, Shinozaki K, editors. Plant Responses to Abiotic Stress. Springer. p. 9–38.
Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 57(1):289–300.
Billings WD. 1974. Adaptations and Origins of Alpine Plants. *Arctic and Alpine Research*. 6(2):129.
Billings WD, Mooney HA. 1968. The ecology of Arctic and alpine plants. *Biological Reviews*. 43(4):481–529.
Birkeland S, Gustafsson ALS, Brysting AK, Brochmann C, Nowak MD. 2020. Multiple Genetic Trajectories to Extreme Abiotic Stress Adaptation in Arctic Brassicaceae. *Molecular Biology and Evolution*. 37(7):2052–2068.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–2120.
Brochmann C, Steen SW. 1999. Sex and genes in the flora of Svalbard - Implications for conservation biology and climate change. In: Nordal I, Razzhivin VY, editors. The Species Concept in the High North - A Panarctic Flora Initiative. The Norwegian Academy of Science and Letters. p. 33–72.
Buti M, Pasquariello M, Ronga D, Milc JA, Pecchioni N, Ho VT, Pucciariello C, Perata P, Francia E. 2018. Transcriptome profiling of short-term response to chilling stress in tolerant and sensitive *Oryza sativa* ssp. Japonica seedlings. *Functional & Integrative Genomics*. 18(6):627–644.
Cai X, Magwanga RO, Xu Y, Zhou Z, Wang X, Hou Y, Wang Y, Zhang Y, Liu F, Wang K. 2019. Comparative transcriptome, physiological and biochemical analyses reveal response mechanism mediated by *CBF4* and *ICE2* in enhancing cold stress tolerance in *Gossypium thurberi*. *AoB PLANTS*. 11(6).
Calixto CPG, Guo W, James AB, Tzioutziou NA, Entizne JC, Panter PE, Knight H, Nimmo HG, Zhang R, Brown JWS. 2018. Rapid and Dynamic Alternative Splicing Impacts the *Arabidopsis* Cold Response Transcriptome. *Plant Cell*. 30(7):1424–1444.
Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.

2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.
Carlsen T, Bleeker W, Hurka H, Elven R, Brochmann C. 2009. Biogeography and Phylogeny of *Cardamine* (Brassicaceae). *Annals of the Missouri Botanical Garden*. 96(2):215–236.
Carlson M. 2018. org.At.tair.db: Genome wide annotation for *Arabidopsis*. https://bioconductor.org/packages/release/data/annotation/html/org.At.tair.db.html.
Chapin FS. 1983. Direct and indirect effects of temperature on arctic plants. *Polar Biology*. 2(1):47–52.
Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 33(18):2938–2940.
Dong MA, Farré EM, Thomashow MF. 2011. Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*. 108(17):7241–7246.
Donoghue MJ. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*. 105 Suppl 1:11549–11555.
Elvebakk A. 1999. Bioclimatic delimitation and subdivision of the Arctic. In: Nordal I, Razzhivin VY, editors. The species concept in the high north: a panarctic flora initiative. The Norwegian Academy of Science and Letters. p. 81–112.
Elven R, Murray DF, Razzhivin VY, Yurtsev BA. 2011. Annotated Checklist of the Panarctic Flora (PAF). Vascular Plants. [accessed 28 of July 2020]. http://panarcticflora.org/
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*. 20(1):238.
Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 32(19):3047–3048.
Fowler SG, Cook D, Thomashow MF. 2005. Low temperature induction of *Arabidopsis CBF1*, *2*, and *3* is gated by the circadian clock. *Plant Physiology*. 137(3):961–968.
Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*. 5(10):R80.
Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. *Developmental Biology*. 332(1):36–47.
Gould BA, Stinchcombe JR. 2017. Population genomic scans suggest novel genes underlie convergent flowering time evolution in the introduced range of *Arabidopsis thaliana*. *Molecular Ecology*. 26(1):92–106.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. 29(7):644–652.

Haake V, Cook D, Riechmann J, Pineda O, Thomashow MF, Zhang JZ. 2002. Transcription Factor *CBF4* Is a Regulator of Drought Adaptation in *Arabidopsis*. *Plant Physiology*. 130(2):639–648.

Hart AJ, Ginzburg S, Xu MS, Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn JL. 2020. EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Molecular Ecology Resources*. 20(2):591–604.

Heino P, Palva ET. 2004. Signal transduction in plant cold acclimation. In: Hirt H, Shinozaki K, editors. *Plant Responses to Abiotic Stress*. p. 151–186.

He L, Nada K, Tachibana S. 2002. Effects of Spermidine Pretreatment through the Roots on Growth and Photosynthesis of Chilled Cucumber Plants (*Cucumis sativus* L.). *Journal of the Japanese Society for Horticultural Science*. 71(4):490–498.

Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, et al. 2016. Resolution of Brassicaceae Phylogeny Using Nuclear Genes Uncovers Nested Radiations and Supports Convergent Morphological Evolution. *Molecular Biology and Evolution*. 33(2):394–412.

Ingram J, Bartels D. 1996. The Molecular Basis of Dehydration Tolerance in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*. 47(1):377–403.

Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF. 2001. Components of the Arabidopsis C-Repeat/Dehydration-Responsive Element Binding Factor Cold-Response Pathway Are Conserved in *Brassica napus* and Other Plant Species. *Plant Physiology*. 127(3):910–917.

Jia Y, Ding Y, Shi Y, Zhang X, Gong Z, Yang S. 2016. The cbfs triple mutants reveal the essential functions of CBFs in cold acclimation and allow the definition of CBF regulons in *Arabidopsis*. *New Phytologist*. 212(2):345–353.

Jónsdóttir IS. 2005. Terrestrial Ecosystems on Svalbard: Heterogeneity, Complexity and Fragility from an Arctic Island Perspective. *Biology & Environment: Proceedings of the Royal Irish Academy*. 105(3):155–165.

Jordon-Thaden I, Hase I, Al-Shehbaz I, Koch MA. 2010. Molecular phylogeny and systematics of the genus *Draba* (Brassicaceae) and identification of its most closely related genera. *Molecular Phylogenetics and Evolution*. 55(2):524–540.

Kazan K. 2015. Diverse roles of jasmonates and ethylene in abiotic stress tolerance. *Trends in Plant Science*. 20(4):219–229.

Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. 2007. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *The Plant Journal*. 50(2):347–363.

Kim SH, Kim HS, Bahk S, An J, Yoo Y, Kim J-Y, Chung WS. 2017. Phosphorylation of the transcriptional repressor *MYB15* by mitogen-activated protein kinase 6 is required for freezing tolerance in *Arabidopsis*. *Nucleic Acids Research*. 45(11):6613–6627.

Koch MA. 2012. Mid-Miocene divergence of *Ionopsidium* and *Cochlearia* and its impact on the systematics and biogeography of the tribe Cochlearieae (Brassicaceae). *Taxon*. 61(1):76–92.
Kolde R. 2019. Pheatmap: Pretty Heatmaps. https://CRAN.R-project.org/package=pheatmap.
Körner C. 2003. Alpine Plant Life. Functional Plant Ecology of High Mountain Ecosystems. Springer.
Korn M, Peterek S, Mock H-P, Heyer AG, Hincha DK. 2008. Heterosis in the freezing tolerance, and sugar and flavonoid contents of crosses between *Arabidopsis thaliana* accessions of widely varying freezing tolerance. *Plant, Cell & Environment*. 31(6):813–827.
Kou S, Chen L, Tu W, Scossa F, Wang Y, Liu J, Fernie AR, Song B, Xie C. 2018. The arginine decarboxylase gene *ADC1*, associated to the putrescine pathway, plays an important role in potato cold-acclimated freezing tolerance as revealed by transcriptome and metabolome analyses. *The Plant Journal*. 96(6):1283–1298.
Kreps JA, Wu Y, Chang H-S, Zhu T, Wang X, Harper JF. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiology*. 130(4):2129–2141.
Lancaster LT, Humphreys AM. 2020. Global variation in the thermal tolerances of plants. *Proceedings of the National Academy of Sciences of the United States of America*.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9(4):357–359.
Lee J, Noh EK, Choi H-S, Shin SC, Park H, Lee H. 2013. Transcriptome sequencing of the Antarctic vascular plant *Deschampsia antarctica* Desv. under abiotic stress. *Planta*. 237(3):823–836.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12(1).
Li X, Ma D, Lu SX, Hu X, Huang R, Liang T, Xu T, Tobin EM, Liu H. 2016. Blue Light- and Low Temperature-Regulated *COR27* and *COR28* Play Roles in the *Arabidopsis* Circadian Clock. *Plant Cell*. 28(11):2755–2769.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15(12):550.
Lütz C. 2010. Cell physiology of plants growing in cold environments. *Protoplasma*. 244(1-4):53–73.
Marín-de la Rosa N, Lin C-W, Kang YJ, Dhondt S, Gonzalez N, Inzé D, Falter-Braun P. 2019. Drought resistance is mediated by divergent strategies in closely related Brassicaceae. *New Phytologist*. 223(2):783–797.
Pai AA, Gilad Y. 2014. Comparative studies of gene regulatory mechanisms. *Current Opinion in Genetics & Development*. 29:68–74.
Park S, Lee C-M, Doherty CJ, Gilmour SJ, Kim Y, Thomashow MF. 2015. Regulation of the *Arabidopsis* CBF regulon by a complex low-temperature regulatory network. *The Plant Journal*. 82(2):193–207.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*. 14(4):417–419.
Rekarte-Cowie I, Ebshish OS, Mohamed KS, Pearce RS. 2008. Sucrose helps regulate cold acclimation of *Arabidopsis thaliana*. *Journal of Experimental Botany*. 59(15):4205–4217.
Reyes-Díaz M, Ulloa N, Zúñiga-Feest A, Gutiérrez A, Gidekel M, Alberdi M, Corcuera LJ, Bravo LA. 2006. *Arabidopsis thaliana* avoids freezing by supercooling. *Journal of Experimental Botany*. 57(14):3687–3696.
Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*. 13(7):505–516.
Ruelland E, Vaultier M-N, Zachowski A, Hurry V. 2009. Cold Signalling and Cold Acclimation in Plants. In: Kader J-C, Delseny M, editors. Advances in Botanical Research: Volume 49. p. 35–150.
Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science*. 364(6435):74–78.
Schubert M, Grønvold L, Sandve SR, Hvidsten TR, Fjellheim S. 2019. Evolution of Cold Acclimation and Its Role in Niche Transition in the Temperate Grass Subfamily Pooideae. *Plant Physiology*. 180(1):404–419.
Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK. 2016. Flavonoids are determinants of freezing tolerance and cold acclimation in *Arabidopsis thaliana*. Scientific Reports. 6:34027.
Shah FA, Wei X, Wang Q, Liu W, Wang D, Yao Y, Hu H, Chen X, Huang S, Hou J, et al. 2020. Karrikin Improves Osmotic and Salt Stress Tolerance via the Regulation of the Redox Homeostasis in the Oil Plant *Sapium sebiferum*. *Frontiers in Plant Science*. 11.
Shamustakimova AO, Leonova TG, Taranov VV, de Boer AH, Babakov AV. 2017. Cold stress increases salt tolerance of the extremophytes *Eutrema salsugineum* (*Thellungiella salsuginea*) and *Eutrema* (*Thellungiella*) *botschantzevii*. *Journal of Plant Physiology*. 208:128–138.
Shi Y, Ding Y, Yang S. 2018. Molecular Regulation of CBF Signaling in Cold Acclimation. *Trends in Plant Science*. 23(7):623–637.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31(19):3210–3212.
Steponkus PL. 1984. Role of the Plasma Membrane in Freezing Injury and Cold Acclimation. *Annual Review of Plant Physiology*. 35(1):543–584.
Stern DL. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics*. 14(11):751–764.
Teige M, Scheikl E, Eulgem T, Dóczi R, Ichimura K, Shinozaki K, Dangl JL, Hirt H. 2004. The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis. *Molecular Cell*. 15(1):141–152.
Thomashow MF. 1999. Plant Cold Acclimation: Freezing Tolerance Genes and

Regulatory Mechanisms. *Annual review of plant physiology and plant molecular biology*. 50:571–599.
Thomashow MF. 2010. Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. *Plant Physiology*. 154(2):571–577.
Toppi LS di, di Toppi LS, Gabbrielli R. 1999. Response to cadmium in higher plants. *Environmental and Experimental Botany*. 41(2):105–130.
Tsutsui T, Kato W, Asada Y, Sako K, Sato T, Sonoda Y, Kidokoro S, Yamaguchi-Shinozaki K, Tamaoki M, Arakawa K, et al. 2009. *DEAR1*, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in *Arabidopsis*. *Journal of Plant Research*. 122(6):633–643.
Tuteja N. 2007. Abscisic Acid and Abiotic Stress Signaling. *Plant Signaling & Behavior*. 2(3):135–138.
Wang J, Zhang Q, Cui F, Hou L, Zhao S, Xia H, Qiu J, Li T, Zhang Y, Wang X, et al. 2017. Genome-Wide Analysis of Gene Expression Provides New Insights into Cold Responses in *Thellungiella salsuginea*. *Frontiers in Plant Science*. 8.
Wang M, Zhao Y, Zhang B. 2015. Efficient Test and Visualization of Multi-Set Intersections. *Scientific Reports*. 5:16923.
Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*. 19(12):639–644.
Wisniewski M, Fuller M. 1999. Ice nucleation and deep supercooling in plants: new insights using infrared thermography. In: Margesin R, Schinner F, editors. Cold Adapted Organisms. Springer. p. 105–118.
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.
Yang L, Wang H-N, Hou X-H, Zou Y-P, Han T-S, Niu X-M, Zhang J, Zhao Z, Todesco M, Balasubramanian S, et al. 2018. Parallel Evolution of Common Allelic Variants Confers Flowering Diversity in *Capsella rubella*. *The Plant Cell*. 30(6):1322–1336.
Yeaman S. 2015. Local Adaptation by Alleles of Small Effect. *The American Naturalist*. 186 Suppl 1:S74–89.
Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science*. 353(6306):1431–1433.
Zhang Z, Huang R. 2010. Enhanced tolerance to freezing in tobacco and tomato overexpressing transcription factor *TERF2/LeERF2* is modulated by ethylene biosynthesis. *Plant Molecular Biology*. 73(3):241–249.
Zhao Z, Tan L, Dang C, Zhang H, Wu Q, An L. 2012. Deep-sequencing transcriptome analysis of chilling tolerance mechanisms of a subnival alpine plant, *Chorispora bungeana*. *BMC Plant Biology*. 12(1):222.

Supplementary material, Paper II

Supplementary figures

Figure S1. UpSet plot of DEGs at 24h with 2 °C, based on putative *A. thaliana* homologs.

Figure S2. UpSet plot of upregulated DEGs at 24h with 2 °C, based on putative *A. thaliana* homologs.

Supplementary tables

Table S1. Sampling information for Arctic focal species

Table S2. DEGs with annotations: **A)** 3h, **B)** 6h, **C)** 24h *Cardamine bellidifolia*. *Separate excel file*.

Table S3. DEGs with annotations: **A)** 3h, **B)** 6h, **C)** 24h *Cochlearia groenlandica. Separate excel file*.

Table S4. DEGs with annotations: **A)** 3h, **B)** 6h, **C)** 24h *Draba nivalis*. *Separate excel file*.

Table S5. Overlapping DEGs among time points in *Cardamine bellidifolia*

Table S6. Overlapping DEGs among time points in *Cochlearia groenlandica*

Table S7. Overlapping DEGs among time points in *Draba nivalis*

Table S8. Results from the supertest.
Intersections of differentially expressed orthogroups.

Table S9. Genes found only in Arctic 24h DEG sets
(based on putative *A. thaliana homologs*)

Table S10. Genes shared by *A. thaliana* and Arctic species in 24h DEG sets (based on putative *A. thaliana homologs*)

Table S11. *Cardamine bellidifolia* DEGs and putative *Arabidopsis thaliana* homologs: BLAST results **A)** 3h, **B)** 6h, **C)** 24h. *Separate excel file*.

Table S12. *Cochlearia groenlandica* DEGs and putative *Arabidopsis thaliana* homologs: BLAST results **A)** 3h, **B)** 6h, **C)** 24h, *Separate excel file*.

Table S13. *Draba nivalis* DEGs and putative *Arabidopsis thaliana* homologs: BLAST results **A)** 3h, **B)** 6h, **C)** 24h, *Separate excel file*.

Table S14. topGO tables for *Cardamine bellidifolia* DEGs, Biological Process. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S15. topGO tables for *Cardamine bellidifolia* DEGs, Cellular Component. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S16. topGO tables for *Cardamine bellidifolia* DEGs, Molecular Function. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S17. topGO tables for *Cochlearia groenlandica* DEGs, Biological Process. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S18. topGO tables for *Cochlearia groenlandica* DEGs, Cellular Component. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S19. topGO tables for *Cochlearia groenlandica* DEGs, Molecular Function. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S20. topGO tables for *Draba nivalis* DEGs, Biological Process. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S21. topGO tables for *Draba nivalis* DEGs, Cellular Component. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S22. topGO tables for *Draba nivalis* DEGs, Molecular Function. *Separate excel file*.
A) upregulated DEGs 3h
B) upregulated DEGs 6h
C) upregulated DEGs 24h
D) downregulated DEGs 3h
E) downregulated DEGs 6h
F) downregulated DEGs 24h

Table S23. Significantly enriched GO-terms shared by all Arctic species (all DEG sets)

Table S24. Positively selected genes from Birkeland et al. 2020 found in Arctic DEG sets. *Separate excel file*.

Table S25. Convergent genes from Birkeland et al. 2020 found in 24h DEG sets. *Separate excel file*.
A) *Cardamine bellidifolia*
B) *Cochlearia groenlandica*
C) *Draba nivalis*

For excel files, see: https://doi.org/10.5061/dryad.7pvmcvdqx

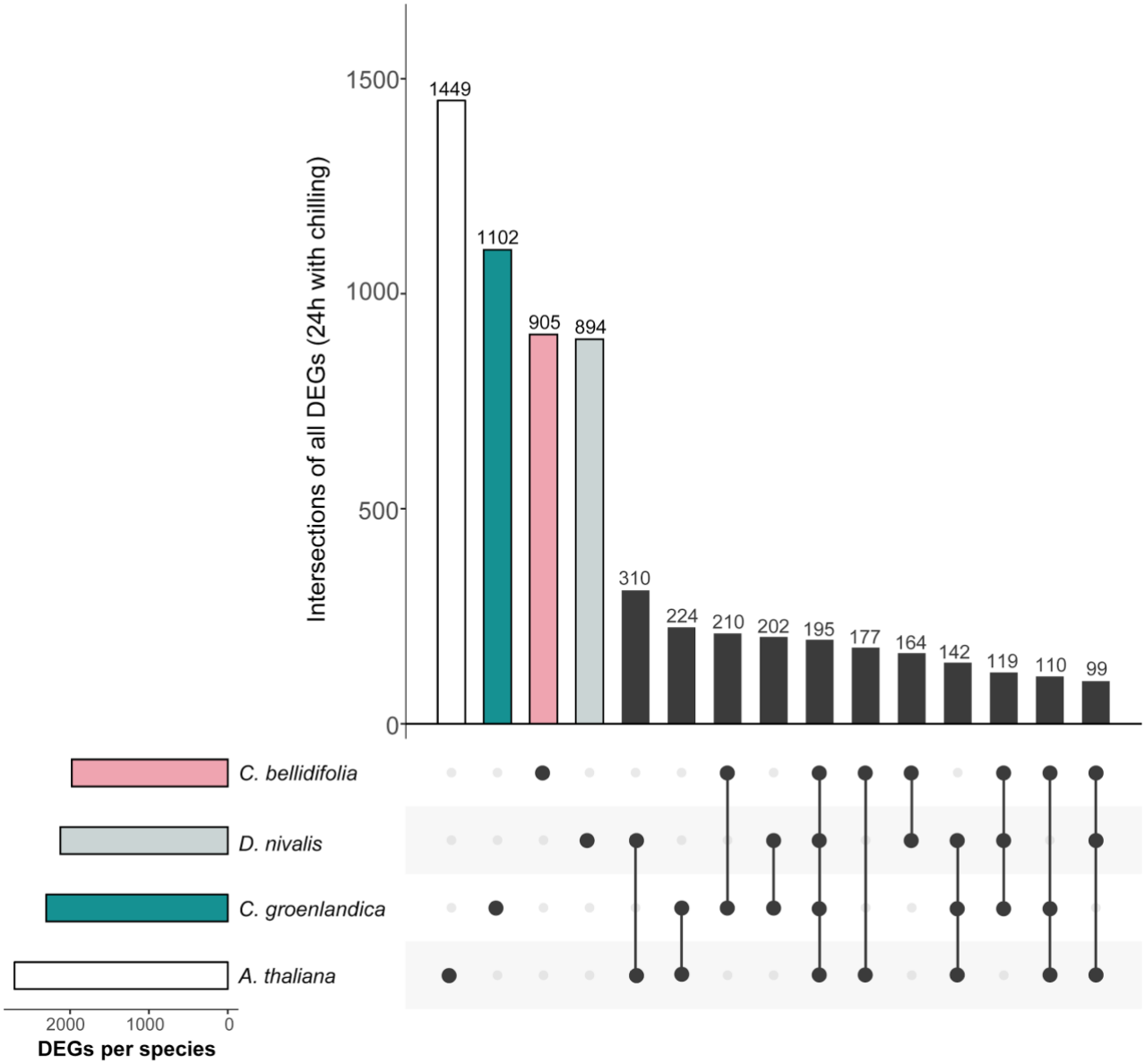
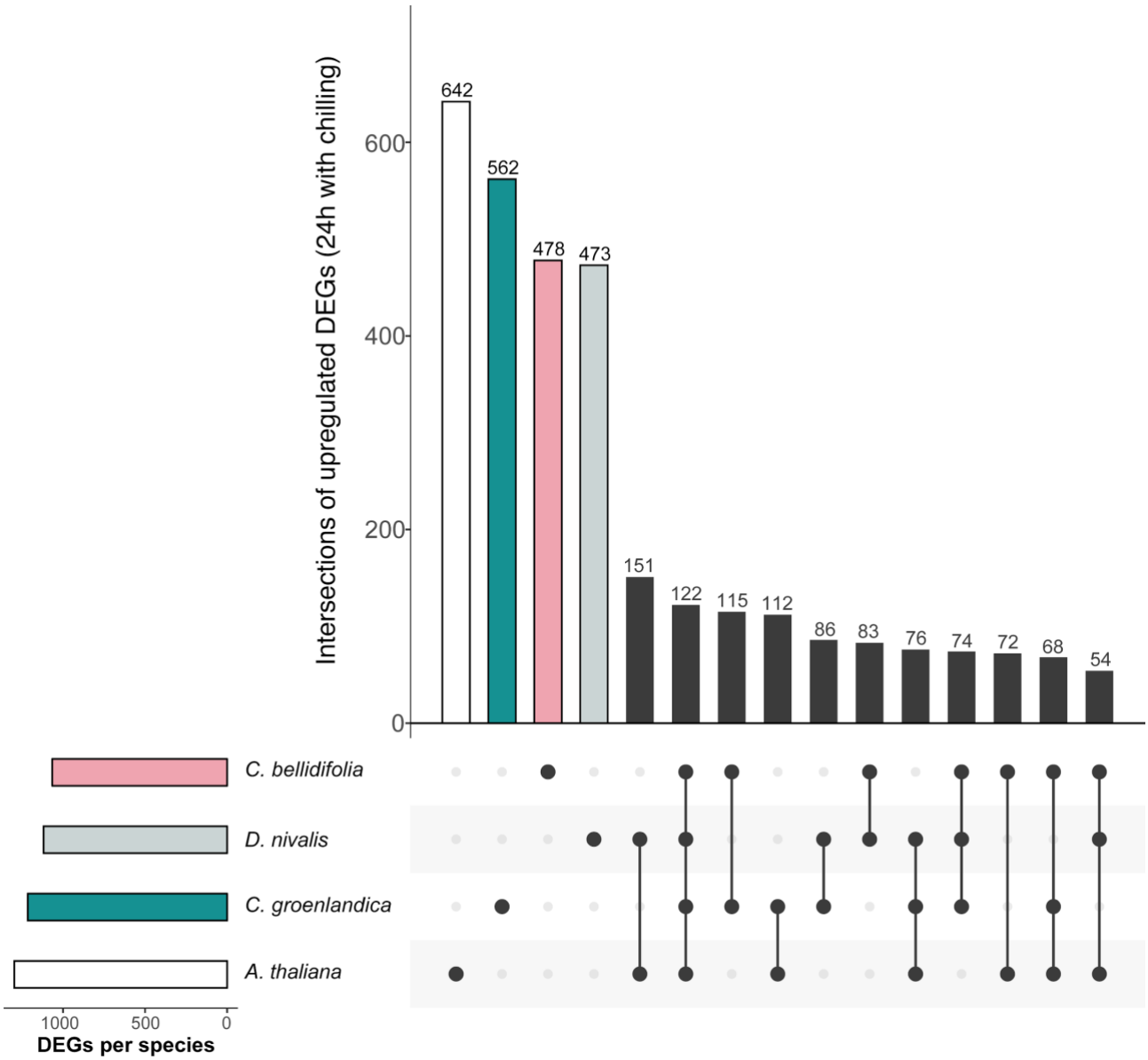


Figure S1. UpSet plot of differentially expressed genes at 24h with 2 ℃, based on putative *A. thaliana* homologs. The plot in the left corner shows total numbers of differentially expressed genes, and the main plot shows the number of unique differentially expressed genes, followed by gene intersections/overlaps between species (connected dots).



Supplementary figure 2. UpSet plot of upregulated differentially expressed genes at 24h with 2 °C, based on putative *A. thaliana* homologs. The plot in the left corner shows total numbers of differentially expressed genes, and the main plot shows the number of unique differentially expressed genes, followed by gene intersections/overlaps between species (connected dots).

Table S1. Sampling information for Arctic focal species

Species	Population ID	Region (Country)	Locality	Coordinates (DMS Latitude, DMS Longitude)
Cardamine bellidifolia	LG09-A-68-06-01	Alaska (US)	Atigun Pass, Brooks Range	N68°8'10.44" W149°28'37.56"
Cochlearia groenlandica	LG09-A-48-09	Alaska (US)	Nome, Seward Peninsula	N64° 29' 4.48" W165° 25' 49.9"
Draba nivalis	HHG-008-7-L-m	Alaska (US)	Waterfall Creek W	N63°2'42' W147°12'3.6"

Table S2. Overlapping DEGs among time points in *Cardamine bellidifolia*

A. Overlapping DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	1007 [0]	760 [247]	623 [384]
6h shared [unique]	760 [278]	1038 [0]	709 [329]
24h shared [unique]	623 [1887]	709 [1801]	2510 [0]

B. Overlapping **upregulated** DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	855 [0]	678 [177]	518 [337]
6h shared [unique]	678 [194]	872 [0]	590 [282]
24h shared [unique]	518 [777]	590 [705]	1295 [0]

C. Overlapping **downregulated** DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	152 [0]	82 [70]	97 [55]
6h shared [unique]	82 [84]	166 [0]	117 [49]
24h shared [unique]	97 [1118]	117 [1098]	1215 [0]

Table S3. Comparison of all DEGs among time points in *Cochlearia groenlandica*

A. Overlapping DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	729 [0]	506 [233]	516 [213]
6h shared [unique]	506 [497]	1003 [0]	685 [318]
24h shared [unique]	516 [2481]	685 [2312]	2997 [0]

B. Overlapping **upregulated** DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	517 [0]	369 [148]	354 [163]
6h shared [unique]	369 [260]	629 [0]	448 [181]
24h shared [unique]	354 [1172]	448 [1078]	1526 [0]

C. Overlapping **downregulated** DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	212 [0]	137 [75]	161 [51]
6h shared [unique]	137 [237]	374 [0]	237 [137]
24h shared [unique]	161 [1310]	237 [1234]	1471 [0]

Table S4. Comparison of all DEGs among time points in *Draba nivalis*

A. Overlapping DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	683 [0]	606 [77]	487 [196]
6h shared [unique]	606 [970]	1576 [0]	982 [594]
24h shared [unique]	487 [2341]	982 [1846]	2828 [0]

B. Overlapping **upregulated** DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	501 [0]	446 [55]	349 [152]
6h shared [unique]	446 [548]	994 [0]	591 [403]
24h shared [unique]	349 [1130]	591 [888]	1479 [0]

C. Overlapping **downregulated** DEGs among time points.

	Compared with 3h:	Compared with 6h	Compared with 24h
3h shared [unique]	182 [0]	160 [22]	137 [45]
6h shared [unique]	160 [422]	582 [0]	389 [193]
24h shared [unique]	137 [1212]	389 [960]	1349 [0]

Table S8. Results from the supertest. Significance of observed intersections between differentially expressed orthogroups in *Draba nivalis* (2167 differentially expressed orthogroups), *Cochlearia groenlandica* (2340 differentially expressed orthogroups), *Cardamine bellidifolia* (1986 differentially expressed orthogroups), and *Arabidopsis thaliana* (2428 differentially expressed orthogroups). Each orthogroup is counted only once per each species.

Intersection	Deg-ree	Back-ground *	Observed overlap**	Expected overlap	Fold Enrich-ment	*p-value*
C. groenlandica, D. nivalis	2	14,717	667	344.55	1.94	8.426537e-81
C. bellidifolia, D. nivalis	2	13,945	604	308.62	1.96	1.043355e-74
C. bellidifolia, C. groenlandica	2	14,052	635	330.72	1.92	1.435591e-160
C. bellidifolia A. thaliana	2	14,268	613	337.96	1.81	1.221776e-61
C. groenlandica A. thaliana	2	14,169	711	400.98	1.77	9.055851e-69
D. nivalis A. thaliana	2	14,182	768	371.00	2.07	2.261915e-114
C. groenlandica, D. nivalis, A. thaliana	3	12,684	367	76.53	4.80	5.176323e-150
C. bellidifolia, C. groenlandica, A. thaliana	3	12,702	334	69.94	4.78	8.524704e-135
C. bellidifolia, D. nivalis, A. thaliana	3	12,748	324	64.30	5.04	2.918931e-137
C. bellidifolia, C. groenlandica, D. nivalis	**3**	**12,429**	**318**	**65.19**	**4.88**	**1.079363e-130**
C. bellidifolia, C. groenlandica, D. nivalis, A. thaliana	4	11,715	212	15.21	13.94	4.218279e-173

*The background was set to the number of orthogroups shared between the species in each comparison. **Observed overlap includes orthogroups also shared with other species in the dataset (i.e. the observed overlap of 318 in *C. bellidifolia*, *C. groenlandica*, *D. nivalis*, includes the 212 orthogroups shared with *A. thaliana*).

Table S9. Genes found only in Arctic 24h DEG sets (*C. bellidifolia*, *C. groenlandica*, and *D. nivalis*) - based on blasting Arctic transcriptomes against Araport 11. ↑ = Upregulated in all species, ↓ = Downregulated in all species, D = different expression patterns in different species. Annotations from www.arabidopsis.org

↑/↓	Locus Identifier	Gene Model Description	Primary Gene Symbol
↑	AT1G01300	Eukaryotic aspartyl protease family protein;(source:Araport11)	
↑	AT1G02860	Encodes a ubiquitin E3 ligase with RING and SPX domains that is involved in mediating immune responses and mediates degradation of PHT1s at plasma membranes. Targeted by MIR827. Ubiquitinates PHT1;3, PHT1;2, PHT1;1/AtPT1 and PHT1;4/AtPT2.	NITROGEN LIMITATION ADAPTATION (NLA)
↑	AT1G04570	Similar to plastid solute transporters.	
↑	AT1G05100	member of MEKK subfamily	MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 18 (MAPKKK18)
↓	AT1G10240	FAR1-related sequence 11;(source:Araport11)	FAR1-RELATED SEQUENCE 11 (FRS11)
↑	AT1G12420	ACT domain repeat 8;(source:Araport11)	ACT DOMAIN REPEAT 8 (ACR8)
↑	AT1G14780	MAC/Perforin domain-containing protein;(source:Araport11)	
↑	AT1G14890	Plant invertase/pectin methylesterase inhibitor superfamily protein;(source:Araport11)	
↑	AT1G15100	Encodes a putative RING-H2 finger protein RHA2a.	RING-H2 FINGER A2A (RHA2A)
D	AT1G16840	hypothetical protein;(source:Araport11)	
↑	AT1G19490	Basic-leucine zipper (bZIP) transcription factor family protein;(source:Araport11)	
↑	AT1G21010	poly polymerase;(source:Araport11)	
↑	AT1G23330	alpha/beta-Hydrolases superfamily protein;(source:Araport11)	
↑	AT1G23550	Encodes a protein with similarity to RCD1 but without the WWE domain. The protein does have a PARP signature upstream of the C-terminal protein interaction domain. The PARP signature may bind NAD+ and attach the ADP-ribose-moiety from NAD+ to the target molecule. Its presence suggests a role for the protein in ADP ribosylation. Its transcript level is up-regulated by tunicamycin (N-linked glycosylation inhibitor causing ER stress).	SIMILAR TO RCD ONE 2 (SRO2)
↑	AT1G24580	RING/U-box superfamily protein;(source:Araport11)	

↑	AT1G26790	Dof-type zinc finger DNA-binding family protein;(source:Araport11)	CYCLING DOF FACTOR 6 (CDF6)
↓	AT1G29500	SAUR-like auxin-responsive protein family;(source:Araport11)	SMALL AUXIN UPREGULATED RNA 66 (SAUR66)
↓	AT1G36940	myotubularin-like protein;(source:Araport11)	
↑	AT1G62180	encodes a adenosine 5'-phosphosulfate reductase, involved in sulfate assimilation. Is a major effect locus for natural variation of shoot sulfate content in Arabidopsis.	5'ADENYLYLPHOSPHOSULFATE REDUCTASE 2 (APR2)
↑	AT1G70640	octicosapeptide/Phox/Bem1p (PB1) domain-containing protein;(source:Araport11)	
↑	AT1G71360	Encodes a member of the mid-SUN subfamily of SUN-domain proteins that is localized to both the nuclear envelope and the ER. It is involved in early seed development and nuclear morphology.	SUN-DOMAIN PROTEIN 4 (SUN4)
↓	AT1G72110	O-acyltransferase (WSD1-like) family protein;(source:Araport11)	
↑	AT1G72210	basic helix-loop-helix (bHLH) DNA-binding superfamily protein;(source:Araport11)	(BHLH096)
D	AT1G73805	Encodes SAR Deficient 1 (SARD1), a key regulator for ICS1 (Isochorismate Synthase 1) induction and salicylic acid (SA) synthesis.	SAR DEFICIENT 1 (SARD1)
D	AT1G74940	cyclin-dependent kinase, putative (DUF581);(source:Araport11)	
↑	AT1G75400	RING/U-box superfamily protein;(source:Araport11)	
↓	AT1G77855	BPS1-like protein;(source:Araport11)	
↓	AT2G01860	Tetratricopeptide repeat (TPR)-like superfamily protein;(source:Araport11)	EMBRYO DEFECTIVE 975 (EMB975)
↑	AT2G02990	Encodes a member of the ribonuclease T2 family that responds to inorganic phosphate starvation, and inhibits production of anthocyanin. Also involved in wound-induced signaling independent of jasmonic acid. Its expression is responsive to both phosphate (Pi) and phosphite (Phi) in roots.	RIBONUCLEASE 1 (RNS1)
↑	AT2G12400	plasma membrane fusion protein;(source:Araport11)	
↑	AT2G15020	hypothetical protein;(source:Araport11)	
↑	AT2G16720	Encodes a member of MYB3R- and R2R3- type MYB- encoding gene family that acts as a repressor of flavonol biosynthesis. AtMYB7 gene expression is induced by salt treatment.	MYB DOMAIN PROTEIN 7 (MYB7)
↑	AT2G18170	MAP kinase 7;(source:Araport11)	MAP KINASE 7 (MPK7)
↓	AT2G23690	HTH-type transcriptional regulator;(source:Araport11)	
↓	AT2G25420	transducin family protein / WD-40 repeat family protein;(source:Araport11)	

↑	AT2G30040	Member of MEKK subfamily. Induced by jasmonic acid and wounding in involved in insectivory response signaling. Iinteracts with At5g40440, and activates At1g59580.	MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 14 (MAPKKK14)
↓	AT2G35260	CAAX protease self-immunity protein;(source:Araport11)	BALANCE OF CHLOROPHYLL METABOLISM 1 (BCM1)
↑	AT2G36270	Encodes a member of the basic leucine zipper transcription factor family, involved in ABA signalling during seed maturation and germination. The Arabidopsis abscisic acid (ABA)-insensitive abi5 mutants have pleiotropic defects in ABA response, including decreased sensitivity to ABA inhibition of germination and altered expression of some ABA-regulated genes. Comparison of seed and ABA-inducible vegetative gene expression in wild-type and abi5-1 plants indicates that ABI5 regulates a subset of late embryogenesis-abundant genes during both developmental stages. Responsible for reducing cadmium uptake, mediated by interaction with MYB49 .	ABA INSENSITIVE 5 (ABI5)
↑	AT2G36320	A20/AN1-like zinc finger family protein;(source:Araport11)	
↑	AT2G36800	Encodes a DON-Glucosyltransferase. The UGT73C5 glucosylates both brassinolide and castasterone in the 23-O position. The enzyme is presumably involved in the homeostasis of those steroid hormones hence regulating BR activity. Transgenic plants overexpressing UGT73C5 show a typical BR-deficient phenotype.	DON-GLUCOSYLTRANSFERASE 1 (DOGT1)
D	AT2G37678	Positive regulator of photomorphogenesis in far-red light. Most abundant in young seedlings in the dark. Downregulated in the light and older as plants develop. Localized in the nucleus and the cytoplasm. Nuclear localization strongest in the dark. Degraded through the 26S proteasome. Regulated by PHYA. It is specifically required for the light-regulated nuclear accumulation of phyA (but not phyB) likely by shuttling PHYA into the nucleus.	FAR-RED ELONGATED HYPOCOTYL 1 (FHY1)
↑	AT2G38250	Homeodomain-like superfamily protein;(source:Araport11)	
↑	AT2G40080	Encodes a novel nuclear 111 amino-acid phytochrome-regulated component of a negative feedback loop involving the circadian clock central oscillator components CCA1 and LHY. ELF4 is necessary for light-induced expression of both CCA1 and LHY, and conversely, CCA1 and LHY act	EARLY FLOWERING 4 (ELF4)

		negatively on light-induced ELF4 expression. ELF4 promotes clock accuracy and is required for sustained rhythms in the absence of daily light/dark cycles. It is involved in the phyB-mediated constant red light induced seedling de-etiolation process and may function to coregulate the expression of a subset of phyB-regulated genes.	
↑	AT2G40130	Encodes a member of an eight-gene family (SMAX1 and SMAX1-like) that has weak similarity to AtHSP101, a ClpB chaperonin required for thermotolerance.	SMAX1-LIKE 8 (SMXL8)
↓	AT2G40150	Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain. TBL gene family has 46 members, two of which (TBR/AT5G06700 and TBL3/AT5G01360) have been shown to be involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers. A nomenclature for this gene family has been proposed (Volker Bischoff & Wolf Scheible, 2010, personal communication).	TRICHOME BIREFRINGENCE-LIKE 28 (TBL28)
↑	AT2G41380	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein;(source:Araport11)	
D	AT2G46950	cytochrome P450, family 709, subfamily B, polypeptide 2;(source:Araport11)	CYTOCHROME P450, FAMILY 709, SUBFAMILY B, POLYPEPTIDE 2 (CYP709B2)
↑	AT2G47460	MYB12 belongs to subgroup 7 of the R2R3-MYB family. It strongly activates the promoters of chalcone synthase (CHS), flavanone 3-hydroxylase (F3H), flavonol synthase (FLS) and - to a lesser extent - chalcone flavanone isomerase (CHI), but cannot activate the promoters of flavonoid-3'hydroxylase (F3'H) and dihydroflavonol 4-reductase (DF). The activation requires a functional MYB recognition element (MRE). Results from the myb12-1f allele indicate that an activation domain might be present in the C-terminus. Overexpression or knock-out plants do not show any obvious phenotype under greenhouse conditions. Young myb12-ko seedlings contain reduced amounts of flavonoids (quercetin and kaempferol), while seedlings as well as leaves of MYB12-OX plants displayed an increased flavonoid content. They did not show any significant difference in anthocyanin content. Expression of CHS and FLS	MYB DOMAIN PROTEIN 12 (MYB12)

		shows a clear correlation to MYB12 expression levels. CHI and F3H show increased transcript levels in the MYB12-OX lines, but no differences in the knock-out. Even in the absence of functional MYB12, flavonol biosynthesis is not completely absent, suggesting functional redundancy. The redundant factors are MYB11 and MYB111 although MYB12 is primarily required for flavonol biosynthesis in roots. Mutations in MYB12 block both auxin and ethylene stimulation of flavonoid synthesis.	
↑	AT3G02910	AIG2-like (avirulence induced gene) family protein;(source:Araport11)	
↑	AT3G09450	fusaric acid resistance family protein;(source:Araport11)	
↑	AT3G10300	Calcium-binding EF-hand family protein;(source:Araport11)	
↑	AT3G14230	encodes a member of the ERF (ethylene response factor) subfamily B-2 of ERF/AP2 transcription factor family (RAP2.2). The protein contains one AP2 domain. There are 5 members in this subfamily including RAP2.2 AND RAP2.12.	RELATED TO AP2 2 (RAP2.2)
D	AT3G14810	mechanosensitive channel of small conductance-like 5;(source:Araport11)	MECHANOSENSITIVE CHANNEL OF SMALL CONDUCTANCE-LIKE 5 (MSL5)
↑	AT3G15760	cytochrome P450 family protein;(source:Araport11)	
↑	AT3G16150	Encodes an asparaginase that catalyzes the degradation of L-asparagine to L-aspartic acid and ammonia. The mRNA is cell-to-cell mobile.	ASPARAGINASE B1 (ASPGB1)
↑	AT3G17520	Late embryogenesis abundant protein (LEA) family protein;(source:Araport11)	
↓	AT3G18773	RING/U-box superfamily protein;(source:Araport11)	ARABIDOPSIS TA^{3}XICOS EN LEVADURA 77 (ATL77)
↑	AT3G20500	purple acid phosphatase 18;(source:Araport11)	PURPLE ACID PHOSPHATASE 18 (PAP18)
↓	AT3G23930	troponin T, skeletal protein;(source:Araport11)	
↓	AT3G25500	Poly-L-proline-containing (PLP) protein that form part of the signal-transduction cascade that leads to rearrangement of the actin cytoskeleton. AFH1 is a nonprocessive formin that moves from the barbered end to the side of an actin filament after the nucleation event.	FORMIN HOMOLOGY 1 (AFH1)
↓	AT3G26490	Phototropic-responsive NPH3 family protein;(source:Araport11)	

↓	AT3G26932	dsRNA-binding protein 3;(source:Araport11)	DSRNA-BINDING PROTEIN 3 (DRB3)
↑	AT3G46640	Encodes a myb family transcription factor with a single Myb DNA-binding domain (type SHAQKYF) that is unique to plants and is essential for circadian rhythms, specifically for transcriptional regulation within the circadian clock. LUX is required for normal rhythmic expression of multiple clock outputs in both constant light and darkness. It is coregulated with TOC1 and seems to be repressed by CCA1 and LHY by direct binding of these proteins to the evening element in the LUX promoter. The mRNA is cell-to-cell mobile.	PHYTOCLOCK 1 (PCL1)
D	AT3G47680	DNA binding protein;(source:Araport11)	
↓	AT3G51950	Contains single CCCH domain.	
↑	AT3G52740	Plant specific protein.BIC1 and BIC2 inhibit cryptochrome function by blocking blue light-dependent cryptochrome dimerization.Light activated transcription of BICs is mediated by cryptochromes.	BLUE-LIGHT INHIBITOR OF CRYPTOCHROMES 1 (BIC1)
↑	AT3G52930	Aldolase superfamily protein;(source:Araport11)	FRUCTOSE-BISPHOSPHATE ALDOLASE 8 (FBA8)
↓	AT3G54880	zinc finger protein;(source:Araport11)	
↓	AT3G62390	Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family containing a plant-specific DUF231 (domain of unknown function) domain. TBL gene family has 46 members, two of which (TBR/AT5G06700 and TBL3/AT5G01360) have been shown to be involved in the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers. A nomenclature for this gene family has been proposed (Volker Bischoff & Wolf Scheible, 2010, personal communication).	TRICHOME BIREFRINGENCE-LIKE 6 (TBL6)
↑	AT4G01120	bZIP (basic leucine zipper) transcription factor that binds to the G-box regulatory element found in many plant promoters. GBF2 nuclear localization is increased by blue light	G-BOX BINDING FACTOR 2 (GBF2)
↑	AT4G02280	Encodes a protein with sucrose synthase activity (SUS3). It appears to be important for sucrose metabolism in developing seeds, especially during the late maturation phase, about 18 days after flowering.	SUCROSE SYNTHASE 3 (SUS3)
↓	AT4G03210	encodes a member of xyloglucan endotransglucosylase/hydrolases (XTHs) that catalyze the cleavage and molecular grafting of xyloglucan chains function in loosening and rearrangement of the cell wall. Gene is expressed in shoot apex region,	XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 9 (XTH9)

		flower buds, flower stalks and internodes bearing flowers.	
↓	AT4G03390	STRUBBELIG-receptor family 3;(source:Araport11)	STRUBBELIG-RECEPTOR FAMILY 3 (SRF3)
↓	AT4G04630	senescence regulator (Protein of unknown function, DUF584);(source:Araport11)	
↑	AT4G12490	Encodes a member of the AZI family of lipid transfer proteins. Contains a PRR domain that appears to be required for localization to the chloroplast.	(AZI3)
↑	AT4G16750	encodes a member of the DREB subfamily A-4 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 17 members in this subfamily including TINY.	ETHYLENE-RESPONSIVE TRANSCRIPTION FACTOR 39 (ERF39)
↑	AT4G17140	pleckstrin homology (PH) domain-containing protein;(source:Araport11)	
↑	AT4G17900	PLATZ transcription factor family protein;(source:Araport11)	(PLATZ11)
↓	AT4G20820	FAD-binding Berberine family protein;(source:Araport11)	(ATBBE18)
↑	AT4G22820	A member of the A20/AN1 zinc finger protein family involved in stress response. Expression is increased in response to water, salt , pathogen and other stressors.SAP9 can pull down both K48-linked and K63- linked tetraubiquitin chains and functions as a E3 ubiquitin ligase suggesting a role in proteasome-dependent protein degradation.	STRESS ASSOCIATED PROTEIN 9 (SAP9)
↓	AT4G26790	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactions with lipid and non-lipid substrates.	
↓	AT4G27440	light-dependent NADPH:protochlorophyllide oxidoreductase B The mRNA is cell-to-cell mobile.	PROTOCHLOROPHYLLIDE OXIDOREDUCTASE B (PORB)
↓	AT4G29310	DUF1005 family protein (DUF1005);(source:Araport11)	
↓	AT4G31510	major centromere autoantigen B-like protein;(source:Araport11)	
↓	AT4G32480	sugar phosphate exchanger, putative (DUF506);(source:Araport11)	
↑	AT4G33625	vacuole protein;(source:Araport11)	
↑	AT4G33985	membrane insertase, putative (DUF1685);(source:Araport11)	
↑	AT4G36900	Encodes a member of the DREB subfamily A-5 of ERF/AP2 transcription factor family (RAP2.10). The protein contains one AP2 domain. There are 16 members in this subfamily including RAP2.9 and RAP2.1.	RELATED TO AP2 10 (RAP2.10)

D	AT5G02180	Transmembrane amino acid transporter family protein;(source:Araport11)	
↑	AT5G02270	member of NAP subfamily	ATP-BINDING CASSETTE I20 (ABCI20)
D	AT5G02890	Encodes a protein with similarity to transferases in plants and fungi.	
↑	AT5G05220	hypothetical protein;(source:Araport11)	
↓	AT5G06790	cotton fiber protein;(source:Araport11)	
↑	AT5G10410	ENTH/ANTH/VHS superfamily protein;(source:Araport11)	(PICALM10B)
↑	AT5G10750	enhanced disease resistance-like protein (DUF1336);(source:Araport11)	
↑	AT5G11090	serine-rich protein-like protein;(source:Araport11)	
↑	AT5G11260	Basic leucine zipper (bZIP) transcription factor. Nuclear localization. Involved in light-regulated transcriptional activation of G-box-containing promoters. Negatively regulated by Cop1. Although cytokinins do not appear to affect the gene's promoter activity, they appear to stabilize the protein. HY5 plays a role in anthocyanin accumulation in far-red light and blue light, but not in red light or in the dark. Mutant studies showed that the gene product is involved in the positive regulation of the PHYA-mediated inhibition of hypocotyl elongation. Binds to G- and Z-boxes, and other ACEs, but not to E-box. Loss of function mutation shows ABA resistant seedling phenotypes suggesting involvement for HY5 in mediating ABA responses. Binds to the promoter of ABI5 and regulates its expression.Involved in the regulation of response to nutrient levels.	ELONGATED HYPOCOTYL 5 (HY5)
↑	AT5G15860	Encodes a protein with prenylcysteine methylesterase activity.	PRENYLCYSTEINE METHYLESTERASE (PCME)
↑	AT5G16360	NC domain-containing protein-like protein;(source:Araport11)	
↓	AT5G19190	hypothetical protein;(source:Araport11)	
↓	AT5G22940	Homolog of FRA8 (AT2G28110), a member of a member of glycosyltransferase family 47; exhibits high sequence similarity to tobacco (Nicotiana plumbaginifolia) pectin glucuronyltransferase.	FRA8 HOMOLOG (F8H)
↑	AT5G37260	Encodes a MYB family transcription factor Circadian 1 (CIR1). Involved in circadian regulation in Arabidopsis.	REVEILLE 2 (RVE2)
↑	AT5G42760	Leucine carboxyl methyltransferase;(source:Araport11)	

↑	AT5G43630	Encodes a zinc knuckle protein that negatively regulates morning specific growth. The role of TZP in hypocotyl elongation was established through a QTL analysis of BayXSha RIL populations. The Bay-0 allele contains a deletion causing a frameshift mutation. TZP is under circadian control and acts to regulate morning-specific hypocotyl growth. The mRNA is cell-to-cell mobile.	TANDEM ZINC KNUCKLE PROTEIN (TZP)
↓	AT5G47370	homeobox-leucine zipper genes induced by auxin, but not by other phytohormones. Plays opposite roles in the shoot and root tissues in regulating auxin-mediated morphogenesis.	(HAT2)
↓	AT5G49690	UDP-Glycosyltransferase superfamily protein;(source:Araport11)	
↑	AT5G51990	encodes a member of the DREB subfamily A-1 of ERF/AP2 transcription factor family (CBF4). The protein contains one AP2 domain. There are six members in this subfamily, including CBF1, CBF2, and CBF3. This gene is involved in response to drought stress and abscisic acid treatment, but not to low temperature.	C-REPEAT-BINDING FACTOR 4 (CBF4)
↑	AT5G52510	SCARECROW-like 8;(source:Araport11)	SCARECROW-LIKE 8 (SCL8)
↓	AT5G52900	Encodes a member of the MAKR (MEMBRANE-ASSOCIATED KINASE REGULATOR) gene family. MAKRs have putative kinase interacting motifs and membrane localization signals. Known members include: AT5G26230 (MAKR1), AT1G64080 (MAKR2), AT2G37380 (MAKR3), AT2G39370 (MAKR4), AT5G52870 (MAKR5) and AT5G52900 (MAKR6).	MEMBRANE-ASSOCIATED KINASE REGULATOR 6 (MAKR6)
↑	AT5G53290	encodes a member of the ERF (ethylene response factor) subfamily B-5 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 7 members in this subfamily. CRF proteins relocalize to the nucleus in response to cytokinin.	CYTOKININ RESPONSE FACTOR 3 (CRF3)
↓	AT5G53730	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family;(source:Araport11)	NDR1/HIN1-LIKE 26 (NHL26)
↑	AT5G54960	pyruvate decarboxylase-2	PYRUVATE DECARBOXYLASE-2 (PDC2)
↑	AT5G55970	RING/U-box superfamily protein;(source:Araport11)	
↑	AT5G57100	Nucleotide/sugar transporter family protein;(source:Araport11)	
↓	AT5G57340	ras guanine nucleotide exchange factor Q-like protein;(source:Araport11)	
D	AT5G61340	transmembrane protein;(source:Araport11)	

↑	AT5G61380	Pseudo response regulator involved in the generation of circadian rhythms. TOC1 appears to shorten the period of circumnutation speed. TOC1 contributes to the plant fitness (carbon fixation, biomass) by influencing the circadian clock period. PRR3 may increase the stability of TOC1 by preventing interactions between TOC1 and the F-box protein ZTL. Expression of TOC1 is correlated with rhythmic changes in chromatin organization. The mRNA is cell-to-cell mobile.	TIMING OF CAB EXPRESSION 1 (TOC1)
↑	AT5G61660	glycine-rich protein;(source:Araport11)	
↑	AT5G62070	Member of IQ67 (CaM binding) domain containing family.	IQ-DOMAIN 23 (IQD23)
↑	AT5G66400	Belongs to the dehydrin protein family, which contains highly conserved stretches of 7-17 residues that are repetitively scattered in their sequences, the K-, S-, Y- and lysine rich segments. ABA- and drought-induced glycine-rice dehydrin protein. The ABA-induced expression of RAB18 was reduced following ACC application, indicating that ethylene inhibits the ABA signaling pathway. RAB18 is also expressed in response to the formation of the phospholipid diacylglycerol pyrophosphate. COR47 and RAB18 double overexpressor plants are cold tolerant. Expressed in guard cells.	RESPONSIVE TO ABA 18 (RAB18)

Table S10. Genes with similar expression patterns found in all 24h DEG sets (*A. thaliana*, *C. bellidifolia*, *C. groenlandica*, and *D. nivalis*) - based on blasting Arctic transcriptomes against Araport 11. ↑ = Upregulated in all species, ↓ = Downregulated in all species, D = different expression patterns in different species. Annotations from www.arabidopsis.org

↑/↓	Locus Identifier	Gene Model Description	Primary Gene Symbol
↑	AT1G01470	Encodes late-embryogenesis abundant protein whose mRNA levels are induced in response to wounding and light stress. Might be involved in protection against desiccation.	LATE EMBRYOGENESIS ABUNDANT 14 (LEA14)
↑	AT1G02270	Calcium-binding endonuclease/exonuclease/phosphatase family;(source:Araport11)	
↓	AT1G03090	MCCA is the biotinylated subunit of the dimer MCCase, which is involved in leucine degradation. Both subunits are nuclear coded and the active enzyme is located in the mitochondrion.	(MCCA)
↓	AT1G03870	fasciclin-like arabinogalactan-protein 9 (Fla9). Possibly involved in embryogenesis and seed development.	FASCICLIN-LIKE ARABINOOGALACTAN 9 (FLA9)
↓	AT1G09750	Eukaryotic aspartyl protease family protein;(source:Araport11)	
↑	AT1G10410	CW14 protein (DUF1336);(source:Araport11)	
↑	AT1G11210	cotton fiber protein, putative (DUF761);(source:Araport11)	
↓	AT1G11380	PLAC8 family protein;(source:Araport11)	
↑	AT1G11390	Protein kinase superfamily protein;(source:Araport11)	
D	AT1G14890	Plant invertase/pectin methylesterase inhibitor superfamily protein;(source:Araport11)	
↑	AT1G16850	transmembrane protein;(source:Araport11)	
↓	AT1G18620	Member of a small gene family in Arabidopsis. Quadruple mutants in this family display defects in cell elongation.	TON1 RECRUITING MOTIF 3 (TRM3)
↑	AT1G18850	PCP2 encodes a novel plant specific protein that is co-expressed with components of pre-rRNA processing complex. Co-localizes with NuGWD1 and SWA1.	LANT-SPECIFIC COMPONENT OF THE PRE- RRNA PROCESSING COMPLEX2 (PCP2)
D	AT1G19530	DNA polymerase epsilon catalytic subunit A;(source:Araport11)	
↑	AT1G20450	Encodes a gene induced by low temperature and dehydration. Inhibits e.coli growth while overexpressed. Belongs to the dehydrin protein family, which contains highly conserved stretches of 7-17 residues that are repetitively scattered in their sequences, the K-, S-, Y- and lysine rich segments.	EARLY RESPONSIVE TO DEHYDRATION 10 (ERD10)

		LTI29 and LTI30 double overexpressors confer cold tolerance. Localized to membranes and cytoplasm.	
↓	AT1G22330	RNA-binding (RRM/RBD/RNP motifs) family protein;(source:Araport11)	
↑	AT1G22770	Together with CONSTANTS (CO) and FLOWERING LOCUS T (FT), GIGANTEA promotes flowering under long days in a circadian clock-controlled flowering pathway. GI acts earlier than CO and FT in the pathway by increasing CO and FT mRNA abundance. Located in the nucleus. Regulates several developmental processes, including photoperiod-mediated flowering, phytochrome B signaling, circadian clock, carbohydrate metabolism, and cold stress response. The gene's transcription is controlled by the circadian clock and it is post-transcriptionally regulated by light and dark. Forms a complex with FKF1 on the CO promoter to regulate CO expression. The mRNA is cell-to-cell mobile.	GIGANTEA (GI)
↓	AT1G27210	ARM repeat superfamily protein;(source:Araport11)	
↑	AT1G27760	Encodes a protein with similarity to human interferon-related developmental regulator (IFRD)that is involved in salt tolerance. Loss of function mutations are hypersensitive to salt stress and have reduced fertility. SAT32 is found in the cytoplasm but appears to translocate to the nucleus when plants are subject to salt stress.	SALT-TOLERANCE 32 (SAT32)
↑	AT1G27910	Encodes a protein containing a UND, a U-box, and an ARM domain. This protein has E3 ubiquitin ligase activity based on in vitro assays.	PLANT U-BOX 45 (PUB45)
↑	AT1G30620	encodes a type-II membrane protein that catalyzes 4-epimerization of UDP-D-Xylose to UDP-L-Arabinose in vitro, the nucleotide sugar used by glycosyltransferases in the arabinosylation of cell wall polysaccharides and wall-resident proteoglycans.	MURUS 4 (MUR4)
↑	AT1G32860	Glycosyl hydrolase superfamily protein;(source:Araport11)	
↑	AT1G43890	ras-related small GTPase	RAB GTPASE HOMOLOG B18 (RAB18)
↑	AT1G48100	Pectin lyase-like superfamily protein;(source:Araport11)	POLYGALACTURONASE INVOLVED IN EXPANSION3 (PGX3)

↓	AT1G49200	RING/U-box superfamily protein;(source:Araport11)	ARABIDOPSIS TA^{3}XICOS EN LEVADURA 75 (ATL75)
↑	AT1G49720	Identified as a protein that binds to abscisic acid response elements. May mediate transcriptional regulation of ABA responses.	ABSCISIC ACID RESPONSIVE ELEMENT-BINDING FACTOR 1 (ABF1)
↑	AT1G51090	Heavy metal associated domain containing protein involved in plant immunity.Mutants show an increase in root length under NO stress and reduction in root length under H2O2 stress conditions. Mutants show increases in defense responses to pathogens including hypersensitive lesions, increased resistance and induction of SAR genes.	(ATHMAD1)
↓	AT1G51940	Encodes a LysM-containing receptor-like kinase. Induction of chitin-responsive genes by chitin treatment is not blocked in the mutant. Based on protein sequence alignment analysis, it has a typical RD signaling domain in its catalytic loop and possesses autophosphorylation activity.It is required for the suppression of defense responses in absence of pathogen infection or upon abscisic acid treatment. Loss-of-function mutants display enhanced resistance to Botrytis cinerea and Pectobacterium carotovorum. Its expression is repressed by pathogen infection and biological elicitors and is induced abscisic acid.Expression is strongly repressed by elicitors and fungal infection, and is induced by the hormone abscisic acid (ABA). Insertional mutants show increased expression of PHYTOALEXIN-DEFICIENT 3 (PAD3), enhanced resistance to Botrytis cinerea and Pectobacterium carotovorum infection and reduced physiological responses to ABA, suggesting that LYK3 is important for the cross-talk between signaling pathways activated by ABA and pathogens (PMID:24639336).	LYSM-CONTAINING RECEPTOR-LIKE KINASE 3 (LYK3)
↓	AT1G52190	Encodes a low affinity nitrate transporter that is expressed in the plasma membrane and found in the phloem of the major veins of leaves. It is responsible for nitrate redistribution to young leaves.	NRT1/ PTR FAMILY 1.2 (NPF1.2)
↓	AT1G54820	Protein kinase superfamily protein;(source:Araport11)	
↓	AT1G56220	Dormancy/auxin associated family protein;(source:Araport11)	
↑	AT1G60190	Encodes PUB19, a plant U-box armadillo repeat protein. Involved in salt inhibition of germination together with PUB18. The mRNA is cell-to-cell mobile.	PLANT U-BOX 19 (PUB19)

↑	AT1G61640	Protein kinase superfamily protein;(source:Araport11)	
↑	AT1G61730	DNA-binding storekeeper protein-related transcriptional regulator;(source:Araport11)	
↑	AT1G61800	glucose6-Phosphate/phosphate transporter 2. Expression is upregulated in the shoot of cax1/cax3 mutant. The mRNA is cell-to-cell mobile.	GLUCOSE-6-PHOSPHATE/PHOSPHATE TRANSLOCATOR 2 (GPT2)
↑	AT1G67360	Encodes a small rubber particle protein homolog. Plays dual roles as positive factors for tissue growth and development and in drought stress responses.	LD-ASSOCIATED PROTEIN 1 (LDAP1)
↓	AT1G68190	B-box zinc finger family protein;(source:Araport11)	B-BOX DOMAIN PROTEIN 27 (BBX27)
↓	AT1G68520	B-box type zinc finger protein with CCT domain-containing protein;(source:Araport11)	B-BOX DOMAIN PROTEIN 14 (BBX14)
↓	AT1G69530	Member of Alpha-Expansin Gene Family. Naming convention from the Expansin Working Group (Kende et al, Plant Mol Bio). Involved in the formation of nematode-induced syncytia in roots of Arabidopsis thaliana.	EXPANSIN A1 (EXPA1)
↑	AT1G69870	Encodes a low affinity nitrate transporter NRT1.7. Expressed in phloem. Responsible for source-to-sink remobilization of nitrate. The mRNA is cell-to-cell mobile.	NRT1/ PTR FAMILY 2.13 (NPF2.13)
↑	AT1G70900	hypothetical protein;(source:Araport11)	
↓	AT1G71030	Encodes a putative myb family transcription factor. In contrast to most other myb-like proteins its myb domain consists of a single repeat. A proline-rich region potentially involved in transactivation is found in the C-terminal part of the protein. Its transcript accumulates mainly in leaves.	MYB-LIKE 2 (MYBL2)
↓	AT1G72300	Encodes a leucine-rich repeat receptor kinase (LRR-RK) involved in the perception of PSY1. PSY1 is an 18-aa tyrosine-sulfated glycopeptide encoded by AT5G58650 that promotes cellular proliferation and expansion.	PSY1 RECEPTOR (PSY1R)
↓	AT1G74670	Gibberellin-regulated family protein;(source:Araport11)	GA-STIMULATED ARABIDOPSIS 6 (GASA6)
↓	AT1G75380	Encodes a nuclease involved in ABA-mediated callose deposition. It has been shown to interact with JAZ proteins, binds to a jasmonic acid-responsive element (JARE) and repress AtJMT expression.	BIFUNCTIONAL NUCLEASE IN BASAL DEFENSE RESPONSE 1 (BBD1)
↑	AT1G77120	Catalyzes the reduction of acetaldehyde using NADH as reductant. Requires zinc for activity. Dimer. Anaerobic response polypeptide (ANP). Fermentation. The protein undergoes thiolation following treatment	ALCOHOL DEHYDROGENASE 1 (ADH1)

		with the oxidant tert-butylhydroperoxide. The mRNA is cell-to-cell mobile.	
↑	AT1G78070	Transducin/WD40 repeat-like superfamily protein;(source:Araport11)	
↑	AT1G78570	Encodes a UDP-L-Rhamnose synthase involved in the biosynthesis of rhamnose, a major monosaccharide component of pectin. Catalyzes the conversion of UDP-D-Glc to UDP-L-Rha. The dehydrogenase domain of RHM1 was shown to catalyze the conversion of UDP-D-Glc to the reaction intermediate UDP-4-keto-6-deoxy-D-Glc using recombinant protein assay but the activity of the full-length protein was not determined as it could not be expressed in E. coli.	RHAMNOSE BIOSYNTHESIS 1 (RHM1)
D	AT1G78600	light-regulated zinc finger protein 1;(source:Araport11)	LIGHT-REGULATED ZINC FINGER PROTEIN 1 (LZF1)
↑	AT1G80130	Tetratricopeptide repeat (TPR)-like superfamily protein;(source:Araport11)	
↓	AT1G80440	Encodes a member of a family of F-box proteins, called the KISS ME DEADLY (KMD) family, that targets type-B ARR proteins for degradation and is involved in the negative regulation of the cytokinin response. Also named as KFB20, a member of a group of Kelch repeat F-box proteins that negatively regulate phenylpropanoid biosynthesis by targeting the phenypropanoid biosynthesis enzyme phenylalanine ammonia-lyase. The mRNA is cell-to-cell mobile.	KISS ME DEADLY 1 (KMD1)
↓	AT1G80920	A nuclear encoded soluble protein found in the chloroplast stroma. Negatively regulated by light and has rapid turnover in darkness.	(J8)
↑	AT2G03760	Encodes a brassinosteroid sulfotransferase. In vitro experiements show that this enzyme has a preference for 24-epibrassinosteroids, particularly 24-epicathasterone, but does not act on castasterone and brassinolide. It also shows sulfating activity toward flavonoids. It is differentially expressed during development, being more abundant in young seedlings and actively growing cell cultures. Expression is induced in response to salicylic acid and methyl jasmonate and bacterial pathogens.	SULPHOTRANSFERASE 12 (SOT12)
↑	AT2G11810	MGD3 is the major enzyme for galactolipid metabolism during phosphate starvation. Does not contribute to galactolipid synthesis under P1-sufficient conditions.	MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE TYPE C (MGDC)

↑	AT2G15970	encodes an alpha form of a protein similar to the cold acclimation protein WCOR413 in wheat. Expression is induced by short-term cold-treatment, water deprivation, and abscisic acid treatment. The mRNA is cell-to-cell mobile.	COLD REGULATED 413 PLASMA MEMBRANE 1 (COR413-PM1)
↑	AT2G16890	UDP-Glycosyltransferase superfamily protein;(source:Araport11)	
↑	AT2G17840	Identified as drought-inducible gene by differential hybridization. Upregulated by high light, drought, cold and salt stress determined by microarray analysis.	EARLY-RESPONSIVE TO DEHYDRATION 7 (ERD7)
↓	AT2G18300	basic helix-loop-helix (bHLH) DNA-binding superfamily protein;(source:Araport11)	HOMOLOG OF BEE2 INTERACTING WITH IBH 1 (HBI1)
D	AT2G19570	Encodes a cytidine deaminase that deaminates cytidine and deoxycytidine and is competitively inhibited by cytosine-containing compounds.	CYTIDINE DEAMINASE 1 (CDA1)
↑	AT2G20370	Encodes a xyloglucan galactosyltransferase located in the membrane of Golgi stacks that is involved in the biosynthesis of fucose. It is also involved in endomembrane organization. It is suggested that it is a dual-function protein that is responsible for actin organization and the synthesis of cell wall materials. The mRNA is cell-to-cell mobile.	MURUS 3 (MUR3)
↓	AT2G20670	sugar phosphate exchanger, putative (DUF506);(source:Araport11)	
↑	AT2G21660	Encodes a small glycine-rich RNA binding protein that is part of a negative-feedback loop through which AtGRP7 regulates the circadian oscillations of its own transcript. Gene expression is induced by cold. GRP7 appears to promote stomatal opening and reduce tolerance under salt and dehydration stress conditions, but, promotes stomatal closing and thereby increases stress tolerance under conditions of cold tolerance. Loss of function mutations have increased susceptibility to pathogens suggesting a role in mediating innate immune response. Mutants are also late flowering in a non-photoperiodic manner and are responsive to vernalization suggesting an interaction with the autonomous flowering pathway. There is a reduction of mRNA export from the nucleus in grp7 mutants. GRP7:GFP fusion proteins can be found in the cytosol and nucleus. A substrate of the type III effector HopU1 (mono-ADP-ribosyltransferase).	COLD, CIRCADIAN RHYTHM, AND RNA BINDING 2 (CCR2)
↑	AT2G22080	transmembrane protein;(source:Araport11)	
↓	AT2G24150	heptahelical transmembrane protein HHP3	HEPTAHELICAL PROTEIN 3 (HHP3)

↑	AT2G25530	AFG1-like ATPase family protein;(source:Araport11)	
↓	AT2G25900	Encodes a protein with two tandem-arrayed CCCH-type zinc fingers that binds RNA and is involved in RNA turnover. The mRNA is cell-to-cell mobile.	(ATCTH)
↑	AT2G29090	Encodes a protein with ABA 8'-hydroxylase activity, involved in ABA catabolism. Member of the CYP707A gene family. This gene predominantly accumulates in dry seeds and is up-regulated immediately following imbibition. CYP707A2 appears to play a major role in the rapid decrease in ABA levels during early seed imbibition.	CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 2 (CYP707A2)
↑	AT2G31360	Encodes a protein homologous to delta 9 acyl-lipid desaturases of cyanobacteria and acyl-CoA desaturases of yeast and mammals. expression up-regulated by cold temperature. It is involved in the synthesis of the 24:1n-9 and 26:1n-9 components of seed lipids, sphingolipids and the membrane phospholipids phosphatidylserine (PS), and phosphatidylethanolamine (PE).	16:0DELTA9 DESATURASE 2 (ADS2)
↓	AT2G32100	ovate family protein 16;(source:Araport11)	OVATE FAMILY PROTEIN 16 (OFP16)
↓	AT2G34620	Mitochondrial transcription termination factor family member.	(MTERF10)
↑	AT2G35480	envelope glycoprotein;(source:Araport11)	
↑	AT2G36010	Member of the E2F transcription factors, (cell cycle genes), key components of the cyclin D/retinoblastoma/E2F pathway.	E2F TRANSCRIPTION FACTOR 3 (E2F3)
↓	AT2G36050	ovate family protein 15;(source:Araport11)	OVATE FAMILY PROTEIN 15 (OFP15)
↑	AT2G37970	SOUL heme-binding family protein;(source:Araport11)	(SOUL-1)
↓	AT2G38120	Encodes an auxin influx transporter. AUX1 resides at the apical plasma membrane of protophloem cells and at highly dynamic subpopulations of Golgi apparatus and endosomes in all cell types. AUX1 action in the lateral root cap and/or epidermal cells influences lateral root initiation and positioning. Shoot supplied ammonium targets AUX1 and inhibits lateral root emergence. The mRNA is cell-to-cell mobile.	AUXIN RESISTANT 1 (AUX1)
↑	AT2G39710	Encodes a Cysteine-rich peptide (CRP) family protein	
↑	AT2G40340	Encodes a member of the DREB subfamily A-2 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are eight members in this subfamily including DREB2A AND DREB2B that are involved in response to drought.	(DREB2C)

↑	AT2G42530	Encodes COR15B, a protein that protects chloroplast membranes during freezing.	COLD REGULATED 15B (COR15B)
↓	AT2G44740	cyclin p4;(source:Araport11)	CYCLIN P4;1 (CYCP4;1)
↑	AT2G47770	Encodes a membrane-bound protein designated AtTSPO (Arabidopsis thaliana TSPO-related). AtTSPO is related to the bacterial outer membrane tryptophan-rich sensory protein (TspO) and the mammalian mitochondrial 18 kDa Translocator Protein (18 kDa TSPO), members of the TspO/MBR domain-containing membrane proteins. Mainly detected in dry seeds, but can be induced in vegetative tissues by osmotic or salt stress or abscisic acid treatment. Located in endoplasmic reticulum and the Golgi stacks. It is degraded through the autophagy pathway.	TSPO (OUTER MEMBRANE TRYPTOPHAN-RICH SENSORY PROTEIN)-RELATED (TSPO)
↑	AT2G47890	B-box type zinc finger protein with CCT domain-containing protein;(source:Araport11)	
↓	AT3G02170	Encodes LONGIFOLIA2 (LNG2). Regulates leaf morphology by promoting cell expansion in the leaf-length direction. The LNG2 homologue LNG1 (At5g15580) has similar function.	LONGIFOLIA2 (LNG2)
↑	AT3G04010	O-Glycosyl hydrolases family 17 protein;(source:Araport11)	
↑	AT3G05800	AtBS1(activation-tagged BRI1 suppressor 1)-interacting factor 1;(source:Araport11)	ATBS1(ACTIVATION-TAGGED BRI1 SUPPRESSOR 1)-INTERACTING FACTOR 1 (AIF1)
↓	AT3G06070	hypothetical protein;(source:Araport11)	
↓	AT3G06130	Heavy metal transport/detoxification superfamily protein;(source:Araport11)	
↓	AT3G06140	Paralog of LOG2 (At3g09770), a ubiquitin ligase that regulates amino acid export.	LOG2-LIKE UBIQUITIN LIGASE4 (LUL4)
↑	AT3G09350	Encodes one of the Arabidopsis orthologs of the human Hsp70-binding protein 1 (HspBP-1) and yeast Fes1p: Fes1A (AT3G09350), Fes1B (AT3G53800), Fes1C (AT5G02150). Fes1A is cytosolic and associates with cytosolic Hsp70. Mutants showed increased heat-sensitive phenotype suggestion the involvement of Fes1A in acquired thermotolerance. Does not have nucleotide exchange factor activity in vitro.	FES1A (Fes1A)
↑	AT3G09540	Pectin lyase-like superfamily protein;(source:Araport11)	

↑	AT3G12320	hypothetical protein;(source:Araport11)	NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED 3 (LNK3)
↑	AT3G12490	Encodes a protein with cysteine proteinase inhibitor activity. Overexpression increases tolerance to abiotic stressors (i.e.salt,osmotic, cold stress).	CYSTATIN B (CYSB)
↑	AT3G12670	Cytidine triphosphate synthase.	EMBRYO DEFECTIVE 2742 (emb2742)
↓	AT3G12710	DNA glycosylase superfamily protein;(source:Araport11)	
↑	AT3G13310	Chaperone DnaJ-domain superfamily protein;(source:Araport11)	DNA J PROTEIN C66 (DJC66)
↓	AT3G13750	beta-galactosidase, glycosyl hydrolase family 35 The mRNA is cell-to-cell mobile.	BETA GALACTOSIDASE 1 (BGAL1)
↑	AT3G14890	Encodes a base excision repair protein that, together with APE2, it plays overlapping roles in the maintenance of epigenome and genome stability in plants.	ZINC 4 FINGER DNA 3'-PHOSPHOESTERASE (ZDP)
↓	AT3G15450	aluminum induced protein with YGL and LRDR motifs;(source:Araport11)	
↑	AT3G17130	Plant invertase/pectin methylesterase inhibitor superfamily protein;(source:Araport11)	
↓	AT3G18050	GPI-anchored protein;(source:Araport11)	
↓	AT3G19930	Encodes a sucrose hydrogen symporter that is induced by wounding. The mRNA is cell-to-cell mobile.	SUGAR TRANSPORTER 4 (STP4)
↑	AT3G21540	transducin family protein / WD-40 repeat family protein;(source:Araport11)	
↑	AT3G21560	Encodes a protein with sinapic acid:UDP-glucose glucosyltransferase activity. Mutants defective in this gene are hyper-fluorescent (which accumulate in their trichomes a compound that is likely to be 3',5'-dimethoxynaringenin chalcone or sinapoyltriacetic acid lactone, potential products of the concerted action of 4-coumarate CoA ligase and chalcone synthase on sinapic acid). Also shown to be required for Arabidopsis nonhost resistance to the Asian soybean rust pathogen Phakopsora pachyrhizi.	UDP-GLUCOSYL TRANSFERASE 84A2 (UGT84A2)
↑	AT3G21890	B-box type zinc finger family protein;(source:Araport11)	B-BOX DOMAIN PROTEIN 31 (BBX31)
↑	AT3G22840	Encodes an early light-inducible protein.	EARLY LIGHT-INDUCABLE PROTEIN (ELIP1)
↑	AT3G23000	Encodes a serine/threonine protein kinase with similarities to CBL-interacting protein kinases, SNF1 and SOS2. The mRNA is cell-to-cell mobile.	CBL-INTERACTING PROTEIN KINASE 7 (CIPK7)

↓	AT3G23730	xyloglucan endotransglucosylase/hydrolase 16;(source:Araport11)	XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 16 (XTH16)
D	AT3G26520	gamma tonoplast intrinsic protein 2 (TIP2). expressed throughout the plant and transcript level is increased upon NaCl or ABA treatments. NaCl stress-sensitive yeast mutant strains exhibit more resistance to salt when expressing this protein.	TONOPLAST INTRINSIC PROTEIN 2 (TIP2)
↑	AT3G27210	hypothetical protein;(source:Araport11)	
↑	AT3G27380	One of three isoforms of the iron-sulfur component of the succinate dehydrogenase complex, a component of the mitochondrial respiratory chain complex II. The product of the nuclear encoded gene is imported into the mitochondrion. Expressed during germination and post-germinative growth.	SUCCINATE DEHYDROGENASE 2-1 (SDH2-1)
↑	AT3G44450	Plant specific protein.BIC1 and BIC2 inhibit cryptochrome function by blocking blue light-dependent cryptochrome dimerization.Light activated transcription of BICs is mediated by cryptochromes.	BLUE-LIGHT INHIBITOR OF CRYPTOCHROMES 2 (BIC2)
↓	AT3G45850	P-loop containing nucleoside triphosphate hydrolases superfamily protein;(source:Araport11)	
↓	AT3G49790	Carbohydrate-binding protein;(source:Araport11)	
↑	AT3G50910	netrin receptor DCC;(source:Araport11)	
↑	AT3G53620	Encodes a soluble protein with inorganic pyrophosphatase activity that is highly specific for Mg-inorganic pyrophosphate. The mRNA is cell-to-cell mobile.	PYROPHOSPHORYLASE 4 (PPa4)
↑	AT3G55120	Catalyzes the conversion of chalcones into flavanones. Required for the accumulation of purple anthocyanins in leaves and stems. Co-expressed with CHS.	TRANSPARENT TESTA 5 (TT5)
↑	AT3G55580	TCF1 encodes a member of the RCC1 gene family and is required for chromatin based gene regulation of cold responsive genes in a CBF-independent manner. It is expressed in response to cold but not ABA.	TOLERANT TO CHILLING AND FREEZING1 (TCF1)
↓	AT3G58120	Encodes a member of the BZIP family of transcription factors. Forms heterodimers with the related protein AtbZIP34. Binds to G-boxes in vitro and is localized to the nucleus in onion epidermal cells.	(BZIP61)
↑	AT3G59820	LETM1-like protein;(source:Araport11)	LEUCINE ZIPPER-EF-HAND-CONTAINING TRANSMEMBRANE PROTEIN 1 (LETM1)

D	AT3G61430	a member of the plasma membrane intrinsic protein subfamily PIP1. localizes to the plasma membrane and exhibits water transport activity in Xenopus oocyte. expressed ubiquitously and protein level decreases slightly during leaf development. The mRNA is cell-to-cell mobile.	PLASMA MEMBRANE INTRINSIC PROTEIN 1A (PIP1A)
↑	AT3G61890	Encodes a homeodomain leucine zipper class I (HD-Zip I) protein. Loss of function mutant has abnormally shaped leaves and stems.	HOMEOBOX 12 (HB-12)
↑	AT3G62260	Protein phosphatase 2C family protein;(source:Araport11)	
↑	AT4G02880	ELKS/Rab6-interacting/CAST family protein;(source:Araport11)	
↑	AT4G11600	Encodes glutathione peroxidase.	GLUTATHIONE PEROXIDASE 6 (GPX6)
↑	AT4G14690	Encodes an early light-induced protein. ELIPs are thought not to be directly involved in the synthesis and assembly of specific photosynthetic complexes, but rather affect the biogenesis of all chlorophyll-binding complexes. A study (PMID 17553115) has shown that the chlorophyll synthesis pathway was downregulated as a result of constitutive ELIP2 expression, leading to decreased chlorophyll availability for the assembly of pigment-binding proteins for photosynthesis.	EARLY LIGHT-INDUCIBLE PROTEIN 2 (ELIP2)
↑	AT4G17090	Encodes a beta-amylase targeted to the chloroplast. Transgenic BMY8 RNAi lines fail to accumulate maltose during cold shock suggesting that maltose accumulation coincides with BMY8 expression. Apart from maltose, the sugar content of the RNAi lines were similar to wildtype (glucose and sucrose unaffected).BAM3 activity declines 2 and 4 days after start of cold stress despite an increase in transcript levels. BAM3 activity has a lower temperature optimum than BAM1 (PMID:25293962).	CHLOROPLAST BETA-AMYLASE (CT-BMY)
↓	AT4G17460	Encodes a class II HD-ZIP protein that regulates meristematic activity in different tissues, and that it is necessary for the correct formation of the gynoecium.	(HAT1)
↑	AT4G17550	Encodes a member of the phosphate starvation-induced glycerol-3-phosphate permease gene family: AT3G47420(G3Pp1), AT4G25220(G3Pp2), AT1G30560(G3Pp3), AT4G17550(G3Pp4) and AT2G13100(G3Pp5).	GLYCEROL-3-PHOSPHATE PERMEASE 4 (G3Pp4)
↑	AT4G17670	senescence-associated family protein (DUF581);(source:Araport11)	
↑	AT4G18070	suppressor;(source:Araport11)	

↓	AT4G19170	chloroplast-targeted member of a family of enzymes similar to nine-cis-epoxycarotenoid dioxygenase The mRNA is cell-to-cell mobile.	NINE-CIS-EPOXYCAROTENOID DIOXYGENASE 4 (NCED4)
↑	AT4G20880	ethylene-responsive nuclear protein / ethylene-regulated nuclear protein (ERT2);(source:Araport11)	
↑	AT4G23630	VIRB2-interacting protein 1;(source:Araport11)	VIRB2-INTERACTING PROTEIN 1 (BTI1)
↑	AT4G24960	Homologous to a eukaryote specific ABA- and stress-inducible gene first isolated from barley. Groups in one subfamily with ATHVA22E. Along with other members of the ATHVA22 family, it may be involved in regulation of autophagy during development. The mRNA is cell-to-cell mobile.	HVA22 HOMOLOGUE D (HVA22D)
↑	AT4G25470	Encodes a member of the DREB subfamily A-1 of ERF/AP2 transcription factor family (CBF2). The protein contains one AP2 domain. There are six members in this subfamily, including CBF1, CBF2, and CBF3. This gene is involved in response to low temperature, abscisic acid, and circadian rhythm. Overexpressing this gene leads to increased freeze tolerance and induces the expression level of 85 cold-induced genes and reduces the expression level of 8 cold-repressed genes, which constitute the CBF2 regulon. Mutations in CBF2 increases the expression level of CBF1 and CBF3, suggesting that this gene may be involved in a negative regulatory or feedback circuit of the CBF pathway.	C-REPEAT/DRE BINDING FACTOR 2 (CBF2)
↑	AT4G27520	early nodulin-like protein 2;(source:Araport11)	EARLY NODULIN-LIKE PROTEIN 2 (ENODL2)
↑	AT4G27940	manganese tracking factor for mitochondrial SOD2;(source:Araport11)	MANGANESE TRACKING FACTOR FOR MITOCHONDRIAL SOD2 (MTM1)
↓	AT4G28080	Encodes REDUCED CHLOROPLAST COVERAGE 2 (REC2). Along with REC1 and REC3 it contributes to establishing the size of the chloroplast compartment.	REDUCED CHLOROPLAST COVERAGE 2 (REC2)
↑	AT4G32800	encodes a member of the DREB subfamily A-4 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 17 members in this subfamily including TINY.	ETHYLENE RESPONSE FACTOR 43 (ERF043)
↑	AT4G33040	Encodes a member of the CC-type glutaredoxin (ROXY) family that has been shown to interact with the transcription factor TGA2.	(ROXY21)

↑	AT4G33905	Peroxisomal membrane 22 kDa (Mpv17/PMP22) family protein;(source:Araport11)	
↑	AT4G33980	hypothetical protein;(source:Araport11)	COLD-REGULATED GENE 28 (COR28)
↓	AT4G34220	Encodes a receptor like kinase involved in ABA-mediated seedling development and drought tolerance.RDK1 is an atypical or pseudokinase and has no phosphorylation activity. Its expression is upregulated in response to ABA.interacts with ABI1 and other PP2C phosphatases.	RECEPTOR DEAD KINASE1 (RDK1)
↓	AT4G34760	SAUR-like auxin-responsive protein family;(source:Araport11)	SMALL AUXIN UPREGULATED RNA 50 (SAUR50)
↑	AT4G36010	Pathogenesis-related thaumatin superfamily protein;(source:Araport11)	
↓	AT4G36540	Encodes the brassinosteroid signaling component BEE2 (BR-ENHANCED EXPRESSION 2). Positively modulates the shade avoidance syndrome in Arabidopsis seedlings.	BR ENHANCED EXPRESSION 2 (BEE2)
↓	AT4G38860	SAUR-like auxin-responsive protein family;(source:Araport11)	SMALL AUXIN UPREGULATED RNA 16 (SAUR16)
↑	AT4G38960	BBX19 is a B-box containing transcriptional regulator involved in photomporphogenesis and flowering.	B-BOX DOMAIN PROTEIN 19 (BBX19)
↑	AT4G39260	Encodes a glycine-rich protein with RNA binding domain at the N-terminus. Protein is structurally similar to proteins induced by stress in other plants. Gene expression is induced by cold. Transcript undergoes circadian oscillations that is depressed by overexpression of AtGRP7. A substrate of the type III effector HopU1 (mono-ADP-ribosyltransferase).	RNA-BINDING GLYCINE-RICH PROTEIN A6 (RBGA6)
↓	AT4G39950	Belongs to cytochrome P450 and is involved in tryptophan metabolism. Converts Trp to indo-3-acetaldoxime (IAOx), a precursor to IAA and indole glucosinolates. The mRNA is cell-to-cell mobile.	CYTOCHROME P450, FAMILY 79, SUBFAMILY B, POLYPEPTIDE 2 (CYP79B2)
↓	AT5G03230	senescence regulator (Protein of unknown function, DUF584);(source:Araport11)	
↑	AT5G04340	Encodes a C2H2 zinc finger transcription factor that coordinately activates phytochelatin-synthesis related gene expression and directly targets GSH1 by binding to its promoter to positively regulate Cd accumulation and tolerance.	ZINC FINGER OF ARABIDOPSIS THALIANA 6 (ZAT6)
↑	AT5G07920	Encodes a putative diacylglycerol kinase that is mainly expressed in roots, shoots and leaves, but its enzyme product was not active in vitro.	DIACYLGLYCEROL KINASE1 (DGK1)

↓	AT5G08520	Duplicated homeodomain-like superfamily protein;(source:Araport11)	(MYBS2)
↑	AT5G12010	nuclease;(source:Araport11)	
↑	AT5G14570	Encodes ATNRT2.7, a nitrate transporter that controls nitrate content in seeds. Expression is detected in reproductive organs and peaks in seeds. Localized to the vacuolar membrane.	HIGH AFFINITY NITRATE TRANSPORTER 2.7 (NRT2.7)
↓	AT5G16030	mental retardation GTPase activating protein;(source:Araport11)	
↑	AT5G16930	AAA-type ATPase family protein;(source:Araport11)	
↑	AT5G17220	Encodes glutathione transferase belonging to the phi class of GSTs. Naming convention according to Wagner et al. (2002). Mutants display no pigments on leaves and stems. Likely to function as a carrier to transport anthocyanin from the cytosol to tonoplasts.	GLUTATHIONE S-TRANSFERASE PHI 12 (GSTF12)
↓	AT5G19120	Eukaryotic aspartyl protease family protein;(source:Araport11)	
↑	AT5G22060	Co-chaperonin similar to E. coli DnaJ	DNAJ HOMOLOGUE 2 (J2)
↑	AT5G23850	O-glucosyltransferase rumi-like protein (DUF821);(source:Araport11)	
↑	AT5G24470	Encodes a pseudo-response regulator whose mutation affects various circadian-associated biological events such as flowering time in the long-day photoperiod conditions, red light sensitivity of seedlings during early photomorphogenesis, and the period of free-running rhythms of certain clock-controlled genes including CCA1 and APRR1/TOC1 in constant white light. Acts as transcriptional repressor of CCA1 and LHY. Acts additively with EC, PRR7 and PRR9 to regulate hypocotyl growth under photoperiodic conditions.	PSEUDO-RESPONSE REGULATOR 5 (PRR5)
↓	AT5G25190	encodes a member of the ERF (ethylene response factor) subfamily B-6 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 12 members in this subfamily including RAP2.11.	ETHYLENE AND SALT INDUCIBLE 3 (ESE3)
↓	AT5G27360	Encodes a sugar-porter family protein that unlike the closely related gene, SFP1, is not induced during leaf senescence.	(SFP2)
↑	AT5G27930	EGR2 functions as a negative regulator of plant growth with prominent effect on plant growth during drought stress. EGR2 regulates microtubule organization and likely affects additional cytoskeleton and trafficking processes along the plasma membrane.	E GROWTH-REGULATING 2 (EGR2)
↑	AT5G41460	transferring glycosyl group transferase (DUF604);(source:Araport11)	

↑	AT5G42900	cold regulated protein 27;(source:Araport11)	COLD REGULATED GENE 27 (COR27)
↓	AT5G44410	FAD-binding Berberine family protein;(source:Araport11)	(ATBBE27)
↑	AT5G44670	glycosyltransferase family protein (DUF23);(source:Araport11)	GALACTAN SYNTHASE 2 (GALS2)
↓	AT5G44680	DNA glycosylase superfamily protein;(source:Araport11)	
D	AT5G48540	receptor-like protein kinase-related family protein;(source:Araport11)	
↑	AT5G49330	Member of the R2R3 factor gene family.	MYB DOMAIN PROTEIN 111 (MYB111)
↓	AT5G49360	Encodes a bifunctional {beta}-D-xylosidase/{alpha}-L-arabinofuranosidase required for pectic arabinan modification. Located in the extracellular matrix. Gene is expressed specifically in tissues undergoing secondary wall thickening. This is a member of glycosyl hydrolase family 3 and has six other closely related members.	BETA-XYLOSIDASE 1 (BXL1)
↓	AT5G49450	Encodes a transcription activator is a positive regulator of plant tolerance to salt, osmotic and drought stresses.	BASIC LEUCINE-ZIPPER 1 (bZIP1)
↑	AT5G49480	AtCP1 encodes a novel Ca2+-binding protein, which shares sequence similarities with calmodulins. The expression of AtCP1 is induced by NaCl. The mRNA is cell-to-cell mobile.	CA2+-BINDING PROTEIN 1 (CP1)
↓	AT5G49740	Encodes a chloroplast ferric chelate reductase. Shows differential splicing and has three different mRNA products. Expressed in the shoot, flower and cotyledon.	FERRIC REDUCTION OXIDASE 7 (FRO7)
↑	AT5G50360	ABA‐induced transcription repressor that acts as feedback regulator in ABA signalling.	ABA-INDUCED TRANSCRIPTION REPRESSOR 5 (AITR5)
↑	AT5G51570	SPFH/Band 7/PHB domain-containing membrane-associated protein family;(source:Araport11)	HYPERSENSITIVE INDUCED REACTION 3 (HIR3)
↑	AT5G52300	Encodes a protein that is induced in expression in response to water deprivation such as cold, high-salt, and desiccation. The response appears to be via abscisic acid. The promoter region contains two ABA-responsive elements (ABREs) that are required for the dehydration-responsive expression of rd29B as cis-acting elements. Protein is a member of a gene family with other members found plants, animals and fungi.	LOW-TEMPERATURE-INDUCED 65 (LTI65)

↑	AT5G52310	cold regulated gene, the 5' region of cor78 has cis-acting regulatory elements that can impart cold-regulated gene expression The mRNA is cell-to-cell mobile.	LOW-TEMPERATURE-INDUCED 78 (LTI78)
↑	AT5G53420	Member of ASML2 family of CCT domain proteins.There is a preferential accumulation of RNA isoforms CCT101.1 and CCT101.2 in response to N-treatment, each isoform has different targets.	CONSTANS, CO-LIKE AND TOC1 MOTIF; (CCT101)
↑	AT5G53970	Encodes a cytosolic tyrosine aminotransferase which is strongly induced upon aging and coronatine treatment. AtTAT1 prefers Tyr as an amino donor but can also use Phe, Trp, His, Met, and Leu. The mRNA is cell-to-cell mobile.	TYROSINE AMINOTRANSFERASE 7 (TAT7)
↑	AT5G54470	B-box type zinc finger family protein;(source:Araport11)	B-BOX DOMAIN PROTEIN 29 (BBX29)
↑	AT5G55180	O-Glycosyl hydrolases family 17 protein;(source:Araport11)	
↓	AT5G56850	hypothetical protein;(source:Araport11)	
↓	AT5G57780	Encodes a atypical member of the bHLH (basic helix-loop-helix) family transcriptional factors.	P1R1 (P1R1)
↑	AT5G59820	Encodes a zinc finger protein involved in high light and cold acclimation. Overexpression of this putative transcription factor increases the expression level of 9 cold-responsive genes and represses the expression level of 15 cold-responsive genes, including CBF genes. Also, lines overexpressing this gene exhibits a small but reproducible increase in freeze tolerance. Because of the repression of the CBF genes by the overexpression of this gene, the authors speculate that this gene may be involved in negative regulatory circuit of the CBF pathway. The mRNA is cell-to-cell mobile.	RESPONSIVE TO HIGH LIGHT 41 (RHL41)
↑	AT5G60790	member of GCN subfamily	ATP-BINDING CASSETTE F1 (ABCF1)
↑	AT5G61030	Encodes a glycine-rich RNA binding protein that is involved in C-> U RNA editing in mitochondria. Gene expression is induced by cold. The mRNA is cell-to-cell mobile.	RNA-BINDING GLYCINE-RICH PROTEIN A7 (RBGA7)
↓	AT5G61440	Encodes a member of the thioredoxin family protein. Located in the chloroplast. The mRNA is cell-to-cell mobile.	ATYPICAL CYS HIS RICH THIOREDOXIN 5 (ACHT5)
↑	AT5G62360	Pectin methylesterase inhibitor expressed throughout the plant.	PECTIN METHYL-ESTERASE INHIBITOR 13 (PMEI13)
↑	AT5G63130	Octicosapeptide/Phox/Bem1p family protein;(source:Araport11)	

↑	AT5G66760	One of two genes in Arabidopsis that encode a flavoprotein subunit of the mitochondrial succinate dehydrogenase complex. The mRNA is cell-to-cell mobile.	SUCCINATE DEHYDROGENASE 1-1 (SDH1-1)
↑	AT5G66780	late embryogenesis abundant protein;(source:Araport11)	
↓	The query for the following terms resulted in no hits: AT2G33050		

Table S23. Shared significantly enriched GO-terms (fisher test in combination with the elim algorithm, $p < 0.05$) within the Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) domains in Arctic species. The 24h DEG sets were also compared with those of *Arabidopsis thaliana* from Park et al. (2015), and **bold** indicates that the GO-term only was shared by Arctic species.

A. Shared GO-terms in upregulated DEG sets

Upregulated DEGs, 3h:
GO:0080167 response to karrikin (BP)
GO:0009873 ethylene-activated signaling pathway (BP)
GO:0007623 circadian rhythm (BP)
GO:0009737 response to abscisic acid (BP)
GO:0006355 regulation of transcription, DNA-templated (BP)
GO:0009409 response to cold (BP)
GO:0009719 response to endogenous stimulus (BP)
GO:0010200 response to chitin (BP)
GO:0009631 cold acclimation (BP)
GO:0009414 response to water deprivation (BP)
GO:0009651 response to salt stress (BP)
GO:0031225 anchored component of membrane (CC)
GO:0080054 low-affinity nitrate transmembrane transporter activity (MF)
GO:0003677 DNA binding (MF)
GO:0003700 DNA-binding transcription factor activity (MF)
Upregulated DEGs, 6h:
GO:0080167 response to karrikin (BP)
GO:0010501 RNA secondary structure unwinding (BP)
GO:0009873 ethylene-activated signaling pathway (BP)
GO:0007623 circadian rhythm (BP)
GO:0009737 response to abscisic acid (BP)
GO:0006355 regulation of transcription, DNA-templated (BP)
GO:0009409 response to cold (BP)
GO:0010378 temperature compensation of the circadian clock (BP)
GO:0010200 response to chitin (BP)
GO:0009631 cold acclimation (BP)
GO:0009753 response to jasmonic acid (BP)
GO:0010017 red or far-red light signaling pathway (BP)
GO:0009414 response to water deprivation (BP)
GO:0048579 negative regulation of long-day photoperiodism, flowering (BP)
GO:0009651 response to salt stress (BP)
GO:0009269 response to desiccation (BP)
GO:0005634 nucleus (CC)
GO:0080054 low-affinity nitrate transmembrane transporter activity (MF)
GO:0003677 DNA binding (MF)
GO:0005509 calcium ion binding (MF)
GO:0003700 DNA-binding transcription factor activity (MF)

Upregulated DEGs, 24h:

GO:0009643 photosynthetic acclimation (BP)
GO:0080167 response to karrikin (BP)
GO:0010501 RNA secondary structure unwinding (BP)
GO:0009873 ethylene-activated signaling pathway (BP)
GO:0009636 response to toxic substance (BP)
GO:0007623 circadian rhythm (BP)
GO:0009737 response to abscisic acid (BP)
GO:0006355 regulation of transcription, DNA-templated (BP)
GO:0009409 response to cold (BP)
GO:0009408 response to heat (BP)
GO:0006121 mitochondrial electron transport, succinate to ubiquinone (BP)
GO:0009631 cold acclimation (BP)
GO:0009813 flavonoid biosynthetic process (BP)
GO:0008295 spermidine biosynthetic process (BP)
GO:0009744 response to sucrose (BP)
GO:0006979 response to oxidative stress (BP)
GO:0046686 response to cadmium ion (BP)
GO:0009414 response to water deprivation (BP)
GO:0009651 response to salt stress (BP)
GO:0051555 flavonol biosynthetic process (BP)
GO:0010286 heat acclimation (BP)
GO:0010224 response to UV-B (BP)
GO:0003677 DNA binding (MF)
GO:0008378 galactosyltransferase activity (MF)
GO:0003700 DNA-binding transcription factor activity (MF)
GO:0015315 organophosphate:inorganic phosphate antiporter activity (MF)
GO:0008792 arginine decarboxylase activity (MF)

B. Shared GO-terms in downregulated DEG sets

Downregulated DEGs, 3h:

None

Downregulated DEGs, 6h:

GO:0009416 response to light stimulus (BP)

Downregulated DEGs, 24h:

GO:0080127 fruit septum development (BP)
GO:0010112 regulation of systemic acquired resistance (BP)
GO:0009734 auxin-activated signaling pathway (BP)
GO:0045490 pectin catabolic process (BP)
GO:0009638 phototropism (BP)
GO:0009828 plant-type cell wall loosening (BP)
GO:0010114 response to red light (BP)
GO:0016021 integral component of membrane (CC)
GO:0005576 extracellular region (CC)
GO:0046658 anchored component of plasma membrane (CC)
GO:0009505 plant-type cell wall (CC)

III

The genome of *Draba nivalis* shows signatures of adaptation to the extreme environmental stresses of the Arctic

Michael D. Nowak[1]*, Siri Birkeland[1], Terezie Mandáková[2], Rimjhim Roy Choudhury[3], Xinyi Guo[2], A. Lovisa S. Gustafsson[1], Abel Gizaw[1], Audun Schrøder-Nielsen[1], Marco Fracassetti[4], Anne K. Brysting[5], Loren Rieseberg[6], Tanja Slotte[4], Christian Parisod[3], Martin A. Lysak[2], Christian Brochmann[1]

Running Title: The genome of the Arctic plant *Draba nivalis*

[1]Natural History Museum, University of Oslo, Norway
[2]CEITEC, Masaryk University, Brno, Czech Republic
[3]Institute of Plant Sciences, University of Bern, Switzerland
[4]Science for Life Laboratory and Department of Ecology, Environment and Plant Science, Stockholm University, Sweden
[5]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Norway
[6]Department of Botany, The University of British Columbia, Vancouver, Canada.

*author for correspondence: michaeldnowak@gmail.com

Abstract

The Arctic is one of the most extreme terrestrial environments on the planet. Here we present the first chromosome-scale genome assembly of a plant adapted to the high Arctic, *Draba nivalis* (Brassicaceae), an attractive model species for studying plant adaptation to the stresses imposed by this harsh environment. We used an iterative scaffolding strategy with data from short-reads, single-molecule long reads, proximity ligation data, and a genetic map to produce a 302 Mb assembly that is highly contiguous with 91.6% assembled into eight chromosomes (the base chromosome number). To identify candidate genes and gene families that may have facilitated adaptation to Arctic environmental stresses, we performed comparative genomic analyses with nine non-Arctic Brassicaceae species. We show that the *D. nivalis* genome contains expanded suites of genes associated with drought and cold stress (e.g. related to the maintenance of oxidation-reduction homeostasis, meiosis, and signaling pathways). The expansions of gene families associated with these functions appear to be driven in part by the activity of transposable elements. Tests of positive selection identify suites of candidate genes associated with meiosis and photoperiodism, as well as cold, drought, and oxidative stress responses. Our results reveal a multifaceted landscape of stress adaptation in the *D. nivalis* genome, offering avenues for the continued development of this species as an Arctic model plant.

Keywords: adaptation, Arctic, Brassicaceae, chromosome-scale assembly, linkage map

1. Introduction

The Arctic accounts for ~10% of Earth's land surface, and the combination of high latitude and regional climate patterns make it one of the harshest terrestrial environments on the planet. Arctic plants must endure an extremely short and unpredictable growing season, with mean July temperatures ≤ 10 °C and up to 24 hours of sunlight. The climatic conditions impose strong selective pressures for managing cellular and physiological responses to environmental stresses, including cold, drought (Lütz, 2010), and high light exposure (Caldwell et al., 2007). Little is known about the molecular basis of Arctic plant adaptations, and an Arctic model species has yet to be developed to facilitate such studies (Wullschleger, et al., 2015; Colella et al., 2020). Exploring the molecular basis of Arctic plant adaptation may reveal novel mechanisms of environmental stress tolerance, potentially offering guidance to agricultural crop improvement.

Here we present the chromosome-scale genome assembly of *Draba nivalis*, a perennial diploid with a circum-Arctic distribution. This species is ideal for studying the evolution of environmental stress tolerance in plants because it occurs in the high Arctic where extremes in temperature, light regime, and low water availability are ever-present. In a recent study of its transcriptome, we identified numerous candidate genes for Arctic adaptation illuminating its potential as an Arctic model species (Birkeland et al. 2020). *Draba nivalis* is also an emerging model for studying the genetics of incipient speciation as intraspecific crosses frequently result in highly sterile hybrids (Grundt et al. 2006; Skrede et al. 2008; Gustafsson et al. 2014). *Draba* is the largest genus in the Brassicaceae with >390 species, which mainly occur in Arctic and alpine regions (Jordon-Thaden et al. 2013). Many of these species, including *D. nivalis*, form dense and hairy cushions as protection from wind and cold, and are strongly autogamous to secure reproduction in their pollinator-poor environments (Brochmann 1993). The Brassicaceae contains numerous species important for agriculture as well as for research in plant ecology, evolution, development, and molecular biology (Gupta 2016). The availability of many Brassicaceae genome assemblies enabled us to conduct comparative analyses of chromosomal evolution and functional genomics, to shed light on the genomic characteristics of a plant adapted to the extreme abiotic stresses of the Arctic.

2. Materials and Methods

2.1 Plant material and DNA sequencing

Seeds of *D. nivalis* accession 008-7 from Alaska (Waterfall Creek W., 63.045 latitude, -147.201 longitude; (see Grundt et al., 2006 for complete locality information) were grown under conditions mimicking the Arctic climate (specified in Brochmann et al., 1992) in the Phytotron at the Department of Biosciences, University of Oslo. Genomic DNA was extracted from young leaf tissue using the Qiagen Plant Mini Kit, and 7.7 µg of DNA was delivered to the Norwegian Sequencing Centre for library preparation and sequencing. Paired-end sequencing libraries with a target insert size of 550 bp were produced using the Illumina TruSeq PCR-free kit. This library was sequenced on one half of each of two Illumina HiSeq 2500 flow cells in rapid run mode with a read length of 250 bp. These data were used to estimate the genome size based on 25mer frequency using GenomeScope (Vurture *et al.* 2017; http://qb.cshl.edu/genomescope/). A total of 8 µg of genomic DNA was used to produce four sequencing libraries for the Oxford Nanopore MinION platform. For each library, 2 µg of genomic DNA was treated with NEBNext Ultra II End-Repair / dA-tailing Module and libraries were prepared using the Oxford Nanopore 1D ligation kit (SQK-LSK108) following the manufacturer's protocol. Each of these four libraries was sequenced on a single Oxford Nanopore MinION flow cell (version R9) for approximately 48 hours. Two whole plants of *D. nivalis* accession 008-7 were flash frozen in liquid nitrogen and shipped to Dovetail Genomics (LLC, Santa Cruz, California 95060, USA) on dry ice for Chicago proximity ligation library preparation and 150 bp paired-end sequencing on the Illumina HiSeq 2500 platform.

2.2 Genome assembly and scaffolding

The first draft assembly of *D. nivalis* accession 008-7 was produced using all (approximately 270 million 250 bp paired-end reads) unfiltered Illumina HiSeq data with the software DISCOVAR *de novo* (release 52488; https://software.broadinstitute.org/software/discovar/blog/) using default parameter settings. This draft assembly (scaffold N50 = 30.083 Kb) was supplied to Dovetail Genomics for scaffolding with the HiRise pipeline using 150 bp Chicago reconstituted chromatin paired-end reads. The resulting scaffolded draft assembly (scaffold N50 = 2.92 Mb) was further improved by scaffolding with approximately

796 Mb of Oxford Nanopore 1D long reads passing the default quality filtration score in the Metrichor base calling pipeline. Long read scaffolding was first conducted using SSPACE-LongRead v1.1 (Boetzer et al., 2014) using a minimum alignment identity of 90. These scaffolds were further improved with the Oxford Nanopore long reads using LINKS v1.8.6 (Warren et al., 2015) with K-mer size set to 21 and using 17 different distances between K-mer pairs (i.e. -d 1 Kb, 2 Kb, 4 Kb, 6 Kb, 7 Kb, 8 Kb, 10 Kb, 12 Kb, 14 Kb, 16 Kb, 18 Kb, 21 Kb, 25 Kb, 30 Kb, 40 Kb, 50 Kb, 60 Kb).

To compare broad patterns of synteny, the *D. nivalis* genome was aligned to the genomes of *A. alpina* (the most closely related species with an assembled genome; Guo et al. 2017) and *A. lyrata* with NUCmer v4.0b2 (Kurtz et al. 2004) using all anchor matches regardless of their uniqueness (--maxmatch) and setting the minimum length of a cluster of matches (-c) to 100.

2.3 Genetic map construction and final map-based scaffolding

We generated an F_2 mapping population by self-pollination of an F_1 hybrid obtained from a cross between a *D. nivalis* plant from Norway (045-5; maternal parent; see Grundt et al., 2006 for complete locality information) and a *D. nivalis* plant from Alaska (008-7; paternal parent). Seeds were gently scarified before sowing and a total of 575 F_2 individuals were grown to maturity under our phytotron conditions (see above). Genomic DNA was extracted from young leaf tissue using the Qiagen Plant Mini Kit, and 96-plex double-digest restriction-associated DNA (ddRAD) libraries were produced. For the ddRAD procedure, the restriction enzymes *NsiI* and *MseI* were used to digest 500 ng of genomic DNA per sample. Indexed P1 and P2 adapters with sticky ends matching the overhangs left by the restriction enzymes were added to the digested DNA. Following adapter ligation, individual indexed libraries were pooled and amplified with an eight cycle PCR. Ampure XP bead cleanup was performed to remove short fragments (i.e. less than ~200 bp), and the multiplexed libraries were visualized on an Advanced Analytical Fragment Analyzer to ensure the libraries were of the correct size (i.e. 300-450 bp). See Supplementary Methods for protocol details. The final multiplexed libraries were sequenced on six Illumina HiSeq 2500 lanes by the Norwegian Sequencing Centre.

Reads were demultiplexed using ipyrad v.0.5.15 (Eaton 2014), and adapters and low-quality reads were removed using Cutadapt and FastQC available in the wrapper script Trim Galore! v0.4.5

(http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; bases with a Phred score less than 20 were trimmed, and reads shorter than 35 bp following trimming were discarded). A total of six F_2 individuals were removed because of low quality reads. Trimmed paired-end sequence reads were each mapped to the Dniv0087_Chicago assembly using BWA-MEM v0.7.8 (Li, Durbin 2010; Li et al., 2013) with default settings, and duplicate reads were filtered using MarkDuplicates in Picard v2.0.1 (http://broadinstitute.github.io/picard/). BAM alignment processing and SNP calling were performed with the Genome Analysis Toolkit v4.beta6 (GATK; McKenna et al., 2010). Briefly, GATK RealignerTargetCreator and IndelRealigner were first used to realign indels, and base quality scores were recalibrated using GATK BaseRecalibrator and PrintReads with default settings. Indels and SNPs were called using GATK UnifiedGenotyper in DISCOVERY mode using default parameters. Indels were discarded with VCFtools v0.1.13 (Danecek et al., 2011) resulting in a VCF file with a total of 166,644 SNPs prior to filtration. VCFtools v0.1.13 was used to isolate biallelic SNPs and exclude any SNPs that mapped to regions of the Dniv0087_Chicago genome assembly annotated as repetitive elements (see below for repetitive element annotation). VCFtools was further used to filter the biallelic SNPs for sites with a minimum and maximum coverage depth of 8 and 200, respectively, sites with a minimum mapping quality of 50, sites with a minor allele frequency greater than 0.0001, and sites that were called in at least 95% of the samples. These filtration steps resulted in a final VCF file containing 13,990 SNPs genotyped in 537 F_2 individuals.

Using these data, we constructed a genetic map using R/qtl v1.42-8 (Broman et al., 2003; Arends et al., 2010) and ASMap v1.0-4 (Taylor et al., 2017). Individuals with more than 5% missing genotypes and those that represented duplicate genotypes were removed. Loci that represented redundant genotypes were removed and an initial genetic map was estimated. Based on this map, 2,086 markers exhibiting significant segregation distortion (p-value < 1e-10) were removed and map distances were recalculated. The cross upon which this map is based is expected to contain biological sources of segregation distortion as hybrid progeny of these genotypes have previously been shown to exhibit both seed and pollen infertility (Grundt et al., 2006; Skrede et al., 2008; Gustafsson et al., 2014). To ensure that we had a genetic map as complete as possible upon which to scaffold the genome assembly, we reintegrated distorted markers into the genetic map using ASMap. The resulting genetic map and

the first map that contained the distorted markers were imported into ASMap and markers exhibiting segregation distortion at $-\log_{10} p$-value < 6 were then pushed back into the map based on the marker order reflected by the initial map, and map distances were estimated again without changing the marker ordering. The final genetic map contained 5055 markers genotyped in 480 F_2 individuals (Supplementary Figure 1, Supplementary Table 1).

Chromonomer v1.08 (http://catchenlab.life.illinois.edu/chromonomer/) was used with default settings to scaffold the assembly based on the genetic map. To create the input files necessary to run Chromonomer, the ddRAD loci containing SNPs constituting the genetic map were first aligned to the scaffolded *D. nivalis* accession 008-7 draft genome assembly using BWA-MEM (Li, Durbin 2010; Li et al., 2013), and an AGP file was then generated for the scaffolded genome assembly using the script fatoagp.pl (https://github.com/sjackman/fastascripts/blob/master/fatoagp).

2.4 Identification and annotation of transposable elements

LTR-RTs were annotated following Choudhury et al., (2017) by identifying full-length LTR-RT copies based on structural features using LTRHARVEST 1.5 (Ellinghaus et al., 2008). After removal of nested as well as overlapping elements, candidate copies with internal regions matching plant non-LTR retrotransposon or DNA transposon consensus sequences from Repbase (www.girinst.org/repbase/; accessed in June 2018; Ellinghaus et al., 2008) were excluded. Internal coding regions and binding sites of remaining candidate full-length copies were annotated and classified into *Gypsy* and *Copia* superfamilies using LTRDIGEST (Steinbiss et al., 2009), using Hidden Markov models based on plant LTR-RT protein (Gag, Reverse transcriptase, Protease, RNaseH, Integrase, Chromodomain, and Envelope; downloaded from www.gydb.org; Llorens et al., 2011) and eukaryotic tRNA entries from the UCSC gtRNA database. Terminal inverted repeat transposons (TIR) were identified based on the presence of TIR sequences flanked by target site duplications, using GenomeTools tirvish (Gremme et al., 2013). After removal of sequences nested with other known transposable elements, putative full-length copies of TIR transposons were confirmed by inspecting internal coding domains of Class II transposons using Hidden Markov models. Helitron sequences were identified with HelitronScanner (Xiong et al., 2014) which searches for upstream and downstream termini of

Helitrons within 200–20,000 bp of each other, in either the direct and reverse complement orientation following Dunning et al. (2019).

Copies of transposable elements identified in *D. nivalis* were classified into families by clustering sequences with at least 80% similarity (i.e. technical definition; Wicker et al., 2009; El-Baidouri, Panaud 2013) using CD-HIT-EST (Li, Godzik 2006). For LTR-RTs, clustering was based on LTR sequences. The full-length copy showing the lowest E-value to an HMM profile in each cluster was selected for the classification of family based on longest significant blastn hits of reverse-transcriptase domains flanked by 800 bp on either side against corresponding reverse-transcriptase sequences from Brassicaceae. Classified Brassicaceae families were further assigned to "tribes" following Choudhury et al., (2017). For TIR transposons, clustering was based on full-length copies for TIR transposons that were further classified into superfamilies (Harbinger, hAT, Mariner/Tc1, MuDR, EnSpm/CACTA) based on longest significant blastn hits against Viridiplantae DNA transposons extracted from Repbase. Helitrons show notoriously diverse internal regions and were classified following Yang, Bennetzen (2009), with clustering based on identity over 30 bp at the 3' end of copies (i.e. the hairpin-forming region crucial for rolling circle replication).

Transposable elements were annotated along the genome assembly using all structurally defined and hierarchically classified copies of LTR-RTs, TIR transposons and helitrons from *D. nivalis* together with non-LTR-RTs (i.e. LINE and SINE from Viridiplantae in Repbase) as a reference. After removal of sequences giving significant blast hit with swissprot protein database for plants and with other transposable element sequences, this reference was used in RepeatMasker (version Open-4; http://www.repeatmasker.org) with RM-BLAST as search engine and divergence set to 20%. Resulting annotations of remnants of transposable element sequences were filtered to remove nested copies and copies with less than 80 bp.

2.5 Transcriptome assembly and gene annotation

Four tissues were sampled from *D. nivalis* accession 008-7 for RNA extraction: young leaves (approximately 2-7 days following leaf blade expansion), mature floral buds (approximately 2-4 days prior to anthesis), open flowers (approximately 1-3 days post-anthesis), and root tissues (thoroughly washed of soil). Total RNA was isolated using the Thermo Fisher RNAqueous-Micro Kit following the manufacturers

standard protocol for fresh tissues. A total of 30.6 μg (leaves), 22 μg (flower buds), 23 μg (open flowers), and 3 μg (roots) of total RNA was provided to the Norwegian Sequencing Centre for Illumina TruSeq Stranded RNA library preparation and sequencing. Each library was sequenced on 1/10th of an Illumina HiSeq 2500 lane to generate 125 bp paired-end reads. A *de novo* transcriptome assembly was generated for *D. nivalis* accession 008-7 using the RNA-seq data from all four tissues with the Trinity v2.4.0 pipeline (Grabherr et al., 2011; Haas et al., 2013) including default read quality trimming and filtration using Trimmomatic v0.32 (Bolger et al., 2014).

Genes were predicted in the *D. nivalis* genome assembly using the Maker v2.31.9 (Holt, Yandell 2011) pipeline. Augustus v3.2.2 (Stanke et al., 2008) and SNAP (Release 2013-11-29; Korf 2004) were used as *ab initio* gene predictors. The Trinity transcriptome assembly (see above) was used as transcript evidence, and protein sequences from the species *Arabis nemorensis, Eutrema salsugineum, Arabis alpina, Arabidopsis lyrata,* and *Arabidopsis thaliana* (version TAIR10) were used as homology-based evidence. The *D. nivalis* repeat library (see above) was included to mask repetitive elements from annotation. The Maker annotation was first run using the *D. nivalis* transcriptome directly to infer gene predictions, and training files for the *ab initio* gene predictors were produced with these results. Maker was run iteratively three additional times using the transcriptome as evidence and providing updated training files for each run. The resulting set of predicted genes was annotated with Pfam domains (El-Gebali et al., 2018) using InterProScan v5.4-47.0 (Jones et al., 2014), and GO terms were annotated using Blast2GO v5.2.5 (Conesa et al., 2005) by searching against the UniProt (https://www.uniprot.org/) database for Viridiplantae. Both the *D. nivalis* genome assembly and the predicted gene set were also evaluated for completeness by searching against a set of 1440 highly conserved plant genes (Embryophyta) using BUSCO v3.0.1 (Simao et al., 2015). The genome assembly and predicted gene set were assessed for completeness by running BUSCO in both 'genome' and 'prot' modes, respectively.

2.6 Comparative chromosome painting

Whole inflorescences of *D. nivalis* were fixed in freshly prepared ethanol:acetic acid fixative (3:1) overnight, transferred into 70% ethanol and stored at -20°C until use. Mitotic and meiotic (pachytene and diakinesis) chromosome preparations were prepared as described by Mandáková, Lysak (2016a) on suitable slides pretreated

with RNase (100 μg/ml, AppliChem) and pepsin (0.1 mg/ml, Sigma-Aldrich). Based on the known chromosome structure of *A. alpina* and other Arabideae species (Willing et al. 2015; Mandáková et al. 2020), representative BAC clones of *A. thaliana* were selected and grouped into contigs for comparative chromosome painting (CCP). Five to ten BAC clones from each tested genomic region of *A. alpina* were used as hybridization probes on mitotic chromosome spreads in *D. nivalis* (Supplementary Figure 2 Dniv1: genomic regions A and B; Dniv2: D and E; Dniv3: Fa and Fb; Dniv4: C, T and Jb; Dniv5: K-L, M-Na and M-Nb; Dniv6: O, V and S; Dniv7: Ua and Ub; Dniv8: R, W and X). The four most reshuffled chromosomes of *A. alpina* and other Arabideae species (Willing et al. 2015; Mandáková et al. 2020) were investigated in detail by hybridization of whole-chromosome paints (i.e. BAC contigs covering whole chromosomes except pericentromere) on pachytene chromosomes of *D. nivalis* (Figure 3; Dniv4, Dniv5, Dniv6 and Dniv7). The *A. thaliana* BAC clone T15P10 (AF167571) bearing 35S rRNA gene repeats was used for *in situ* localization of nucleolar organizer regions, and the clone pCT4.2 (M65137), corresponding to a 500-bp 5S rDNA repeat, was used to localize 5S rDNA loci (Supplementary Figure 3). All BACs and rDNA probes were labeled with biotin-dUTP, digoxigenin-dUTP, or Cy3-dUTP by nick translation as described by Mandáková, Lysak (2016b). The labeled BACs were pooled together, ethanol precipitated, dissolved in 20 μl of hybridization mixture (50% formamide and 10% dextran sulfate in 2× SSC) per slide and pipetted to a microscopic slide containing chromosome spreads. The slide was heated at 80°C for 2 min and incubated in a moist chamber at 37°C overnight. Hybridized probes were visualized either as the direct fluorescence of Cy3-dUTP or through fluorescently labeled antibodies against biotin-dUTP and digoxigenin-dUTP following Mandáková, Lysak (2016b). Chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI, 2 μg/ml) in Vectashield antifade. Fluorescent signals were analyzed and photographed using a Zeiss Axioimager epifluorescence microscope and a CoolCube camera (MetaSystems). Individual images were merged and processed using Photoshop CS6 software (Adobe Systems).

2.7 Analyses of Gene Family Evolution

To compare the *D. nivalis* genome assembly with other Brassicaceae species whose genomes have been sequenced, whole genome assemblies and associated gene annotations were downloaded from public databases (Supplementary Table 2) for the

following nine species (representing three Brassicaceae clades, Guo et al. 2017): *Arabis alpina* (clade B), *Arabidopsis lyrata* (clade A), *Arabidopsis thaliana* (clade A), *Capsella rubella* (clade A), *Raphanus raphanistrum* (clade B), *Brassica oleracea* (clade B), *Thellungiella parvula* (clade B), *Thlaspi arvense* (clade B), and *Aethionema arabicum* (clade F). Among these species, *A. alpina* is thought to be the most closely related to *D. nivalis* (clade B) with an assembled genome (Guo et al. 2017). As a first analysis of gene content, we annotated Pfam domains (El-Gebali et al., 2018) for the predicted genes of each assembly using InterProScan (Jones et al., 2014). Pfam domains were quantified for each species, and domains with a Z-score above 1.96 or below -1.96 in *D. nivalis* were considered significantly enriched or contracted, respectively.

To estimate gene family composition and membership Orthofinder v2.2.7 (Emms, Kelly 2015; Emms, Kelly 2019) was run using the proteins annotated in the *D. nivalis* genome and the nine Brassicaceae genomes. OrthoFinder was run with default settings using MMseqs2 (Steinegger, Söding 2017) to cluster proteins by sequence similarity. Tests for significant contractions and expansions of gene families (defined as 'orthogroups' by OrthoFinder) were performed with CAFE v4.2 (Han et al., 2013). The species tree used for the CAFE analysis was generated by OrthoFinder using STAG (Emms, Kelly 2018) and rooted using STRIDE (Emms, Kelly 2017). This species tree was transformed into an ultrametric tree using r8s (Sanderson 2003) by fixing the age of the most recent common ancestor of *A. arabaicum* and the remaining nine Brassicaceae species to 35.2 Ma based on the divergence times reported by Guo et al., (2017). We classified gene duplications within the *D. nivalis* genome using the Dup_GenFinder pipeline (Qiao et al., 2019). This pipeline used the results of an all-versus-all BLASTp of the *D. nivalis* gene set to itself, and BLASTp results comparing the *D. nivalis* gene set to a closely related species (here the *A. alpina* v.4 genome assembly was used; Willing et al., 2015) to identify homologous gene pairs. MCSCANX (Wang et al., 2012) was used to identify patterns of synteny and collinearity in the duplicated genes, both within the *D. nivalis* genome and between *D. nivalis* and *A. alpina*. Dup_GenFinder synthesizes the outputs of these analyses to classify gene duplications into one of five categories using methods that are described in detail in Qiao et al., (2019) and Wang et al., (2012). Tandem duplications (TD) are gene pairs that are located next to one another on the same *D. nivalis* chromosome, likely resulting from unequal crossing over. Proximal

duplications (PD) represent gene pairs that are located on the same *D. nivalis* chromosome and separated from one another by ten or fewer genes, likely resulting from local transposon activity or ancient TD events. Transposed duplications (TRD) are gene pairs comprised of ancestral and novel copies, which are defined based on intra- and inter-species collinearity (Qiao et al., 2019), and are likely the product of distant RNA or DNA transposon activity. Whole genome duplication (WGD) is inferred for genes that reside in relatively large collinear chromosomal regions (collinear blocks) shared by *D. nivalis* and *A. alpina* (also called segmental duplicates by Wang et al., 2012). Dispersed duplications (DSD) are gene pairs that do not fulfill the criteria for classification into one of the previous four categories, and while the mechanisms responsible for their proliferation remain unclear, such distant single gene translocations may be mediated by several types of DNA transposons (Wang et al., 2012).

2.8 Positive selection tests

We used the branch-site model (Zhang et al., 2005) implemented in codeml of PAML v.4.9i (Yang 1997) to test for site-wise positive selection happening on the branch leading to *D. nivalis*. Briefly, *D. nivalis* was defined as the foreground branch on a predefined phylogeny consisting of eight Brassicaceae species (*T. parvula*, *B. oleraceae*, *R. raphanistrum*, *A. alpina*, *C. rubella*, *A. thaliana*, *A. lyrata*, and *D. nivalis*; Supplementary Table 2), and two models were compared with a likelihood ratio test (LRT): an alternative model that allowed positive selection on the foreground branch, and a null model that did not allow positive selection on the foreground branch (omega fixed to 1). The alternative model was accepted if $p < 0.05$ (using χ^2 with one degree of freedom), implying that positive selection has acted on a subset of sites along the branch leading to *D. nivalis*. The test was run on orthologous gene-alignments with one gene copy from all eight species, constructed from orthogroups identified with OrthoFinder (orthogroups are genes descended from a single gene in the last common ancestor of the eight species; Emms, Kelly 2015; Emms, Kelly 2019). Due to the low number of single-copy orthogroups, multiple copy orthogroups were divided into subsets based on the smallest genetic distance to each of the *D. nivalis* gene copies. This was achieved by i) aligning all orthogroups based on protein sequence using MAFFT (Katoh et al., 2005), ii) calculating Kimura protein distances (Kimura 1983) with the distmat algorithm in EMBOSS v.6.6.0 (Rice et al.,

2000), and iii) extracting one gene copy from all Brassicaceae species based on the smallest protein distance to each *D. nivalis* gene copy. The resulting orthogroup subsets were realigned using PRANK (Löytynoja, Goldman 2005) in GUIDANCE v. 2.02 (Sela et al., 2015) with ten bootstraps. GUIDANCE enables identification and filtration of unreliable alignment regions and sequences, and has been shown to improve positive selection inference on simulated data when used in combination with a phylogeny aware aligner like PRANK (Jordan, Goldman 2012; Privman et al., 2012). All alignments containing sequences scoring < 0.6 and all alignment columns scoring < 0.8 in GUIDANCE were removed from the dataset. Codeml was run 3-4 times for each model with different initial parameter values, and the run with the highest likelihood score was used in the final LRT (see e.g. Wong et al., 2004). Sites with ambiguity data were removed within codeml, and the species phylogeny inferred in OrthoFinder was used in all runs.

2.9 Gene ontology enrichment tests

The positively selected gene set and the sets of expanded and contracted gene families were tested for overrepresented gene ontology (GO) terms using the Bioconductor package topGO v.2.34 (Gentleman et al. 2004; Alexa et al. 2006). We used a Fisher's exact test in combination with the "classic", "elim" and "weight" algorithms to test for GO-term overrepresentation within the three domains: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). The three algorithms differ in that the "classic" algorithm processes each GO-term independently without considering the GO-graph, the "elim" algorithm processes the GO-graph bottom-up while discarding genes that have already been mapped to significant GO-terms, and the "weight" algorithm is weighing genes annotated to a GO-term based on the scores of neighboring GO-terms (Alexa et al. 2006). Based on simulated data, the "weight" algorithm has been shown to produce less false positives than the "classic" algorithm, whereas the "elim" algorithm further reduces false-positive rate, but with a higher risk of discarding true positives (Privman et al. 2012). The *D. nivalis* annotated gene set was used as a custom background for all GO term enrichment tests. The significance level was set to $p < 0.05$, and the results were not corrected for multiple testing following the recommendations of the creators of the topGO package (Gentleman et al. 2004).

3. Results

3.1 Genome Assembly

Based on the 25mer frequency distribution we estimated the genome size of *D. nivalis* to 278.48 Mb (Supplementary Figure 4; flow cytometry estimates report 254-308 Mb; Grundt et al. 2005). The initial *de novo* draft assembly (based on the Illumina paired-end HiSeq data) had a scaffold N50 of 30.083 Kb and a length of 280.94 Mb. Scaffolding this assembly with 1.57 Gb of Chicago proximity ligation data using the HiRise pipeline resulted in a scaffold N50 of 2.92 Mb and a length of 300.29 Mb. Scaffolding with 1.33 Gb of Oxford Nanopore MinION long read data (207,896 reads ranging in length from 1 Kb to 158.14 Kb with a mean read length of 3.8 Kb) further improved the scaffold N50 to 4.44 Mb and the length to 301.71 Mb (Table 1; Supplementary Table 3). We also produced a linkage map using 480 F_2 individuals genotyped with 5055 SNPs (Supplementary Figure 1) to order scaffolds into eight pseudomolecules, referred to as chromosomes (see Methods). The final assembly is 301.64 Mb (scaffold N50 = 31.02 Mb), of which 276.24 Mb is anchored to chromosomes varying from 29.2 Mb to 43.1 Mb (Figure 1a).

3.2 Chromosome evolution

To examine how the *D. nivalis* genome conforms to broader patterns of genome evolution in the Brassicaceae, we compared pairwise synteny between chromosomes of *D. nivalis* and those of *Arabidopsis lyrata* and *Arabis alpina*, and performed comparative chromosome painting (CCP) experiments to identify genomic blocks of the Ancestral Crucifer Karyotype (ACK, Schranz et al. 2006, Lysak et al. 2016; represented by the *A. lyrata* genome; Figures 1b, 2, 3; Supplementary Figures 2, 3). By synthesizing these results, we inferred the structure of the *D. nivalis* chromosomes. We identified several rearrangements and extensive centromere repositioning relative to the ACK. The structure of the *D. nivalis* genome is very similar to that of *A. alpina* (Willing et al. 2015), the closest relative of *D. nivalis* for which a chromosome-scale genome assembly is available, and consistent with genome structures determined for other Arabideae species, including three *Draba* species (Mandáková et al. 2020).

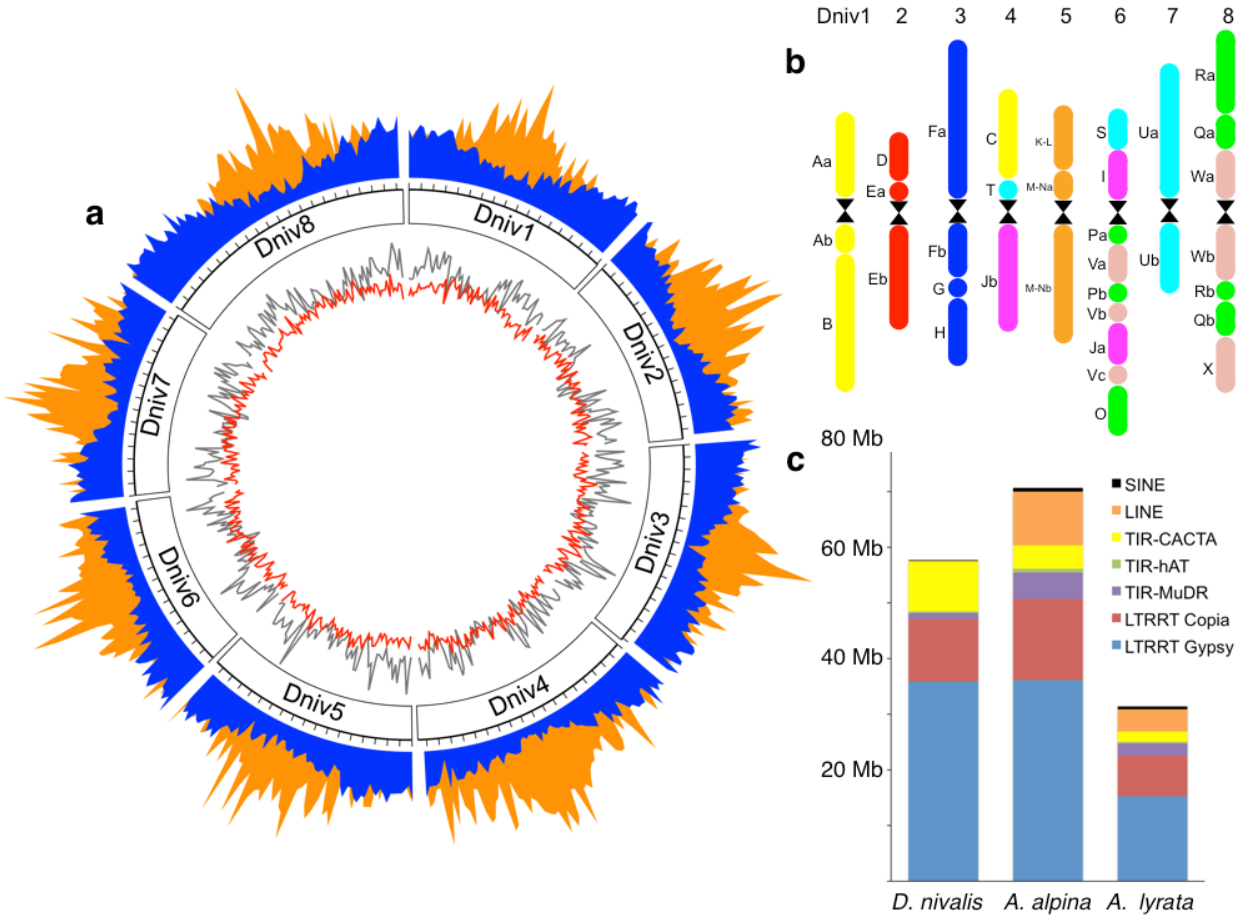


Figure 1. **The *D. nivalis* genome assembly**. **a**, Circos plot of the eight chromosomes (Dniv1-Dniv8) showing the distribution of gene annotations (blue) and LTR-RT elements (orange) in 500 Kb windows. Ticks represent 2 Mb intervals. The inner tracks show the distribution of TIR (grey) and Helitron (red) elements in 500 Kb windows. **b**, Organization of ancestral Brassicaceae genomic blocks in the eight *D. nivalis* chromosomes (Dniv1-8) based on CCP and comparative analyses relative to *A. lyrata* and *A. alpina*. Centromere positions of chromosomes 1-3 and 8 are tentative, but supported by results in Supplementary Figures 2, 3 and the structure of other Arabideae species (Mandáková et al. 2020). Genomic blocks are colored to match eight colors corresponding to eight chromosomes of the Ancestral Crucifer Karyotype (Lysak et al. 2016). **c**, Relative abundance of TE superfamilies in selected species (see Supplementary Table 5; Willing et al. 2015). Whereas LTR-RT abundance is similarly elevated in *D. nivalis* and *A. alpina* relative to *A. lyrata*, LINEs appear to be reduced and TIR-CACTA elements enriched in *D. nivalis* relative to both species.

Table 1. **Assembly statistics for the *D. nivalis* genome**. Each row shows a different stage in the scaffolding process of the genome assembly. For a more complete table, see Supplementary Table 3.

Assembly	# of Scaffolds (> 1Kb)	N50 (Kb)	L50	Longest Scaffold (Kb)	Length (Including Ns; Mb)	#Ns per 100 Kb
DDN	28,339	30	2,663	203	281	51.58
DDN_Chicago	7,857	2,918	30	9,273	300	6,486
DDN_Chicago_ONT	6,765	4,437	21	20,017	302	6,874
DDN_Chicago_ONT_map	6,682	31,019	5	43,070	302	6,853

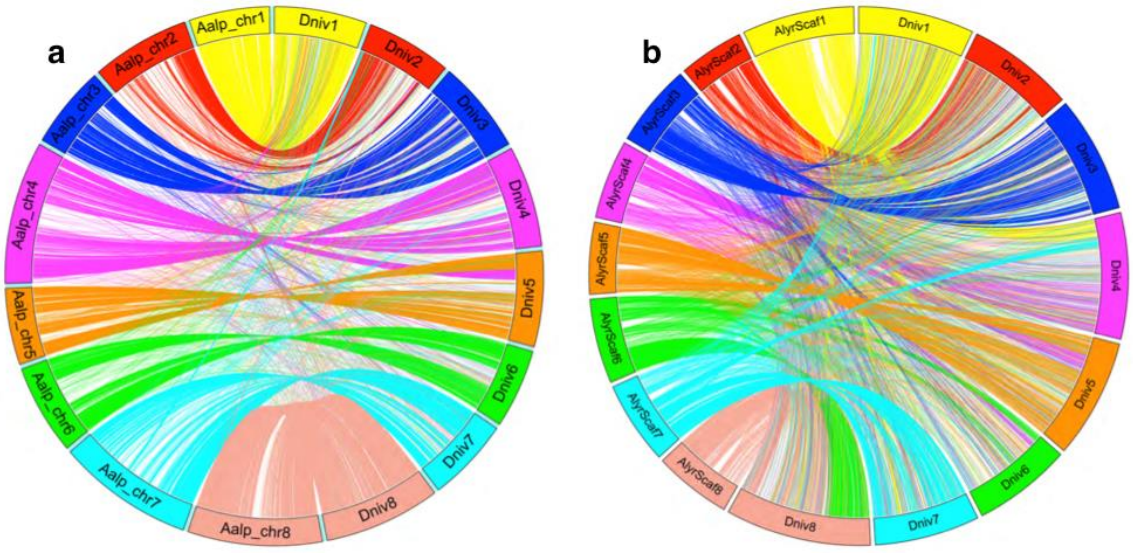


Figure 2. **Syntenic relationships between *D. nivalis* and *Arabis alpina* (a) and *Arabidopsis lyrata* (b).** The *D. nivalis* genome was aligned to the genomes of *A. alpina* and *A. lyrata* with NUCmer. Chromosomes are color-coded to match the Ancestral Crucifer Karyotype (ACK; Lysak et al. 2016) structurally resembling the *A. lyrata* genome (b).

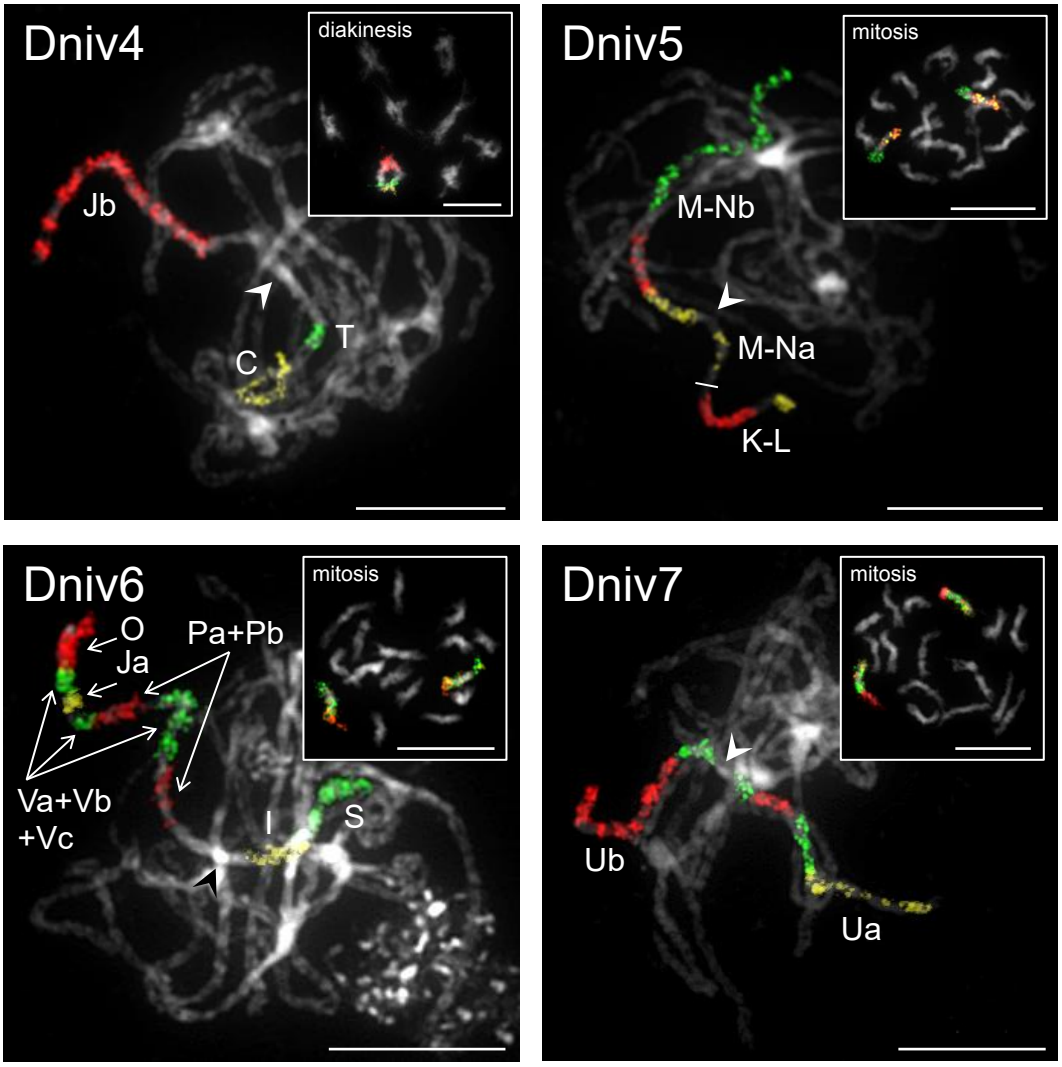


Figure 3. Detailed CCP analysis of *D. nivalis* chromosomes Dniv4, Dniv5, Dniv6, and Dniv7. *Arabidopsis thaliana* BAC contigs were *in situ* hybridized to pachytene chromosome spreads in *D. nivalis* (insets show the same probes localized on diakinetic chromosomes or mitotic metaphase chromosomes). Arrowheads indicate the position of centromeres. Green, yellow, and red color corresponds to fluorescence of Alexa 488, Cy3 and Texas Red, respectively. Chromosomes were counterstained with DAPI. Scale bars 10 μm.

3.3 Repetitive element annotation

We annotated 94.8 Mb of the genome as direct remnants of repetitive elements, dominated by long terminal repeat retrotransposons (LTR-RT, 60.1 Mb), terminal inverted repeat transposons (TIR, 20.9 Mb), and Helitrons (13.4 Mb; Figure 1; Supplementary Tables 4, 5, 6). Consistent with *A. alpina*, *A. lyrata*, and *A. thaliana*, LTR-RT density increases in pericentromeric regions of each chromosome, TIR density decreases in pericentromeric regions, and Helitron density is stable along chromosomes. Abundance of LTR-RT *Copia* and *Gypsy* elements is similar to that of

A. alpina (Choudhury et al., 2017), whereas TIR-CACTA elements and Helitrons seem to be particularly abundant in *D. nivalis* (Hu et al., 2019). Nucleotide divergence among LTR-RTs identifies several *Copia* and *Gypsy* LTR-RTs showing recent transposition bursts across the genome of *D. nivalis*. Some abundant LTR-RTs (e.g. ALYCopia74, ATLANTYS2) have very similar copies (most >98%) and thus seem to have proliferated more recently than in *A. alpina* (Figure 4). These results show that *Copia* elements, including the heat-activated ATCOPIA78 (Ito et al., 2011) and tribes preferentially transposing across the gene space of Brassicaceae (Quadrana et al., 2016), have specifically contributed to the evolution of the *D. nivalis* genome.

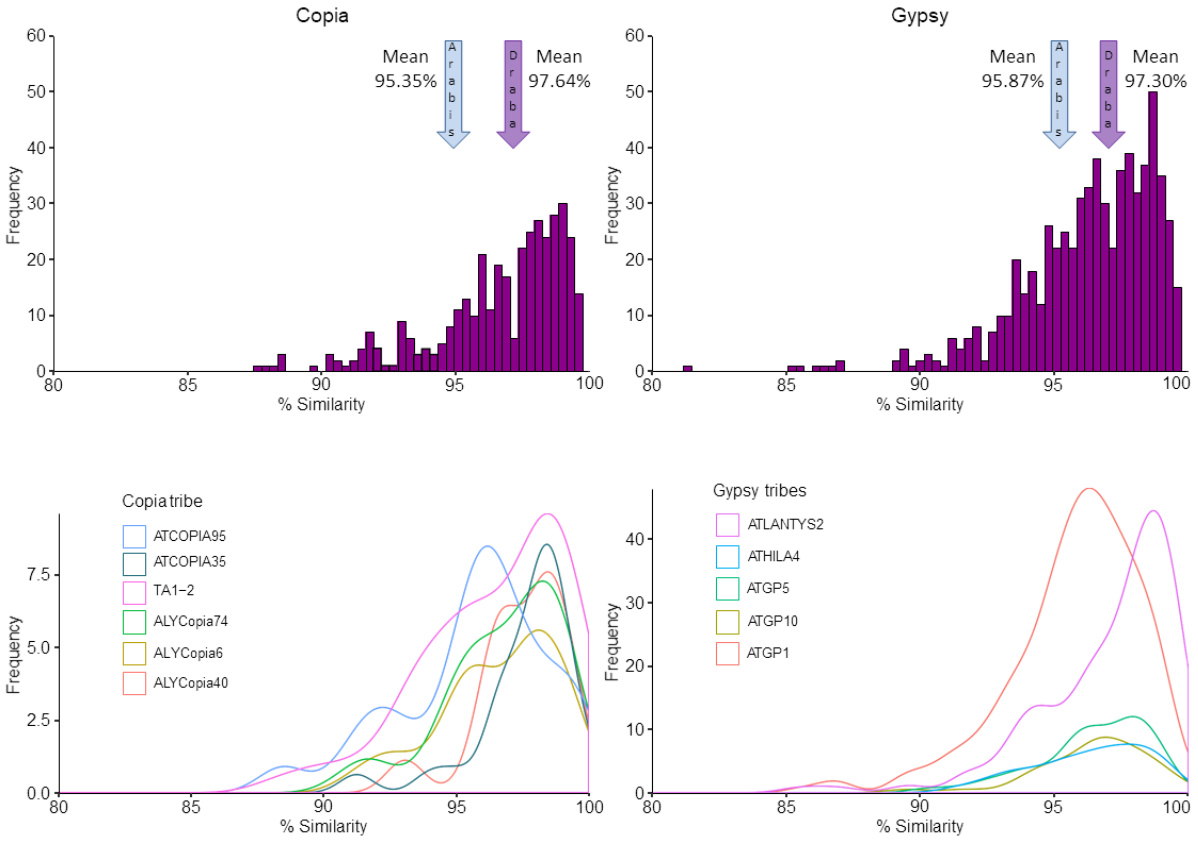


Figure 4. Evolutionary dynamics of abundant LTR-RT tribes in *D. nivalis*. Results are based on the percent nucleotide similarity of LTR sequences among full-length copies in the genome. Selected tribes show peaks indicative of transposition bursts that were above the average of all tribes for *Copia* and *Gypsy* (97.64% and 97.30%, respectively), distinguishing retrotransposons such as ATCOPIA95 showing ancestral proliferation from tribes such as ATCOPIA35 with a majority of recent transposition events. Summary statistics of all LTR-RT tribes in *D. nivalis* and *A. alpina* are provided in Supplementary Table 6.

3.4 Gene annotation

We predicted gene models with the Maker2 pipeline using BLAST homology to five Brassicaceae genomes and a *de novo* transcriptome assembly of *D. nivalis* based on RNA-seq data from leaves, roots, flowers, and flower buds (see Methods). We identified 33,557 gene models, and 74% of the genes were functionally annotated based on similarity to UniProtKB entries, and 70% were annotated with InterPro domains. Approximately 58% of the 33,557 gene models had an annotation edit distance less than or equal to 0.25, suggesting a relatively high degree of agreement between predicted gene models and external evidence. This gene set is somewhat larger than that of *A. thaliana* (27,654), but consistent with those of closely related species with similar genome size (Supplementary Figure 5), and BUSCO analyses indicate 95.2% completeness of conserved embryophyte genes (Supplementary Table 7). The average gene density in *D. nivalis* is approximately one gene per 9 Kb, and similar to *A. thaliana*, *A. lyrata*, and *A. alpina*, gene density decreases towards the centromeres (Figure 1a; Willing et al., 2015).

3.5 Gene and gene family evolution

To explore specialization in the *D. nivalis* gene set, we compared the abundance of protein family (Pfam) annotations with those of nine Brassicaceae genomes representing broad phylogenetic sampling. We found 226 Pfam domains to be significantly enriched and 32 to be significantly depleted relative to the other species (Supplementary Table 8). To summarize functional associations of the enriched Pfam domains, we extracted gene ontology (GO) terms from their corresponding InterPro entries. Amongst these GO terms, the most common biological process (BP) GO term is “oxidation-reduction process”. Numerous environmental stimuli and stresses can lead to the production of reactive oxygen species (ROS), which can damage cell membranes, nucleic acids, proteins, and metabolites (Apel, Hirt 2004). Regulation of ROS metabolism is essential for maintaining cellular oxidation-reduction (redox) homeostasis and is an integral part of the intracellular signal transduction networks evoked by external stimuli (Mittler 2017), particularly for responses to environmental stresses induced by light, drought, and cold (Neill et al., 2002). The significant increase in Pfam annotations involved in redox processes in *D. nivalis* may indicate that it has evolved novel ways to cope with ROS accumulation associated with Arctic environmental stress (Figure 5). The salt tolerant *Eutrema salsugineum* (syn.

Thellungiella salsuginea), for example, responds to salt stress by expressing an aldehyde dehydrogenase, a scavenger of toxic aldehydes produced as a byproduct of ROS accumulation (Hou, Bartels 2014). Consistent with this, the *D. nivalis* annotated gene set contains 26 genes containing the significantly enriched molybdopterin-binding domain of aldehyde dehydrogenase.

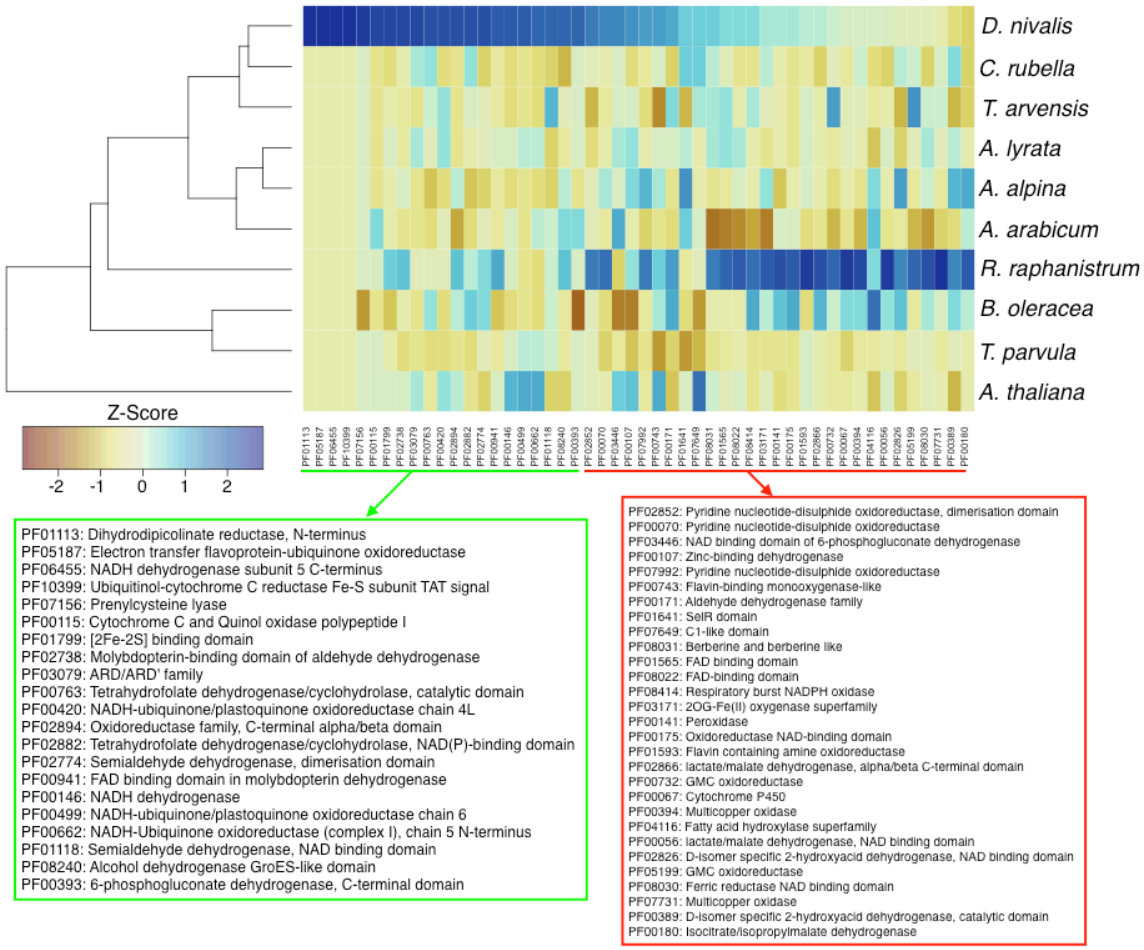


Figure 5. **Pfam domains associated with oxidation-reduction processes are enriched in *D. nivalis***. Heatmap comparing Pfam domains annotated with the BP GO term "oxidation-reduction process" (GO:0055114) in selected Brassicaceae species. Cells are colored by Z-score. The dendrogram on the left represents groupings based on similar counts of selected Pfam domains. Pfam domains detailed in green are significantly enriched in *D. nivalis* relative to the other nine species (Z-score > 1.96). Pfam domains detailed in red are not significantly enriched, but have >10 genes annotated in *D. nivalis*. The genomes of both *D. nivalis* and *R. raphinistrum* contain relatively abundant and significantly enriched Pfam domains associated with oxidation-reduction processes, which are important in stress response signaling pathways.

To compare the diversity and abundance of *D. nivalis* gene families relative to the nine other Brassicaceae species, we estimated gene family (orthogroup) membership using OrthoFinder. A total of 29,194 (87%) *D. nivalis* genes were assignable to one of the 21,635 gene families identified, and 10,401 of the gene families contain at least one gene copy in all ten species. Genome-wide classification of gene duplications in *D. nivalis* using the Dup_GenFinder pipeline (Qiao et al., 2019) resulted in similar patterns across all species (Supplementary Table 9). Gene duplications in *D. nivalis* are dominated (22,989 gene pairs, 89.6%) by transposed (TRD, 7,308 pairs) and dispersed (DSD, 15,681 pairs) duplicates (Figure 6a), both of which are likely the product of transposition events that can be mediated by transposable elements (Qiao et al., 2019).

Relative to the nine other species, *D. nivalis* contains 198 significantly expanded and 31 significantly contracted gene families (Figure 6a). Exploring the functional annotations of the 2,958 genes of the expanded gene families (EGFs), we found 158 significantly enriched BP GO annotations including several functions highlighting how this species was able to adapt to Arctic habitats (Figure 6b; Supplementary Tables 10, 11, 12). Functions associated with stress signaling include both the abscisic acid (ABA) activated and brassinosteroid-mediated pathways, involved in cellular functions including abiotic stress signaling (Planas-Riverola et al., 2019). We also see functional enrichment for heat acclimation associated with three EGFs, which also are associated with defense responses to fungal pathogens. While fungal pathogens are not expected to be particularly virulent in the Arctic, stress can make plants more susceptible to pathogens. The EGFs are also enriched for functions associated with meiosis, specifically the assembly of the synaptonemal complex. The efficiency and fidelity of recombination is sensitive to temperature (Bomblies et al., 2015), and these results may indicate adaptation in *D. nivalis* to facilitate meiosis in cold habitats. Gene families associated with desiccation resistance are also expanded in *D. nivalis*, consistent with its occurrence in extremely dry, so-called polar deserts. While BP terms related to redox homeostasis were not enriched in the EGFs (Figure 6b), redox activity was prominent among the 64 enriched MF terms (Figure 6c), consistent with the overrepresentation of redox Pfam domains observed in the genome (Figure 5). Functions associated with protein modification and ubiquitination were enriched in the EGFs, consistent with previously published results for the salt tolerant *Thellungiella salsuginea* (Yang et al., 2013). Finally, the

D. nivalis genome contains several EGFs that function in histone binding and methylation, integral parts of epigenetic regulatory mechanisms that can play important roles in numerous abiotic stress signaling and response pathways (Ueda, Seki 2020). Patterns of duplication inferred for the 2,958 genes that constitute the EGFs in *D. nivalis* are broadly consistent with genomic patterns in that TRD and DSD duplications dominate (79.9%), but TRD duplications are less frequent, and proximal (PD) and tandem (TD) duplications are more frequent in EGFs than would be expected by chance (Figure 6a, Supplementary Table 9). This suggests that the activity of prevalent LTR-RTs, TIR transposons, and Helitrons likely played important roles in the expansion of *D. nivalis* gene families, but processes of proximal and tandem duplication also appear to have been important in the expansion of gene families associated with protein modification, stress signaling, desiccation resistance, and defense responses to fungal pathogens (Figure 6b,c).

3.6 Tests of positive selection

To search for further evidence of Arctic adaptation in *D. nivalis*, we performed genome-wide positive selection tests to identify genes that likely evolved under positive selection in this lineage relative to seven related species (see Methods). We found 1,307 positively selected genes (PSGs). These include several candidate genes with functions directly relevant to typical environmental stresses of the Arctic, associated with “response to cold”, “response to water deprivation”, “photoperiodism”, “response to oxidative stress”, and “meiosis I” (Figures 7a,b; Supplementary Tables 13, 14). Patterns of functional enrichment of the PSGs also highlight several significant BP GO terms likely connected to Arctic adaptation, including “vernalization response”, “drought recovery”, “short-day photoperiodism”, and “oxidation-reduction process” (Figure 7b; Supplementary Table 14). We also found four PSGs associated with meiosis I, including two *D. nivalis* homologs to *A. thaliana ZYP1A*, which is one of three synaptonemal complex transverse filament proteins whose function is disrupted by temperature stress (Bomblies et al., 2015). These results provide evidence for the likely adaptive evolution of core meiosis genes reflected both in EGFs (Figure 6b) and in positive selection acting on specific components of the synaptonemal complex.

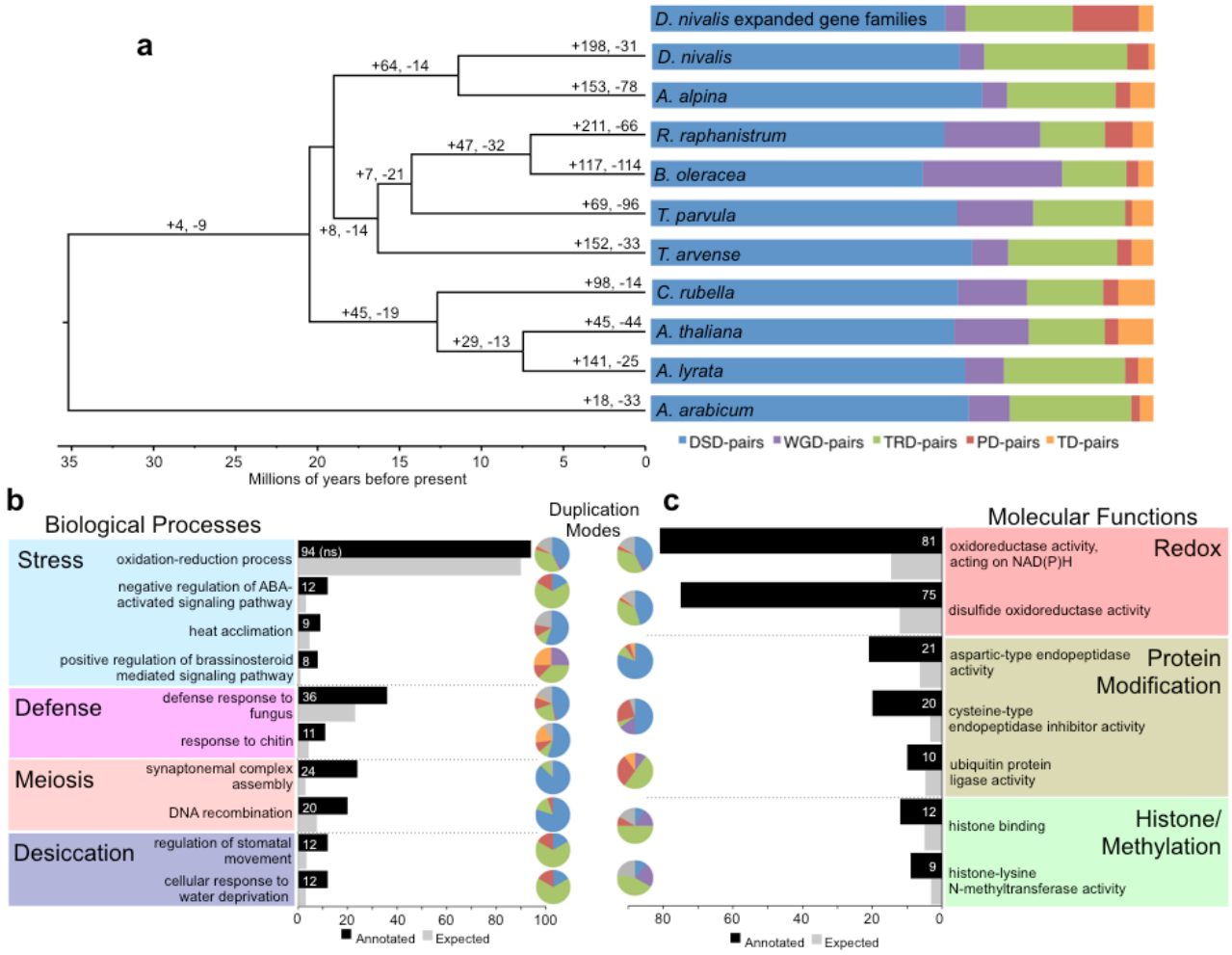


Figure 6. **Gene family evolution in *D. nivalis***. **a,** Time-calibrated species tree of ten Brassicaceae genomes showing branches labeled with significant gene family expansions and contractions (+gene families gained, -gene families lost). To the right we show the percentage of gene pairs derived from different modes of duplication in the ten species. For *D. nivalis,* gene duplication modes are shown for both the whole genome (labeled *D. nivalis*) and for the 2,645 genes that constitute 198 significantly expanded gene families (EGFs). Patterns of gene duplication are broadly consistent between the whole genome and the EGFs, except that transposed (TRD) duplications are less frequent (*p*-value = 1.8e-15), and proximal (PD, *p*-value < 2.2e-16) and tandem (TD, *p*-value = 2.432e-10) duplications are more frequent in EGFs than would be expected by chance (Fisher's exact test). Values for all species other than *D. nivalis* are from Qiao et al., (2019), and since their analysis did not include *R. raphanistrum*, the results shown are from the closely related *Raphanus sativa*. DSD = dispersed duplication, WGD = whole-genome duplication, TRD = transposed duplication, PD = proximal duplication, TD = tandem duplication. **b** and **c,** Select biological process (**b**) and molecular function (**c**) GO terms significantly enriched (Fisher's exact test *p*-value < 0.05) in *D. nivalis* EFGs (in **b**, oxidation-reduction process was not significant, indicated by "ns"). Terms are grouped into broad categories to simplify interpretation. Pie charts for each term show the modes of gene duplication inferred for genes annotated with these terms (DSD, WGD, TRD, PD, TD; color scheme follows **a**, with unclassified in grey).

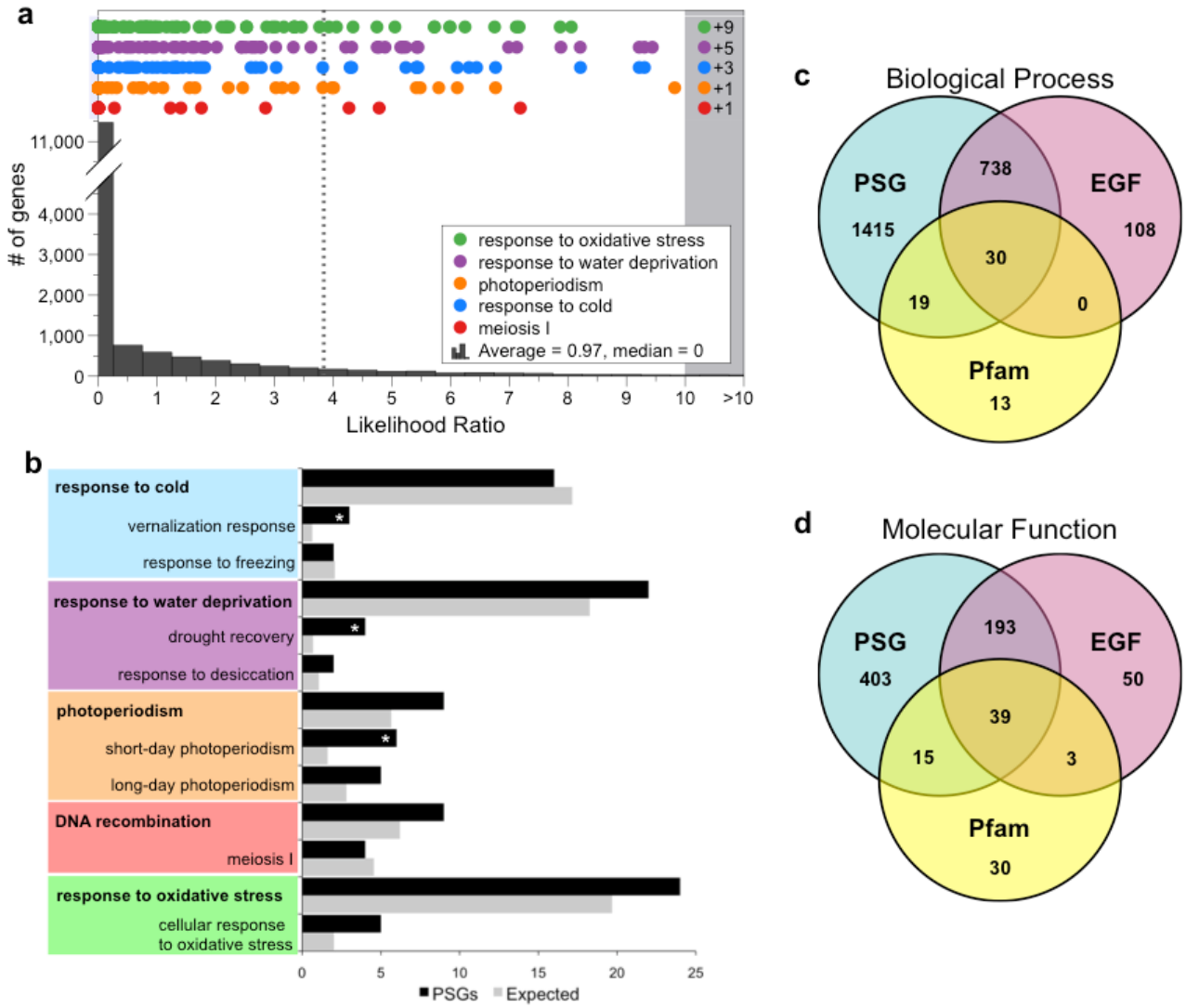


Figure 7. **Genes under positive selection in *D. nivalis***. **a,** Distribution of the ratios of ln likelihoods (lnL) from tests for positive selection in 15,828 *D. nivalis* genes. Genes with a higher proportion of nonsynonymous to synonymous substitutions have a higher lnL ratio, and those with an lnL ratio above the X^2 critical value (3.84, dashed line, p value < 0.05, d.f. = 1) are considered significantly likely to contain codons that evolved under positive selection in *D. nivalis* (PSGs; see Methods). Colored dots represent genes that are annotated with biological process (BP) GO terms of particular interest for Arctic adaptation. **b,** Summary of key BP GO terms in the *D. nivalis* PSGs. Asterisks (*) indicate significantly enriched terms relative to the genomic background. Parent terms are in bold. **c** and **d,** Venn diagrams showing the overlap of BP (**c**) and molecular function (**d**) GO terms resulting from analysis of Pfam domains, expanded gene families (EGF), and PSGs in *D. nivalis* (see also Supplementary Table 10).

4. Discussion

Summarizing the results of our comparative genomic analyses, we observe some similarities in functional patterns among enriched Pfam domains, gene family expansions, and genes under positive selection (Figure 7c,d). Our results reveal a multifaceted landscape of stress adaptation in the *D. nivalis* genome, and highlight the important roles that genes involved in stress signaling/response, redox homeostasis, light sensing, and meiosis likely play in plant adaptation to the extreme Arctic environment. The numerous genes that we have identified represent good candidates for future studies of functional validation in various stress responses. If such studies are successful, they could provide guidance for various approaches to crop improvement. The highly contiguous genome assembly of *D. nivalis* that we have produced provides numerous avenues for the continued development of this species as the first Arctic specialist model plant. Future uses of this resource could include, e.g. studies of the adaptive potential of Arctic species to future climate change.

Acknowledgments

We thank Filip Kolár for helpful discussions. The main work was funded by grant 240223/F20 to CB from the Research Council of Norway; additional support was obtained from the Czech Science Foundation (grant 15-18545S) and the CEITEC 2020 project (grant LQ1601). Computational analyses were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway, and on the Abel Cluster, owned by the University of Oslo and UNINETT Sigma2, and operated by the Department for Research Computing at USIT, the University of Oslo IT-department (http://www.hpc.uio.no/).

Author Contributions

AKB, CB, MDN, LR conceived and initiated the study.

MDN performed the assembly and annotation of the *D. nivalis* genome.

CP, RRC carried out the repeat analysis.

MAL, TM, XG conducted the CCP experiments.

AG, ALSG, MDN, SB carried out experimental plant work for genome sequencing.

AG, ASN, ALSG, MDN, SB produced and genotyped the F_2 mapping population.

ALSG, MF, MDN, TS constructed the genetic map.

SB conducted the analyses of molecular evolution.

MDN, SB carried out the comparative genomics analyses.

MDN wrote the manuscript with comments from all coauthors.

AKB, CB, LR, MDN, TS, supervised the study.

Data Accessibility Statement

The raw data (shotgun sequence data, MinION long reads, Chicago Linked Reads, and RNA-seq) have been deposited in the NCBI SRA with Bioproject number PRJNA657155. The final chromosome-scale assembly, gene annotation, repeat library and annotation, transcriptome assembly, ddRAD sequence data of the F2 mapping population, and vcf file of variants used in the construction of the genetic map are available on Dryad (DOI: https://doi.org/10.5061/dryad.pg4f4qrm4).

References

Alexa, A., Rahnenführer, J., Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, *22*(13), 1600–1607.

Apel, K., Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annual Review of Plant Biology*, 55, 373–399.

Arends, D., Prins, P., Jansen, R. C., Broman, K. W. (2010). R/qtl: high-throughput multiple QTL mapping. *Bioinformatics*, 26, 2990–2992.

Birkeland, S., Gustafsson, A. L. S., Brysting, A. K., Brochmann, C., Nowak, M. D. (2020). Multiple genetic trajectories to extreme abiotic stress adaptation in Arctic Brassicaceae. *Molecular Biology and Evolution*, accepted manuscript, https://doi.org/10.1093/molbev/msaa068.

Boetzer, M., Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics,* 15, 211.

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Bomblies, K., Higgins, J. D., Yant, L. (2015). Meiosis evolves: adaptation to external and internal environments. *New Phytologist*, 208, 306–323.

Brochmann C., Soltis P. S., Soltis, D. E. (1992). Multiple origins of the octoploid Scandinavian endemic *Draba cacuminum*: electrophoretic and morphological evidence. *Nordic Journal of Botany*, 12, 257–272.

Brochmann, C. (1993). Reproductive strategies of diploid and polyploid populations of arctic *Draba* (Brassicaceae). *Plant Systematics and Evolution*, 185, 55–83.

Broman, K. W., Wu, H., Sen, S., Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19, 889–890.

Caldwell, M. M., Bornman, J. F., Ballaré, C. L., Flint, S. D., Kulandaivelu, G. (2007). Terrestrial ecosystems, increased solar ultraviolet radiation, and interactions with other climate change factors. *Photochemical, Photobiological Sciences*, 6, 252–266.

Choudhury, R. R., Neuhaus, J.-M., Parisod, C. (2017). Resolving fine-grained dynamics of retrotransposons: comparative analysis of inferential methods and genomic resources. *The Plant Journal*, 90, 979–993.

Colella, J. P., Talbot, S. L., Brochmann, C., Taylor, E. B., Hoberg, E. P., Cook, J. A. (2020). Conservation genomics in a changing Arctic. *Trends in Ecology, Evolution*, 35, 149–162.

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

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., Dionorad, J., Quick, W. P., Parke, M., Bennetzen, J. L., Besnardf, G.,

Nosil, P., Osborne, C. P., Christina, P-A. (2019). Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences*, 116, 4416–4425.
Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30, 1844–1849.
El-Baidouri, M., Panaud, O. (2013). Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biology and Evolution*, 5, 954–965.
El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto S. C. E., Finn, R. D. (2018). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47, D427–D432.
Ellinghaus, D., Kurtz, S., Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, 9, 18–14.
Emms, D. M., Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16, 157.
Emms, D. M., Kelly, S. (2017). STRIDE: Species tree root inference from gene duplication events. *Molecular Biology and Evolution*, 34, 3267–3278.
Emms, D., Kelly, S. (2018). STAG: Species tree inference from all genes. Preprint at https://www.biorxiv.org/content/10.1101/267914v1.
Emms, D. M., Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20, 238.
Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. H, Zhang J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5(10), R80.
Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644–652.
Gremme, G., Steinbiss, S., Kurtz, S. (2013). GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 10(3), 645–656.
Grundt, H. H., Obermayer, R. Borgen, L. (2005). Ploidal levels in the arctic-alpine polyploid *Draba lactea* (Brassicaceae) and its low-ploid relatives. *Botanical Journal of the Linnean Society*, 147, 333–347.

Grundt, H. H., Kjølner, S., Borgen, L., Rieseberg, L. H., Brochmann, C. (2006). High biological species diversity in the arctic flora. *Proceedings of the National Academy of Sciences*, 103, 972–975.

Gupta, S. K. (2016). *Biology and Breeding of Crucifers*. Boca Raton USA: CRC Press.

Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X., Zhang, D., Ma, T., Hu, Q., Al-Shehbaz, I. A., Koch, M. A. (2017). Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics*, 18, 176.

Gustafsson, A. L. S., Skrede, I., Rowe, H. C., Gussarova, G., Borgen, L., Rieseberg, L. H., Brochmann, C., Parisod, C. (2014). Genetics of cryptic Speciation within an Arctic mustard, *Draba nivalis*. *PLoS ONE*, 9, e93834.

Haas B., J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8, 1494–1512.

Han, M. V., Thomas, G. W. C., Lugo-Martinez, J., Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, 30, 1987–1997.

Holt, C., Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12, 491.

Hou, Q., Bartels, D. (2014). Comparative study of the aldehyde dehydrogenase (ALDH) gene superfamily in the glycophyte *Arabidopsis thaliana* and *Eutrema* halophytes. *Annals of Botany*, 115, 465–479.

Hu, K., Xu, K., Wen, J., Yi, B., Shen, J., Ma, C., Fu, T., Ouyang, Y., Tu, J. (2019). Helitron distribution in Brassicaceae and whole genome Helitron density as a character for distinguishing plant species. *BMC Bioinformatics*, 20, 354–20.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., Paszkowski, J., (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, 472, 115–120.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240.

Jordan, G., Goldman, N. (2012). The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Molecular Biology and Evolution*, 29(4), 1125–1139.

Jordon-Thaden, I. E., Al-Shehbaz, I. A., Koch, M. A. (2013). Species richness of the globally distributed, arctic–alpine genus *Draba* L. (Brassicaceae). *Alpine Botany*, 123, 97–106.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, 110, 462–467.

Katoh, K., Kuma, K., Toh, H., Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, 33(2), 511–518.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, 5, 59.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, 5, R12.

Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997.

Li, H., Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26, 589–595.

Li, W., Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658–1659.

Llorens, C., Futami, R., Covelli, L., Domínguez-Escribá, L., Viu, J. M., Tamarit, D., Aguilar-Rodríguez, J., Vicente-Ripolles, M., Fuster, G., Bernet, G. P., Maumus, F., Munoz-Pomer, A., Sempere, J. M., Latorre, A., Moya, A. (2011). The Gypsy Database (GyDB) of mobile genetic elements: Release 2.0. *Nucleic Acids Research*, 39, D70–D74.

Löytynoja, A., Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences*, 102(30), 10557–10562.

Lütz, C. (2010). Cell physiology of plants growing in cold environments. *Protoplasma*, 244, 53–73.

Lysak, M.A., Mandáková T., Schranz M.E. (2016). Paleogenomics of Crucifers: Ancestral Genomic Blocks Revisited. *Current Opinion in Plant Biology*, 30, 108-115.

Mandáková T., Lysak M. A. (2016a). Chromosome preparation for cytogenetic analyses in Arabidopsis. *Current Protocols in Plant Biology*, 1, 1-9.

Mandáková T., Lysak M. A. (2016b). Painting of Arabidopsis chromosomes with chromosome-specific BAC clones. *Current Protocols in Plant Biology*, 1, 359-371.

Mandáková, T., Hloušková, P., Koch, M. A., Lysak, M. A. (2020). Genome evolution in Arabideae was marked by frequent centromere repositioning. *Plant Cell*, tpc.00557.2019–43.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.

Mittler, R. (2017). ROS are good. *Trends in Plant Science*, 22, 11–19.

Neill, S. J., Desikan, R., Clarke, A., Hurst, R. D., Hancock, J. T. (2002). Hydrogen peroxide and nitric oxide as signalling molecules in plants. *Journal of Experimental Botany*, 53, 1237–1247.

Planas-Riverola, A., Gupta, A., Betegón-Putze, I., Bosch, N., Ibañes, M., Caño-Delgado, A. I. (2019). Brassinosteroid signaling in plant development and adaptation to stress. *Development*, 146, dev151894–11.

Privman, E., Penn, O., Pupko, T. (2012). Improving the performance of positive selection inference by filtering unreliable alignment regions. *Molecular Biology Evolution*, 29(1), 1–5.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., Paterson, A. H. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biology*, 20, 1–23.

Quadrana, L., Bortolini Silveira, A., Mayhew, G. F., LeBlanc, C., Martienssen, R. A., Jeddeloh, J. A., Colot, V. (2016). The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife*, 5, 6919.

Rice, P., Longden, I., Bleasby, A. (2000). EMBOSS : The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(6), 2–3.

Sanderson, M. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, 19, 301–302.

Schranz, M., Lysak, M., Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends in Plant Science*, 11, 535–542.

Sela, I., Ashkenazy, H., Katoh, K., Pupko, T. (2015). GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, 43, W7–W14.

Simao, 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, 3210–3212.

Skrede, I., Brochmann, C., Borgen, L., Rieseberg, L. H. (2008). Genetics of intrinsic postzygotic isolation in a circumpolar plant species, *Draba nivalis* (Brassicaceae). *Evolution*, 62, 1840–1851.

Stanke, M., Diekhans, M., Baertsch, R., Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, 24, 637–644.

Steinbiss, S., Willhoeft, U., Gremme, G., Kurtz, S. (2009). Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Research*, 37, 7002–7013.

Steinegger, M., Söding, J. M. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35, 1026–1028.

Taylor, J., Butler, D. (2017). RPackage ASMap: Efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software*, 79, 1–29.

Ueda, M., Seki, M. (2020). Histone modifications form epigenetic regulatory networks to regulate abiotic stress response. *Plant Physiology*, 182, 15–26.

Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40, e49.

Warren, R. L., Yang, C., Vandervalk, B. P., Behsaz, B., Lagman, A., Jones, S. J., Birol, I. (2015). LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *GigaScience*, 4, 35.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E, SanMiguel, P, Schulman, A. H. (2009). Reply: A unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nature Reviews Genetics*, 10, 276.

Willing, E. M., Rawat, V., Mandáková, T., Maumus, F., James, G. V., Nordström, K. J., Becker, C., Warthmann, N., Chica, C., Szarzynska, B., Zytnicki, M., Albani, M. C., Kiefer, C., Bergonzi, S., Castaings, L., Mateos, J. L., Berns, M. C., Bujdoso, N., Piofczyk, T., de Lorenzo, L., Barrero-Sicilia, C., Mateos, I., Piednoël, M., Hagmann, J., Chen-Min-Tao, R., Iglesias-Fernández, R., Schuster, S. C., Alonso-Blanco, C., Roudier, F., Carbonero, P., Paz-Ares, J., Davis, S. J., Pecinka, A., Quesneville, H., Colot, V., Lysak, M. A., Weigel, D., Coupland, G., Schneeberger, K. (2015). Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants*, 1, 14023–7.

Wong, W. S. W., Yang, Z., Goldman, N., Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*, 168(2), 1041–1051.

Wullschleger, S. D., Breen, A. L., Iversen, C. M., Olson, M. S., Näsholm, T., Ganeteg, U., Wallenstein, M. D., Weston, D. J. (2015). Genomics in a changing arctic: critical questions await the molecular ecologist. *Molecular Ecology*, 24, 2301–2309.

Xiong, W., He, L., Lai, J., Dooner, H. K., Du, C. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences*, 111, 10263–10268.

Yang Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, 13(5), 555–556.

Yang, L., Bennetzen, J. L. (2009). Distribution, diversity, evolution, and survival of Helitrons in the maize genome. *Proceedings of the National Academy of Sciences*, 106, 19922–19927.

Yang, R., Jarvis, D. E., Chen, H., Beilstein, M. A., Grimwood, J., Jenkins, J., Shu, S., Prochnik, S., Xin, M., Ma, C., Schmutz, J., Wing, R. A., Mitchell-Olds, T.,

Schumaker, K. S., Wang, X. (2013). The reference genome of the halophytic plant *Eutrema salsugineum*. *Frontiers in Plant Science*, 4, 46.
Zhang, J., Nielsen, R., Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, 22(12), 2472–2479.

Supplementary Material, Paper III

Supplementary tables

Supplementary Table 1. Summary of marker positions in the genetic map

Supplementary Table 2. Sources of other Brassicaceae genomes used in this study

Supplementary Table 3. Genome assembly statistics

Supplementary Table 4. Summary of repetitive element annotation

Supplementary Table 5. Repetitive element comparison between *D. nivalis* and *A. alpina*

Supplementary Table 6. Dynamics of repetitive element insertions in both *D. nivalis* and *A. alpina*

Supplementary Table 7. BUSCO statistics of the *D. nivalis* genome assembly

Supplementary Table 8. Significant enrichment and depletion of PFAM annotations in *D. nivalis*

Supplementary Table 9. Modes of gene duplication in Brassicaceae genomes

Supplementary Table 10. GO terms functional enrichment tests for significantly expanded gene families in *D. nivalis*

Supplementary Table 11. Annotations of the 2,958 genes that compose significantly expanded gene families in *D. nivalis*

Supplementary Table 12. GO terms functional enrichment tests for significantly contracted gene families in *D. nivalis*

Supplementary Table 13. Annotations of positively selected genes (PSGs) identified in the *D. nivalis* genome

Supplementary Table 14. GO terms functional enrichment tests for genes under positive selection in *D. nivalis*

All supplementary tables can be found at https://doi.org/10.5061/dryad.7pvmcvdqx

Supplementary figures

Supplementary Figure 1. Genetic linkage map of *D. nivalis*

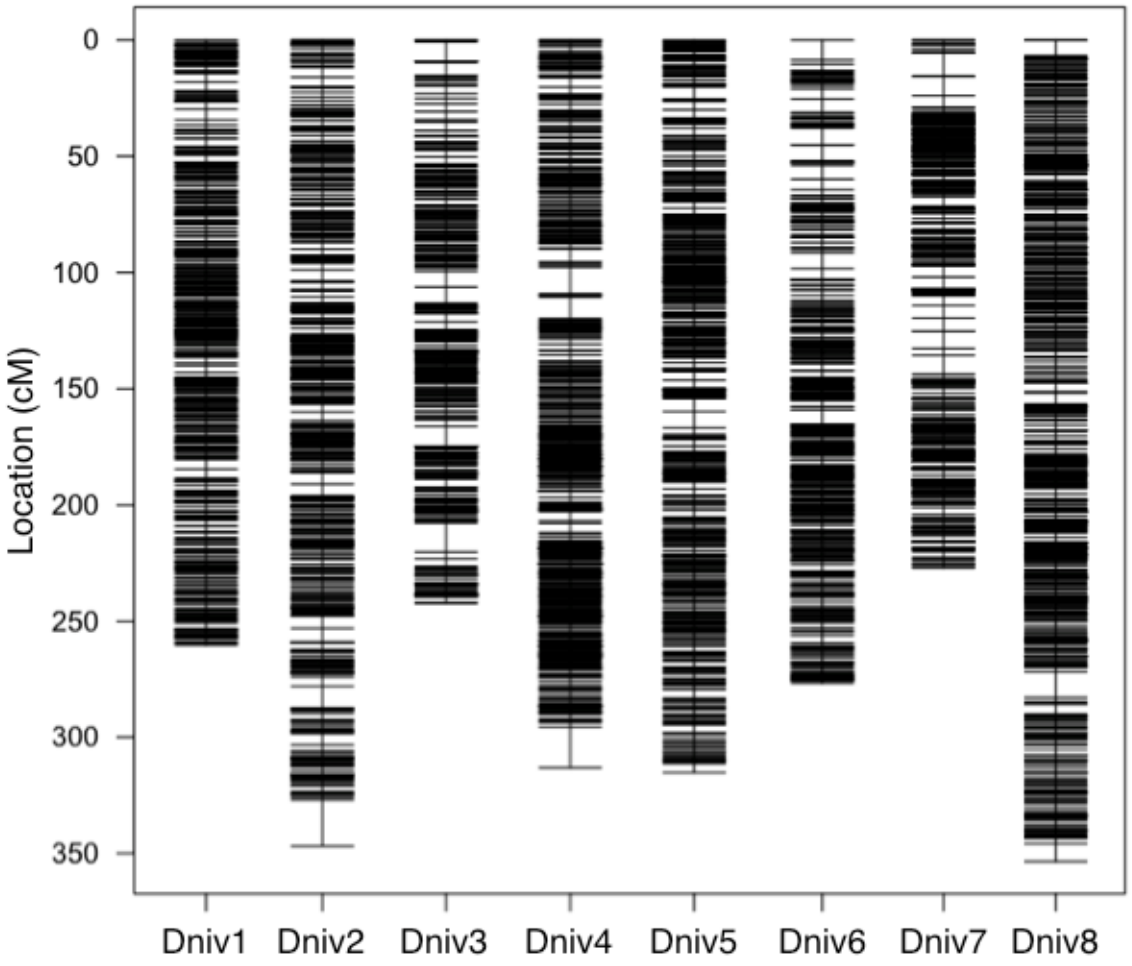
Supplementary Figure 2. Comparative chromosome painting (CCP) in *D. nivalis*

Supplementary Figure 3. Detailed CCP analysis of *D. nivalis* chromosome Dniv4.

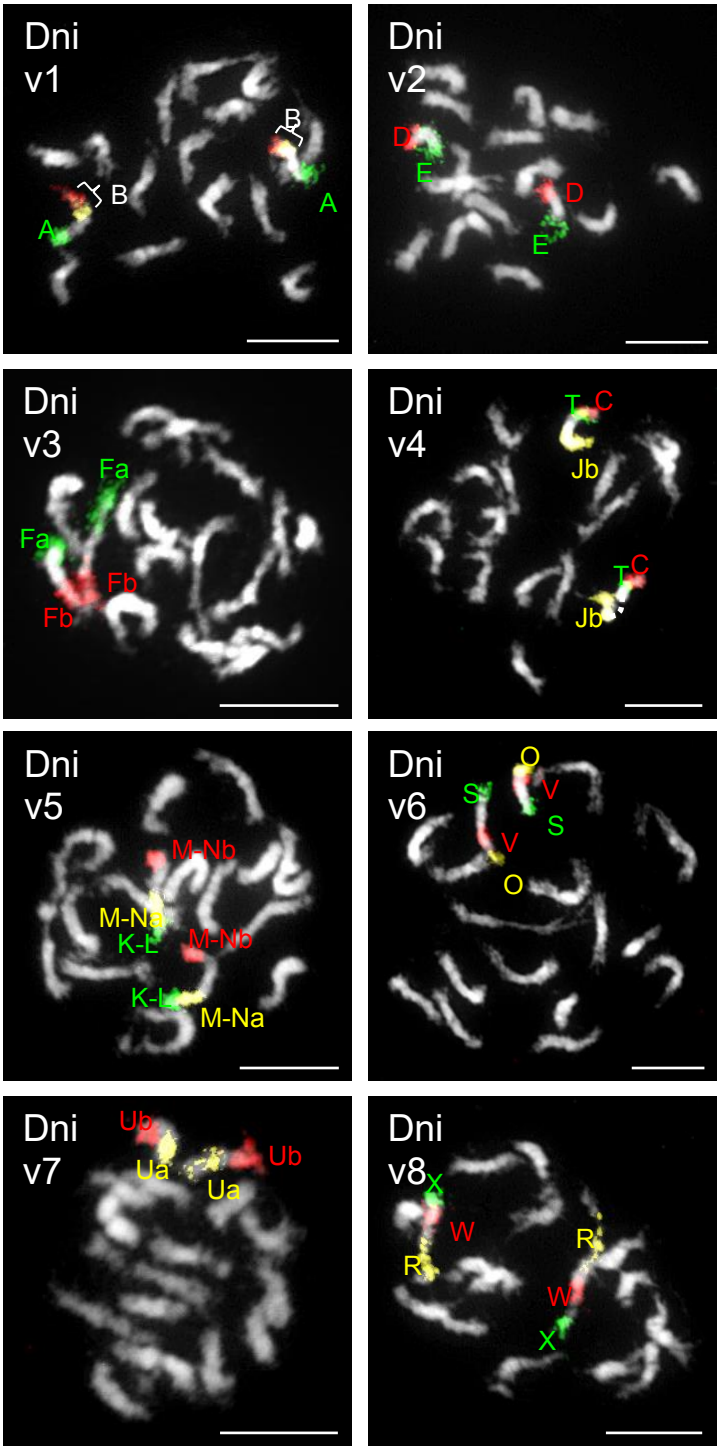
Supplementary Figure 4. Distribution of 25mer frequencies in the *D. nivalis* genome assembly

Supplementary Figure 5. Venn diagram comparing gene family membership in Brassicaceae Clade B species

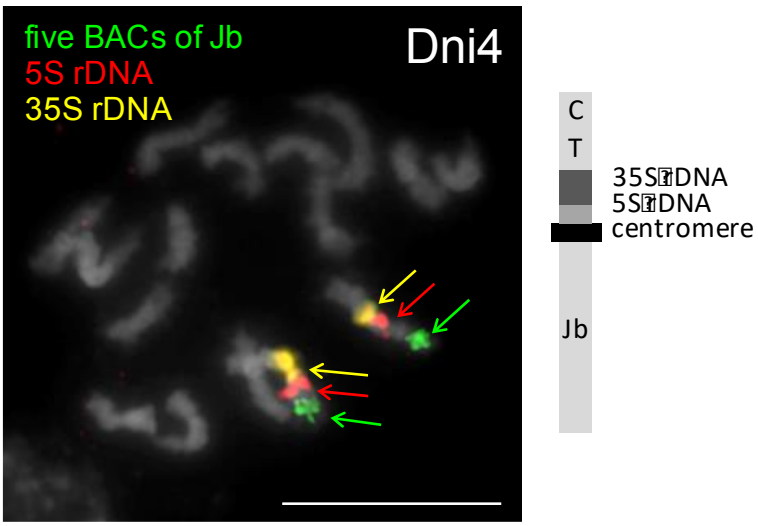
Supplementary Figures



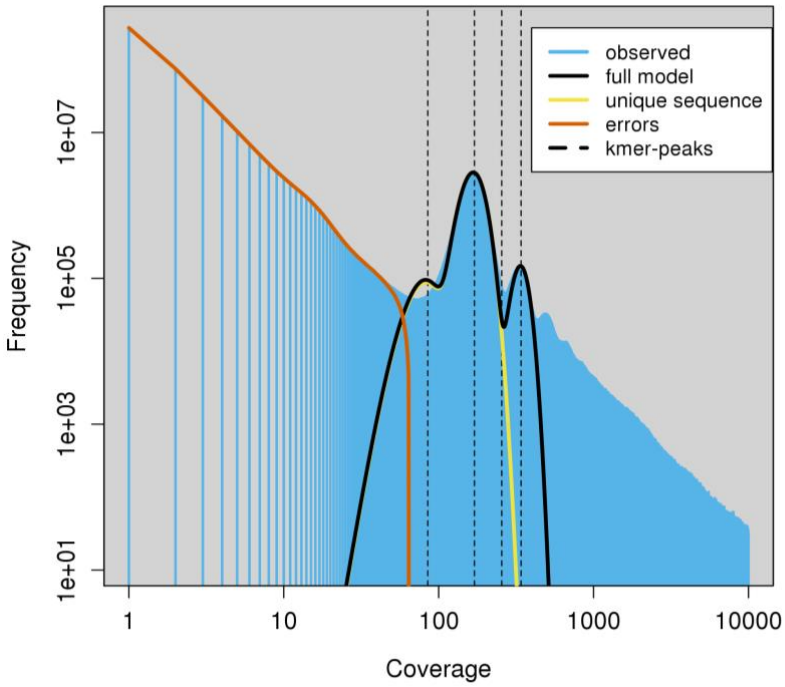
Supplementary Figure 1. Genetic linkage map of *D. nivalis*. Based on genotyping with 5,055 biallelic SNPs in 480 F_2 progeny resulting from self-pollination of an F_1 hybrid between *D. nivalis* accession 045-5 from Norway (maternal parent) and *D. nivalis* accession 008-7 from Alaska (paternal parent; this is the plant from which the reference genome assembly was produced; see methods). Numbering of chromosomes follows the final *D. nivalis* genome assembly (Figure 1). Map locations are provided in Supplementary Table 2.



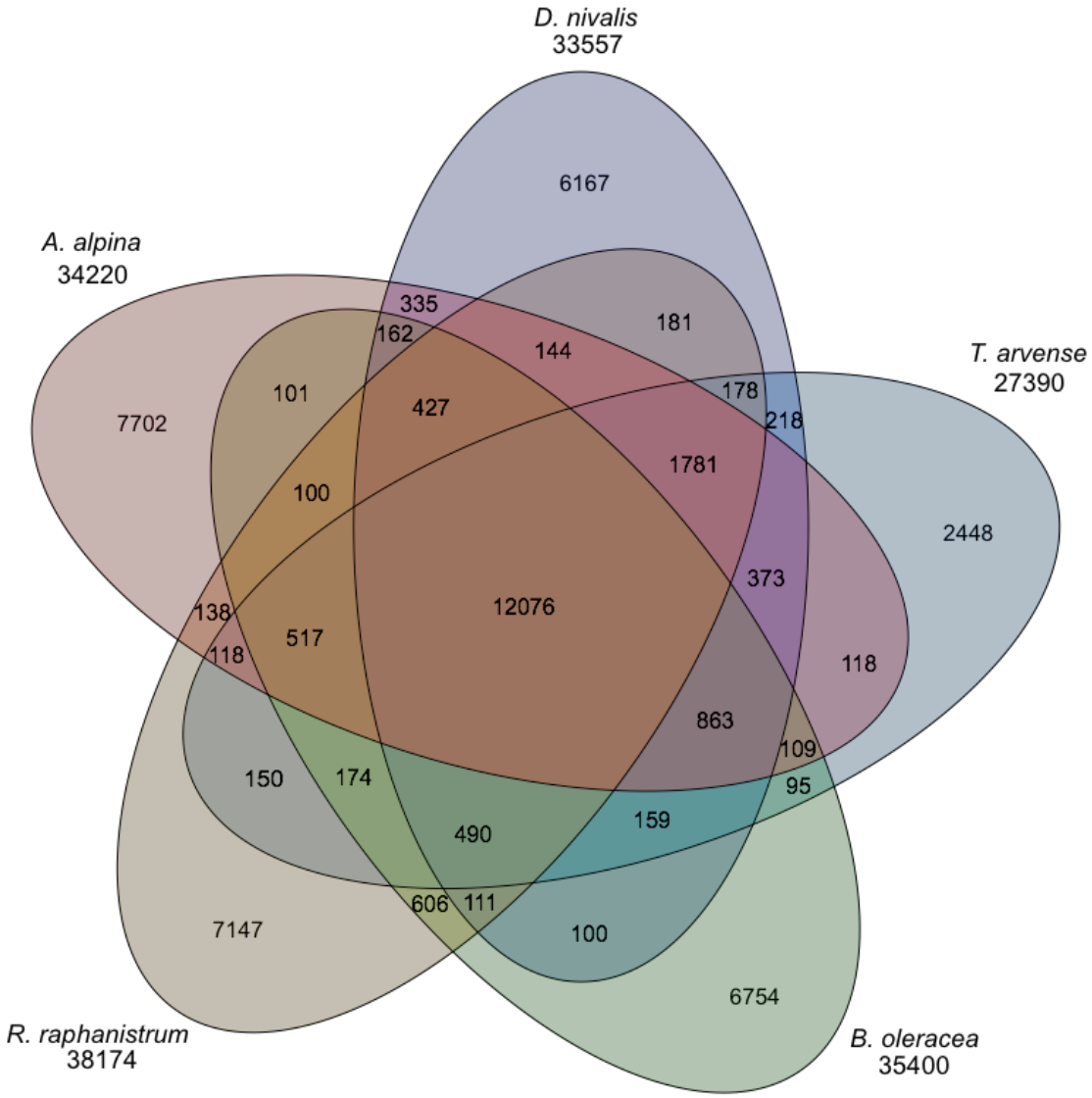
Supplementary Figure 2. Comparative chromosome painting (CCP) in *D. nivalis*. The structure of the eight *D. nivalis* chromosomes (Dniv1-Dniv8) reconstructed by *in situ* hybridization of probes for conserved Brassicaceae genomic blocks to mitotic prometaphase/metaphase chromosomes. Chromosomes were counterstained with DAPI. Scale bars 10 μm.



Supplementary Figure 3. Detailed CCP analysis of *D. nivalis* chromosome Dniv4.
Localization of rDNA loci is shown on mitotic chromosomes. FISH probes for 5S rDNA, 35S rDNA and five *A. thaliana* BAC clones corresponding to the genome block Jb revealed both rDNA loci adjacently positioned on the upper arm of Dniv4. A graphical representation of the structure of Dniv4 is shown to the right. Chromosomes were counterstained with DAPI. Scale bar: 10 µm.



Supplementary Figure 4. Distribution of 25mer frequencies from approximately 134 million reads of 250 bp length from the whole genome shotgun sequence dataset. A genome size of 278.484 Mb was estimated. Figure produced using GenomeScope (Vurture *et al.* 2017; http://qb.cshl.edu/genomescope/).



Supplementary Figure 5. **Venn diagram comparing gene family membership in Brassicaceae Clade B species.** Following the clade definition of Guo et al. (2017). The total annotated gene count for each species is shown below the species name.

Supplementary Methods

Protocol for ddRAD genotyping of F_2 mapping population. The digestion step was performed in 40 µl reactions containing 500 ng DNA, 1X CutSmart buffer, 10 U NsiI-HF (New England Biolabs [NEB], R3127) and 5 U MseI (NEB, R0525) in a thermal cycler at 37°C for 2 hrs. The DNA was cleaned using 1.2 volumes Ampure XP (Beckman Coulter) and eluted in 25 µl H_2O. Adapter ligation was performed in 30 µl reactions containing 1x T4 DNA ligase buffer (NEB), 400 U T4 DNA ligase (NEB, M0202), 2.66 µM indexed P1 and P2 adapter mix, and 24 µl digested DNA in a thermal cycler set to 25°C - 30 min, 65°C - 10 min and 4°C - hold. The samples to be sequenced together on an Illumina HiSeq lane were pooled, cleaned using 1.2 volumes Ampure XP and eluted in 240 µl 10 mM Tris-HCl. The libraries were amplified in a PCR reaction with 1x Q5 HiFi Master Mix (NEB, M0492), 0.5 µM of each primer, template DNA and nuclease-free water. We did not exceed the recommended DNA input amount stated by NEB for the Q5 polymerase, which is 1 µg template DNA per 50 µl PCR reaction. PCR conditions were 98°C – 30 sec, (98°C – 10 sec, 60°C – 15 sec, 72°C – 15 sec) x 8, 72°C – 2 min. The libraries were cleaned using 0.8 volumes Ampure XP. The DNA had to be cleaned up to three times to remove all short fragments (<200 bp).

References

Guo, X. *et al.* Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* **18,** 176 (2017).

Schranz, M. E., Lysak, M. A. & Mitchell-Olds, T. The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends in Plant Science* **11,** 535–542 (2006).

Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).