Transcriptome Analysis of Perennial Ryegrass (Lolium perenne L.) during an Early Stage Infection to Pink Snow Mould

anil kunapareddy

Norwegian University of Life Sciences
Department of Plant and Environmental Sciences
Master Thesis 60 credits 2013

Acknowledgements

The present master thesis was carried out at Department of Plant and Environmental sciences under the supervision of Professors Odd Arne Rognli, Anne Marte Tronsmo and Post Doctorates Åshild Ergon, Mallikarjuna Rao Kovi.

I would like to thank all the supervisors and co-supervisors for their supervision and support from the beginning until to end of my master thesis. I am very grateful to Åshild and Mallikarjuna Rao for their professional guidance throughout the whole work and always having the time to help me despite of their busy schedule. I am also thankful to Mohamed, PhD student and Øyvind for their help during green house work.

Finally, I would like to thank all the members of the group and special thanks to all the supervisors for giving me the oppurtunity to join ryegrass group.

Ås, December 2013.

Abbreviations

PAL	Phenylalanine ammonia lyase
NGS	Next generation sequencing
GA	Genome analyzer
PCR	Polymerase chain reaction
RT-PCR	Reverse transcriptase PCR
PDA	Potato dextrose agar
PDB	Potato dextrose broth
RIN	RNA integrity value
GO	Gene Ontology
DAG	Directed acyclic graphs
ANOVA	Analysis of variance
FDR	False discovery rate
CT	Cycle Threshold
ΔΔCT	Delta-Delta CT
RNA	Ribonucleic acid
PR-5	Pathogenesis related gene 5
EdgeR	Empirical analysis of digital gene expression data in R
HR-PCD	Hypersensitive response programmed cell death

Abstract

Microdochium nivale causes pink snow mould disease that damages forage grasses (especially perennial ryegrass) covered by snow, particularly in late winter. This makes the survival of perennial ryegrass (*Lolium perenne L.*), which is an economically important forage grass in Europe, difficult. Resistance mechanisms against pink snow mould disease is not fully understood. Thus, to gain a better understanding of resistance mechanisms in perennial ryegrass, we investigated variation in resistance among *L. perenne* genotypes, measured as relative regrowth after inoculation with several *M. nivale* isolates, and conducted a global transcriptome study of resistant and susceptible genotypes using an RNA sequencing approach. Ten genotypes, 8 from the Norwegian cultivar genotypes Fagerlin and 2 from the European cultivar Picaro, were used in this study. Clones of ten genotypes were inoculated with *M. nivale* isolates and incubated in darkness at 2°C for 5-12 weeks. Relative regrowth was calculated as the ratio of the regrowth of inoculated and non-inoculated plants after 5, 8 and 12 weeks of incubation. Based on relative regrowth one 'resistant' genotype (M) and one 'susceptible' genotype (F) were selected for transcriptome studies. Total RNA was extracted from the leaf blade tissue of plants exposed to three different treatments: **non-inoculated and non-incubated plants (non-incubated control), non-inoculated plants after 4 days of incubation (incubated control), and inoculated plants after 4 days of incubation (inoculated)**. There were two biological replicates of each genotype and treatment (totally 12 samples). cDNA libraries were prepared and paired-end sequencing was performed using Illumina Hiseq 2000. Before sequencing, qPCR was performed with the candidate gene *PR-5* (*thaumatin-like protein*), a gene which is expressed in response to pathogen attack. The results indicate that there is significant variation among M and F genotypes during 0 day and the 4th day. Higher transcript levels of *PR-5* were observed in both non-incubated control (0 day) and inoculated (4th day) plants of the 'resistant' M genotype than in the 'susceptible' F genotype. This variation in the expression of *PR-5* indicates that the two selected genotypes are different in respect to resistance against pink snow mould, and that they are suitable for being used for global RNA sequencing. *De-novo* assembly was done using the Trinity software platform. A total number of 188,355 and 261,978 assembled contigs with N50 values of 1,672 and 1,784 bp were generated from the "M" and "F" genotypes, respectively. Mapping was done by aligning the reads back to the assemble contigs using the Bowtie program and then transcript abundance was estimated by RSEM. Using EdgeR, differential expression analysis of read counts was done. A total number of 7,282 and 19,055 significantly differentially

expressed (DE) genes were detected between non-incubated control and inoculated (incubated at 2°C), 6,227 and 19,832 DE genes were detected between non-incubated control and non-inoculated plants, and 275 and 83 DE genes were detected between inoculated and non-inoculated plants of the “M” and “F” genotypes, respectively. In this study we mainly focused on gene expression in inoculated and non-inoculated incubated plants as they were treated under the same conditions, i.e. incubation in darkness. DE genes from the ‘resistant’ (M) genotype and from the ‘susceptible’ (F) genotype were blasted and annotated against the NCBI non-redundant database using Blast2go program. The annotation results gave us about various genes that are differentially expressed at various stress factors. Further detailed studies on these genes will help us to understand the interactions between perennial ryegrass plants and *M. nivale* and also provied information that can be used to develop cultivars with improved resistance against pink snow mould by studying expressions of various defence related genes.

Sammendrag

Snømugg (*Microdochium nivale*) forårsaker snømugg som skader gras, spesielt flerårig raigras (*Lolium perenne* L.) når det er snødekke seint på våren. Dette gjør overvintring av flerårig raigras, som er en økonomisk svært viktig fôrgrasart i Europa, usikker. Mekanismer som bidrar til resistens mot snømuggsoppen er lite kjent. For å få en bedre forståelse av resistens mot snømugg i flerårig raigras har vi undersøkt resistensen hos genotyper av *L. perenne*, målt som relativ gjenvektsevne etter inokulering med flere isolater *M. nivale*, og studert global transkripsjon hos en 'resistent' og en 'mottakelig' genotype etter inokulering med et aggressivt isolat ved bruk av RNA sekvensering. Ti genotyper, 8 fra den norske sorten 'Fagerlin' og 2 fra den europeiske sorten 'Picaro', ble brukt i dette studiet. Kloner av de 10 genotypene ble inokulert med *M. nivale* isolater og inkuberrt i mørke ved 2°C i 5-12 uker. Relativ gjenvekst ble beregnet som forholdet mellom gjenvekst hos inokulerte og ikke-inokulerte planter etter 5, 8 og 12 uker inkubering. Basert på relativ gjenvekst ble en 'resistant' genotype (M) og en 'mottakelig' genotype (F) valgt ut for transkripsjonsstudier. Totalt RNA ble ekstrahert fra bladmateriale hos planter som hadde fått følgende 3 behandlinger: ikke-inokulert og ikke-inkubert planter (ikke-inkubert kontroll); inokulerte planter etter 4 dagers inkubasjon (inkubert kontroll); og inokulerte planter etter 4 dagers inkubering (inokulert). Det var to biologiske replikater av hver genotype og behandling, total 12 prøver. cDNA libraries ble laget og 'paired-end' sekvensering utført ved bruk av Illumina HiSeq2000. Før sekvensering utførte vi også kvantitativ real-time PCR (qRT-PCR) på kandidatgenet *PR-5* (*thaumatin-like protein*), et gen som er uttrykt soom response på patogenangrep. Resultatene indikerte at det var signifikant variasjon mellom genotype M og F ved starten av forsøket (dag 0) og ved dag 4. Høyere transkripsjonsnivå av *PR-5* ble observert i den 'resistente' genotypen M enn i den 'mottakelige' genotypen F både i ikke-inkubert kontroll (dag 0) og i inokulerte og inkuberte planter ved dag 4. Denne variasjonen i uttrykk av *PR-5* indikerer at de to utvalgte genotypene er forskjellig med hensyn på respons på tidlig angrep av snømugg, og at de var egnet til å studere global transkripsjon vha. RNA sekvensering. *De-novo* assembly ble utført ved bruk av bioinformatikkpakken Trinity. Totalt antall contiger som ble satt sammen var 188,355 og 261,978 for henholdsvis genotype M og F, med N50 verdier på henholdsvis 1,672 og 1,784 bp. Differensiell ekspresjon (antall reads) ble estimert ved bruk av EdgeR. Total ble det funnet 7,282 og 19,055 signifikant differensielt uttrykte gener (DEG) mellom ikke-inkubert kontroll og inokulert (inkubert ved 2°C i mørke), 6,227 og 19832 DEG mellom ikke-inkubert kontroll og

ikke-inokulerte, inkuberte planter, og 275 og 83 DEG mellom inokulerte, inkuberte og ikke-inokulerte, inkuberte planter av henholdsvis genotype M og F. I dette studiet fokuserte vi i hovedsak på sammenlikning mellom genekspresjon i inokulerte, inkuberte og ikke-inokulerte, inkuberte planter siden disse ellers hadde like forhold, dvs. de var begge inkubert i mørke. DEG gener fra den 'resistente' genotypen M (275) og fra den 'mottakellige' genotypen F ble 'blasta' og annotert mot NCBI 'non-redundant' database ved bruk av programmet Blast2GO. Annoteringen ga oss informasjon om funksjonen til de ulike genene vi hadde identifisert som differensielt uttrykte. Framtidige detaljerte studier av disse genene vil hjelpe oss til å bedre forstå interaksjoner mellom flerårig raigrasplanter og *M. nivale*, og også gi informasjon som kan benyttes til utvikle nye sorter med forbedret resistens mot snømugg. Detaljerte studier av ekspresjon av ulike gener knyttet til motstandsevne mot snømugginfeksjon vil bidra til dette.

Table of Contents

1. Introduction

1.1 Perennial ryegrass (*Lolium perenne* L.)

Perennial ryegrass (*Lolium perenne* L.,) belongs to the Poaceae (grass and cereal) family and is a diploid species (2n=2x=14) (Shinozuka et al. 2012). It is native to Europe, Asia and Northern Africa and is the most important forage grass in the temperate regions of the world. Perennial ryegrass is widely used as a forage crop because of its high forage quality and yield (Comont et al. 2013; Yu et al. 2013). Out of 52 million ha of grasslands available in Europe, 23% is cultivated with *Lolium* species with perennial ryegrass being the most widespread species (Humphreys et al. 2010).

Perennial ryegrass is used for forage primarily in the coastal Northwest, the Midwest and Northeast of United States. Because of its high palatability, digestibility it is the preferred forage grass species for dairy and sheep forage systems (Hannaway et al. 1999; Lee et al. 2010). It is primarily grown for pasture and silage in USA. Also, it is an important turfgrass species because of its fast establishment rate and good tolerance. However, perennial ryegrass has low winter-hardiness (ability to survive low temperature) compared to other perennial cool season grasses (Hannaway et al. 1999). Because of these characteristics of perennial ryegrass researchers choice of interest has been in developing new cultivars that show resistance to both biotic and abiotic stress factors in different genotypes, as winter survival and snow mould infection of *L. perenne* are the major problematic factors reducing the grass production. A study conducted on variation in resistance to different species and varieties of grasses showed that low temperature and good light conditions promotes resistance to freezing, ice encasement and snow mould fungi especially in northern and alpine regions (Hofgaard et al. 2003). Thus, these studies on resistance to various winter abiotic stress factors help breeders to improve the quality of forage grasses and to understand more about resistance to fungal diseases and pests for the development of new resistant cultivars (Ruttink et al. 2013).

1.1.1 Effects of various stress factors on non-acclimated turf grasses

Plants have evolved to live in an environment where they are exposed to different kinds of environments and attacked by various kinds of pathogens. To protect themselves they have developed special mechanisms to sense these stresses and adapt themselves (Atkinson & Urwin 2012). Abiotic stresses such as cold, salinity, drought, heat, nutrient stress and biotic stress like

attack by pests, fungi, bacteria leads to physiological, morphological, biochemical changes that has large effect on agriculture by reducing the yields to >50% for major crops (Wang et al. 2003). Current climate predictions shows that the average surface temperatures on earth will rise by 3ºC-5ºC in the next 100 years which increases drought, flooding and heat incidences causing yield reductions in mid-continental regions like Central-Europe and Central Africa (Mittler & Blumwald 2010). Also, some studies have shown that abiotic stresses like salt-induced osmotic stress (drought stress) were directly correlated with resistance to powdery mildew in barley and to the fungus *Botrytis cinerea* in tomato (Achuo et al. 2006; Wiese et al. 2004). In temperate regions where Turf grasses are major crops, they could not escape these stresses thereby giving reduced yields. Also perennial ryegrass is a type of cool-season turf grass susceptible to low-temperature kill. Inspite of poor low temperature tolerance, its popularity and use continues because of newly improved cultivars (Ebdon et al. 2002). So, the non-hardened grasses become hardened to various stresses by obtaining freezing tolerance when exposed to low temperatures during early winter/autumn.

1.2 Snow moulds

Snow mould fungi are psychrophilic or psychrotrophic fungi that attack the plants at low temperature under snow cover. During winter, the plants become less resistant to disease because of depletion of carbohydrate reserves which makes the plant weak. As a result of depletion it is being attacked by the snow mould fungi (Chang et al. 2006b). These fungi are taxonomically diverse and several species can thrive on plants during the growing season in the summer. Matsumoto (2009) classified them into obligate and facultative snow moulds. Obligate snow moulds grow in winter with or without snow, facultative snow moulds have higher optimum temperature than obligate snow moulds and can damage growing plants during the growing season in summer (Matsumoto 2009). The environmental conditions like depth, duration of snow cover and availability of disease inoculum are important factors influencing snow mould damage to forage crops and winter cereals (Gaudet et al. 1999). High humidity, constant temperatures at or below 0° C and darkness under snow cover are highly favourable to snow mould fungi, that have adapted to grow at low temperatures between 1° C and -8° C (Gaudet et al. 2000). The word “Snow moulds” is a general name containing diverse fungi belonging to various taxa like oomycetes, ascomycetes, basidiomycetes and major snow mould fungal species causing snow moulds are *Pythium iwayamai*, *Microdochium nivale*, *Sclerotinia borealis*, *Coprinus psychromorbidus*. However, *Typhula incarnata* and *T. ishikariensis* have been widely studied in

the northern hemisphere from the aspect of plant protection (Hoshino et al. 2009). At low temperatures the growth of *Typhula* spp. is facilitated by lipolytic enzyme activity and by the production of antifreeze-like proteins while *M.nivale* cannot grow at subfreezing temperatures *in vitro* (Espevig 2011). Generally snow moulds can tolerate low temperature where other antagonists and competitors of snow mould cannot grow giving them an advantage to dominate the nutrient-rich and too weakened plant tissues. Usually fungi adapted to low temperature can be found in polar, temperate and mountain regions all over the world but how many species of snow mould are growing under snow? (Hsiang et al. 1999). The effect of snow mould damage depends on environmental factors like temperature, humidity and snow cover. Prolonged periods of snow cover and low temperature decrease the metabolic rate of host plants increasing the disease severity by a pathogen (Chang et al. 2006a).

1.2.1 *Microdochium nivale*

In temperate and cooler regions, snow mould disease caused by *Microdochium nivale* (Fr) Samuels & Hallet is one of the important diseases causing winter damage to grasses. It is a serious, opportunistic species causing pink snow mould on turf and forage grasses during low temperatures (Tronsmo et al. 2001). Apart from snow mould, this fungus also causes leaf blotch in oat, foot rot and head blight in cereals (Hofgaard et al. 2006a). This fungus can be saprophytic growing on dead organic matter. Growth of mycelia from soil penetrates lower stem parts of plant causing infection (Posselt & Altpeter 1994). Pronczuk and Messyasz (1991) showed that inoculation of *L. perenne* with conidia did not give any symptoms while mycelial inoculation provoked severe disease (Prończuk & Messyasz 1991; Tronsmo et al. 2001). The nomenclature of *M. nivale* has been changed several times, before it was called as *Fusarium nivale* but because of lack of conidia foot cell it was called as *Gerlachia* and finally called as *Microdochium nivale*. Infection in grasses is caused by three types of propagules which include mycelia, conidia and ascospores. *M. nivale* is categorized into two subgroups based on conidial morphology. Gam and Muller (1980) illustrated that *M. nivale var. majus* differs from *M. nivale var. nivale* in having larger conidia which was later confirmed by Gerlach and Nirenberg (Gams & Müller 1980; Lees et al. 1995). Molecular studies using restriction digests of the internal transcribed spacer region of ribosomal DNA and RAPD analysis have separated the isolates of *M. nivale* into two sub-groups. *Microdochium nivale* and *M. majus* targets different hosts and the variation of pathogenicity is found between different isolates (Hofgaard et al. 2006a). Hyphae of

M. nivale does not show any significant morphological changes when grown at temperatures ranging from 4 to 20°C (Tronsmo et al. 2001).

Fig.1. Pink snow mould disease on perennial ryegrass (Photo: Odd Arne Rognli).

1.3 Cold Acclimation

Exposure of plant to cold triggers a process known as cold-acclimation or hardening (Christie et al. 1994) which induces the expression of cold-responsive genes in response to low and non-freezing temperatures (Rudi et al. 2011). During this process the plant undergoes numerous physiological and bio-chemical changes which are essential to acquire snow mould resistance and frost tolerance. These physiological and bio-chemical processes include reduction in growth rate, reduction of tissue water content, and changes in membrane lipid composition (Gaudet et al. 2000). The activation of phenylalanine ammonia-lyase(PAL), which is a key enzyme for the phenyl-propanoid pathway, results in accumulation of phenolic compounds indicating that it is a physiological representation of winter stress. Pociecha and Płażek (2010) performed an experiment to know whether the heat production rate induced during pre-hardening and PAL activity induced during cold acclimation process are related to resistance to snow mould disease in forage grasses. They tested in the four forage grasses Festulolium, meadow fescue, tall fescue and Italian ryegrass and showed that accumulation of phenolic compounds and heat emission

during metabolic process is related to resistance to snow mould disease (Pociecha & Płażek 2010). Even though it was shown that plants acquire resistance through heat emission process and PAL activity, many studies also shown that carbohydrate storage particularly fructans, and the expression of PR-genes help in development of snow mould resistance during winter (Gaudet et al. 2000).

Some of the PR-proteins have an anti freeze activity which is assumed to play an important role in frost tolerance (Hon et al. 1994). Studying about PR-proteins having antifreeze activity makes sensible, as they are accumulated during cold acclimation which are indirectly useful for snow mould resistance.

1.4 Anti-Freeze proteins similarity with pathogen-induced PR-proteins

Hon and Griffith have shown that ice formation takes place within the tissues of freezing-tolerant plants to survive the temperatures below zero and this formation takes place especially in intercellular spaces but not inside the cells as it could be lethal to the organism by damaging cell membranes (Hon et al. 1995). They also showed that the specific proteins called antifreeze proteins, which accumulate during cold acclimation are similar to the members of pathogenesis-related proteins in their amino-terminal sequences and enzyme activity assays (Hon et al. 1995). Other researchers suggested that the cold induced PR proteins may be the isoforms of PR-proteins produced during pathogen infection which has antifreeze activity that function to modify the growth of ice by attaching to the surface of ice crystals and also prevents the crystallization which takes place during the fluctuation of temperatures in subzero range (Hiilovaara-Teijo et al. 1999). Here comes the concept of cross-adaptation where one kind of stress provides tolerance to other stress factors indicating that cold-induced PR-proteins are involved both in freezing tolerance and resistance to snow moulds as the levels of cold induced PR-proteins accumulate when plants been subjected to cold acclimation (Hiilovaara-Teijo et al. 1999).

1.5 Hexose-sensing activation of disease resistance

Sugars are the most important substrates for growth and development of the organism. Sugars are found to be important in respiration metabolism, cell cycles and physiology as regulatory molecules. Sugar sensing is defined as the interaction between a sugar molecule and a sensor protein giving the signal which initiates the signalling pathways that result in cellular responses

(Jang et al. 1997; Rolland et al. 2002; Smeekens 2000). Recently plant hexokinase has been shown to be involved in sugar sensing and signalling having a dual function with both catalytic and regulatory functions (Xiao et al. 2000). Gaudet et al. (1999) gave a hypothetical model for hexose-sensing activation of disease resistance in hardening winter cereals. According to this model hexokinase signal transduction is induced by the hexose sugars effecting the expression of many genes. This hexokinase signal transduction in turn activates the plant defence resistance genes like PR-proteins and thaumatin like proteins (Gaudet et al. 1999).

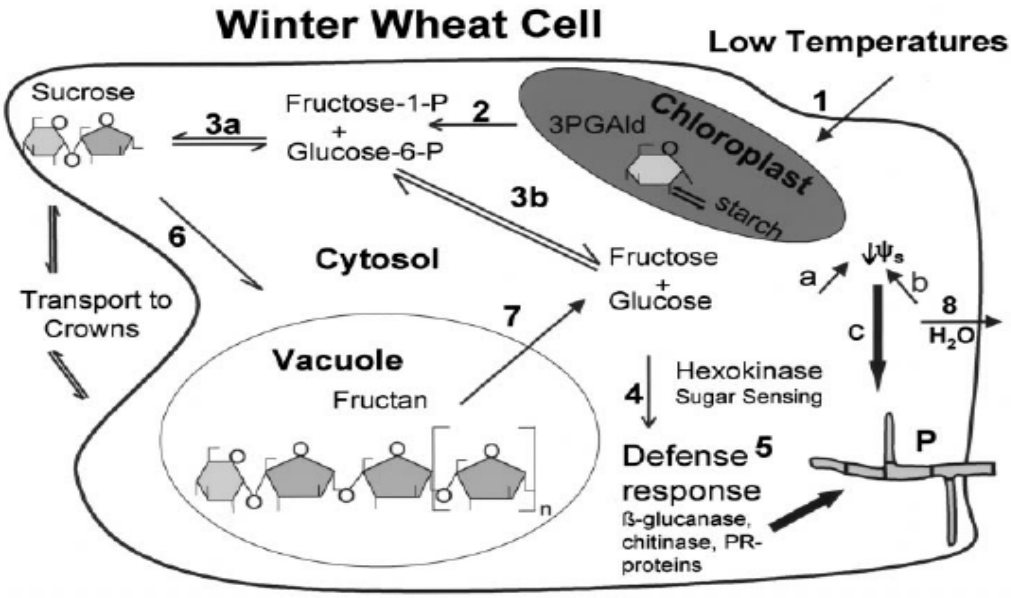

Fig.2. Hexose-sensing pathway model of disease resistace in hardening cereals(Gaudet et al. 1999).

Researchers also proposed other possible mechanism in limiting the growth of snow mould fungi which in turns helps in snow mould resistance by decreasing the water potentials of the plant as fungal pathogens have a charateristic range of water potentials for growth (Bruehl & Cunfer 1971).

Thus, for studying about the genes expressed during pink snow mould attack we have utilized an RNA sequencing approach.

1.6 Transcriptome Studies

Development of strategies for disease control relies on understanding the responses of hosts (plants) to various infections caused by bacteria, fungus or virus. As plant responses are complex and include various physiological processes, transcriptome studies are important in

understanding the plant responses (Lu et al. 2012). Earlier microarrays have been used to study the differential gene expression changes with respect to specific conditions, but because of less information provided by microarrays, mostly next-generation sequencing (NGS) methods are being used (Soneson & Delorenzi 2013). RNA-sequencing (RNA-Seq) is a whole transcriptome sequencing method that can measure gene expression at the transcriptional level thus giving lots of information about non-coding regions, identified genes and determines the structure of transcripts (Lu et al. 2012). This transcriptome analysis has two approaches, align-than-assemble and assemble-then align. Align-than-assemble completely depends on the reference genome while assemble then align uses de novo assembly as the reference genome is not available. Fig.3 gives the overview of the two approaches and There is no clear recommendation which method is most appropriate for any nonmodel species (Ward et al. 2012).

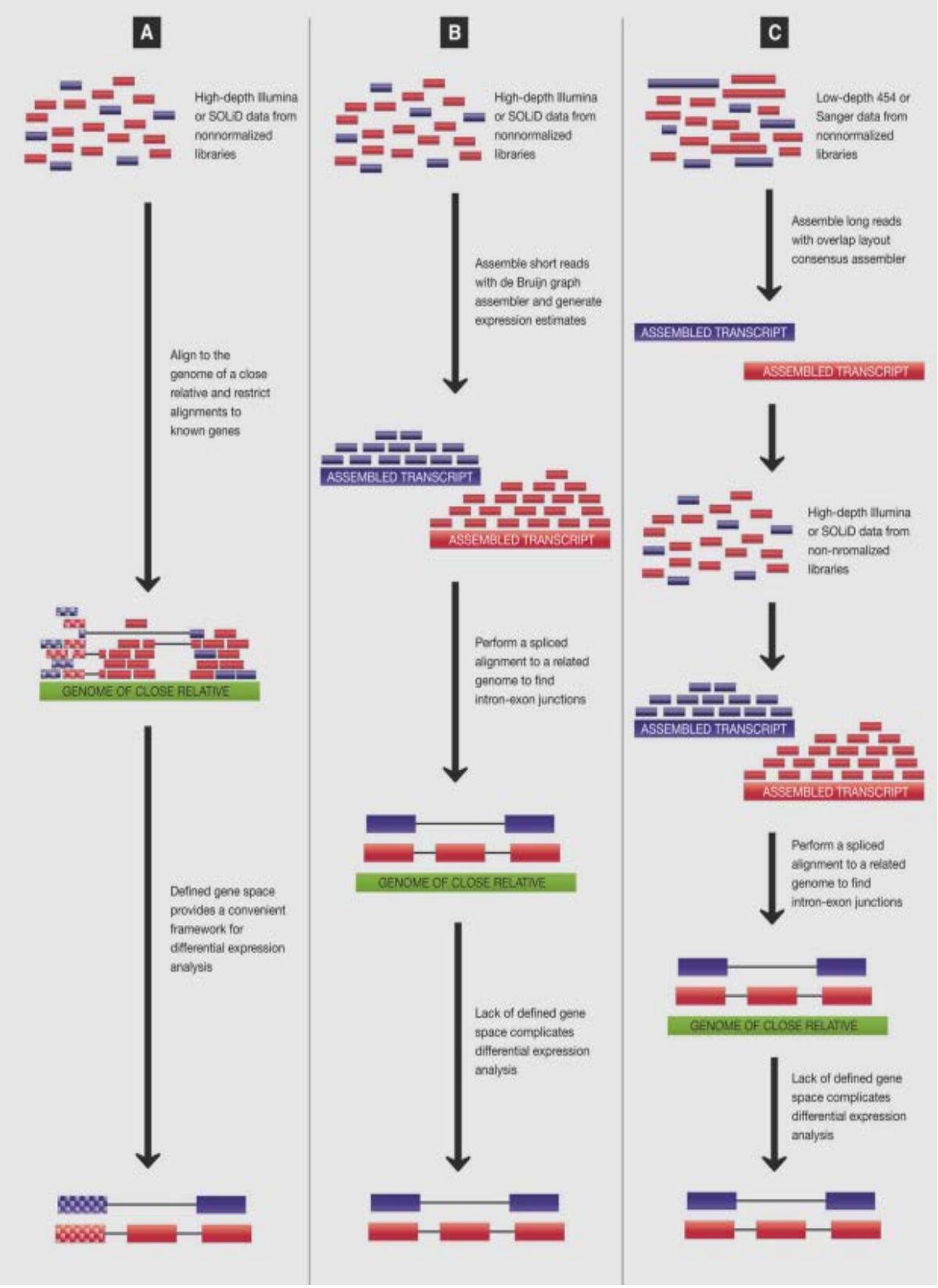

Fig.3. A : Align-then- assemble approach when reference genome is available; B : Assemble-then-align approach by de novo assembly; C : Assemble first approach by low-depth 454 (Ward et al. 2012).

1.6.1 Sequencing methodologies

Since the beginning of the cultivation of plants, plant breeding has been successful in developing the modern cultivars using conventional genetic tools and approaches leading to new genomics-based plant breeding. One of the main basis of genomic based plant breeding is the development of high-throughput DNA sequencing technologies which are collectively known as next generation sequencing (NGS) methods (Pérez-de-Castro et al. 2012).

First generation sequencing (Sanger sequencing) has been using for the past 30 years but came to an end with the introduction of Roche/454 platform in 2005, Solexa/Illumina system in 2006, Applied Biosystems SOLID system in 2007 and Ion Torrent system in 2010 (Ozsolak 2012). These technologies obtained many features from Sanger sequencing like the use of enzymes for synthesis, fluorescent detection etc. As they are continually developing, it is hard to refer the old and new sequencing methods as belonging to certain generations (Stranneheim & Lundeberg 2012). Of all the three technologies Illumina sequencing is the most popular as it is less expensive and give more reads when compared to the Roche/454 platform.

1.6.2 Illumina GA/Hiseq System

Solexa introduced the Genome Analyzer(GA) in 2006 which was bought by Illumina in 2007. It works on the principle of sequencing-by-synthesis in which all the four nucleotides are added to the single stranded DNA fragments with the adapters on both ends. These library fragments are attached to the surface of flow cells which consists of 8 separate lanes. The surface of a flow cell has dense adapters complementary to the specific adapters that are ligated to the library fragments which are then amplified to produce multiple DNA copies. These amplified clusters or multiple copies are provided with polymerase and four nucleotide bases which carry a base-unique fluorescent label and 3’-OH group that is chemically blocked, making each base incorporation a unique event. During incorporation of a nucleotide, the nucleotide with fluorescent dye is excited by laser so that the light is emitted from the fluorescent dye which is detected and recorded by the CCD camera before incorporation of the next nucleotide. At the end of sequencing step the sequence of each cluster is subjected to quality filtering and eliminates the low-quality reads. This sequencer can yield an average of approximately 40-50 million reads per run which takes around 4 days (Ansorge 2009; Mardis 2008a).

There was an upgrade of the Genome analyser introduced by Illumina in 2008 which tripled the output compared to the previous instrument. The sequencer produces about 1.5GB of single-read

data per run, at least 3Gb of data in a paired-end run and the run time for a paired end was reduced to 4 days (Ansorge 2009).

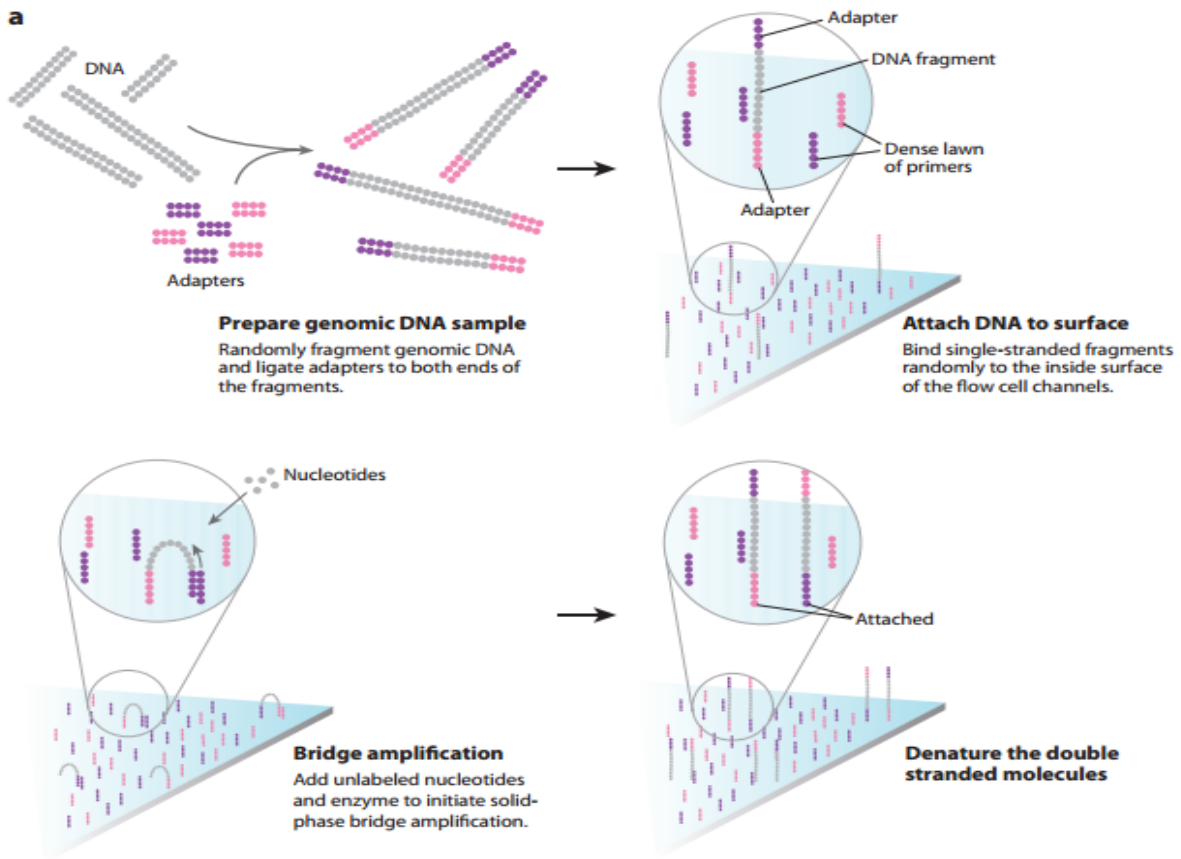

Fig.4 Illumina sequencing workflow (Mardis 2008b) .

The above described sequencing approach is the most popular method that have been used in all the major fields which provided routes for novel discoveries. For instance, in the plant science field all the three technologies are used in breeding applications and crop genetics. Of these three technologies Roche/454 give longer sequence reads compared with Solexa and AB SOLiD, but the output data obtained per run is lower. Also is it very expensive when compared to other technologies (Varshney et al. 2009).

1.7 RNA Sequencing

Microarrays had been the choice for gene expression studies during past years, but the genome-wide sequencing made the biologists to leave microarrays and use the next-generation sequencing because of improvements in efficiency, quality and cost (Bullard et al. 2010). In the past two decades automated Sangers method is treated as a first generation technology and the new technologies are considered as Next Generation Sequencing(NGS) (Metzker 2009). Next-generation sequencing (NGS) have become a revolutionary tool creating outstanding possibilities for understanding the complex eukaryotic transcriptomes. The most widely used next–generation sequencing platforms are Roche 454 Life Sciences, the Illumina Genome Analyzer, and Applied Biosystems SOLiD. These technologies are able to handle the *de novo* sequencing of large genomes, revealing individual genome differences within the same species, and can create genome-wide profiles of epigenetic modifications. These technologies are different from the traditional hybridization-based approaches(microarrays), referred to as RNA-Seq, has many applications like identifying transcript sequence polymorphisms and novel trans-splicing and splice isoforms. It can also quantify gene expression as the number of mapped reads to the given transcript gives an estimation of the level of expression (Tarazona et al. 2011; Xu et al. 2011).

1.8 RNA sequencing data analysis

We received the raw sequencing data from the GATC company in Germany. The initial quality check of the raw sequencing data was performed using the software FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Though there are many bioinformatics tools available, the usage of software tools depends on the researchers aim. The data analysis starts with the transcriptome assembly.

1.8.1 *De novo* transcriptome Assembly

De novo transcriptome assembly is performed for those species which does not have sequenced genomes or reference genomes. Sequencing of such species are carried out by using NGS technologies which produce millions of reads that are to be analysed further.

The first data analysis step of transcriptome assembly will be the mapping of the reads to reference genomes or transcriptome databases. However, many species have not been sequenced, so reference genomes are lacking. As a result *de novo* transcriptome assembly has become an important approach for species that are not sequenced and various softwares have been

developed for *de novo* assembly. Software packages used are Velvet, ABySS, Trans-ABySS and Trinity which use the de Bruijn graphs algorithms to assemble short reads into contigs which are then assembled into transcripts (Chen et al. 2011). Trinity consists of three individual software tools: Inchworm, Chrysalis, and butterfly which are used to produce transcripts. Inchworm (Fig. 5) cuts the reads into k-mers (which are short nucleotide bases around 25) and assembles the most frequent k-mer into unique sequences of transcripts or contigs that extend in both directions of k-mer until it cannot be elongated anymore and then report the linear contig. This process of extension is continued with the next frequently or next abundantly available k-mer until all the k-mers has been completely utilised. Then all the contigs are collected by chrysalis into sets of clusters or components which are acquired from alternative isoforms or paralogous genes and then the de Bruijn graphs are constructed for each cluster. Each component represents the set of genes that share sequences in common. Butterfly then process the de Bruijn graphs and reconstructs the full-length, linear transcripts for spliced isoforms and paralogous genes (Grabherr et al. 2011; Iyer & Chinnaiyan 2011).

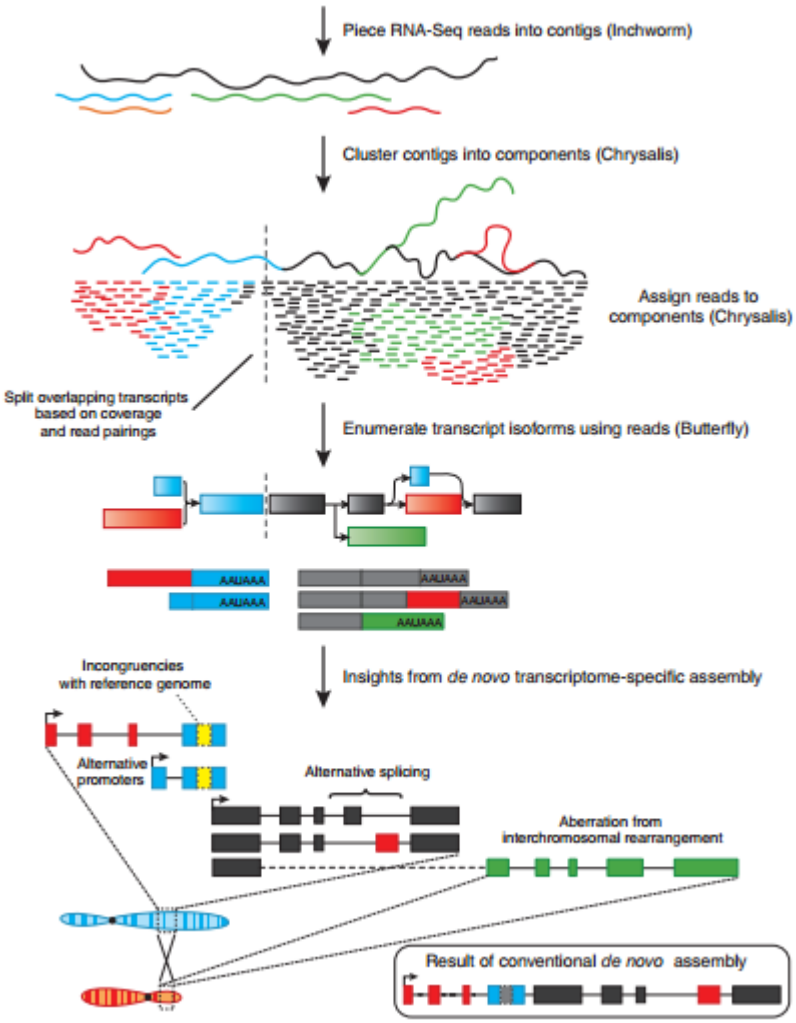

Fig.5 The process of constructing the *de novo* transcriptome assembly by Trinity software (Iyer and Chinnaiyan 2011).

After assembly of transcripts the reads are mapped back by using Bowtie program which is integrated with in the Trinity software (Langmead et al. 2009) and then use of RSEM for abundance transcripts (Li & Dewey 2011) and finally edgeR produces plots with differentially expressed genes (Robinson & Smyth 2007).

1.9 Advantages and Disadvantages of RNA Sequencing

Unlike hybridization-based approaches, RNA-Seq is not limited to detecting transcripts that correspond to the existing genomic sequence. For instance, it has been used for sequencing the transcriptome of Arabidopsis (Meyers et al. 2004). This makes RNA-Seq attractive for non-model organisms. It also discloses the particular location of transcription boundaries, to a single-base resolution and also short reads from RNA-Seq gives information how the exons are

connected and longer reads reveals the connection between multiple exons. These factors make RNA-Seq useful for studying complex transcriptomes, revealing sequence variations in transcribed regions. It has been shown to be highly accurate for quantifying expression levels, as determined using qPCR, and also shown high levels of reproducibility for both technical and biological replicates. It also has a wide range of expression levels over which transcripts can be detected (Wang et al. 2009). Depending on the availability of a reference genome transcriptome assembly strategies falls into three categories: a reference-based strategy, *de novo* strategy or combined strategy that merges the two. The reference-based strategy has several advantages and this strategy can transform a large assemblies (millions of reads) into smaller assemblies (thousands of reads or less) and more importantly this method is very sensitive and can assemble transcripts of low abundance. This strategy tends to generate longer UTRs, which usually have a low sequencing coverage. Owing to the high sensitivity of this approach, it allows to discover novel transcripts. Compared to the reference-based strategy, *de novo* transcriptome assembly does not depend on a reference genome; it can provide an initial set of transcripts for organisms which do not have reference genomes, allowing for RNA-Seq expression studies. Sometimes *de novo* assembly is performed even though a reference genome is available because it can recover the transcripts that are missing from genome assembly, or it can detect transcripts from an unknown exogenous source. Trans-spliced transcripts, originating from chromosomal rearrangements can be assembled by the *de novo* approach. It does not depend on the correct alignment of reads to known splice sites as required for reference based assemblers. A combined approach strategy starts by assembling the data using the reference genome, followed by *de novo* assembly of reads that fail to map to the reference genome.

Apart from advantages in reference based strategies there are few drawbacks. The success of the reference-based strategy is dependent on the quality of the reference genome used. Many genome assemblies, except for a few model organisms, contain many mis-assemblies and large genomic deletions. Reference-based strategies cannot assemble trans-spliced genes. Reference-based assembler can reconstruct full length transcripts with <10x sequencing coverage where as *de novo* assembly requires more than 30 X coverage for same transcript assembly (Martin & Wang 2011).

Aims of the study

During autumn most of the perennial ryegrass is infected with *M. nivale* causing snow mould disease. As a result there is huge economical loss every year. Thus developing strategies for developing cultivars to snow mould disease have become important in temperate regions. So, studying the transcriptional responses of perennial ryegrass to infection by *M. nivale* could contribute to the development of disease resistant cultivars and to improved productivity. Therefore the aims of this study were i) to select resistant and susceptible genotypes basing on relative regrowth after inoculation of non-hardened plants with pink snow mould; ii) to characterize transcriptional changes in pathogenesis related candidate genes (PR-5 and PR-3) by measuring relative expression levels of candidate genes using real-time quantitative PCR (RT-qPCR) (Zhang et al. 2009); and iii) process and Analyze RNA-Seq data obtained from sequencing samples collected from plants of genotypes with variable degree of resistance subjected to an inoculation experiment simulating pink snow mould attack under snow cover.

2. Materials and methods

2.1 Plant Materials and growth conditions

Non-hardened clonal plants of perennial ryegrass (*Lolium perenne* L.) were in the experiment called 'snow mould resistance test' in which two *L. perenne* cultivars varieties were used, namely 'Fagerlin' (owner Graminor, Norway) and 'Picaro' (owner Eurograss, Germany). A total number of 10 genotypes were selected of which 8 genotypes were from 'Fagerlin' and 2 genotypes from 'Picaro'. The 10 genotypes were cloned, from which approx. 45 ramets (tillers) were transplanted into pots. A total of 450 pots were distributed into 8 trolleys which were then placed in a growth chamber at 18/20°C day/night temperature, at a light intensity of 220-240 µmol m^{-2} s^{-1} for four weeks. The trolleys were rotated anti-clockwise for every 3 or 4 days while watering the plants, so that all the tillers would receive similar light and temperature conditions. After four weeks of growth of tillers, half of the plants (240 pots, first set) were randomly picked from 4 trolleys and moved into four small trolleys, and the remaining 210 pots (second set) were used for sampling tissue for the RNA Sequencing experiment (for gene expression studies between resistant and susceptible genotypes based on the snow mould resistance test).

2.2 Inoculation of *M. nivale*

Inoculum was prepared from *M. nivale* var. *nivale* isolate 3/98 inoculum, originally isolated from perennial ryegrass and stored at -80°C. Inoculum preparation was done as previously described by Tronsmo (Tronsmo 1993). On potato dextrose agar (PDA) the fungus was cultured at 9°C for two weeks and then inoculated in Erlenmeyer flasks containing 100 ml sterile potato dextrose broth(PDB). Then the PDB were inoculated with four agar plugs containing actively growing mycelium and further incubated at 9°C for two weeks. Mycelium was then filtered by using cheese cloth and the harvested mycelium was homogenized in distilled water containing 0.2% gelatin using Ultra Turrax. Then the final inoculum was adjusted to an optical density of 0.5 at 430 nm. Then the plants were inoculated with 0.5-1ml inoculum per plant and the control plants were inoculated with 0.2%gelatin. After inoculation the plants were incubated, i.e., covered with wet cellulose papers and then with black plastic sheets to maintain humidity at 2°C in darkness (Hofgaard et al. 2006b).

2.3 Tissue Sampling

The control samples (non-incubated) were collected after four weeks from a growth chamber having the same conditions as described above. Then three pots from each genotype were picked and, randomly placed in 4 different trolleys and three biological replicates (leaf samples) were collected from each genotype (inoculated and incubated =3 and non-inoculated and incubated =3) after 1 day of inoculation. In the same way both inoculated and non-inoculated samples were collected after the 4th day of inoculation. The collected samples were immediately placed in liquid nitrogen and stored at -80°C until used for RNA extraction.

2.4 Snow mould resistance test

After inoculation with *M. nivale* the plants were incubated for 5-12 weeks at 2 °C in darkness. The first set of plants was taken out of the incubation room after 5 weeks and transferred to the green house. Then the tiller survival rate was calculated using the number of live tillers divided by the total number of tillers including dead tillers (e.g. if a plant has 16 tillers in total, the number of live tillers = 16 and the dead tillers = 0 then the survival proportion = 1). Then the plants were cut 3 cm above the soil level and left for 2 weeks to regrow. After 2 weeks of regrowth the shoot were cut, placed in paper bags and dried in an air dryer at 60 °C for 2 days. Then the dry weights of shoot samples were recorded. Snow mould resistance was expressed as the relative regrowth (dry weights of inoculated, incubated plants divided by the dry weights of

non-inoculated, incubated plants). In the same way the second set and third set of plants were taken out after 8 and 12 weeks and the relative regrowth was measured in the same way as described above. The plants were taken at different incubation periods because the plants seem to be healthier and no symptoms were observed on leaves after 5 weeks of incubation. So, the incubation periods were prolonged to 8 weeks and 12 weeks during which infections on leaves were observed.

2.5 RNA extraction

The frozen leaf samples were crushed with a pestle and mortar and the powdered sample used for extraction of total RNA using the Pure Link RNA Mini Kit (Life technologies) plus plant RNA Isolation Aid (Life technologies). To remove DNA from the samples we added 80 µl of DNase to the column which degrades the DNA. Thus, total RNA was eluted from the 30 µl from RNase-free water. The concentration and quality were checked using the Nanodrop (Nanodrop technologies, Wilmington, DE, USA) and measured using the Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) to check the RNA Integrity Value (RIN) which should be above 7 for better quality sequencing using high throughput technologies.

The extracted RNA was used for performing qRT-PCR and further for RNA Sequencing.

2.6 qRT-PCR

The EXPRESS Two-Step qRT-PCR kit which includes the SuperScript VILO cDNA Synthesis kit which was used for generating the single-stranded cDNA that was later used for quantifying the amount of specific gene expression using forward and reverse primers, following the manufacturer's instructions. cDNA synthesis was done using up to 2.5 µg total RNA in 20 µl reaction. 5 µl of cDNA was used in each well of Fast Optical 96 well plate along with other components making the total volume 20 µl. The fast cycling program was then set at 95°C for 20 sec, 40 cycles of 95°C for 3 sec(denaturation) and 60°C for 35 sec (annealing). Then each plate(with samples) for each gene with a bar code was placed in ABI7500 qRT-PCR machine. The SYBR® Green dye was used to detect the amplified products. The expression of the specific genes PR-3 and PR-5 were normalized by the house-keeping gene or internal control gene GAPDH (EC 1.2.1.12). The fold change in expression of the target gene relative to the house-keeping gene at various time points have been studied by using the 2*-ΔΔCT method where the ΔΔCT is calculated by the equation ΔΔCT = (CT of target – CT of GAPDH) time x – (CT of target – CT of

GAPDH) time zero which gives the mean fold change in expression of target genes at each time point (Livak & Schmittgen 2001).

The primer sequences used for measuring expression of PR-5 were (Zhang et al. 2011):

Forward Primer : GCAGCTGAACAGCGGCGAGACGTGGAAC

Reverse Primer : GCCGGTGCTGCAGGAGAAGCCCATGC

2.7 RNA Sequencing

A total of 12 samples, selected based on having an RNA concentration above 200 ng/µl (having 260/280, 260/230 ratios ≥1.8) with RIN values above 6.5, were sent to the GATC company (Germany) for preparation of cDNA libraries and paired-end sequencing using an Illumina HiSeq2000 sequencer. At first cDNA was digested using a restriction enzyme NlaIII and then ligated with two adapters. The sequencing primers based on two adapters will generate millions of raw reads with sequencing length of 100bp (Xu et al. 2011). Out of these 12 samples only 11 samples were sequenced as GATC faced problems during library preparation of one replicate of the control sample of genotype F. All these 11 samples belong to the F and M genotypes of which 5 samples belongs to F and 6 samples to the M genotype. Of the 5 samples of the F genotype, one sample was the non-incubated control, 2 biological replicates each belonging to inoculated and non-inoculated samples. The sample which was not sequenced was the non-incubated control of genotype F. In the same way the M genotype also had 2 biological replicates from each of the 3 treatments (non-incubated control, inoculated and incubated, and non-inoculated and incubated) as shown in Table.1

Table.1 The 12 samples with treatments each having 2 biological replicates for both genotype F and M.

Sample number	Treatment	Sample Name
1	Non-incubated control	F21

2	Non-incubated control	F16
3	Non-incubated control	M31
4	Non-incubated control	M39
5	Incubated, inoculated	F31
6	Incubated, inoculated	F42
7	Incubated, inoculated	M40
8	Incubated, inoculated	M43
9	Incubated, non-inoculated	F19
10	Incubated, non-inoculated	F29
11	Incubated, non-inoculated	M27
12	Incubated, non-inoculated	M37

2.8 RNA Sequencing data analysis

The raw sequencing data was received as compressed files from GATC company and the data analysis was initiated by uploading the data into the server, which requires high computational power. This data was processed by various steps. The first analysis step was Quality checking and trimming.

2.8.1 Quality check, filtering and Trimming

The compressed files of raw sequencing data are decompressed and fastqc was run which gives the output files in html format. These output files are useful for checking the initial quality; they give us information on raw sequencing data which is useful for further analysis. For example, if we take an average read of a sample; it is judged to be of good quality if the quality score is above 20. Initially we got around 155-170 million reads including adapters which are used during cDNA synthesis. These adapter sequences have to be removed by trimming the raw sequence data; this was done using the program sickle. Filtering helps in removing the lower

quality reads and trimming of adapters helps in increasing the mapping efficiency https://github.com/najoshi/sickle/blob/master/README.md.

After trimming the raw sequence data is assembled by using the program Trinity.

2.8.2 De novo assembly

The trimmed raw sequencing data was assembled separately for genotypes “M” and “F”. The 5 samples from genotype “F” and the 6 samples from genotype “M” were assembled by the three independent software modules Inchworm, Chrysalis and Butterfly (see description in the introduction).

2.8.3 Read alignment

After assembly further analysis was carried out by aligning back the raw sequence data of each sample to the assembled transcripts. Read alignment was done by Bowtie which is an ultrafast, memory-efficient alignment program for aligning short DNA sequence reads to large genomes (Langmead et al. 2009).

2.8.4 Transcript abundance estimation using RSEM

RSEM is a software package for quantifying gene and isoform abundances from single-end or paired-end RNA seq data. This software does not require any reference genome. So, this gives accurate transcript abundance for species without sequencing genome (Li & Dewey 2011). By running this program, two output files are generated. They are RSEM.isoforms.results and RSEM.genes.results. An example of an isoform result file is presented in Table. 2

Table. 2 The output file of RSEM.isoform.results which gives us information about transcript abundance taken from (http://trinityrnaseq.sourceforge.net/analysis/abundance_estimation.html). Transcript_id is the transcript identifier, gene_id is the component to which reconstructed transcript was derived, length is the length of reconstructed transcript , TPM is the number of transcripts per million reads , FPKM is the number of fragments per kilobase of transcript per millionfragments mapped to all transcripts, IsoPct is the percentage of expression for a given transcript.

transcript_id	gene_id	length	effective_length	expected_count	TPM	FPKM	IsoPct
comp128_c0_seq1	comp128_c0	209	1.73	0.00	0.00	0.00	0.00
comp13_c0_seq1	comp13_c0	235	7.16	1.00	12561.51	5282.75	100.00
comp22_c0_seq1	comp22_c0	215	2.62	0.00	0.00	0.00	0.00
comp28_c0_seq1	comp28_c0	329	54.60	4.00	6591.85	2772.21	100.00
comp33_c0_seq1	comp33_c0	307	40.30	3.00	6697.56	2816.66	100.00
comp35_c0_seq1	comp35_c0	219	3.33	0.00	0.00	0.00	0.00
comp35_c1_seq1	comp35_c1	204	1.19	1.00	75295.99	31665.75	100.00
comp39_c0_seq1	comp39_c0	348	68.20	1.00	1319.32	554.84	100.00
comp39_c0_seq2	comp39_c0	255	13.97	0.00	0.00	0.00	0.00
comp41_c0_seq1	comp41_c0	592	295.77	12.00	3650.37	1535.16	100.00
comp44_c0_seq1	comp44_c0	361	78.10	1.00	1151.96	484.46	100.00
comp44_c1_seq1	comp44_c1	280	25.22	1.00	3568.05	1500.54	100.00

2.8.5 Identifying Differentially Expressed Transcripts

EdgeR and DESeq bioconductor tools are used for differential expression analysis. These bioconductor tools are supported by the Trinity pipeline which generates the transcripts.counts.matrix file and gene.counts.matrix file from which they generate MA plot (M= log ratios and A = mean values) and Volcano plots for visualization of differentially expressed genes.

2.8.6 Functional annotation by Blast2GO

Functional annotation and analysis of gene or protein sequences was done by the bioinformatics tool called Blast2GO as it was developed to supply user-friendly interface for Gene Ontology (GO). Blast2GO consists of 5 steps: blasting, mapping, annotation, statistical analysis and visualisation (Conesa et al. 2005; Conesa & Götz 2009). Blasting is the first step of functional annotation to find similar sequences to the query set which is in fasta format. Those similar sequences are obtained by blast search to public or private databases (Altschul et al. 1990), followed by mapping which is the process of obtaining GO terms to the hits obtained from a blast search which is performed in four different ways. The blast results are directly used to get gene names or symbols or getting UniProt ids or accessions of blast hits that are searched directly in GO database or in the gene-product table of GO database (Conesa & Götz 2009). Thereafter mapping annotation was run by keeping the E-value hit filter to default = 1.0E-6 then only GO terms obtaining hits with a greater e-value than given will be used for annotation and

the annotation cutoff was by default 55 as the annotation rule selects the lowest term per branch that lies over this threshold (Conesa & Götz 2009) .

The statistical analysis collects a number of charts that are generated during BLAST, mapping and annotation. BLAST statistical E-value distribution gives the histogram of the number of hits with a given e-value. Mapping statistics produce a histogram of the number of GO terms obtained from each possible database source of annotations. Annotation gives a histogram with the number of sequences having a given number of annotations or pie chart generates the multilevel pie with the modest node per branch of DAG (directed acyclic graphs) that satisfies the filter conditions (minimal number of sequences a GO node have assigned) (Conesa & Götz 2009).

3. Results

3.1 Snow Mould Resistance test

A snow mould resistance test was done in order to identify genotypes with varying resistance levels to infection by *M. nivale*.

After inoculation the plants were incubated under artificial snow cover at 2°C in darkness for 5-12 weeks and the graph below (Fig. 6) shows the relative regrowth rate after the 1st incubation period.

3.1.1 Relative regrowth of 1st set of incubation period

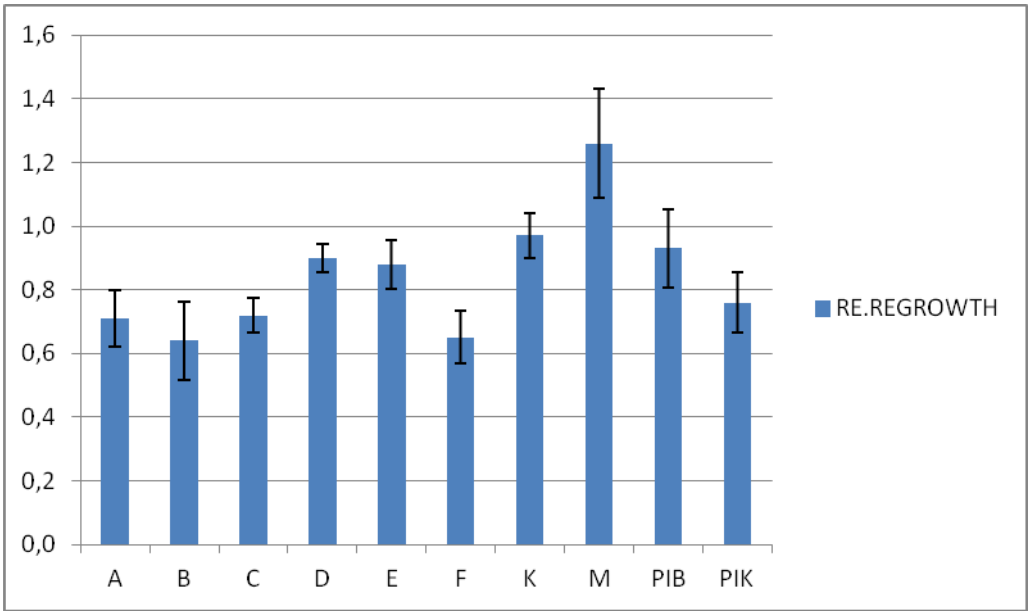

Fig. 6. Average relative regrowth after 5 weeks of incubation with standard error. In this graph the X-axis indicates the genotypes and the Y-axis indicates the regrowth values of tillers. Here the average relative regrowth shown for each genotype is the average of six values obtained from six biological replicates.

After the first incubation period the M genotype from cultivar 'Fagerlin' had the highest relative regrowth value while the A, B and F genotypes of 'Fagerlin' had the lowest relative regrowth values.

3.1.2 Relative regrowth of 2nd set of incubation period

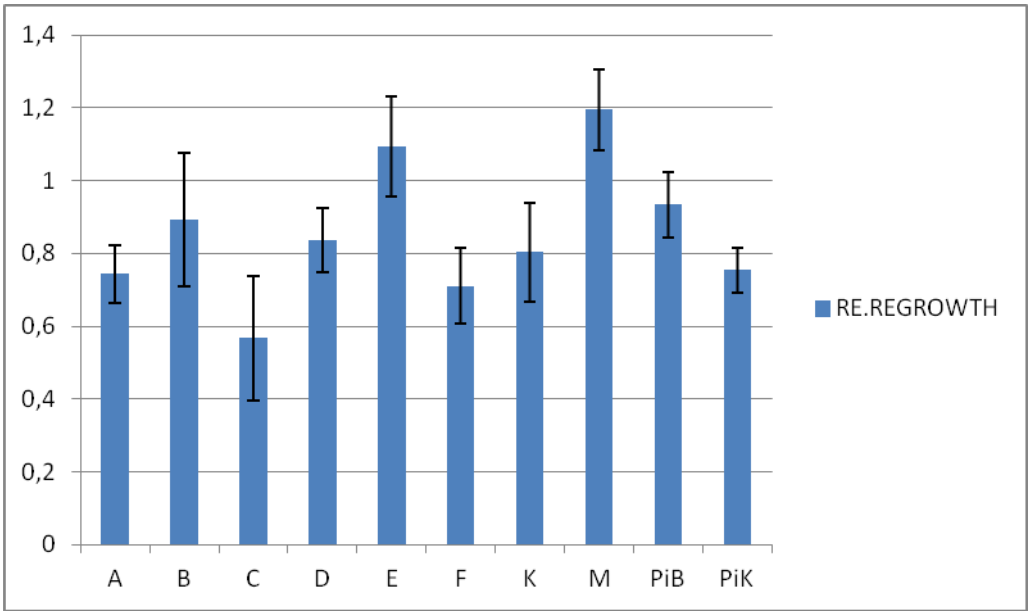

Fig. 7. Average relative regrowth after 8 weeks of incubation period. Here the average relative regrowth shown for each genotype is the average of three values obtained from three biological replicates.

The results from 8 weeks of incubation (Fig. 7) show that genotype C had the lowest regrowth while genotype M had among the highest regrowth values. A lot of variation between the genotypes was observed.

3.1.3 Relative regrowth of 3rd set of incubation period

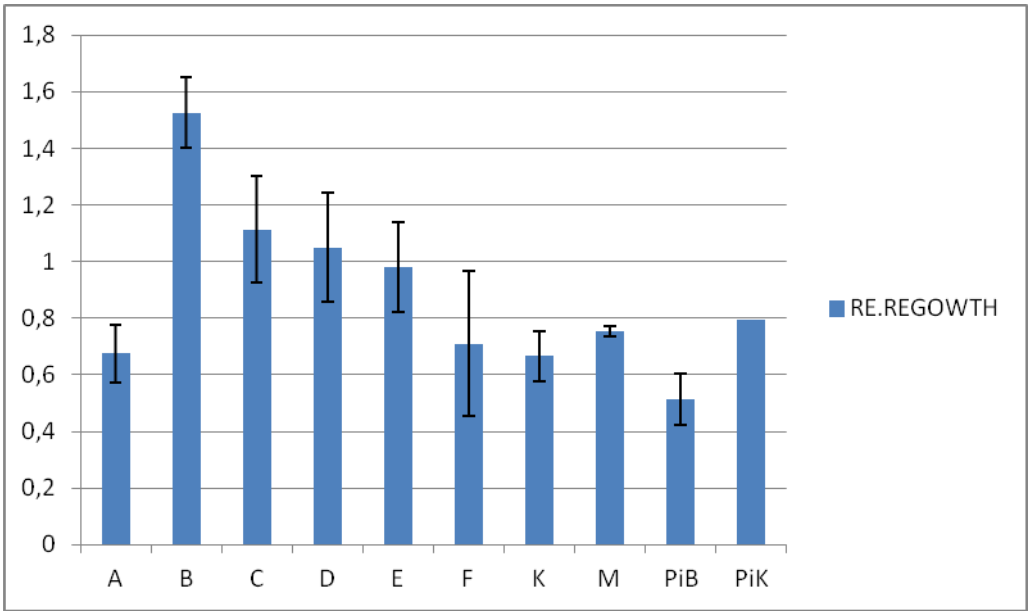

Fig. 8. Average relative regrowth after 12 weeks of incubation period. Here the average relative regrowth shown for each genotype is the average of three values obtained from three biological replicates.

Eight weeks of incubation (Fig. 8) showed a lot of variation in relative regrowth values. Genotype PiB (Picaro) genotype had the lowest relative regrowth value while genotype E had the highest relative regrowth value. This was surprising, and might be due to manual or technical errors while measuring the relative regrowth values.

3.1.4 Relative regrowth of all three incubation periods

The average relative regrowth values across genotypes (Fig. 9) did not really decrease with the incubation time and there was too much error within each incubation time to measure relative regrowth of individual genotypes accurately. Therefore incubation periods are used as replicates.

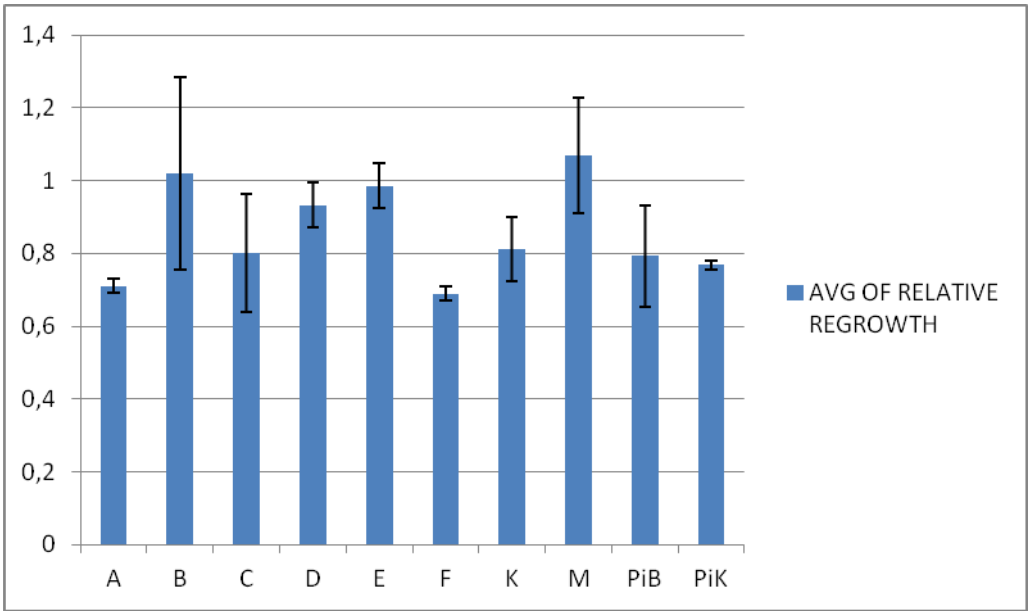

Fig. 9. Average regrowth of 10 genotypes after 5-12 weeks of incubation periods. Here the average relative regrowth shown for each genotype is the average of three values obtaioned from each of three incubation periods.

The average regrowth values (Fig. 9) over the 3 incubation periods show that genotypes A and F had the lowest relative regrowth values while genotype E and M had the highest relative regrowth values.

Initially, genotypes A and E was selected as ‘resistant’ and ‘susceptible’, however, the total RNA extracted from samples of these genotypes gave low quality and concentration of RNA not suited for next-generation sequencing. So, we choose M and F as the ‘resistant’ and ‘susceptible’ genotype, respectively.

3.1.5 Analysis of Variance (ANOVA)

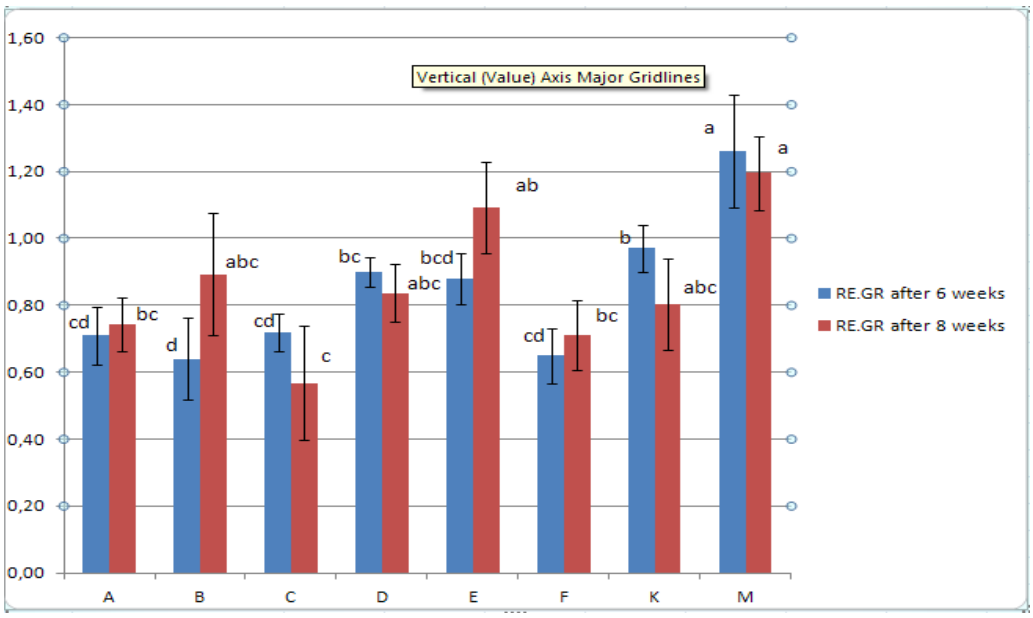

Fig. 10. Analysis of variance among the genotypes after 6 and 8 weeks of incubation. Entries with different letters are significantly different (P≤0.05).

Analysis of Variance (ANOVA) was used to analyse the variation between the genotypes. This analysis helps us to find the genotypes which are significant different. Fig. 10 shows that the two selected genotypes, M and F have significantly different regrowth values (they have different letters).

3.2 Expression profiles of specific genes

3.2.1 Thaumatin-like protein (PR-5)

To study the gene expression levels in the ‘resistant’ M and the ‘susceptible’ F genotype during snow mould infection we performed qRT-PCR for the gene PR-5 which encodes a defence related protein involved during infection. The fold change was observed to be higher in genotype M (inoculated) than in F (inoculated) both at day 1 and at day 4 after incubation (Fig. 11). Also,

the expression levels were higher in genotype M (non-inoculated) than in genotype F (non-inoculated). In contrast, we found expression in the non-incubated controls also where there should not be any expression in 0 days as the plants were not inoculated with the fungus.

Though we observed higher expression at the 1 day of incubation, we used incubated for 4 days as the RNA quality and concentration were higher at that time point.

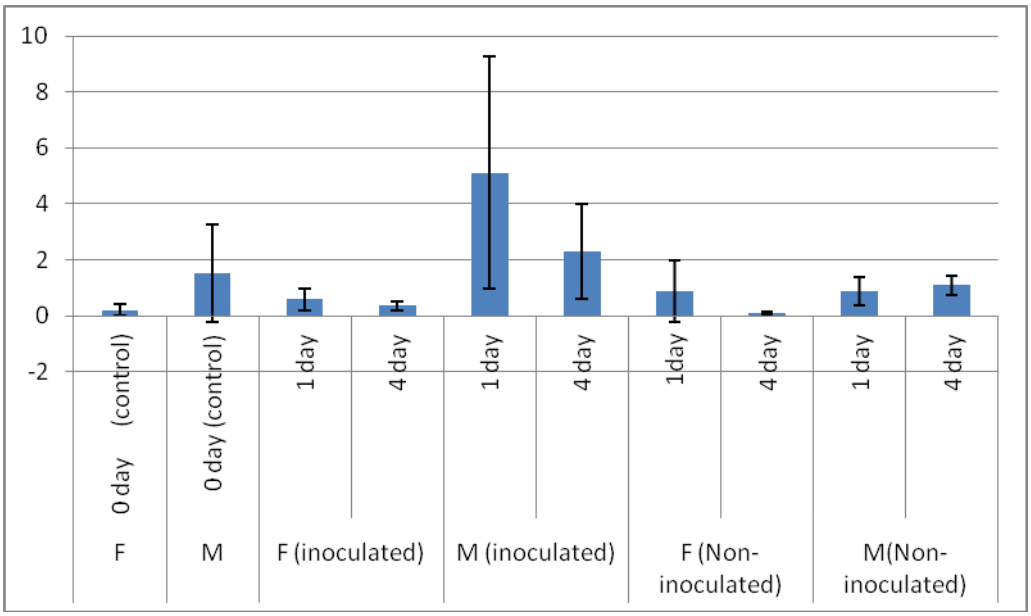

Fig. 11**.** Expression levels of the PR-5 gene in genotypes F and M genotypes during snow mould infection with *M. nivale*. qRT-PCR analysis was done using total RNA extracted from leaf samples of plants that were treated as control, inoculated and non-inoculated by the comparative CT method. The X-axis represents the different duration of snow mold infection and the Y-axis represents the fold change in expression of respective genotypes. Error bars represent the standard errors of the means calculated from two biological replicates.

3.2.2 Chitinase like protein (PR-3)

Gene expression studies of PR-3 using qRT-PCR did not work as the primer used for this gene failed to amplify the gene.

3.3 RNA sequencing data analysis

The RNA sequencing data analysis of the 11 samples was carried out in the following steps.

3.3.1 Quality check, filtering and trimming

The first step of data analysis was to check the quality of receiving sequencing data by running the FastQC program which generates an output file of raw sequencing data with quality scores across all bases.

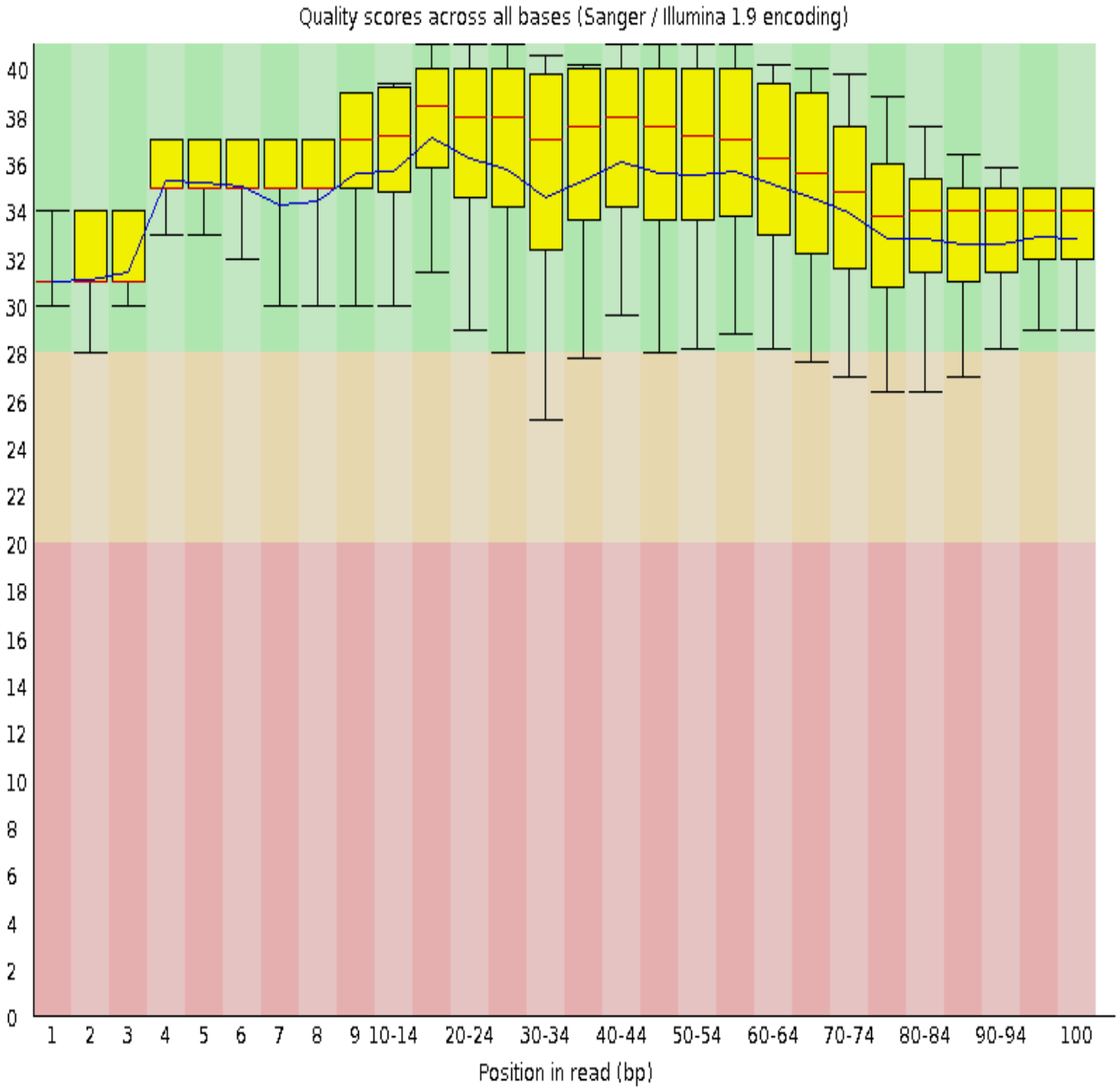

Fig. 12. Example of good quality of an average read in the sample. In this picture X-axis represents the position of each base in a read and the Y-axis represents the quality score. A score above 20 is considered as good quality.

A total of 155,765,940 and 170,258,124 raw sequence reads were obtained from genotypes ”F” and ”M” respectively, before filtering. Total reads obtained after filtering were 150,620,126 from genotype F and 165,833,628 from genotype M (Table. 3).

Table. 3 showing the total reads of F and M genotypes produced from the Illumina sequencer.

	F genotype		**M genotype**	
	Raw reads	After QC	Raw reads	After QC
Control: Rep1	28,930,920	28,364,598	30,654,426	33,424,958
Rep2	No replicate	-	34,162,768	30,024,114
Inoculated: Rep1	27,645,344	26,683,518	27,275,614	26,649,236
Rep2	30,632,570	29,990,476	21,430,630	21,013,258
Non inoculated:Rep1	35,656,348	34,069,392	32,900,758	31,386,042
Rep2	32,900,758	31,512,142	23,833,928	23,336,020
Total reads	155,765,940	150,620,126	170,258,124	165,833,628

3.3.2 *De novo* assembly

De novo assembly by the Trinity software produced 261,978 assembled contigs from genotype F and 188,355 assembled contigs from genotype M (Table. 4). The N50 value is an important statistics as this value assess the quality of the sequence assembly and the higher the value of N50 is, the better is the assembly (Annadurai et al. 2012).

Table. 4 Statistics of the Trinity assembly., n = total number of assembled contigs, n:200 = the number of contigs below minimum length, n:N50= the number of contigs with lenght less than the N50 value, min = minimum length of contigs, N80 = 80% of assembled bases are in contigs of this length or longer, N50 = 50% of assembled bases are in contigs of this length or longer, N20 = 20% of assembled bases are in contigs of this length or longer, max = length of longest contig, sum= total number of assembled bases.

	F genotype	**M genotype**
n	261,978	188,355
n:200	261,978	188,355
n:N50	52,740	38,482
min	201	201
N80	853	794
N50	1,784	1,672
N20	3,064	2,845
max	17,632	12,882
sum	292 million	199 million

3.3.3 Read alignment

83% of the sequencing reads were mapped back to the assembly using Bowtie in the Trinity pipeline.

3.3.4 Analysis of differential gene expression

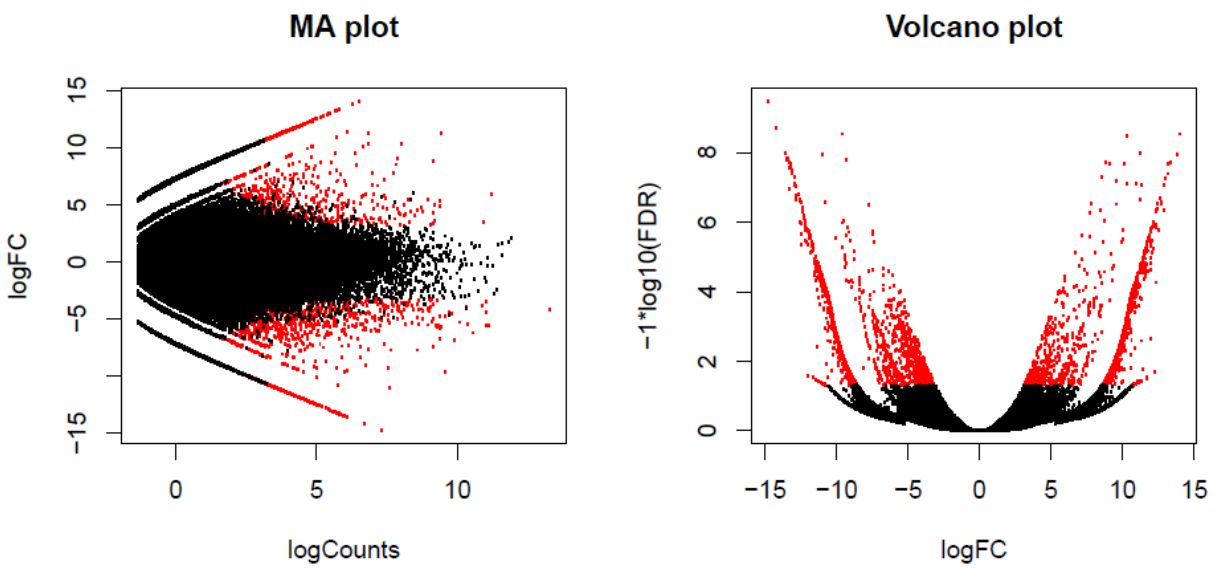

Fig. 13. Plots produced for non-incubated control vs non-inoculated and incubated of genotype M. The red coloured dots in both plots represents genes that are differentially expressed. Log fold changes of above zero indicates genes that are upregulated and negative fold changes indicate down-regulated genes.

In the MA plot of ggenotype M (Fig. 13) there are many genes that are differentially expressed.

The volcano plot is a type of scatter-plot which is constructed between two treatments, in this case (Fig. 13, right) between non-incubated control and non-inoculated and incubated plants of genotype M. The X-axis represents the log of fold changes in which we see datapoints represented in black and red colour in two directions (up and down-regulation) from the center. These plots are used to identify variation between large data sets of replicates. Dots that are far to either the left or the right hand side represents genes that are highly up or down-regulated. The X-axis represents the biological impact of the change and the Y-axis represents the reliability of change (statistical evidence).

Transcripts judged to be differentially expressed at **0.05 FDR** are coloured in red. In the above plot (Fig. 13, right) 6,227 genes are significantly differentially expressed in genotype M.

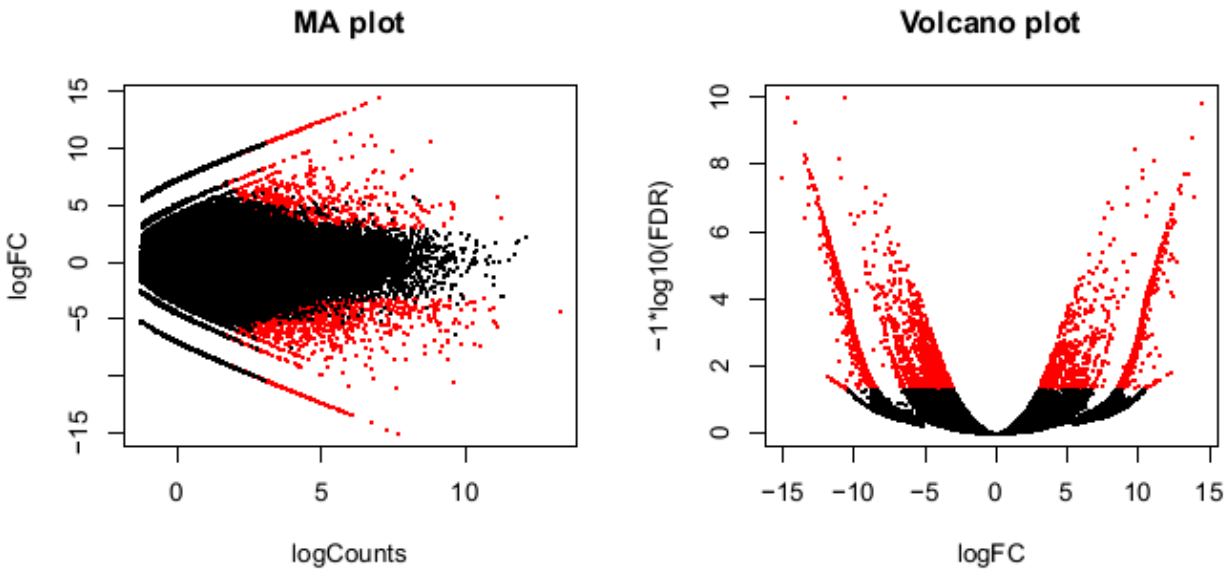

Fig. 14. Plots produced for non-incubated control vs inoculated and incubated genotype M. The red coloured dots in both plots represents genes that are expressed differentially. Log fold changes above zero indicate genes that are upregulated and negative fold changes indicate down-regulated genes.

In these plot between non-incubated control plant and inoculated and incubated plants of genotype M around 7,282 genes are differentially expressed.

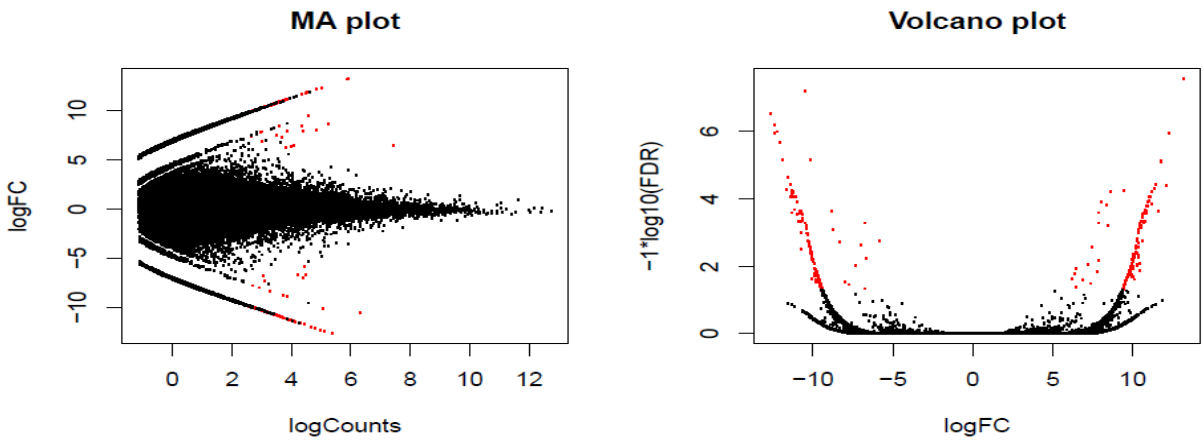

Fig. 15. Plots produced for inoculated vs non-inoculated, incubated plants of genotype M. The red colour dots in both plots represents genes that are expressed differentially. Log fold changes above zero indicates genes that are upregulated and negative fold changes indicate down-regulated genes.

The results displayed in Fig. 15 are more interesting since fewer genes, only 275 are significantly differentially expressed, and biologically this is the most correct comaprison. When comparing non-incubated control plants with incubated plants, a large number of differentially expressed genes identified are related to the incubation in darkness.

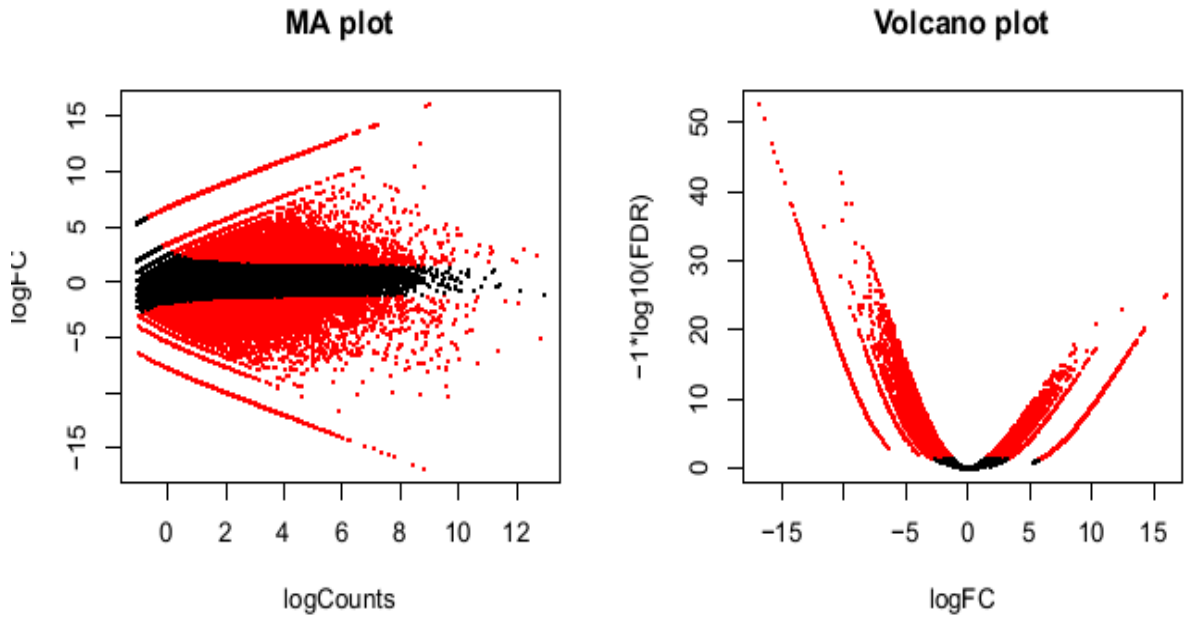

Fig. 16. Plots produced for non-incubated control vs inoculated and incubated of genotype F. The red colour dots in both plots represents genes that are expressed differentially. Log fold changes above zero indicates genes that are upregulated and negative fold changes indicate down-regulated genes.

Fig. 16 shows that many genes (19,055) are differentially expressed between the non-incubated control and inoculated and incubated plants of genotype F.

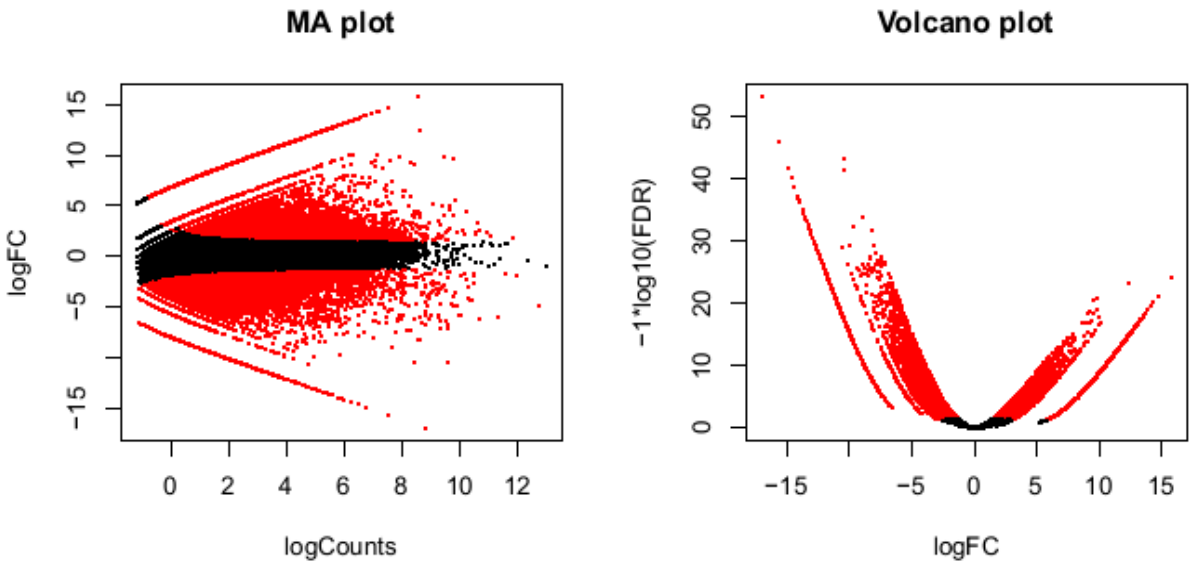

Fig. 17. Plots produced for non-incubated control vs non-inoculated and incubated genotype F. The red colour dots in both plots represents genes that are expressed differentially. Log fold changes above zero indicates genes that are upregulated and negative fold changes indicate down-regulated genes.

In genotype F around 19,832 genes are significantly differentially expressed between non-incubated control and non-incubated and incubated plants.

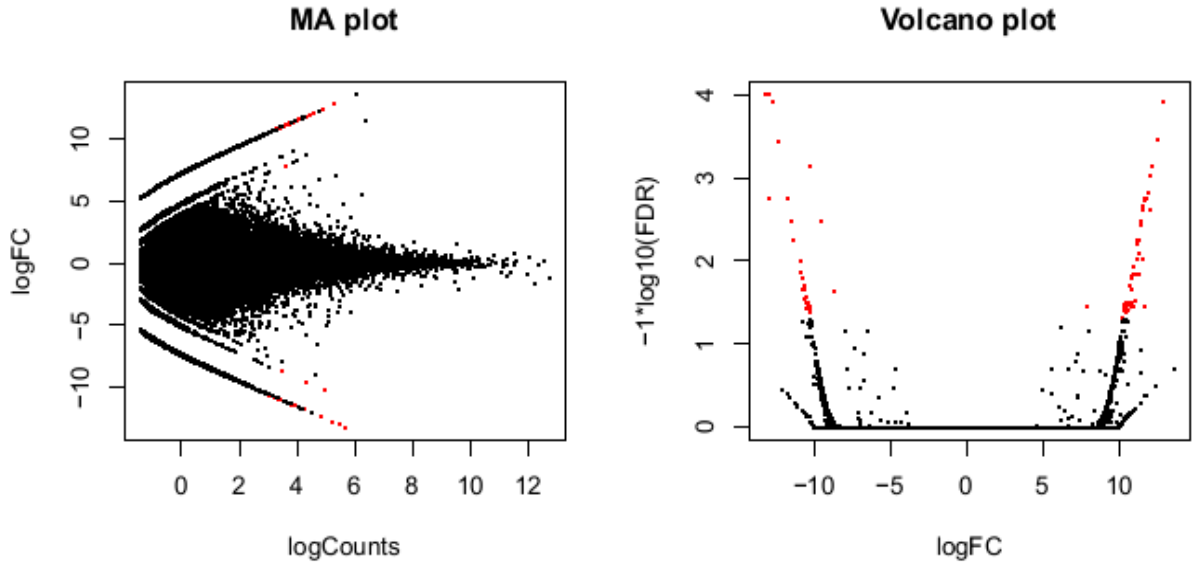

Fig. 18. Plots produced for inoculated vs non-inoculated of genotype F. The red colour dots in both plots represents genes that are expressed differentially. Log fold changes above zero indicates genes that are upregulated and negative fold changes indicate down-regulated genes.

From the Volcano plot (Fig. 18, right) we can see that around 83 genes are significantly differentially expressed (at FDR ≤0.05) when inoculated and non-inoculated incubated plants are compared.

The total number of genes that are differentially expressed are presented in Table. 5. The F genotype has more differentially expressed genes as we are lacking one replicate from control because of which became difficult to compare overall between one replicate in control and 2 replicates in inoculated or non-inoculated. As a result the comparisons are not possible however, Table.5 showed many differentially expressed genes in the F genotype except in comparison between Inoculated vs Non-Inoculated indicating that there is a biased result in non-incubated control vs Inoculated, non-incubated control vs Non-inoculated.

Table.5 A summary of significantly different expressed genes at FDR ≤0.05

	FDR = 0.05	
	F genotype	M genotype
Control vs Inoculated	19,055	7,282
Control vs Non-inoculated	19,832	6,227
Inoculated vs Non-inoculated	83	275

3.3.6 Blast2GO results

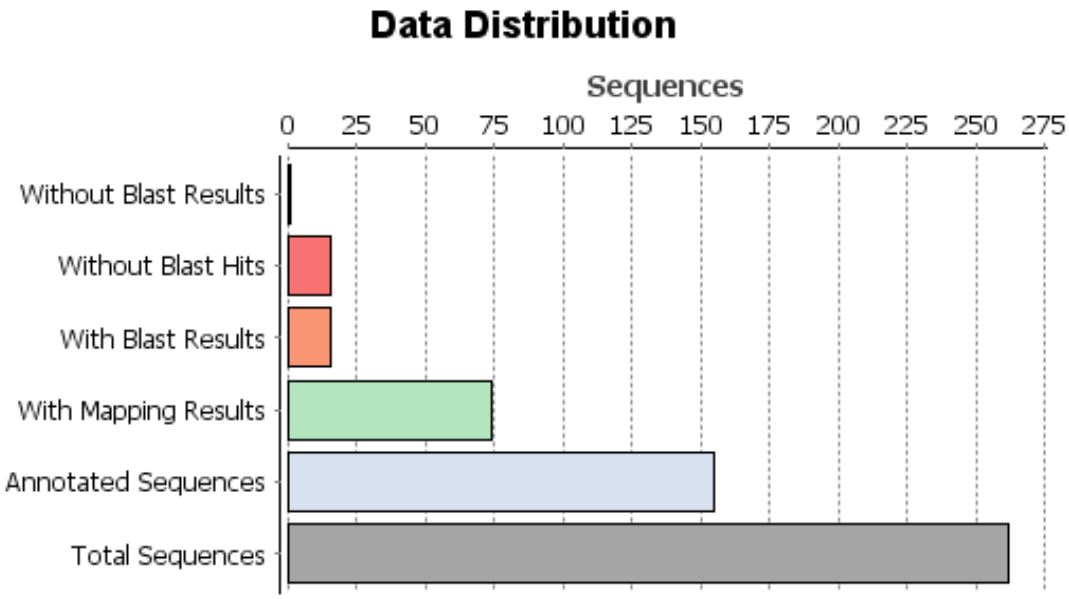

Fig. 19. Blast, mapping and annotated results of inoculated vs non-inoculated incubated plants of genotype M.

The histogram shown in Fig. 19 is the statistical report of the annotation process which uses three main steps. The first step is to Blast to find the homologous sequences, the second one is to map to collect the Gene ontology terms associated with blast hits, and finally the annotation step to give truthful information to query sequences. Our data shows how many sequences are Blasted, Mapped and Annotated from the total sequences. In this histogram about 260 sequences are available in which 20 of them were without and with blast hits and around 75 and 155 sequences are available with mapping results and annotated results.

The Blast search of our data for inoculated vs non-inoculated,incubated plants of genotype M gave 275 differentially expressed genes against public databases (Fig. 19).

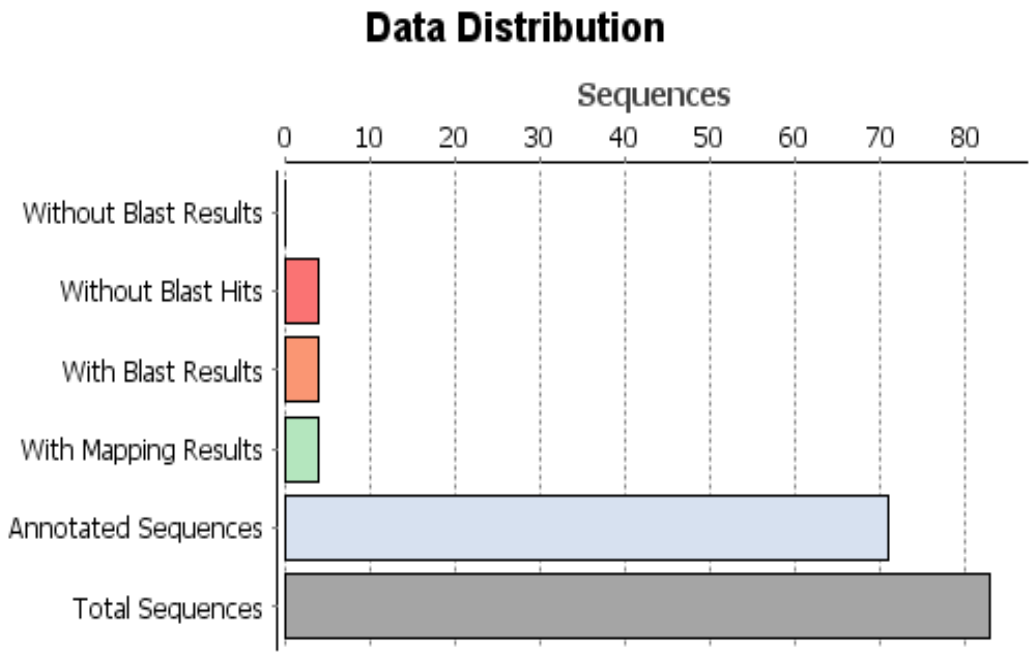

Fig. 20. Blast, mapping and annotation results for inoculated vs non-inoculated incubated plants of genotype F.

3.3.7 Blast result

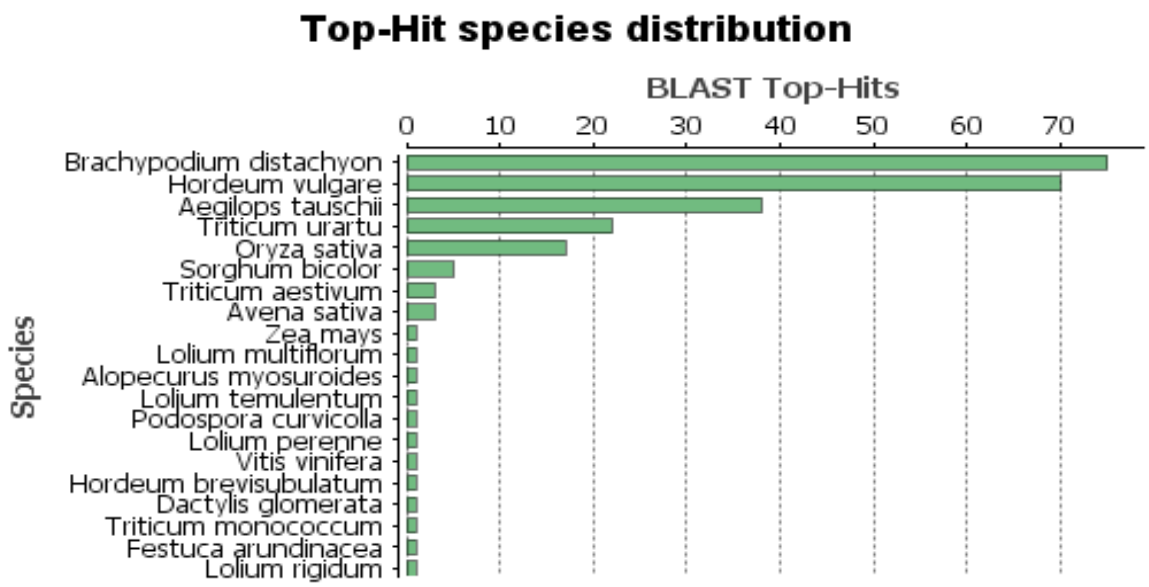

Fig. 21. Top-hits species distribution chart of genotype M.

Top-hit species distribution is obtained as the first step of the annotation process performed by Blastx to NCBI, when our sample information is uploaded in the Blast2go program in the FASTA format. Fig. 21 shows the species distribution of our sample information where our sample data shows top-hits with grass species. *Brachypodium distachyon, Hordeum vulgare, Aegilops tauschii, Triticum urartu, Oryza sativa, Sorghum bicolor, Triticum aestivum, Avena sativa, Lolium multiflorum, Alopecurus myosuroides, Lolium temulentum, Lolium perenne, Hordeum brevisubulatum, Dactylis glomerata, Triticum monococcum, Festuca aruninacea, Lolium rigidum,* and *Zea mays* belongs to the grass family (Poaceae), and *Podospora curvicolla* is a fungal species, and *Vitis vinifera* is a dicot belonging to the Vitaceae family.

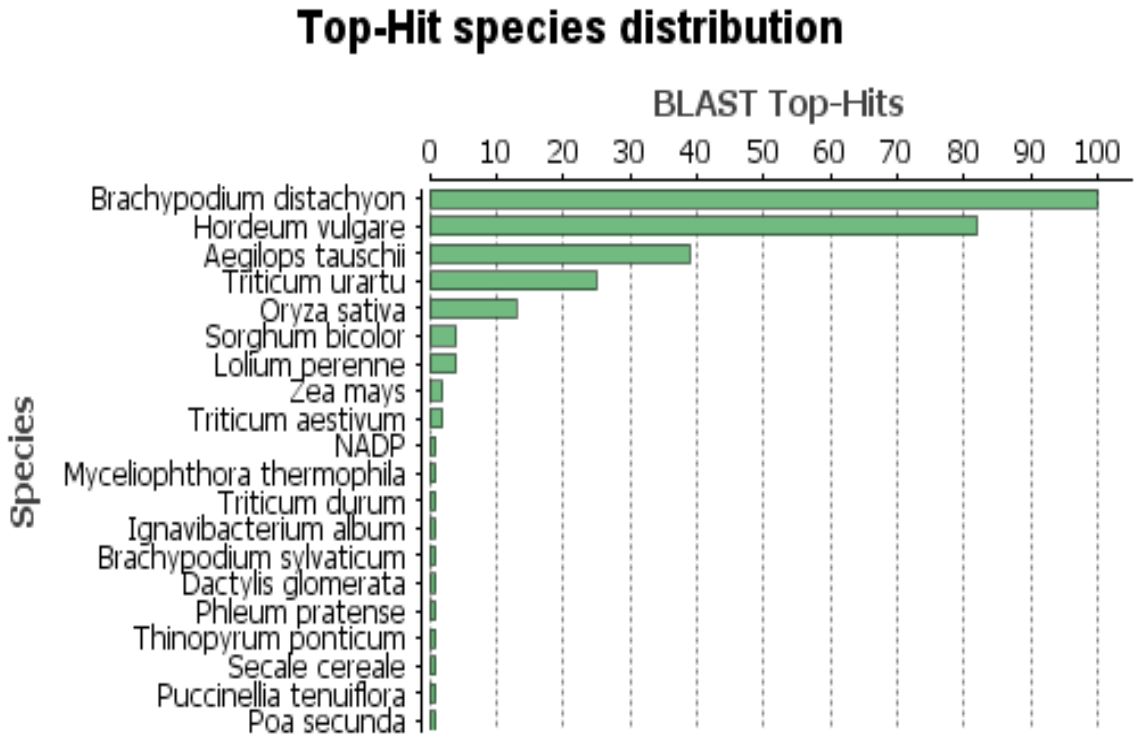

Fig. 22. Top-hits species distribution chart of genotype F.

It is no surprise that *B. distachyon* and *H. vulgare* are the top-hit species. They are the closest relatives of L. perenne with a fully sequenced genomes present in the databases.

3.3.8 Mapping

The second step of the annotation process is the mapping to collect the GO-terms obtained from different source databases associated with Blast hits.

3.3.9 Annotation result

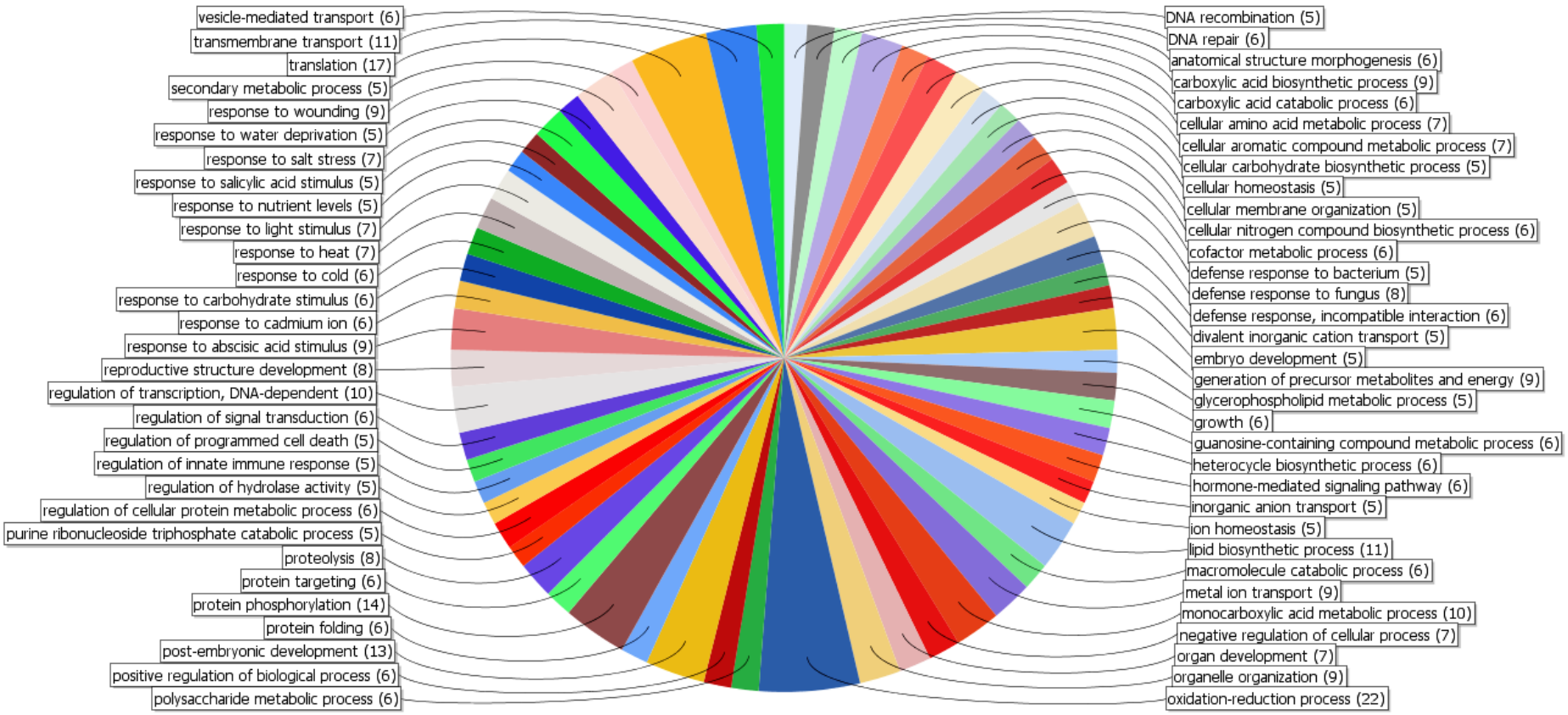

Fig. 23. Sequence distribution of differentially expressed genes in the M genotype (inoculated vs non-inoculated).

Fig. 23 shows the summary of the annotation results represented in a pie chart. This chart shows the lowest Gene ontology terms per branch i.e., sequence abundance based on the specified cutoff by the user. For instance this chart shows 8 sequences that are related to defence response to fungus which means 8 genes are differentially expressed. Also 6 sequences are related to abiotic stress like cold, 7 sequences for heat, 9 sequences related to abscisic acid and 6 sequences are in response to cadmium ion.

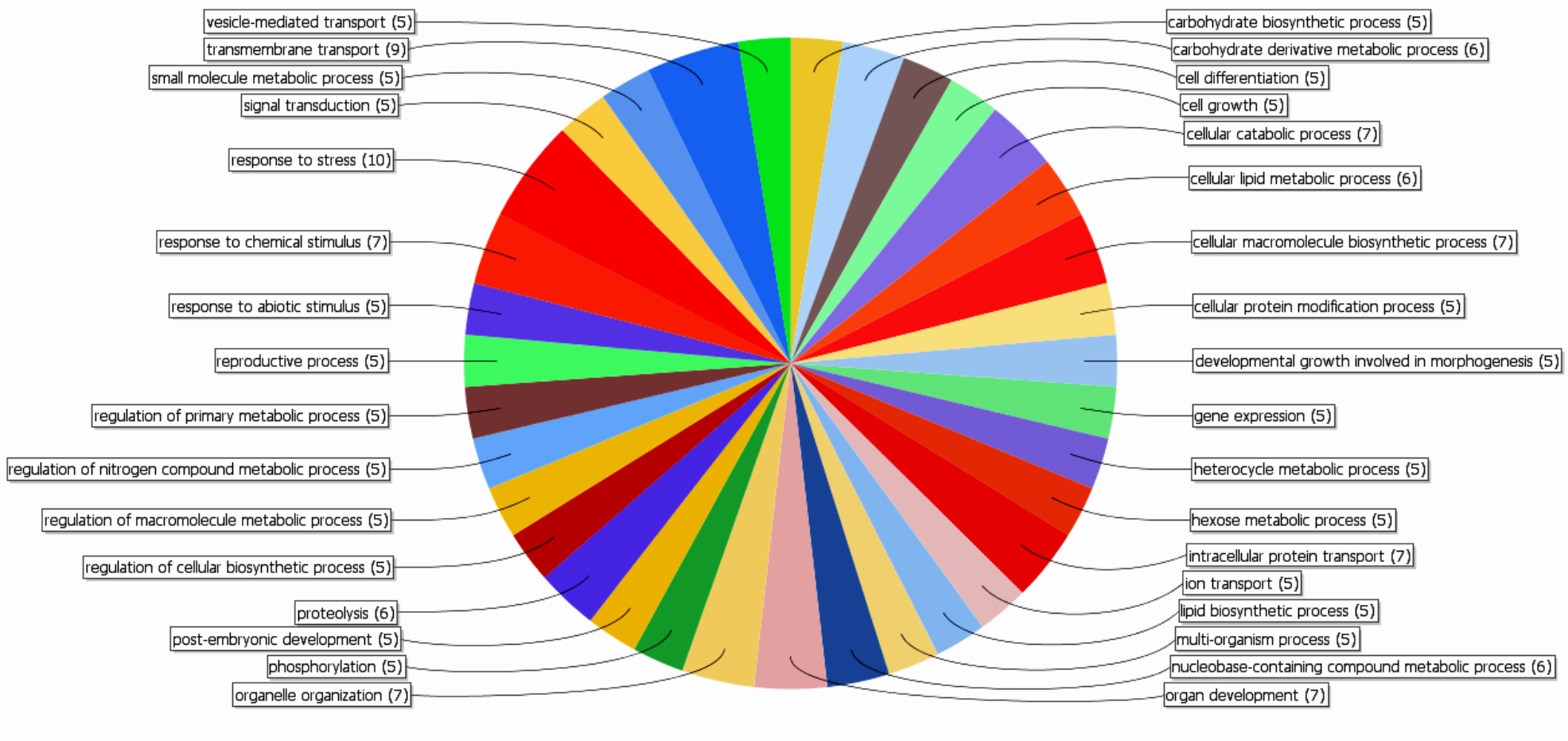

Fig. 24. Sequence distribution of differentially expressed genes in the F genotype (inoculated vs non-inoculated).

GO terms of above sequence distribution for F genotype (Fig. 24) shows that there are 10 sequences that are responding to stress, 7 sequences are responding to chemical stimulus, 5 sequences are responding to abiotic stress, and 5 sequences are involved in carbohydrate biosynthetic processes.

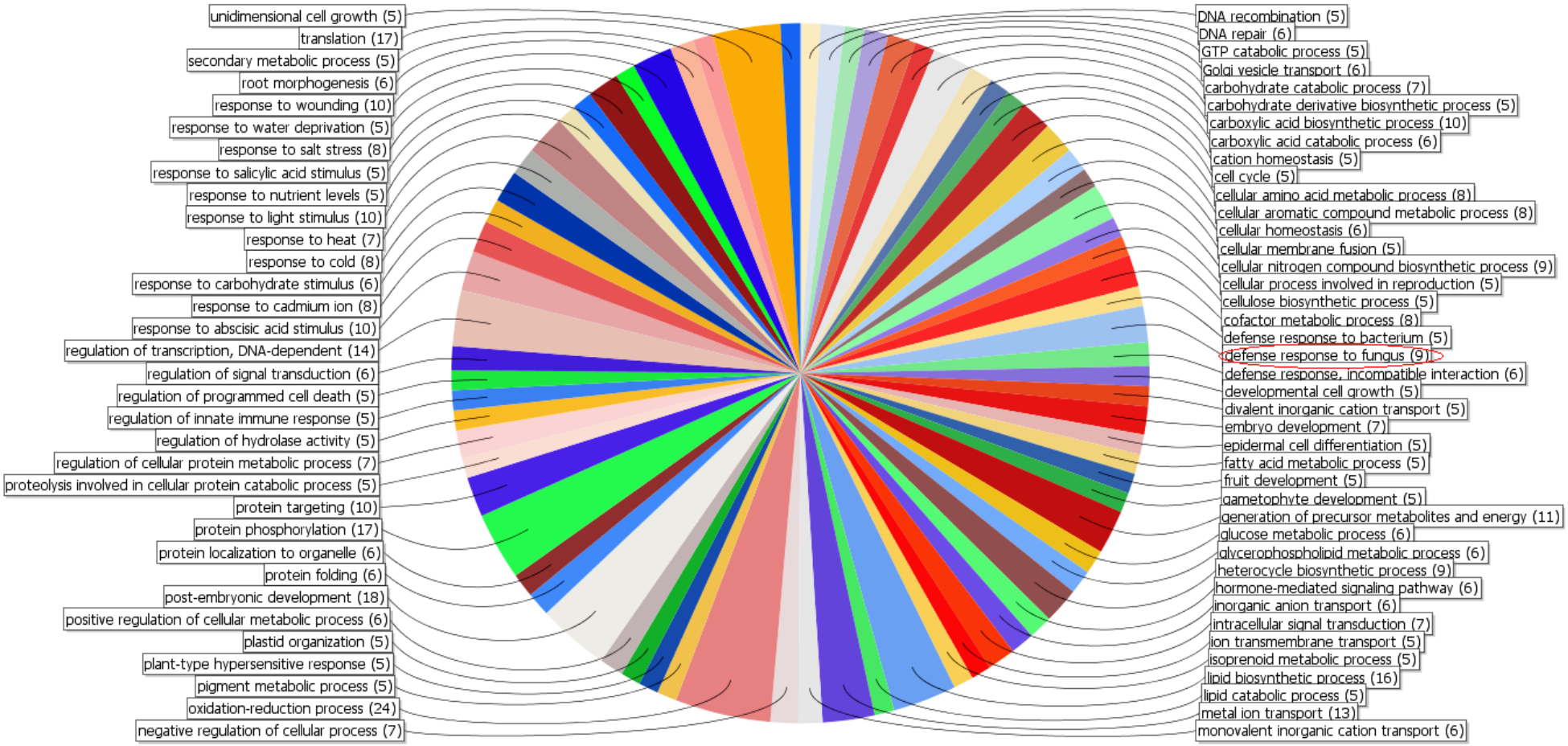

Fig. 25. Over all sequence distribution of differentially expressed genes in the M and F genotypes (inoculated vs non-inoculated).

Here the sequence distribution shows that the sequences that are related to defence responses are 9, i.e., 8 belonging to genotype M and one belongs to genotype F. Ten sequences are related to abscisic acid stimulus (9 belongs to genotype M and one to F), 8 sequences are related to cold (6 belong to genotype M and 2 to F), and 8 sequences are related to cadmium ion responses (6 belong to genotype M and 2 to F).

In Table. 6 the 9 sequences related to defence responses to the fungus that are extracted from the sequence distribution of differentially expressed genes in genotypes M and F are listed.

Table. 6. GO terms for defence response genes to fungus. The GO terms are: C = cellular component, P = biological process, F = molecular function.

Sequence name	**Sequence description**	**GO terms**
F_95080_c0_seq1	ubiquitin-likemodifier-activating enzyme atg7-like	P: leaf senescence; P:defense response to fungus; P:autophagy
M_32060_c0_seq1	tyrosine-proteinphosphatase tpte-like	P:indole glucosinolate catabolic process; P:salicylic acid biosynthetic process; P:tetracyclic triterpenoid biosynthetic process; P:regulation of plant-type hypersensitive response; P:glucosinolate biosynthetic process; P:response to chitin; P:indoleacetic acid biosynthetic process; P:cellular cation homeostasis; P:negative regulation of

		defenseresponse; P:phosphatidylinositol dephosphorylation; P:divalent metal ion transport; P:jasmonic acid mediated signaling pathway; P:photosynthesis, light reaction; P:detection of biotic stimulus; P:defense response to bacterium; P:protein targeting to membrane; P:tryptophan catabolic process; P:response to nitrate; P:regulation of response to biotic stimulus; P:regulation of hydrogen peroxide metabolic process; P:pentacyclic triterpenoid biosynthetic process; P:cellular membrane fusion; P:defense response by callose deposition in cell wall; P:defense response to fungus, incompatible interaction; P:negative regulation of MAP kinase activity; P:systemic acquired resistance, salicylic acid mediated signaling pathway; P:Golgi vesicle transport;

		P:response to cold; P:MAPK cascade; P:nitrate transport; P:peptidyl-tyrosine dephosphorylation; P:regulation of multi-organism process;
M_45673_c1_seq2	bbd1_orysi ame: full=bifunctional nuclease 1	P:response to wounding; P:nucleic acid phosphodiester bond hydrolysis; P:defense response to fungus
M_57028_c0_seq1	asparticproteinase nepenthesin-2-like	P:response to chitin; P:defense response to fungus; P:proteolysis; P:translation
M_64472_c0_seq1	proteasome subunit beta type 1	P:defense response to fungus, incompatible interaction;
M_68750_c0_seq1	mlo protein homolog 1-like	P:negative regulation of defense response; P:cell death; P:cellular response to organic substance; P:response to bacterium; P:signal transduction; P:defense response to fungus, incompatible interaction; P:regulation of innate immune response; P:cellular metabolic

		process;
M_68949_c0_seq1	o-glycosyl hydrolase superfamily protein	P:carbohydrate metabolic process; P:response to cold; P:response to bacterium; P:defense response to fungus,
M_71061_c0_seq2	peroxisomal acyl-coenzyme a oxidase 1-like	P:response to wounding; P:fatty acid beta-oxidation; P:regulation of plant-type hypersensitive
M_81582_c0_seq1	cytochrome p450 superfamily protein	P:electron transport chain; hydroxylase activity; P:response to water deprivation; P:abscisic acid catabolic process; P:secretion by cell; P:response to chitin; P:response to deep water; P:response to abscisic acid stimulus; P:release of seed from dormancy; P:defense response to fungus; P:response to red or far red light

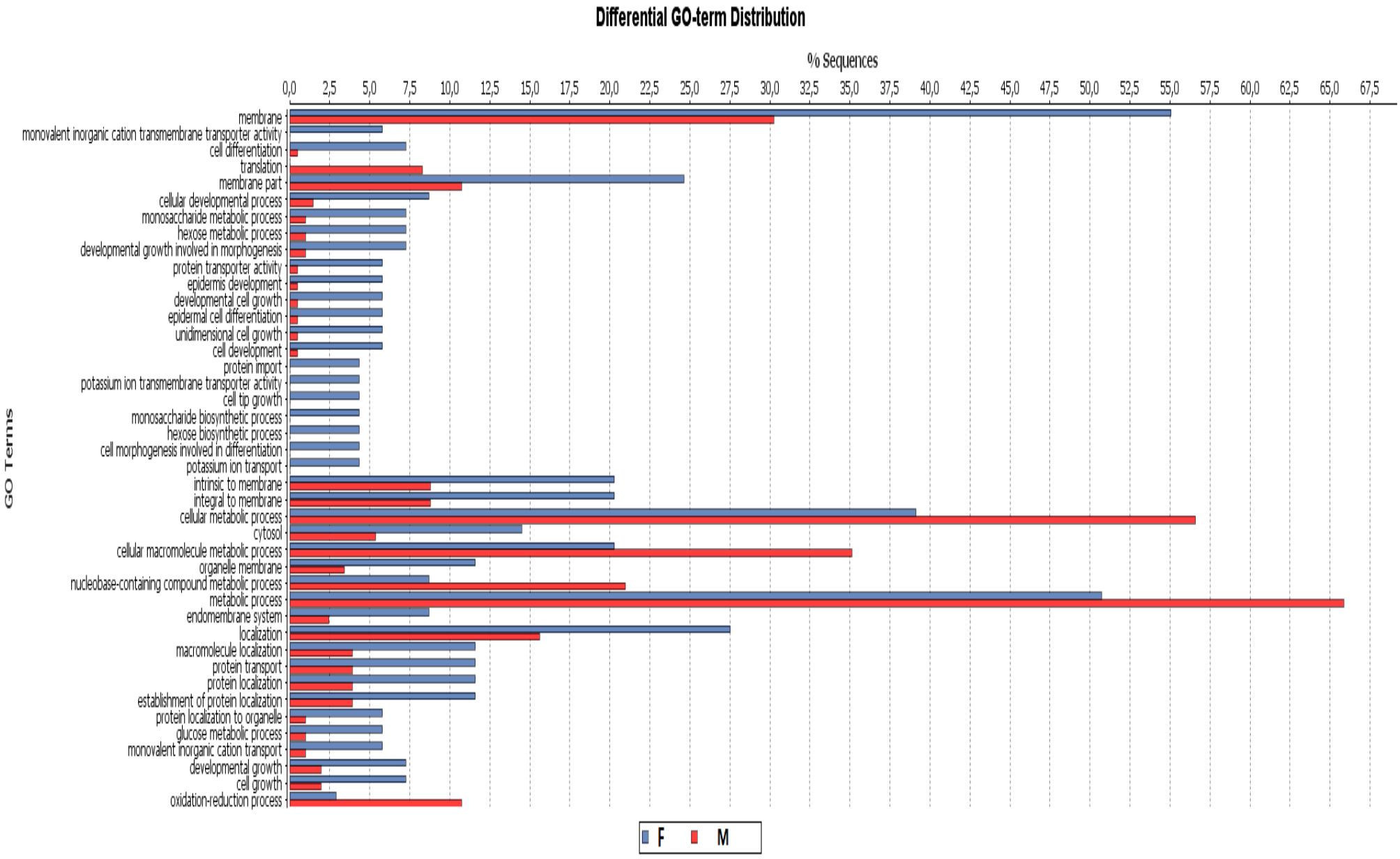
Differential GO-term Distribution
% Sequences
0,0
2,5
5,0
7,5
10,0
12,5
15,0
17,5
20,0
22,5
25,0
27,5
30,0
32,5
35,0
37,5
40,0
42,5
45,0
47,5
50,0
52,5
55,0
57,5
60,0
62,5
65,0
67,5
GO Terms
membrane
monovalent inorganic cation transmembrane transporter activity
cell differentiation
translation
membrane part
cellular developmental process
monosaccharide metabolic process
hexose metabolic process
developmental growth involved in morphogenesis
protein transporter activity
epidermis development
developmental cell growth
epidermal cell differentiation
unidimensional cell growth
cell development
protein import
potassium ion transmembrane transporter activity
cell tip growth
monosaccharide biosynthetic process
hexose biosynthetic process
cell morphogenesis involved in differentiation
potassium ion transport
intrinsic to membrane
integral to membrane
cellular metabolic process
cytosol
cellular macromolecule metabolic process
organelle membrane
nucleobase-containing compound metabolic process
metabolic process
endomembrane system
localization
macromolecule localization
protein transport
protein localization
establishment of protein localization
protein localization to organelle
glucose metabolic process
monovalent inorganic cation transport
developmental growth
cell growth
oxidation-reduction process
F
M

Fig. 26. Fisher's Exact Test result of genotypes F and M.

The annotation results can be visualised by various Enrich GO Graphs, a bar chart is one way of visualizing the results in which the X-axis represents the frequency of each term and the Y-axis represents the enriched GO terms (Fig. 26). Red bars corresponds to the M genotype and blue bars correspond to the F genotype. To perform an enriched bar chart, Fisher's exact test results are necessary as it is a statistical significance test used in the analysis of various terms between the genotypes.

4. Discussion

In this present study we aimed at investigating the expression of genes in relation to different treatments by transcriptome analysis using an RNA Sequencing approach. Before performing the transcriptome analysis we had screened for selecting genotypes from which the extracted RNA is further processed for studying the expression of genes.

4.1 Selection of genotypes using snow mould resistance test

Selection of two genotypes, one having higher relative regrowth values (resistance) while the other having low relative regrowth values (suceptible) was done from the 10 genotypes of non-hardened *L. perenne* based on a snow mould resistance test. Initially all the 10 genotypes were grown in growth chamber for 4 weeks at 18ºC/20ºC (day/night temperatures) followed by inoculation and incubation(plants) for testing the plant varieties resistance to snow mould disease. A related study was conducted on winter wheat to study the effect of prehardening growth on resistance to snow mould which helps in screening test for resistant cultivars(Gaudet & Kozub 1991). Thus for screening of snow mould using a resistance test, extended incubation periods are required to estimate the level of resistance in plants which is usually conducted at incubation temperatures at or near to 0ºC which favours the snow mould fungi where high infection and plant mortality rate can be observed (Bruehl et al. 1966; Nakajima & Abe 1990). A similar experimental setup was used in our experiment where all the 10 genotypes were incubated at 2ºC in darkness for 3 different incubation periods and the results showed varying degree of resistance to snow mould infection under controlled conditions. The tiller survival rate and the relative regrowth values of the first and the second incubation period showed consistent relative regrowth values for genotypes M (high regrowth value) and F (low regrowth value),

while the third incubation period showed more variation among genotypes. This varying resistance could be associated with the rapid accumulation of high levels of carbohydrates in some cultivars (Kiyomoto & Bruehl 1977) and also could be associated with increasing plant size, older or larger plants (Bruehl 1967; Gaudet & Chen 1987). The study conducted by Nakajima & Abe showed that wheat cultivars under snow with temperatures around 0ºC and high humidity had reduced snow mould resistance because of the metabolism of carbohydrate food reserves stored during autumn (Kiyomoto & Bruehl 1977; Nakajima & Abe 1990). Based on the survival rate and relative regrowth values, genotype M (resistant) and genotype F (suceptible) were selected for further transcriptome studies using an RNA sequencing approach.

4.2 Selection of RNA samples based on high quality and concentration

Many problems were faced in extracting total RNA. We couldn't get high yield and quality may be because of high accumulation of phenolic compounds during prehardening growth conditions at 18ºC/20ºC. The selection of RNA samples was based on Nanodrop and Bioanalyzer results. The quality of RNA above 1.8 is considered as good RNA and the RIN value above 7 is considered as best for sequencing. Based on these parameters we selected 12 samples for sequencing.

4.3 Expression of specific genes PR-5 and PR-3 in response to snow mould disease

The relative gene expression of PR-5 varies between genotypes during different treatments. During 0 days (non-incubatedcontrol) there should be no expression but surprisingly we observed higher expression of gene PR-5 in genotype M than in the F genotype. This may be due to presence of some bacteria, viruses or fungi in soil which induce expression of this defence gene. The relative expression of PR-5 in M (inoculated) genotype is higher during the 1st day than in at day 4. Also we found the expression in non-inoculated plants of the M and F genotypes. This expression indicates that the PR-5 genes are induced by low temperature under dark conditions.

4.4 Transcriptome Analysis

Till date many plant transcriptomes have been sequenced and transcriptome studies on non-model plants will grow in the coming years because the sequencing costs goes down. More transcriptome studies help the researchers to choose the best approach of analysis based on their interest, but still there is a lot of work to be done. So for studying the transcritpomes of non-model organisms, *de novo* assembly is the best approach which helps in studying unique transcripts and expression levels in multiple organisms in parallel (Ward et al. 2012). In the present thesis, we studied the transcriptome of *L. perenne* using the Trinity platform (Haas et al. 2013) as the sequence data is very limited for this species so far in public databases. Our results showed that the N50 length generated using TRINITY assembly program with k-mer length 25 (as default) was 1,784 bp for genotype F and 1,672 bp for genotype M at which the best assembly was possible. The N50 value varies among the two genotypes since the total sum of assembled bases varied (292 million bases for F and 199 million for M). Thus, trimming of low quality bases at the ends of sequence reads and modifying the k-mer length parameters gives the best assembly output (like in our case K-mer =25)(Garg et al. 2011). The processed sequence read data of each biological replicate were then aligned back to the contigs from the assembly using Bowtie where we got 83% alignment and 94% alignment with the Bowtie-2 program. Though the read alignment with Bowtie-2 program was very good we were not be able to use that program as it is not supported by the Trinity pipeline. Then we generated output files from RSEM which tells us about transcript abundance and did pairwise comparisons of transcript abundance between samples using EdgeR to generate MA and volcano plots. Transcripts that are differentially expressed are seen in red colour at 0.05 FDR (False Discovery Rate) (Haas et al. 2013). A total of 275 genes were significantly differentially expressed in genotype M and 83 genes in genotype F.

Transcripts or genes are said to be significantly expressed basing on Fold change and FDR. Fold change is the method of identifying differentially expressed genes and FDR is the expected number of false positives within a group of positives(Benjamini & Hochberg 1995). Controlling of increased error while comparing multiple groups is done by FDR to find real significant differences(Benjamini et al. 2001). For example, an FDR of 0.05 means that in a set of 1000 genes that are predicted to be positive, 50 genes are said to be False positives.

4.4.1 Functional Annotation

Functional genomics is the most widely adopted in applied Bioinformatics in which functional interpretation is the key step for the analysis of data. An efficient functional annotation of DNA is the major requirement for the successful application of these approaches. Therefore Blast2go is an integrated, and biologist oriented solution for high-throughput sequence analysis based on gene ontology (Götz et al. 2008). So, in this study we used Blast2GO that regains GO terms based on BLAST to assign gene ontology annotation to the significantly expressed genes (Argout et al. 2008). These GO terms are standardized terms for annotating gene functions. All GO terms are categorized into three main groups: “Biological Process” (P), “Cellular component” (C) and “Molecular function” (F) (Argout et al. 2008). In my study I got around 83 genes that were significantly differentially expressed in the F genotype and around 275 genes in the M genotype. Out of the 275 genes that are retained by running Blast2GO program we are interested in 8 genes as they relate to responses to the fungus. Also there were few genes that are responding to carbohydrate stimulus, and surprisingly we have not found any genes in the F genotype which is the susceptible genotype. This could be due to the sequence distribution having a cutoff value as 5. However, in the overall sequence distribution pie chart we found that 9 genes are related to fungal responses. Also Fischer Exact test is useful to determine the specific GO terms affected by the disease (Martinelli et al. 2012). In our Fisher test there was no GO term related to fungal response, but there were GO terms related to monosaccharide metabolic process and hexose metabolic pathways. Between these pathways genotype F is more affected than genotype M.

4.5 Role of sequences related proteins to defence responses to fungus

There are about 9 sequences related to defence responses to fungus. They are: mlo protein homolog1-like, ubiquitin-like modifier-activating enzyme atg-7 like, tyrosine-protein phosphatase tpte-like, bbd1_orysi ame: full=bifunctional nuclease 1 protein, aspartic proteinase nepenthesin-2-like protein, proteasome subunit beta type 1 protein, o-glycosyl hydrolase superfamily protein, peroxisomal acyl-coenzyme a oxidase 1-like, and cytochrome p450 superfamily protein.

The sequence coding for mlo protein homolog 1-like is related to the defence response against powdery mildew fungus. It was shown that mlo protein was successful in controlling the powdery mildew fungus (*Blumeria graminis* f. sp. Hordei; Bgh) (Jørgensen 1992). In animals these proteins are known as G-protein-coupled receptors. It has been identified that a domain of MLO mediates a Ca2+-dependent interaction with calmodulin (Kim et al. 2002). The mlo protein of barley is the founder of a novel class of plant integral membrane proteins. Lack of the wild type protein of mlo leads to resistance to plant against pathogenic powdery mildew fungus which has been successfully implemented in agriculture (Devoto et al. 1999; Kim et al. 2002). Resistant mlo mutants against powdery-mildew were identified more than 60 years ago and had been used for more than 20 years (Elliott et al. 2002). In the same way this protein may be helpful in providing resistance against pink snow mould.

Also the other protein, ubiquitin-like modifier-activating enzyme ATG7, is mainly involved in autophagy. Autophagy is the process of degradation and nutrient elimination that occurs in all eukaryotes and in plants it is very important for the regulation of hypersensitive response programmed cell death (HR-PCD) during the plant innate immune response. This HR-PCD is initiaited upon pathogen recognition at site of infection by immune receptors, preventing the spread of HR-PCD to healthy tissues (Hayward et al. 2009). Although the importance of autophagy in innate immunity in mammals is well documented; it is not clear how autophagy contributes to plant innate immunity and cell death.

5 Conclusion

Pink snow mould is a difficult fungal disease to study experimentally, illustrated by the challenges of identifying genotypes with different resistant levels and separating the genes expressed as a result of fungal infection from those expressed due to low temperatures and incubation in darkness. New strategies have to be developed for improving resistance against pink snow mould in perennial ryegrass cultivars. In order to do this, it is very important to understand plant-fungus interactions and also one should identify the genes that are making the plant resistant to various stress factors. For this reason we have selected a next generation sequencing approach called RNA Sequencing that helps in studying the complete transcriptome of perennial ryegrass during early stages of inoculation with an aggressive *M. nivale* strain. RNA sequencing gave us a large number of sequences that are differentially expressed both as an effect of low temperature and incubation in darkness to simulate snow cover, and as an effect of

inoculation by pink snow mould. Detailed studies of these sequences can help us to understand the expression levels of particular sequences or genes during different treatments and conditions. Understanding of expression levels of various genes in turn helps us to identify genes that are actually making the plants resistant to fungal infections.

6 Future work

The present results do not permit us to understand completely the genes expressed during pink snow mould infection of non-hardened perennial ryegrass plants. However, a few interesting candidate genes have been identified and this is promising. Detailed studies of these defence related genes would be interesting and might lead to better understanding of the infection process and the development of resistance against pink snow mould in grasses.

References

Achuo, E., Prinsen, E. & Höfte, M. (2006). Influence of drought, salt stress and abscisic acid on the resistance of tomato to Botrytis cinerea and Oidium neolycopersici. *Plant Pathology*, 55 (2): 178-186.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3): 403-410.

Annadurai, R. S., Jayakumar, V., Mugasimangalam, R. C., Katta, M. A., Anand, S., Gopinathan, S., Sarma, S. P., Fernandes, S. J., Mullapudi, N. & Murugesan, S. (2012). Next generation sequencing and de novo transcriptome analysis of Costus pictus D. Don, a non-model plant with potent anti-diabetic properties. *BMC genomics*, 13 (1): 663.

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *New biotechnology*, 25 (4): 195-203.

Argout, X., Fouet, O., Wincker, P., Gramacho, K., Legavre, T., Sabau, X., Risterucci, A. M., Da Silva, C., Cascardo, J. & Allegre, M. (2008). Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of Theobroma cacao L. generated from various tissues and under various conditions. *BMC genomics*, 9 (1): 512.

Atkinson, N. J. & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of experimental botany*, 63 (10): 3523-3543.

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)*: 289-300.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural brain research*, 125 (1): 279-284.

Bruehl, G. (1967). Effect of plant size on resistance to snow mold of winter wheat. *Plant Dis. Rep*, 51: 815-819.

Bruehl, G. & Cunfer, B. (1971). Physiologic and environmental factors that affect the severity of snow mold of wheat. *Phytopathology*, 61: 792-799.

Bruehl, G. W., Sprague, R., Fischer, W., Nagamitsu, M., Nelson, W. & Vogel, O. (1966). *Snow molds of winter wheat in Washington*: Washington Agricultural Experiment Station, College of Agriculture, Washington State University.

Bullard, J., Purdom, E., Hansen, K. & Dudoit, S. (2010). *Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments*, BMC bioinformatics. 94 pp.

Chang, S., Chang, T., Tredway, L. & Jung, G. (2006a). Aggressiveness of Typhula ishikariensis isolates to cultivars of bentgrass species (Agrostis spp.) under controlled environment conditions. *Plant disease*, 90 (7): 951-956.

Chang, S., Scheef, E., Abler, R., Thomson, S., Johnson, P. & Jung, G. (2006b). *Distribution of Typhula spp. and Typhula ishikariensis varieties in Wisconsin, Utah, Michigan, and Minnesota*. 0031-949X. 926-933 pp.

Chen, G., Yin, K., Wang, C. & Shi, T. (2011). De novo transcriptome assembly of RNA-Seq reads with different strategies. *Science China Life Sciences*, 54 (12): 1129-1133.

Christie, P. J., Alfenito, M. R. & Walbot, V. (1994). Impact of low-temperature stress on general phenylpropanoid and anthocyanin pathways: enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. *Planta*, 194 (4): 541-549.

Comont, D., Winters, A., Gomez, L. D., McQueen-Mason, S. J. & Gwynn-Jones, D. (2013). Latitudinal variation in ambient UV-B radiation is an important determinant of Lolium perenne forage production, quality, and digestibility. *Journal of experimental botany*, 64 (8): 2193-2204.

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 (18): 3674-3676.

Conesa, A. & Götz, S. (2009). Blast2GO Tutorial.

Devoto, A., Piffanelli, P., Nilsson, I., Wallin, E., Panstruga, R., von Heijne, G. & Schulze-Lefert, P. (1999). Topology, subcellular localization, and sequence diversity of the Mlo family in plants. *Journal of Biological Chemistry*, 274 (49): 34993-35004.

Ebdon, J., Gagne, R. & Manley, R. (2002). Comparative cold tolerance in diverse turf quality genotypes of perennial ryegrass. *HortScience*, 37 (5): 826-830.

Elliott, C., Zhou, F., Spielmeyer, W., Panstruga, R. & Schulze-Lefert, P. (2002). Functional conservation of wheat and rice Mlo orthologs in defense modulation to the powdery mildew fungus. *Molecular plant-microbe interactions*, 15 (10): 1069-1077.

Espevig, T. (2011). *Winter hardiness and management of velvet bentgrass (Agrostis canina) on golf greens in the Nordic climate*: PhD dissertation. Norwegian University of Life Sciences.

Gams, W. & Müller, E. (1980). Conidiogenesis of Fusarium nivale and Rhynchosporium oryzae and its taxonomic implications. *Netherlands Journal of Plant Pathology*, 86 (1): 45-53.

Garg, R., Patel, R. K., Tyagi, A. K. & Jain, M. (2011). De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA research*, 18 (1): 53-63.

Gaudet, D. & Chen, T. (1987). Effects of hardening and plant age on development of resistance to cottony snow mold (Coprinus psychromorbidus) in winter wheat under controlled conditions. *Canadian journal of botany*, 65 (6): 1152-1156.

Gaudet, D. & Kozub, G. (1991). Screening winter wheat for resistance to cottony snow mold under controlled conditions. *Canadian Journal of Plant Science*, 71 (4): 957-965.

Gaudet, D., Laroche, A., Frick, M., Davoren, J., Puchalski, B. & Ergon, Å. (2000). Expression of plant defence-related (PR-protein) transcripts during hardening and dehardening of winter wheat. *Physiological and molecular plant pathology*, 57 (1): 15-24.

Gaudet, D. A., Laroche, A. & Yoshida, M. (1999). Low temperature-wheat-fungal interactions: A carbohydrate connection. *Physiologia Plantarum*, 106 (4): 437-444.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. & Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29 (7): 644-652.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, 36 (10): 3420-3435.

Hannaway, D., Fransen, S., Cropper, J. B., Teel, M., Chaney, M., Griggs, T., Halse, R. R., Hart, J. M., Cheeke, P. R. & Hansen, D. E. (1999). Perennial ryegrass (Lolium perenne L.).

Hayward, A. P., Tsao, J. & Dinesh-Kumar, S. (2009). *Autophagy and plant innate immunity: defense through degradation*. Seminars in cell & developmental biology: Elsevier. 1041-1047 pp.

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

Hofgaard, I. S., Vollsnes, A. V., Marum, P., Larsen, A. & Tronsmo, A. M. (2003). Variation in resistance to different winter stress factors within a full-sib family of perennial ryegrass. *Euphytica*, 134 (1): 61-75.

Hofgaard, I. S., Wanner, L. A., Hageskal, G., Henriksen, B., Klemsdal, S. S. & Tronsmo, A. M. (2006a). Isolates of Microdochium nivale and M-majus differentiated by pathogenicity on perennial ryegrass (Lolium perenne L.) and in vitro growth at low temperature. *Journal of Phytopathology*, 154 (5): 267-274.

Hofgaard, I. S., Wanner, L. A. & Tronsmo, A. M. (2006b). The effect of age and cold hardening on resistance to pink snow mould (Microdochium nivale) in perennial ryegrass (Lolium perenne L). *Acta Agriculturae Scandinavica Section B-Soil and Plant Science*, 56 (4): 315-323.

Hon, W.-C., Griffith, M., Chong, P. & Yang, D. S. (1994). Extraction and isolation of antifreeze proteins from winter rye (Secale cereale L.) leaves. *Plant Physiology*, 104 (3): 971-980.

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

Hoshino, T., Xiao, N. & Tkachenko, O. B. (2009). Cold adaptation in the phytopathogenic fungi causing snow molds. *Mycoscience*, 50 (1): 26-38.

Hsiang, T., Matsumoto, N. & Millett, S. M. (1999). Biology and management of Typhula snow molds of turfgrass. *Plant Disease*, 83 (9): 788-798.

Humphreys, M., Feuerstein, U., Vandewalle, M. & Baert, J. (2010). Ryegrasses. In *Fodder crops and amenity grasses*, pp. 211-260: Springer.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B. & Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8 (8): 1494-1512.

Iyer, M. K. & Chinnaiyan, A. M. (2011). RNA-Seq unleashed. *Nature biotechnology*, 29 (7): 599.

Jang, J.-C., León, P., Zhou, L. & Sheen, J. (1997). Hexokinase as a sugar sensor in higher plants. *The Plant Cell Online*, 9 (1): 5-19.

Jørgensen, I. H. (1992). Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica*, 63 (1-2): 141-152.

Kim, M. C., Panstruga, R., Elliott, C., Müller, J., Devoto, A., Yoon, H. W., Park, H. C., Cho, M. J. & Schulze-Lefert, P. (2002). Calmodulin interacts with MLO protein to regulate defence against mildew in barley. *Nature*, 416 (6879): 447-451.

Kiyomoto, R. & Bruehl, G. (1977). Carbohydrate accumulation and depletion by winter cereals differing in resistance to Typhula idahoensis. *Phytopathology*, 67: 206-211.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10 (3): R25.

Lee, J. M., Roche, J. R., Donaghy, D. J., Thrush, A. & Sathish, P. (2010). Validation of reference genes for quantitative RT-PCR studies of gene expression in perennial ryegrass (Lolium perenne L.). *BMC Molecular Biology*, 11 (1): 8.

Lees, A., Nicholson, P., Rezanoor, H. & Parry, D. (1995). Analysis of variation within< i> Microdochium nivale</i> from wheat: evidence for a distinct sub-group. *Mycological Research*, 99 (1): 103-109.

Li, B. & Dewey, C. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12 (1): 323.

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2< sup>− ΔΔCT</sup> Method. *methods*, 25 (4): 402-408.

Lu, J., Du, Z.-X., Kong, J., Chen, L.-N., Qiu, Y.-H., Li, G.-F., Meng, X.-H. & Zhu, S.-F. (2012). Transcriptome analysis of Nicotiana tabacum infected by Cucumber mosaic virus during systemic symptom development. *PloS one*, 7 (8): e43447.

Mardis, E. R. (2008a). The impact of next-generation sequencing technology on genetics. *Trends in genetics*, 24 (3): 133.

Mardis, E. R. (2008b). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 9: 387-402.

Martin, J. A. & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, 12 (10): 671-682.

Martinelli, F., Uratsu, S. L., Albrecht, U., Reagan, R. L., Phu, M. L., Britton, M., Buffalo, V., Fass, J., Leicht, E. & Zhao, W. (2012). Transcriptome profiling of citrus fruit response to huanglongbing disease. *PloS one*, 7 (5): e38039.

Matsumoto, N. (2009). Snow molds: a group of fungi that prevail under snow. *Microbes and Environments*, 24 (1): 14-20.

Metzker, M. L. (2009). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11 (1): 31-46.

Meyers, B. C., Tej, S. S., Vu, T. H., Haudenschild, C. D., Agrawal, V., Edberg, S. B., Ghazal, H. & Decola, S. (2004). The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. *Genome research*, 14 (8): 1641-1653.

Mittler, R. & Blumwald, E. (2010). Genetic engineering for modern agriculture: challenges and perspectives. *Annual review of plant biology*, 61: 443-462.

Nakajima, T. & Abe, J. (1990). A method for assessing resistance to the snow molds Typhula incarnata and Microdochium nivale in winter wheat incubated at the optimum growth temperature ranges of the fungi. *Canadian journal of botany*, 68 (2): 343-346.

Ozsolak, F. (2012). Third-generation sequencing techniques and applications to drug discovery. *Expert opinion on drug discovery*, 7 (3): 231-243.

Pérez-de-Castro, A., Vilanova, S., Cañizares, J., Pascual, L., Blanca, J., Díez, M., Prohens, J. & Picó, B. (2012). Application of genomic tools in plant breeding. *Current genomics*, 13 (3): 179.

Pociecha, E. & Płażek, A. (2010). Cold acclimation of forage grasses in relation to pink snow mould (Microdochium nivale) resistance. *Acta Physiologiae Plantarum*, 32 (1): 37-43.

Posselt, U. K. & Altpeter, F. (1994). Improvement of snow mould resistance by conventional and in vitro techniques. *Euphytica*, 77 (3): 251-255.

Prończuk, M. & Messyasz, M. (1991). Infection ability of mycelium and spores ofMicrodochium nivale (Fr) samuels hallett toLolium Perenne L. *Mycotoxin research*, 7 (2): 136-139.

Robinson, M. D. & Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 23 (21): 2881-2887.

Rolland, F., Moore, B. & Sheen, J. (2002). Sugar sensing and signaling in plants. *The Plant Cell Online*, 14 (suppl 1): S185-S205.

Rudi, H., Sandve, S. R., Opseth, L. M., Larsen, A. & Rognli, O. A. (2011). Identification of candidate genes important for frost tolerance in< i> Festuca pratensis</i> Huds. by transcriptional profiling. *Plant Science*, 180 (1): 78-85.

Ruttink, T., Sterck, L., Rohde, A., Bendixen, C., Rouzé, P., Asp, T., Van de Peer, Y. & Roldan-Ruiz, I. (2013). Orthology Guided Assembly in highly heterozygous crops: creating a reference transcriptome to uncover genetic diversity in Lolium perenne. *Plant biotechnology journal*.

Shinozuka, H., Cogan, N. O., Spangenberg, G. C. & Forster, J. W. (2012). Quantitative Trait Locus (QTL) meta-analysis and comparative genomics for candidate gene prediction in perennial ryegrass (Lolium perenne L.). *BMC genetics*, 13 (1): 101.

Smeekens, S. (2000). Sugar-induced signal transduction in plants. *Annual review of plant biology*, 51 (1): 49-81.

Soneson, C. & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC bioinformatics*, 14 (1): 91.

Stranneheim, H. & Lundeberg, J. (2012). Stepping stones in DNA sequencing. *Biotechnology Journal*, 7 (9): 1063-1073.

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome research*, 21 (12): 2213-2223.

Tronsmo, A. M. (1993). Resistance to Winter Stress Factors in Half-Sib Families of Dactylis glomerata, Tested in a Controlled Environment. *Acta Agriculturae Scandinavica, Section B - Soil & Plant Science*, 43 (2): 89-96.

Tronsmo, A. M., Hsiang, T., Okuyama, H. & Nakajima, T. (2001). Low temperature diseases caused by Microdochium nivale. *Low temperature plant microbe interactions under snow. Hokkaido National Agricultural Experiment Station, Sapporo*: 75-86.

Varshney, R. K., Nayak, S. N., May, G. D. & Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in biotechnology*, 27 (9): 522-530.

Wang, W., Vinocur, B. & Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta*, 218 (1): 1-14.

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10 (1): 57-63.

Ward, J. A., Ponnala, L. & Weber, C. A. (2012). Strategies for transcriptome analysis in nonmodel plants. *American Journal of Botany*, 99 (2): 267-276.

Wiese, J., Kranz, T. & Schubert, S. (2004). Induction of pathogen resistance in barley by abiotic stress. *Plant Biology*, 6 (5): 529-536.

Xiao, W., Sheen, J. & Jang, J.-C. (2000). The role of hexokinase in plant sugar signal transduction and growth and development. *Plant molecular biology*, 44 (4): 451-461.

Xu, L., Zhu, L., Tu, L., Liu, L., Yuan, D., Jin, L., Long, L. & Zhang, X. (2011). Lignin metabolism has a central role in the resistance of cotton to the wilt fungus Verticillium dahliae as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *Journal of experimental botany*, 62 (15): 5607-5621.

Yu, X., Bai, G., Liu, S., Luo, N., Wang, Y., Richmond, D. S., Pijut, P. M., Jackson, S. A., Yu, J. & Jiang, Y. (2013). Association of candidate genes with drought tolerance traits in diverse perennial ryegrass accessions. *Journal of experimental botany*.

Zhang, C., Fei, S.-z., Warnke, S., Li, L. & Hannapel, D. (2009). Identification of genes associated with cold acclimation in perennial ryegrass. *Journal of plant physiology*, 166 (13): 1436-1445.

Zhang, N., Zhang, S., Borchert, S., Richardson, K. & Schmid, J. (2011). High levels of a fungal superoxide dismutase and increased concentration of a PR-10 plant protein in associations between the endophytic fungus Neotyphodium lolii and ryegrass. *Molecular Plant-Microbe Interactions*, 24 (8): 984-992.

Appendixes:

Materials used in the study

Chemicals and solutions	**Supplier**
1-kb ladder	
2-mercaptoethanol	invitrogen
Ethanol	Kemetyl
Liquid nitrogen	
RNA Extraction Buffers	Ambion, Life technologies
Lysis Buffer RNA Isolation Aid Wash Buffer I Wash Buffer II	,, ,, ,, ,,
RNase-free water	Ambion
DEPC treated water	Invitrogen
dH20	

MilliQ water	Millipore
Enzymes	
Taq DNA polymerase	Invitrogen
DNase I	Invitrogen
Dyes	
SyBr dye	Invitrogen
Kits	
Express two-step qRt-PCR universal kit	Invitrogen
PureLink® RNA Mini Kit	Ambion
RNA 6000 Nano reagent and supplies	Agilent Technology
Equipments	
7500 Fast Real-Time PCR instrument	Applied Biosystems
Fast Optical 96 well plate with barcode	Applied Biosystems
Agilent 2100 Bioanalyzer	Agilent technologies
Collection tubes	Ambion
Eppendorf tubes	Ambion
Spin Cartridge columns	Ambion
Freezer (-20)	Bosch
Freezer (-80)	Sanyo
Laminar air flow chamber	
Micro centrifuge	
NanoDrop® ND 1000	NanoDrop® technologies
Thermal cycler	Applied Biosystems
Software	
Primer express software	Applied Biosystems
Minitab	

Softwares for *De novo* assembly	
FastQC	
Sickle	
Trinity Assembler	
Bowtie	
RSEM	
EdgeR	
Blast2GO program	

Appendix 1: Isolation of Total RNA from Plant Cells

Procedure:

1. Add liquid nitrogen to an RNase-free mortar and grind fresh leaves or tissue thoroughly using an RNase-free pestle.
2. Weigh the grinded plant material ≤ 100 mg and transfer to an RNase-free microcentifuge tube that has cooled on liquid nitrogen.
3. Immediately add 0.5 ml of lysis buffer and 80 µl of RNA Isolation Aid to the weighed leaf powder (approx. 80 mg).
4. Vortex the lysate and incubate for 3 minutes at room temperature.
5. Transfer 0.5 ml of the lysate to a homogenizer inserted in RNase-free tube and centrifuge at 12,000 x g for 2 minutes. Remove the Homogenizer when done.
6. Then add 0.5 volume 96-100% ethanol to each volume of tissue homogenate and mix thorougly to scatter any visible precipitate that may be formed after adding ethanol.
7. Transfer up to 700 µl of our sample to Spin Cartridge.
8. Centrifuge at 12,000 x g for 15 sec at room temperature. Discard the flow-through, and re-insert the Spin Cartidge in the same collection tube.
9. Repeat steps 7-8 until the entire sample is processed.

Optional: On-column PureLink® DNase Treatment

To degrade the genomic DNA using DNase enzyme

10. Add 350 µl Wash Buffer I to the Spin Cartridge containing the bound RNA, and centrifuge at 12,000 x g for 15 sec at room temperature. Discard the flow-through and the collection tube and insert the Spin Cartridge into a new collection tube.
11. Add 80 µl PureLink® DNase mixture directly onto the surface of the Spin Cartridge membrane and incubate at room temperature for 15 minutes.
12. Repeat the step 10.
13. Add 500 µl Wash Buffer II with ethanol to the Spin Cartridge.
14. Centrifuge at 12,000 x g for 15 sec at room temperature and discard the flow-through snd reinsert the Spin Cartridge into the same Collection tube.
15. Repeat steps 10 and 11, once.
16. Centrifuge the Spin Cartridge at 12,000 x g for 1 minute to dry the membrane with bound RNA and discard the collection tube and reinsert the Spin Cartridge into recovery tube.
17. Add 30 µl-100 µl RNase-free water to the center of the Spin Cartridge
18. Incubate at room temperature for 1 minute.
19. Centrifuge Spin Cartridge and recovery tube for 1 minute at 12,000 x g at room temperature.
20. Store the purified RNA until further usage.

Appendix 2: Analyzing RNA yield and Quality using Agilent RNA 6000 Nano Assay

Procedure:

A. Loading the Gel-Dye Mix

1. Allow the gel-dye mix to equilibrate to room temperature for 30 min before use and protect the gel-dye mix from light during this time.

2. Take a new RNA Nano chip out and place on the chip priming station.
3. Pipette 9 µl of the gel-dye mix at the bottom well marked as **G** and dispense the gel-dye mix.
4. Set the timer to 30 sec, make sure that the plunger is positioned at 1 ml and then close the chip priming station. The lock of the latch will click when the Priming Station is closed correctly.
5. Press the plunger of the syringe down until it is held by the clip.
6. Wait for exactly 30 sec and then release the plunger with the clip release mechanism.
7. Visually inspect that the plunger moves back at least to the 0.3 ml mark.
8. Wait for 5 sec, and then slowly pull back the plunger to the 1 ml position.
9. Open the chip priming station.
10. Pipette 9 µl of the gel-dye mix in each of the wells marked **G**

B. Loading the RNA 6000 Nano Marker

1. Pipette 5 µl of the RNA 6000 Nano marker (green) into the Well marked with the ladder symbol and eaxh of the 12 sample well.

C. Loading the ladder and samples

1. Before use, thaw ladder aliquots and keep them on ice (avoid extensive warming upon thawing process)
2. To minimize secondary structure, heat denature (70°C, 2 minutes) the samples before loading on the chip.
3. Pipette 1 µl of the RNA ladder into the well marked with the ladder symbol.
4. Pipette 1 µl of each sample into each of the 12 sample wells.
5. Set the timer to 60 sec.
6. Place the chip horizontally in the adapter of the IKA vortex mixer and make sure not to damage the buldge that fixes the chip during vortexing.
7. Vortex for 60 sec at 2400 rpm.
8. Insert the chip in the Agilent 2100 bioanalyzer. Make sure that the run is started within 5 minutes.

Appendix 4: Primer sequence used for qPCR

Name	**Sequence 5'-3'**
LpPR3-F	ACCCCGACTCTTCCCTT
LpPR3-R	ATGATGTTGGTGATCACGCCG
LpPR5-F	GCAGCTGAACAGCGGCGAGACGTGGAAC
LpPR5-R	GCCGGTGCTGCAGGAGAAGCCCATGC
LpGAPDH-F	CATCACCATTGTCTCCAACG
LpGAPDH-R	AACCTTCAACGATGCCAAAC

Appendix 5: cDNA Synthesis Protocol

cDNA synthesis has been carried out by using SuperScriptR VILO™ cDNA Synthesis kit.

Procedure:

1. For a single reaction, combine the following components in a tube on ice.
   a. 4 µl of 5X VILO™ Reaction Mix.
   b. 2 µl of 10X super ScriptR Enzyme Mix.
   c. 2.5 µg of total RNA (template).
   d. 14 µl of DEPC-treated water.
2. Gently mix the contents and incubate at 25°C for 10 min.
3. Incubate tube at 42°C for 60 min.
4. Terminate the reaction at 85°C at 5 min.
5. Store the cDNA at -20°C until use.

Appendix 6: qPCR

Procedure:

1. Set up the reaction on ice.
2. Prepare master mix containing the components (table. A).
3. Add 18 µl of master mix in to 96 well-PCR plate and add 2 µl of cDNA template of the gene of interest.
4. Prepare non-template control (NTC) reactions to test for DNA contamination of the enzyme/primer mixes.
5. Cap or seal each PCR plate, and gently mix. Make sure that all components are at the bottom of the tube/plate; centrifuge briefly at 1500 rpm for 2 min.
6. Place reaction in a real-time instrument programmed as described in methods.
7. Collect data and analyze results.

Table. A. Master mix protocol

Concentration	**Master mix components**	**Amount per reaction** (µl)
	cDNA	5
	Sybr Green	10
	water	4
10 µM (final 250nM)	Forward primer	0,5
10 µM (final 250nM)	Reverse primer	0,5
	sum	20

Appendix 7: Trinity Pipeline

The following commands were given during the processing and analysis of data:

Processing of raw sequencing data:

tar files (receive raw sequencing data)

↓

extract.tar files

↓

bunzip or gzip files

↓

decompress gzip files

↓

fastq files

↓

/usr/ryegrass/bin/fastqc file name (intial quality check)

↓

Sickle pe –f file name.fastq –r file name.fastq –t sanger –o filename.fastq-QT –p file name.fastq-QT –s *_singles –l 25 & (Trimming)

Assembly of processed raw sequencing data:

Trinity to generate transcriptome assembly

$TRINITYDIR/Trinity.pl–seqType fq –left all_samples.R1.fastq.gz –right all_samples.R2.fastq.gz –no_cleanup –JM 100G –CPU 8 –inchworm CPU 8 –bflyHeapSpaceMax 20G –bfly CPU 4

Output of Trinity

trinity_out_dir/Trinity.fasta

Will get the output of transcript assembly which is then further processed.

Mapping of reads to transcript assembly (Bowtie)

$TRINITYDIR/util/alignReads.pl –seqType fq –left sampleX.R1.fastq –right sampleX.R2.fastq –target trinity_out_dir/Trinity.fasta –aligner bowtie –o sampleX.bowtie_output -- -p 16

Bowtie output (BAM files)

sampleX.bowtie_output

Transcript abundance with RSEM

$TRINITYDIR/util/RSEM_util/run_RSEM.pl--transcriptstrinity_out_dir/Trinity.fasta--name_sorted_bam sampleX.bowtie_output/sample.bowtir_output_nameSorted.bam --paired

Output tables with transcript abundance

RSEM.genes.results

RSEM.isoforms.results

Identifying differentially Expressed Transcripts by edgeR

$TRINITY_HOME/Analysis/DifferentialExpression/run_DE_analysis.pl --matrix SP2.rnaseq.counts.matrix --method edgeR --samples_file samples_described.txt

Analyzing the differentially expressed genes

$TRINITY_HOME/Analysis/DifferentialExpression/analyze_diff_expr.pl --matrix matrix.TMM_normalized.FPKM -P 1e-3 -C 2

After processing of the sequencing data, the edgeR files are imported to Blast2GO program where we get different bar graphs, pie charts.