**MARMARA UNIVERSITY**
**INSTITUTE FOR GRADUATE STUDIES IN PURE AND APPLIED SCIENCES**

# TRANSCRIPTOME ANALYSIS OF MORPHINE BIOSYNTHESIS IN OPIUM POPPY (*Papaver somniferum* L.)

KUAYBE YÜCEBİLGİLİ KURTOĞLU

**Ph.D. THESIS**
Department of Biology

**Thesis Supervisor**
Prof. Dr. İbrahim İlker ÖZYİĞİT

**Thesis CO- Supervisor**
Assist. Prof. Dr. Tuğba GÜRKÖK

ISTANBUL, 2016

**MARMARA UNIVERSITY**
**INSTITUTE FOR GRADUATE STUDIES IN PURE AND APPLIED SCIENCES**

# TRANSCRIPTOME ANALYSIS OF MORPHINE BIOSYNTHESIS IN OPIUM POPPY (*Papaver somniferum* L.)

KUAYBE YÜCEBİLGİLİ KURTOĞLU

(720111701)

**Ph.D. THESIS**
Department of Biology

**Thesis Supervisor**
Prof. Dr. İbrahim İlker ÖZYİĞİT

**Thesis CO- Supervisor**
Assist. Prof. Dr. Tuğba GÜRKÖK

ISTANBUL, 2016

# ACKNOWLEDGMENT

I would like to express my special appreciation and thanks to my advisor Prof. Dr. İbrahim İlker ÖZYİĞİT, he has been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. I would also express my thanks to Asst. Prof. Dr. Tuğba GÜRKÖK for her guidance, assistance and sincere friendship as a co-advisor during my PhD thesis.

I would also like to thank all my committee members; Assoc. Prof. Dr. Turgay ÜNVER for his support on my thesis research which have been priceless, and Assoc. Prof. Dr. Filiz VARDAR for her brilliant comments and suggestions. In addition, I also want to express my gratitude to the Jury members; Prof. Dr. Nermin GÖZÜKIRMIZI for her valuable evaluation on my thesis and Assoc. Prof. Dr. Yıldız AYDIN for her polite and lovely attitude towards me and being so helpful in both my academic career and personal life.

I also thank The Scientific and Technological Research Council of Turkey (TÜBİTAK), BİDEB-2211 National Graduate Student Scholarship Program for supporting me financially during my education.

A very special thanks to my precious lifetime best friend and love. Words cannot express how grateful I am to my beloved husband for his infinite support, patience and encouragement throughout these years. Without him this thesis would never be possible. I also express my thanks and appreciation to our families for their understanding, motivation and patience.

Lastly, but in no sense the least, I am thankful to all colleagues and friends who made my stay at the university a memorable and valuable experience.

This work was supported by Marmara University through Scientific Research Project (FEN-C-DRP-080415-0120).

**June, 2016** **Kuaybe YÜCEBİLGİLİ KURTOĞLU**

# TABLE OF CONTENTS

**ACKNOWLEDGMENT** ..... **i**

**TABLE OF CONTENTS** ..... **ii**

**ABSTRACT** ..... **vi**

**ÖZET** ..... **vii**

**CLAIM FOR ORIGINALTY** ..... **viii**

**SYMBOLS** ..... **ix**

**ABBREVIATIONS** ..... **x**

**LIST OF FIGURES** ..... **xiv**

**LIST OF TABLES** ..... **xvi**

**1. INTRODUCTION** ..... **1**

1.1. General Characteristics of *Papaver somniferum* L. ..... 1

1.2. History of Opium Poppy and Its Cultivation ..... 2

1.3. Secondary Metabolites ..... 5

1.3.1. Alkaloids ..... 5

1.3.2. Benzylisoquinoline alkaloids (BIA) and opium poppy ..... 6

1.3.3. Biosynthesis of the major alkaloids in opium poppy ..... 7

1.3.4. (S)-norcoclaurine to (S)-reticuline ..... 8

1.3.4.1. Papaverine biosynthesis ..... 8

1.3.4.2. Protoberberine, protopine and benzophenanthridine biosynthesis ..... 10

1.3.4.3. Noscapine biosynthesis ..... 11

1.3.4.4. Morpine biosynthesis ..... 12

1.3.5. MeJA and its effect on plants ..... 13

1.3.6. Approaches to establish genomic resources for specialized metabolisms ... 14

1.3.7. Omics studies on opium poppy ..... 16

1.3.8. Regulation of gene expression ..... 18
1.3.9. Metabolic engineering in opium poppy ..... 19
1.3.10. Compartmentalization and trafficking of BIA metabolism ..... 23
1.3.10.1. Cellular localization and transport ..... 23
1.3.10.2. Subcellular compartmentalization of BIA enzymes ..... 26
*1.3.10.2.1.* Metabolic channeling ..... 28
1.4. State-of-the-art Sequencing Technologies Used in Plant Omics Research ..... 29
1.4.1. Sanger sequencing ..... 29
1.4.2. Next generation sequencing ..... 30
1.4.2.1. Roche 454 sequencing ..... 30
1.4.2.2. ABI/Solid sequencing ..... 32
1.4.2.3. Illumina/Solexa platform ..... 33
*1.4.2.3.1.* Comparison of Illumina HiSeq and 454 GS-FLX platforms ..... 36
1.4.3. Third-Generation Sequencing ..... 37
1.4.4. Transcriptome sequencing for gene expression profiling in plants ..... 38
1.4.4.1. Microarray technology ..... 39
1.4.4.2. cDNA libraries, ESTs, SAGE and RNAseq ..... 39
1.4.4.3. Advantages of RNAseq and comparison with other methods ..... 41
1.4.4.4. Challenges and limitations of RNAseq ..... 43
1.4.4.5. RNAseq data analysis ..... 44
**2. MATERIAL AND METHODS ..... 47**
2.1. Plant Material and Growth Conditions ..... 47
2.2. MeJA Treatment ..... 48
2.3. Analysis of Alkaloid Contents by ToF LC/MS ..... 49
2.4. RNA Isolation and Characterization ..... 50

2.5. Library Construction for Illumina Sequencing ..... 51
2.6. Cluster Generation and Transcriptome Sequencing ..... 53
2.7. Raw Data Processing ..... 53
2.8. *De novo* Transcriptome Assembly ..... 54
2.9. Reference Transcriptome Dataset ..... 56
2.10. Transcriptome Annotation ..... 56
2.10.1. Functional annotation of transcripts: GO and KEGG analysis ..... 56
2.10.2. Unigene CDS prediction ..... 58
2.10.3. Detection of gene expression levels ..... 58
2.10.4. Detection of differentially expressed genes (DEGs) ..... 59
2.10.5. Gene ontology analysis of DEGs ..... 59
2.10.6. Pathway analysis of DEGs ..... 59
**3. RESULTS AND DISCUSSION ..... 60**
3.1. Alkaloid Content Analysis of Capsule and Stem Upon MeJA Treatment ..... 60
3.2. RNA Quality Assessment and Quantitation ..... 65
3.3. Transcriptome Sequencing & Analysis ..... 67
3.3.1. Illumina HiSeq RNA-Seq results ..... 67
3.3.2. *De novo* transcriptome assembly ..... 70
3.4. Transcriptome Annotation ..... 73
3.4.1.1. MRCT/MRCC comparison ..... 76
3.4.1.2. MRST/MRSC comparison ..... 79
3.4.1.3. MPCT/MPCC comparison ..... 82
3.4.1.4. MPST/MPSC comparison ..... 85
3.4.1.5. MRST/MRCT comparison ..... 88
3.4.1.6. MPST/MPCT comparison ..... 91

3.4.1.7. MPCT/MRCT comparison ..... 94
3.4.1.8. MPST/MRST comparison ..... 97
3.5. KEGG Pathway Analysis ..... 101
3.6. Integrative Correlation of Metabolic and RNA-Seq Transcriptomic Data ..... 115
3.6.1. Digital gene expression analysis ..... 116
3.6.2. DEGs in MeJA applied opium poppy cultivars and tissues ..... 117
**4. CONCLUSION ..... 126**
**REFERENCES ..... 128**
**SUPPLEMENTARY ..... 148**
**RESUME ..... 180**

# ABSTRACT


**TRANSCRIPTOME ANALYSIS OF MORPHINE BIOSYNTHESIS**

**IN OPIUM POPPY (*Papaver somniferum* L.)**

Opium poppy (*Papaver somniferum* L.) is one of the most ancient medicinal plants of the world. Its medicinal and economic importance is highly dependent on the benzylisoquinoline alkaloid (BIA) production. Morphine is one of the most important and abundant BIA commercially obtained opium poppy. It is mainly used as narcotic analgesic in medicine. In this thesis, a total of 36,16 Gb sequence data was generated by Illumina HiSeq platform and transcriptome profiles of elicitor (MeJA) induced capsule and stem tissue from morphine rich (MR) and morphine poor (MP) cultivars are compared with an integrative metabolite and transcriptome approach. Alkaloid analysis confirmed the induction of morphine levels in both MR and MP capsule tissues. Based on GO and KEGG pathway analysis, biologic processes related to stress response, signal transduction and seconder metabolite biosynthesis are found to be prominent. Aquaporin TIP, 2ODD, bHLH113, Calmodulin, (+)-neomenthol dehydrogenase, (S)-tetrahydroprotoberberine N-methyltransferase, major latex protein 146, NRT1/PTR family, F-box/LRR-repeat protein coding transcripts are found to be the most differentially expressed transcripts. Besides, morphine biosynthesis route-specific transcripts corresponding to SalSyn, SalAT, SalR, T6ODM, COR and CODM enzymes were found to be differentially expressed in different tissues and cultivars. This thesis produced the largest transcriptome data obtained so far and is an important source for further functional genomic studies to investigate BIA biosynthesis mechanism in opium poppy.

**KUAYBE YÜCEBİLGİLİ KURTOĞLU**

# ÖZET


## HAŞHAŞ (*Papaver somniferum* L.)'TA TRANSKRİPTOM ANALİZİ İLE MORFİN SENTEZ MEKANİZMASININ ARAŞTIRILMASI

Haşhaş (*Papaver somniferum* L.) dünyanın en eski tıbbi bitkilerinden biridir. Tıbbi ve dolayısıyla ekonomik değeri haşhaşın benzilizokuinolin (BIA) üretimine dayanmaktadır. Morfin, haşhaşta yoğun olarak bulunan önemli bir BIA alkaloidi olup tıpta narkotik ağrı kesici ve uyuşturucu olarak kullanılmaktadır. Bu tezde Illumina HiSeq plantformu kullanılarak toplam 36,16 Gb dizi verisi elde edilmiş ve MeJA elisitörü ile indüklenen morfince zengin (MR) ve fakir (MP) haşhaş varyetelerinin kapsül ve gövde dokularındaki metabolit profilleriyle birlikte bütüncül bir yaklaşımla kıyaslanarak değerlendirilmiştir. Alkaloid analizleri her iki varyetede de kapsül dokularında morfin düzeylerinin indüklendiğini göstermiştir. GO ve KEGG analizlerinde, stres yanıtı, sinyal iletimi ve sekondar metabolit biyosentezi ile ilgili biyolojik işlemlerin öne çıktığı tesbit edilmiştir. Aquaporin TIP, 2ODD, bHLH113, calmodulin, (+)-neomenthol dehydrogenase, (S)-tetrahydroprotoberberine N-methyltransferase, major latex protein 146, NRT1/PTR Family proteins, F-box/LRR-repeat proteinlerini kodlayan transkriptlerin MeJA uygulamasıyla ifade düzeyleri en çok değişen genler olduğu belirlenmiştir. Ayrıca, BIA biyosentezine özgün SalSyn, SalAT, SalR, T6ODM, COR ve CODM enzimlerine ait transkriptlerde dokuya ve varyeteye göre değişim olduğu saptanmıştır. Bu tezde bugüne kadar elde edilmiş en kapsamlı transkriptom verisi üretilmiş olup bu bilgiler haşhaşta BIA biyosentez mekanizmasının araştırılacağı ileri işlevsel genomik çalışmalar için kaynak oluşturacak niteliktedir.

**KUAYBE YÜCEBİLGİLİ KURTOĞLU**

# CLAIM FOR ORIGINALTY

I hereby declare that this thesis comprises my original work. No material in this thesis has been previously published and written by another person, except where due reference is made in the text. I further declare that this thesis contains no material which has been submitted for a degree or diploma or other qualifications at any other university.

**Kuaybe YÜCEBİLGİLİ KURTOĞLU**

**June, 2016**

# SYMBOLS

| | |
|---|---|
| °C | Celsius |
| µg | Microgram |
| µl | Microliter |
| µm | Micrometer |
| µM | Micromolar |

# ABBREVIATIONS

| | |
|---|---|
| 4-HPAA | 4-hydroxyphenylacetaldehyde |
| 4OMT | 3-hydroxy-N-methylcoclaurine 4′methyltransferase |
| 6OMT | Norcoclaurine 6-O-methyltransferase |
| 7OMT | Reticuline7-O-methyltransferase |
| BBE | Berberine bridge enzyme |
| BIA | Benzylisoquinoline alkaloids |
| BLAST | Basic local alignment search tool |
| bp | Base pair |
| CAS | Canadine syhthase |
| CDS | Coding DNA sequence |
| CFS | Cheilanthifoline synthase |
| CheSyn | Cheilanthifoline synthase |
| CNMT | Coclaurine N-methyltransferase |
| CODM | Codeine-O-demethylase |
| CoOMT | Columbamine O-methyltransferase |
| COR | Codeinone reductase |
| CYP82Y1 | N-methylcanadine 1-hydroxylase |
| DBOX | Dihydrobenzo-phenanthridine oxidase |
| ddNTP | Dideoxynucleotide |
| DEG | Differentially expressed genes |
| DEPC | Diethylpyrocarbonate |
| DNase | Dioxyribonuclease |
| dNTPs | Deoxynucleotide-triphosphate |
| DRR | 1,2-dehydroreticuline reductase |
| DRS | 1,2-dehydroreticuline synthase |
| emPCR | Emulsion PCR |
| EMS | Ethyl methanesulfonate |
| ER | Endoplasmic reticulum |

| | |
|---|---|
| ESI | Electro spray ionization |
| ESTs | Expressed sequence tags |
| eTMs | Endogenous target mimics |
| FDR | False discovery rate |
| FPKM | Fragments per kilo base per million of mapped reads |
| FT-ICR-MS | Fourier transform ion cyclotron resonance mass spectrometry |
| GO | Gene ontology |
| GS | Genome sequencer |
| HPLC | High-pressure liquid chromatography |
| HQ | High quality |
| KEGG | Kyoto encyclopedia of genes and genomes |
| lncRNAs | Long noncoding rnas |
| MALDI | Matrix-assisted laser desorption/ionization |
| MeJA | Methyl jasmonate |
| miRNAs | Micrornas |
| MP | Morphine poor |
| MPCC | Mophine poor capsule control |
| MPCT | Mophine poor capsule treatment |
| MPSC | Mophine poor stem capsule |
| MPSS | Massively parallel signature sequencing |
| MPST | Mophine poor stem treatment |
| MR | Morphine rich |
| MRCC | Mophine rich capsule control |
| MRCT | Mophine rich capsule treatment |
| MRSC | Mophine rich stem control |
| MRST | Mophine rich stem treatment |
| MS | Mass spectroscopy |
| MSH | N-methylstylopine 14-hydroxylase |
| N7OMT | 7-O-methyltransferase |
| NCS | Norcoclaurine synthase |

| | |
|---|---|
| NGS | Next generation sequencing |
| NMCH | N-methylcoclaurine-3’-hydroxylase |
| NOS | Noscapine synthase |
| Nr | Non-redundant |
| P6H | Protopine 6-hydroxylase |
| PCA | Principal component analysis |
| PDS | Phytoene desaturase |
| PPi | Pyrophosphate |
| PTGS | Post-transcriptional gene silencing |
| PTP | Picotiter plate |
| qRT-PCR | Qualitative real time PCR |
| REPI | Reticuline epimerase |
| REPI | Reticuline epimerase fusion |
| RIN | RNA integrity |
| RNAi | RNA interference |
| RNAseq | RNA sequencing |
| RPKM | Reads per kilo base per million of mapped reads |
| SAGE | Serial analysis of gene expression |
| SalAT | Salutardinol 7-O-acetyltransferase |
| SalR | Salutardine reductase |
| SalSyn (Cyp719B1) | Salutardine synthase |
| SanR | Sanguinarine reductase |
| SBS | Sequencing-by-synthesis |
| siRNAs | Small interfering RNA |
| SNP | Single nucleotide polymorphism |
| SOMT | Scoulerine 9-O-methyltransferase |
| ssDNA | Single-stranded DNA |
| SSRs | Short sequence repeats |
| STORR | *S* to *r*-reticuline |
| STOX | Tetrahydroprotoberine oxidase |
| STS | Stylopine synthase |

| | |
|---|---|
| StySyn | Stylopine synthase |
| T6ODM | Thebaine 6-O-demethylase |
| TF | Transcription factor |
| TGICL | Tigr gene indices clustering tools |
| TMO | Turkish grain board |
| TNMT | Tetrahydroprotoberberine cis-N-methyltransferase |
| ToF-LC/MS | Time-of-flight liquid chromatography/ Mass spectroscopy |
| TRV | Tobacco rattle virus |
| TYDC | Tyrosin/dopa decarboxylase |
| TyrAT | Tyrosine aminotransferase |
| VIGS | Virus-induced gene silencing |

# LIST OF FIGURES

Figure 1.1 *Papaver somniferum* L. ..... 1
Figure 1.2. Legal opium poppy cultivation areas ..... 3
Figure 1.3. Cities allowed to cultivate opium poppy in Turkey ..... 4
Figure 1.4. Production of opiate raw materials rich in morphine ..... 4
Figure 1.5. Major alkaloids in opium poppy ..... 6
Figure 1.6. Biosynthesis pathway of BIA alkaloids in opium poppy ..... 9
Figure 1.7. Cellular localization of morphine biosynthetic enzymes ..... 26
Figure 1.8. Subcellular localization of BIA pathway enzymes in sanguinarine biosynthesis ..... 28
Figure 1.9. Sequencing-by-synthesis in 454 GS-FLX Platform ..... 32
Figure 1.10. Overview of Illumina sequencing ..... 35
Figure 1.11. Steps of a typical RNAseq experiment ..... 41
Figure 1.12. *De novo* transcriptome analysis flow chart ..... 45
Figure 2.1. Opium poppy grown in green house ..... 47
Figure 2.2. MeJA treatment and further analysis ..... 48
Figure 2.3. Agilent 1260 infinity LC and 6210 ToF-LC/MS Sytems (www.agilent.com) ..... 50
Figure 2.4. TruSeq Total RNA sequencing library preparation workflow ..... 52
Figure 2.5. Schematic overview of the assembly process. ..... 55
Figure 3.1. Morphine contents of a) capsule and b) stem in different time points. ..... 61
Figure 3.2. Comparative alkaloid contents in capsule tissues of a) MR and b) MP ..... 62
Figure 3.3. Comparative alkaloid contents in stem tissues of a) MR and b) MP ..... 63
Figure 3.4. Bioanalyzer electropherograms for all RNA samples ..... 66
Figure 3.5. Distribution of base composition on clean reads. ..... 68
Figure 3.6. Distribution of base quality on clean reads. ..... 69
Figure 3.7. Unigene length distribution. ..... 72
Figure 3.8. Changes in differentially expressed transcript numbers ..... 74
Figure 3.9. Number of the transcripts expressed in all control and treatment groups ..... 75

Figure 3.10. MA plot of MRCT/MRCC comparison library ..... 78
Figure 3.11. Most abundant biological processes in MRCT/MRCC comparison..... 79
Figure 3.12. MA plot of MRST/MRSC comparison library ..... 80
Figure 3.13. Most abundant biological processes in MRST/MRSC comparison..... 82
Figure 3.14. MA plot of MPCT/MPCC comparison ..... 83
Figure 3.15. Most abundant biological processes in MPCT/MPCC comparison..... 85
Figure 3.16. MA plot of MPST/MPSC comparison ..... 86
Figure 3.17. Most abundant biological processes in MPST/MPSC comparison..... 88
Figure 3.18. MA plot of MRST/MRCT comparison ..... 89
Figure 3.19. Most abundant biological processes in MRST/MRCT comparison ..... 91
Figure 3.20. MA plot of MPST/MPCT comparison..... 92
Figure 3.21. Most abundant biological processes in MPST/MPCT comparison ..... 94
Figure 3.22. MA plot of MPCT/MRCT comparison..... 95
Figure 3.23. Most abundant biological processes in MPCT/MRCT comparison ..... 97
Figure 3.24. MA plot of MPST/MRST comparison..... 98
Figure 3.25. Most abundant biological processes in MPST/MRST comparison ..... 100
Figure 3.26. A representative BIA pathway KEGG map ..... 102
Figure 3.27. Pathway functional enrichment of DEGs for MRCT/MRCC* ..... 103
Figure 3.28. Pathway functional enrichment of DEGs for MRST/MRCT*..... 104
Figure 3.29. Pathway functional enrichment of DEGs for MPCT/MRCT*..... 105
Figure 3.30. Pathway functional enrichment of DEGs for MRST/MRSC*..... 106
Figure 3.31. Pathway functional enrichment of DEGs for MPST/MRST*..... 107
Figure 3.32. Pathway functional enrichment of DEGs for MPCT/MPCC*..... 108
Figure 3.33. Pathway functional enrichment of DEGs for MPST/MPCT*..... 109
Figure 3.34. Pathway functional enrichment of DEGs for MPST/MPSC*..... 110
Figure 3.35. Heatmap of top 50-mostly expressed transcripts among libraries ..... 117
Figure 3.36. Statistics of differentially expressed transcripts..... 118
Figure 3.37. Statistics of BIA-related transcripts ..... 125

# LIST OF TABLES

Table 1.1. Comparison of 2nd generation sequencing platforms ..... 37
Table 1.2. Advantages of RNASeq compared with other transcriptomics methods ..... 43
Table 2.1. Control and MeJA-treated tissue samples ..... 47
Table 3.1. RNA concentration and quality check results ..... 65
Table 3.2. Summary of sequencing after filtering ..... 67
Table 3.3. Quality metrics of transcripts ..... 71
Table 3.4. Quality metrics of unigenes ..... 71
Table 3.5. KEGG Pathway enrichment analysis in MR capsule tissues ..... 111
Table 3.6. FPKM values of selected transcripts remarkable in DEG analysis ..... 123
Table 3.7. Morphine biosynthesis specific genes compared at tissue level ..... 124
Table 3.8. Morphine biosynthesis specific genes compared at cultivar level ..... 125

# 1. INTRODUCTION

## 1.1. General Characteristics of *Papaver somniferum* L.

Opium poppy (*Papaver somniferum* L., 2n=22) systematically belongs to the *Papaveraceae* family of Ranunculales order (Figure 1.1). The species name was inspired from a Latin word meaning “sleep-inducing”, and the genus was named as papaver which means poppy in Greek.



**Figure 1.1** *Papaver somniferum* L. (by the courtesy of Özlem Korkmaz, Science Illustrator)

The plant is grown in varying climate conditions like tropical, subtropical, and warm regions of the world. It is an annual herbaceous plant, a source of opium that has been cultivated and utilized as a food and drug since the beginning of civilizations (Brownstein, 1993; Huang and Kutchan, 2000). So, its medicinal properties have been known to mankind for thousands of years. The opium poppy remains the one and only commercial source to manufacture special metabolites including narcotic analgesics morphine and codeine, cough-suppressant noscapine, semi-synthetic precursor thebaine, anti-microbial agent sanguinarine and vasodilator papaverine (Ye et al., 1998; Desgagné-Penix et al., 2010; Desgagné-Penix et al., 2012). Hence, it does affect the countries' economic, political and social relations as being one of the most important crops of the world (Wijekoon and Facchini, 2012). In addition to its medicinal use, opium poppy is rich in oil content in seeds and thus is a valuable food source. And also it is known as "common garden poppy" and widely grown for ornamental purposes in Europe. The illegal cultivation of the plant for the heroin production is a serious problem with negative consequences worldwide.

### 1.2. History of Opium Poppy and Its Cultivation

The opium poppy is known as one of the most ancient cultivated plants in the world (Frick et al., 2007). It was grown in lower Mesopotamia around 3400 B.C., however the first written reference to our knowledge is Sumerian texts (dated 4000 B.C.) which names opium poppy as *Hul Gil*, -the joy plant. Egyptians were then cultivated the plant and exported to the Middle East and Europe. Hippocrates (400s B.C.) was the first physician to notice its painkiller effect. Persia and India met opium by Alexander the Great (330 B.C.) first, later became large producers. Around the same century China was also introduced opium. Laudanum (an alcohol mixture with Paracelsus) was first started to be used to treat some disorders in 1527, the addictive characteristic was not known yet. In conjunction with the growing number of people became opium addicts; opium use was banned in China except medical purposes. The British East India Company then dominated the opium market. In 1753, the botanical classification as *Papaver somniferum* was first done by Linnaeus, a father of botany. Furthermore, Friedrich Sertuerner extracted its morphine content for the first time. Since then, morphine is utilized as the main painkiller in extreme medical cases.

England and China could not find a way for opium trading since China was insisting on opium traffic ban. In 1841, First Opium War was broken out followed by the second, in 15 years. As a result, opium import became legal in China again. In the US, opium poppy cultivation is not allowed, except for culinary and ornamental purposes.

Today, world legal production of opium poppy is maintained by six countries (Australia, France, Spain, Turkey, Hungary and India) under the supervision of United Nations (UN). Turkey and India are particularly accepted as traditional cultivators. U.S., as being the largest importer of the opiates, purchases 80% of its imports from Turkey and India following a rule known as "80/20".

According to the data from the Turkish Grain Board (TMO), Turkey has the 45% of all opium poppy cultivated area in the world with a maximum limit of 70.000 hectares of annual available land as shown in Figure 1.2. Opium poppy can grow in the whole country except for ever-raining East Black Sea and South-East Anatolia with high temperatures. However, the cultivation, which is under the control of TMO, is restricted to following cities; Afyon, Denizli, Burdur, Isparta, Corum, Tokat, Amasya, Kutahya, Usak, and specific counties of Balikesir, Eskisehir, Konya and Manisa (Figure 1.3).

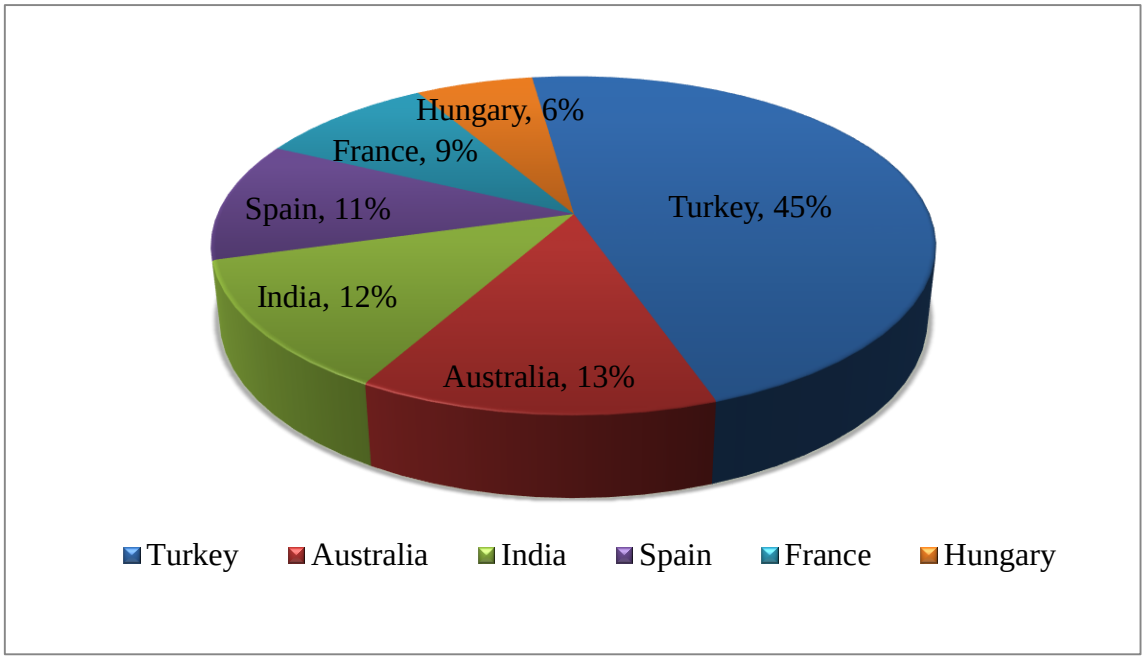



**Figure 1.2**. Legal opium poppy cultivation areas




**Figure 1.3.** Cities allowed to cultivate opium poppy in Turkey

In 2012, right after the year Turkey was the largest producer, there has been a drastic decrease in yield due to the inconvenient climate conditions. Based on the International Narcotics Control Board (INCB) report, expected amount of global production of morphine-rich opiate raw material is 715 tons in 2015. Of which Australia will be the largest producer with 32%, followed by France (19%), Spain (17%), Turkey (11%), India (9%) and Hungary (9%) as depicted in Figure 1.4.

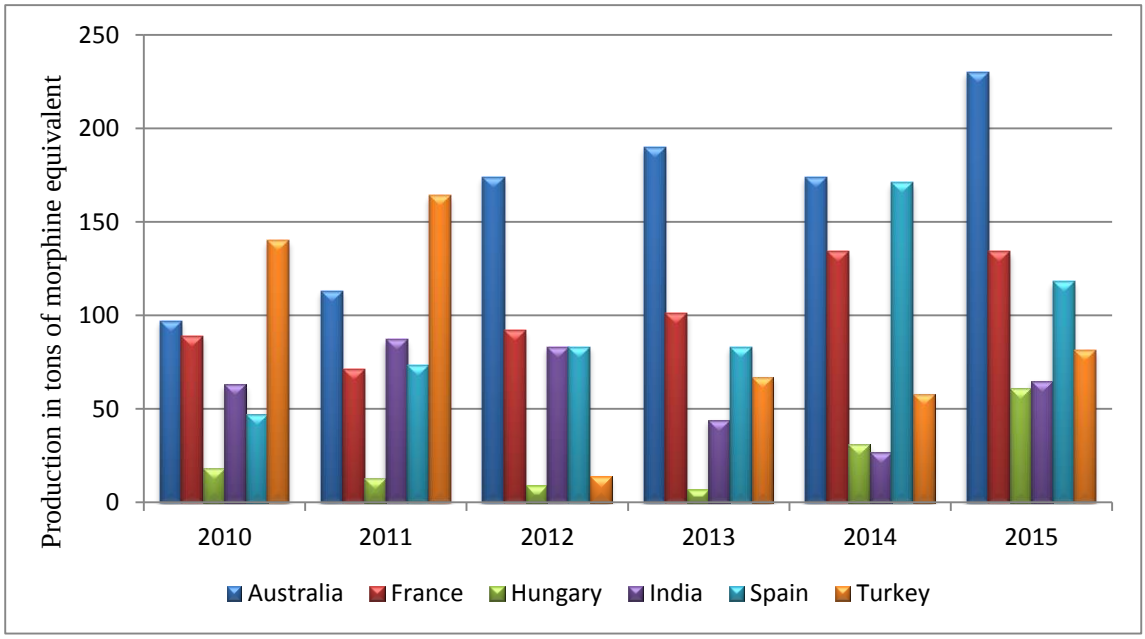



**Figure 1.4.** Production of opiate raw materials rich in morphine (based on INCB report)

Legal producers are supposed to produce about 96% of global morphine-rich opiate production in 2015 (http://www.incb.org/documents/Narcotic-Drugs/Technical-Publications/2014/ND_TR_2014_3_SD_EN.pdf ).

### 1.3. Secondary Metabolites

Plants produce diverse set of compounds that can be grouped as primary and secondary metabolites. First group is called primary metabolites, since they are essential for survival and comprised of carbohydrates, proteins and lipids. On the other hand, secondary metabolites are not required for the plant's survival, but might have a key role in defense mechanism. Plants accumulate a bunch of secondary metabolites in response to environmental conditions such as biotic/abiotic stresses, various elicitors or signal molecules which are also unique sources for human use as pharmaceuticals, food supplement, flavors, and fragrances. Thus, it is important to understand biochemical pathways to optimize their commercial production (Samanani et al., 2005; Zhao et al., 2005). Secondary metabolites are mainly classified into four groups based on the pathway by which they are synthesized: terpenoids, phenolics, glicosides and alkaloids (Wu and Chappell, 2008).

#### 1.3.1. Alkaloids

Alkaloids are a large group of secondary metabolites with low molecular weight that contain mostly basic nitrogen atoms. The first alkaloid isolated from plants was coined as "morphine" -derived from the name of Morpheus, Greek god of dreams (Hamilton and Baskett, 2000). Sarturner's method was then used to obtain several alkaloids in the following years. The name "alkaloid" was given by a German chemist Meißner, because of the metabolites' alkaline-like nature. At least 25% of higher plants possess alkaloids in great amounts in specific organs such as seeds, leaves and fruits. They are much more soluble in organic solvents than in water. Known alkaloid types are monoterpenoid indole alkaloids, tropane alkaloids, pyrrolizidine alkaloids, purine alkaloids, quinolizidine alkaloids and benzylisoquinoline alkaloids (Samanani et al., 2005; Gurkok, 2013; Inal, 2015). In opium poppy, alkaloids are mainly in the latter group (Ziegler et al., 2005; Gurkok et al., 2010).

### 1.3.2. Benzylisoquinoline alkaloids (BIA) and opium poppy

Benzylisoquinoline alkaloids (BIA) comprise a group of about 2500 products, some of which is of great pharmacological importance such as narcotic analgesics morphine and codeine, antimicrobial sanguinarine and berberine, cough-suppressant and anticancer drug noscapine and vasodilator papaverine (Facchini and Park, 2003; Samanani et al., 2005; Hagel and Facchini, 2013; Onoyovwe et al., 2013) (Figure 1.5).

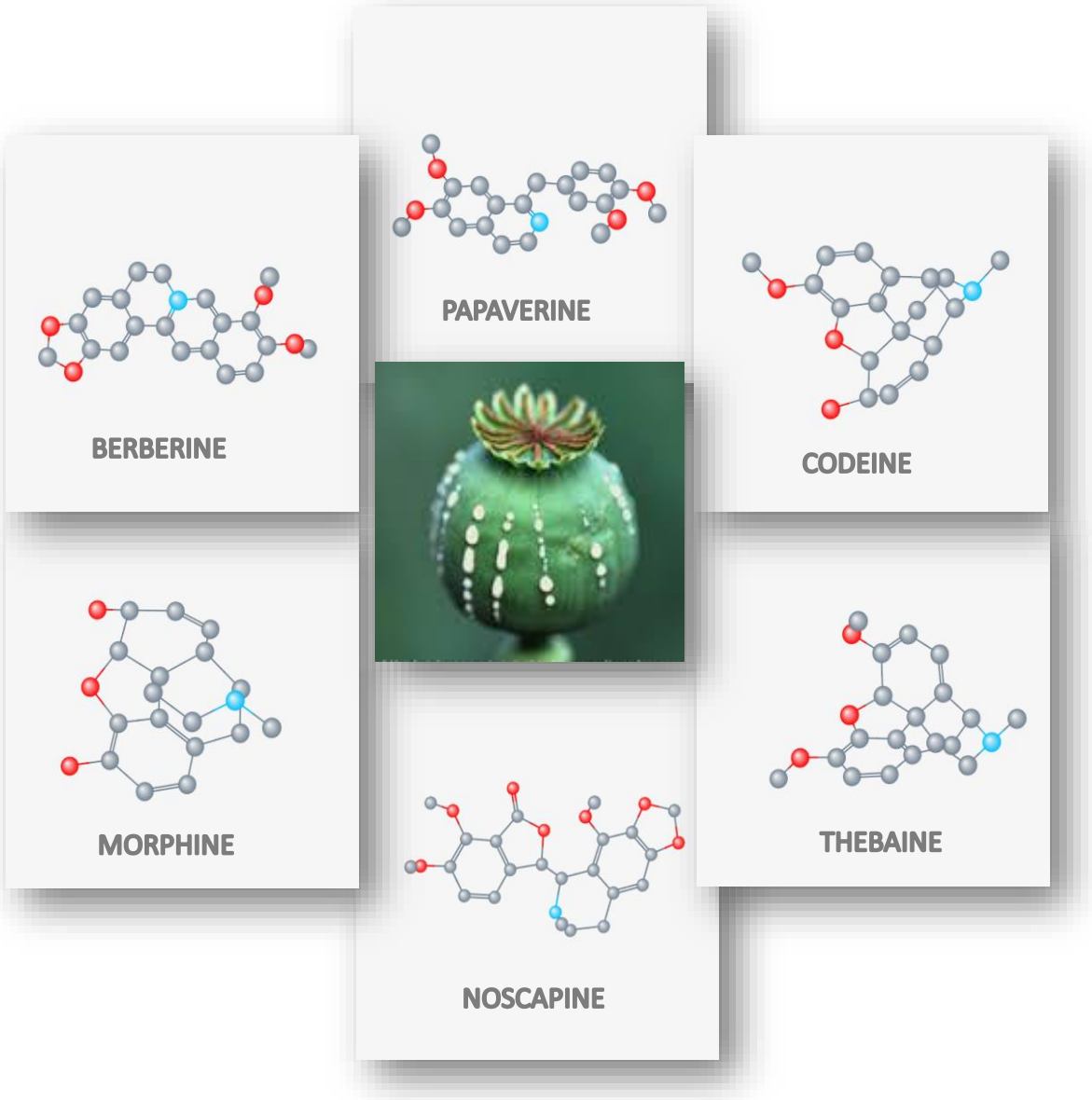



**Figure 1.5.** Major alkaloids in opium poppy (Photo by the courtesy of Prof. Toni Kutchan, Washington University, St Louis)

These alkaloids are mostly found in Papaveraceae, Berberidaceae, Ranunculaceae, and Menispermaceae families of Ranunculales order (Liscombe et al., 2005; Beaudoin and Facchini, 2014).

Opium poppy produces vast amount of benzylisoquinoline alkaloids belonging to morphinan, aporphine, protopine, phthalideisoquinoline, benzophenanthridine, protoberberine, pavine, and rhoeadine subgroups (Hagel and Facchini, 2013). The morphinan alkaloids (morphine as the most abundant and relatively low levels of thebaine and codeine), benzylisoquinoline papaverine and phthalideisoquinoline noscapine are accumulated largely in opium latex, the cytoplasm of laticifers, which are in contact with phloem channels. Conversely, roots are rich in benzophenanthridine sanguinarine (Frick et al., 2005). The traditional method of opiate raw material isolation involves the lancing step at the beginning to collect the latex leaking out of the unripe seed capsule while the mechanized (modern) method employs solvent extraction from field-dried plants (Bryant, 1988; Pilatzke-Wunderlich and Nessler, 2001; Onoyovwe et al., 2013).

### 1.3.3. Biosynthesis of the major alkaloids in opium poppy

BIA metabolism in terms of biochemistry and physiology has long been studied. In this respect, radiotracer technique was adapted in 1960s, enzyme extraction and characterization methods developed around 1980s and recombinant DNA era started in 1990s (Beaudoin and Facchini, 2014). Along with recent applications of these methods in opium poppy, the knowledge of BIA metabolism is rapidly increased.

Biosynthesis of BIAs begins with the conversion of two L-tyrosine molecules to dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA), via tyrosine/DOPA decarboxylase (TYDC) (Facchini and De Luca, 1994) and tyrosine aminotransferase (TyrAT) enzymes respectively (Liscombe et al., 2005; Lee and Facchini, 2010). The Picted-Spengler condensation reaction of these tyrosine derivatives coupled by norcoclaurine synthase (NCS) (Samanani et al., 2005) to yield a central BIA precursor: (*S*)-norcoclaurine (Lee and Facchini, 2010) (Figure 1.6).

### 1.3.4. (S)-norcoclaurine to (S)-reticuline

(*S*)-norcoclaurine is further converted to the last common intermediate (*S*)-reticuline in the biosynthesis of most BIAs such as morphinans (*e.g.* thebaine, codeine, oripavine, morphine), phthalideisoquinolines (*e.g.* noscapine), benzophenanthridines (*e.g.* sanguinarine) simple benzylisoquinolines (*e.g.* papaverine, laudanine), and protoberberines (*e.g.* berberine). This reaction involves four sequential steps including *O*-methylation at 6$^{th}$ position by norcoclaurine 6-*O*-methyltransferase (6OMT) and *N*-methylation of norcoclaurine by coclaurine *N*-methyltransferase (CNMT) respectively to produce (*S*)-*N*-methylcoclaurine which will be further hydroxylated at 3' by a P450-dependent monooxygenase enzyme (*S*)-*N*-methylcoclaurine-3'-hydroxylase (NMCH) and lastly 4'-*O*-methylation step of (*S*)-3'-hydroxy-*N*-methylcoclaurine with 3-hydroxy-*N*-methyl coclaurine 4′ methyltransferase (4OMT) produces (*S*)-reticuline (Pauli and Kutchan, 1998; Huang and Kutchan, 2000; Morishige et al., 2000; Choi et al., 2002; Facchini and Park, 2003; Ounaroon et al., 2003; Ziegler et al., 2005).

Four pathways diverge from (*S*)-reticuline, leading to benzophenanthridine, protoberberine, phthalideisoquinolines as well as morphinan alkaloids. Papaverine biosynthesis partially involves (*S*)-reticuline (Beaudoin and Facchini, 2014).

#### 1.3.4.1. Papaverine biosynthesis

Papaverine biosynthesis remain controversial which two different metabolic routes have been proposed. The first proposed pathway, (NH pathway) involves demethylation of (S)-norreticuline while the second (NH3 pathway), proceeds through *N*-methylation of (*S*)-reticuline and unspecified intermediates (Mishra et al., 2013). In the first one, (*S*)-coclaurine is converted to (*S*)-norreticuline utilizing 3'-hydroxylase and 7-O-methyltransferase (N7OMT) (Pienkny et al., 2009). On the other side, NH3 pathway is proposed to commence with the activity of 7-O-methyltransferase (7OMT) yielding (*S*)-laudanine from (*S*)-reticuline (Ounaroon et al., 2003). Both routes have successive methylation, demethylation and dehydrogenation steps to form papaverine at the end. (Han et al., 2010).

In a recent transcriptome sequencing and VIGS study, intermediates of both pathways were supported to take place in papaverine biosynthesis (Dang and Facchini, 2012).

However another study of the same group later reported that not (*S*)-reticuline but (*S*)-norreticuline is involved in the formation of papaverine (Desgagné-Penix et al., 2012). In another recent article, comparative transcriptomics were performed on high papaverine mutants (*pap1*) of opium poppy suggesting (*S*)-coclaurine as the primary route for papaverine biosynthesis (Pathak et al., 2013).

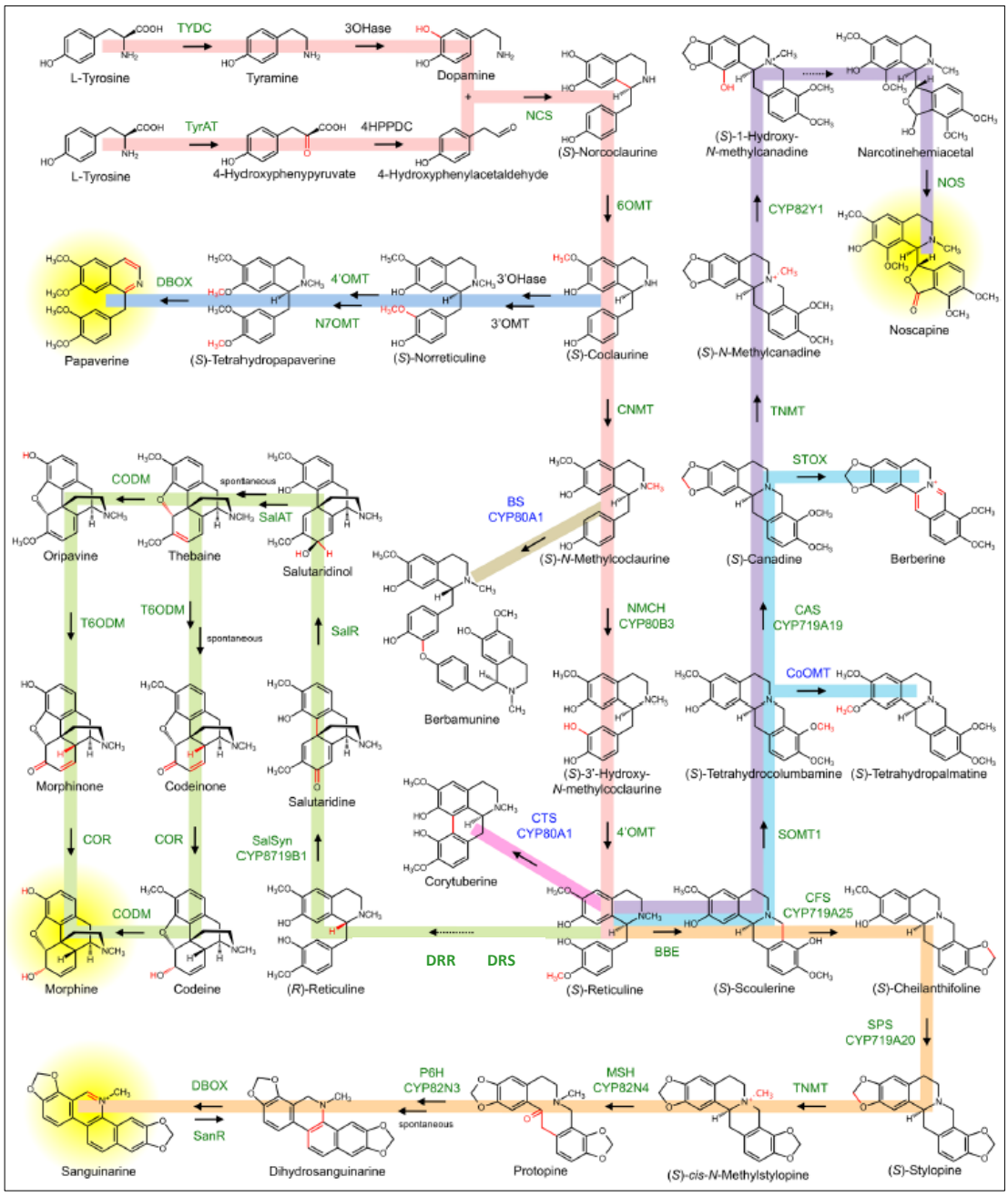



**Figure 1.6.** Biosynthesis pathway of BIA alkaloids in opium poppy(Beaudoin and Facchini, 2014)

### 1.3.4.2. Protoberberine, protopine and benzophenanthridine biosynthesis

(*S*)-scoulerine is a branch-point intermediate in the biosynthesis of several protoberberines, protopine and benzophenanthridine alkaloids. The formation of (*S*)-scoulerine from (*S*)-reticuline is catalyzed by berberine bridge enzyme (BBE) through a cyclization reaction involving oxidation of (*S*)-reticuline to form a C-C bond between the phenolic ring and *N*-methyl group. BBE encoding cDNAs have been isolated from opium poppy and California poppy (Dittrich and Kutchan, 1991; Facchini, 2006). The enzyme has also been well characterized (Winkler et al., 2008).

Through methylenedioxy bridging and O-methylation, (*S*)-scoulerine can lead to the formation of several protoberberines such as (*S*)-canadine and (*S*)-stylopine. In berberine patway, (*S*)-scoulerine is first converted to (*S*)-tetrahydrocolumbamine by a FAD-linked enzyme scoulerine 9-*O*-methyltransferase (SOMT1) (Takeshita et al., 1995). Successively (*S*)-canadine is formed by a methylenedioxy bridging enzyme canadine syhthase (CAS) belonging to CYP subfamily. Ultimately, aromatization by (*S*)-tetrahydroprotobererine oxidase (STOX) enzyme gives rise to **berberine**. Alternatively, (*S*)-tetrahydrocolumbamine can be methylated to yield (*S*)-tetrahydropalmatine with a protoberberine-specific enzyme columbamine *O*-methyltransferase (CoOMT) (Morishige et al., 2002).

The initial steps in the biosynthesis of benzophenanthridine type alkaloids such as sanguinarine were elucidated by tracer experiments and advanced at molecular and biochemical levels (Battersby et al., 1975; Takao et al., 1983; Hagel and Facchini, 2013). In **sanguinarine** pathway, conversion of (*S*)-scoulerine to (*S*)-stylopine is the first commited step. Cytochrome P450 dependent mono-oxygenases cheilanthifoline synthase (CFS or CheSyn) and stylopine synthase (SPS or StySyn) are the two enzymes responsible for the formation of methylenedioxy bridges that yield (*S*)-stylopine (Bauer and Zenk, 1989; Bauer and Zenk, 1991). cDNAs that encode SPS have been isolated in *E. californica* (Ikezawa et al., 2007). (*S*)-stylopine is then converted to (*S*)-*cis*-*N*-methylstylopine by the enzyme tetrahydroprotoberberine *cis*-*N*-methyltransferase (TNMT) (Liscombe and Facchini, 2007) and further hydroxylated by another P450 dependent *N*-methylstylopine 14-hydroxylase (MSH) to yield protopine (Ruffer and Zenk, 1987). TNMT encoding cDNAs were isolated in opium poppy and characterized

(Liscombe and Facchini, 2007). Benzophenanthridine alkaloid dihydrosanguinarine (which is known to be root-specific) synthesis from protopine is catalyzed by a CYP82 member protopine 6-hydroxylase (P6H) (Tanahashi and Zenk, 1990; Takemura et al., 2013). The resulting 6-hydroxyprotopine is spontaneously rearranged to dihydrosanguinarine and which is successively oxidized to sanguinarine by oxygen-dependent oxidoreductase dihydrobenzo-phenanthridine oxidase (DBOX) (Hagel et al., 2012). Furthermore, a sanguinarine reductase (SanR) enzyme is also involved in the benzophenanthridine alkaloid metabolism which catalyzes the reverse conversion to dihydrosanguinarine (Weiss et al., 2006; Vogel et al., 2010).

#### 1.3.4.3. Noscapine biosynthesis

As mentioned above, phthalideisoquinoline alkaloids such as noscapine is known for its cough suppressing and anticancer activity (Ye et al., 1998; Facchini et al., 2007; Barken et al., 2008; Dang and Facchini, 2012). Although noscapine is one of the first isolated alkaloids, its biosynthesis has not been investigated thoroughly. However, a metabolic scheme was proposed to include (*S*)-*N*-methylcanadine in the pathway. Distinguishing from berberine route, TNMT converts a common intermediate (*S*)-canadine to (*S*)-*N*-methylcanadine (Liscombe and Facchini, 2007). In a recent research, a developed scheme has been revealed with all proposed enzymes identified in opium poppy. Accumulation of compounds as intermediates in response to suppression of a specific gene by VIGS suggested the functions of the gene clusters (Dang and Facchini, 2012; Winzer et al., 2012). Additionally, three more enzymes have been isolated and characterized in opium poppy: N-methylcanadine 1-hydroxylase (CYP82Y1), CAS and noscapine synthase (NOS) (Chen and Facchini, 2014; Dang and Facchini, 2014; Dang and Facchini, 2014). NOS was shown to be a NAD+-dependent, dehydrogenase/reductase yielding noscapine through irreversible conversion of narcotinehemiacetal (Beaudoin and Facchini, 2014).

### 1.3.4.4. Morpine biosynthesis

The epimerization of (*S*)-reticuline to (*R*)-reticuline is required to initiate the biosynthesis of morphinan branch alkaloids. This is a two-step process that starts with the dehydrogenation of (*S*)-reticuline to 1,2-dehydroreticuline followed by the reduction to (*R*)-reticuline employing 1,2-dehydroreticuline synthase (DRS) and NADPH-dependent 1,2-dehydroreticuline reductase (DRR), respectively. These two enzymes have been isolated and characterized in poppy seeds (De-Eknamkul and Zenk, 1992; Hirata et al., 2004). In a recent study, reticuline epimerase (REPI) has been detected in opium poppy and genes encoding REPI, DRS and DRR have been isolated and characterized in the plant (Farrow et al., 2015). The first promorphinan alkaloid, salutardine is formed by intramolecular C-C phenol coupling of (*R*)-reticuline driven by a P450 monooxygenase salutardine synthase (SalSyn) (Gesell et al., 2009) which is subsequently reduced by NADPH-dependent salutardine reductase (SalR) to yield salutardinol (Ziegler et al., 2006). The following step in the pathway is the *O*-acetylation of salutardinol by acetyl-CoA-dependent enzyme salutardinol 7-*O*-acetyltransferase (SalAT) (Grothe et al., 2001). The conversion of salutardinol 7-*O*-acetate to the first pentacyclic morphinan alkaloid thebaine is proposed to happen either spontaneously (Lenz and Zenk, 1995) or by a deacetylation enzyme isolated from protein extracts of opium poppy (Fisinger et al., 2007). Following to thebaine formation, the morphine pathway is divided into two routes. In the major route, thebaine is first converted to neopinone by thebaine 6-*O*-demethylase (T6ODM), and then spontaneously rearranged to yield codeinone (Hagel and Facchini, 2010). Subsequent to this step, codeinone is reduced to codeine by a NADPH-dependent codeinone reductase (COR) (Unterlinner et al., 1999). As a final step, morphine is produced by demethylation of codeine by codeine-*O*-demethylase (CODM) (Hagel and Facchini, 2010). As an alternative route in morphine synthesis, the CODM catalyzes 3-*O*-demethylation of thebaine to oripavine and further 6-*O*- demethylation to morphinone by T6ODM prior to be reduced to morphine by COR (Brochmann-Hanssen, 1984). Genes for SalSyn (Cyp719B1), SalR, SalAT, and COR enzymeshave been isolated in opium poppy by different studies (Unterlinner et al., 1999; Grothe et al., 2001; Wijekoon and Facchini, 2012).

In opium poppy, although the enzymatic synthesis of morphine has been almost completely revealed, the regulation remains largely unknown (Kutchan et al., 2008; Kempe et al., 2009).

### 1.3.5. MeJA and its effect on plants

MeJA is a volatile derivative of Jasmonic Acid (JA) and first identified from *Jasminum grandiflorum*. MeJA with its free-acid are accepted as plant hormones that play regulatory roles in a diverse set of developmental processes such as germination and growth, as well as plant defense responses against wounding, pathogen attack and abiotic stresses (Reymond and Farmer, 1998; Wasternack and Hause, 2002). Several gene expression studies have been conducted upon MeJA treatment to identify jasmonate-responsive genes in plants. Upregulated genes are found to encode defensive proteins and those involved in JA and secondary metabolite biosynthesis. On the contrary, downregulated genes included Rubisco and chlorophyll a/b binding proteins which are active in photosynthesis (Lorenzo et al., 2003).

Furthermore, jasmonates could function synergistically or antagonistically with the other plant hormones such as ethylene, auxin and abscisic acid (ABA) (Cheong and Choi, 2003).

The effect of MeJA on signal transduction commences with the perception of the elicitor signals. Once the receptors are stimulated, several ion channels and protein/protein kinases are also activated (Blume et al., 2000). Subsequently, $Ca^{+2}$ concentration is increased and NADPH oxidase is activated to produce reactive oxygen species (ROS). Afterwards, following responses are induced; ethylene and jasmonate production, expression of response genes and accumulation of secondary metabolites (Zhao et al., 2005). For instance, analysis of Arabidopsis mutants that lack of a proper jasmonate biosynthesis analysis revealed that *coi1* functions in JA signaling and pathogen response. The gene is found to encode an LRR/F-box motif with that calls regulatory proteins for ubiqutination. So that the protein has an important role in JA signaling pathway activating JA-responsive gene expression (Xie et al., 1998; Devoto et al., 2002).

Additionally, exogenous MeJA treatment causes to enhance secondary metabolism by inducing secondary metabolite accumulation through signal transduction in stress response. Elicitors are also known to trigger endogenous JA and MeJA accumulation which causes an increase in secondary metabolite biosynthesis. (Memelink et al., 2001).

Previous studies indicated that MeJA elicitor treatment effects a set of reactions in opium poppy and *Catharanthus roseus* that cause alkaloid accumulation (Huang and Kutchan, 2000; Ruiz-May et al., 2009; Holkova et al., 2010). Secondary metabolite accumulation is also triggered by elicitor treatment through altering the gene expression responsible for their biosynthesis (Memelink et al., 2001; Zhang et al., 2011).

Moreover, it has been shown that exogenous MeJA treatment increases the release of signal molecules such as flavonoids and indole as a result of activation of JA signaling pathway (Badri et al., 2008). Another study indicates that plants have different jasmonate signalling cascade in different organs, thus responses may vary based on the wounded site (Tytgat et al., 2013).

### 1.3.6. Approaches to establish genomic resources for specialized metabolisms

The discovery of new genes active in alkaloid biosynthesis following forward genetics approach involves purification and sequencing of the protein-of-interest, primer design for the cloning of related gene and further functional analysis in an appropriate host. In opium poppy, a newer strategy for gene discovery is applied and expressed sequence tags (ESTs) are generated by Sanger sequencing. The EST database is derived from stem, seedlings, roots, laticifers, and cell cultures of opium poppy (Pilatzke-Wunderlich and Nessler, 2001; Ziegler et al., 2005; Ziegler et al., 2006; Zulak et al., 2007). Thus, a bunch of homology based enzymes are identified. However, both methods are time-consuming, expensive and labor-dependent while the efficiency of EST method is mostly dependent on the representation of candidate genes in the current database since it cannot cover all the transcriptome.

Along with the recent advances in “omics” technologies, the integration of metabolomics, proteomics and transcriptomics data become useful for candidate gene discovery in opium poppy BIA metabolism.

High throughput **transcriptomics** is used to determine candidate genes producing BIA enzymes and/or measure relative transcript levels in cDNA libraries derived from different organs or cells with varying alkaloid contents. Hybridization-based DNA macro and microarray studies, in this sense, are important to understand gene expression profiles in different parts of the plant. EST libraries of relevant tissues should be utilized to construct the cDNA microarray. (Zulak et al., 2007; Hagel and Facchini, 2010; Gurkok et al., 2015). In addition to that, next generation sequencing (NGS) platforms, GS-FLX, Illumina and SOLID, offer a successful technology to overcome the limitations of previous methods. For instance, cDNA redundancy in the databases decrease and no further bacterial cloning step is required which reduce the cost and labor.

**Proteomics** is of great importance to complete gene discovery studies with the isolation of the protein and its further analysis using sensitive and fast mass spectrometers (Desgagné-Penix et al., 2010). Protein separation can be performed either by one or two dimensional (2D) SDS-PAGE. Subsequent to proteolysis, fragmentation of peptides or collision-induced dissociation type or mass spectroscopy (MS/MS) is utilized to determine the molecular weight and amino acid (aa) sequence. Furthermore, peptide-mass fingerprint of the protein is generated by using matrix-assisted laser desorption/ionization (MALDI) or electro spray ionization (ESI).

Lastly, the **metabolomics** is necessary to the discovery of BIA pathway genes. Profiling targeted metabolites with the use of analytical and chromatographical methods like high-pressure liquid chromatography (HPLC) have been conducted in opium poppy (Frick et al., 2005; Hagel and Facchini, 2008; Hagel et al., 2008; Zulak et al., 2009; Gurkok et al., 2015).

The integration of the -omics data holds the advantage to select gene candidates that encode enzymes responsible in BIA metabolism. More specifically, correlating targeted metabolite profiling data with gene transcripts and corresponding proteins can result in selection of a reasonable number of potential genes for functional characterization including enzyme assays as well as virus-induced gene silencing (VIGS) (Dang and Facchini, 2012; Dang et al., 2012; Desgagné-Penix et al., 2012).

Post-transcriptional gene silencing (PTGS) is a defense mechanism that targets foreign genome sequences by producing dsRNAs to be degraded in both plants and animals (Wesley et al., 2003). Inspired from this endogenous phenomenon, researchers developed a powerful gene silencing technique that is used in functional genomics. Once a cDNA-insert carrying viral vector enters the plant cell, PTGS is stimulated to degrade corresponding endogenous plant mRNA. Hence, temporary loss-of-function of a specific gene can provide insights for its function which makes the technique valuable as a reverse genetics tool (Hagel and Facchini, 2010; Dang et al., 2012). VIGS has also been shown to silence specific gene transcripts of BIA enzymes and elucidate its function in opium poppy as well as in other plants. (Hileman et al., 2005; Hagel and Facchini, 2010; Lee and Facchini, 2010; Dang and Facchini, 2012; Hagel et al., 2012; Wijekoon and Facchini, 2012; Farrow and Facchini, 2013).

### 1.3.7. Omics studies on opium poppy

Functional genomics in opium poppy has been first performed by Decker and colleagues in which latex proteins have been isolated using 2D SDS-PAGE method and their sequences identified (Decker et al., 2000). Furthermore, a group of researchers have isolated gene transcripts of three enzymes (BBE, COR and NMCH(CYP80B1)) in different tissues of the plant utilizing RNA gel blot analysis upon elicitor treatment (Huang and Kutchan, 2000).

In another study, molecular characterization of salutardinol 7-O-acetyltransferase (SalAT) has been achieved through cloning of the cDNA using amino acid sequence of the purified enzyme which catalyzes the previous step of thebaine. Opium poppy genome has shown to contain only a single copy of the gene by genomic DNA blotting experiment (Grothe et al., 2001).

Identification of novel genes in these alkaloid biosynthesis pathways is also possible with mutant analysis. There are two main reports using this approach that characterize mutants and further perform transcriptome screening in opium poppy. The first one describes “top1” poppy mutant which was established by ethyl methane sulfonate (EMS) mutagen treatment and screened for alkaloid content of latex. The mutant was found to accumulate thebaine and oripavine as major alkaloids instead of codeine and

morphine. This study provides an insight to grow morphine-free but thebaine/oripavine-producing crops for opioid addiction treatment. Global gene expression was explored using DNA microarrays (Millgate et al., 2004). In the second report, mutant poppy "pap1" which accumulates high papaverine in the latex compared to wild-type opium poppy (BR086) has been characterized and 454 pyrosequencing of the transcriptome enabled researchers to identify candidate genes in papaverine biosynthesis (Pathak et al., 2013).

Comparative gene expression analysis has been performed in nine Papaver species only one with morphine producing capability. AFLP-based DNA macroarray was used to reveal differential gene expression profile among inspected species. In morphine-containing opium poppy, *4OMT* gene transcripts were shown to have high expression levels indicating a role in BIA biosynthesis (Ziegler et al., 2005). Subsequently the same group has identified SalR gene using the same approach (Ziegler et al., 2006). More recently gene transcripts of N7OMT (Pienkny et al., 2009) and SalSyn were also identidied using array-based transcriptomics (Gesell et al., 2009). Zulak *et al*. (2007) have conducted a DNA microarray study for a large-scale transcriptome analysis of sanguinarine-accumulation in opium poppy upon fungal elicitor treatment (Zulak et al., 2007). At first, an EST database was generated from random clones derived from cDNA library. ESTs related to sanguinarine pathway enzymes were identified and further hybridized with RNA samples belonging to different time points. In comparison to control RNAs, BIA enzyme and defense protein transcripts demonstrated a fast and significant upregulation in response to elicitor treatment. Besides, many other transcripts belonging to plant primary metabolism were also stimulated. Metabolite profiling was screened by using fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). Recently, another microarray study has been conducted on MeJa treated opium poppy capsules. Expression levels of BIA metabolism specific genes such as CNMT, NCS, StySyn and COR were found to be altered (Gurkok et al., 2015).

Taking together the deep transcriptomics, proteomics and targeted metabolomics Desgagné-Penix and his colleagues published two comprehensive studies that reveal the components of BIA metabolism in opium poppy cell cultures. In the first one, upon MeJA treatment 454 sequencing was used to generate deep transcriptome data and SDS-

PAGE coupled with advanced LC-MS/MS-based protein profiling has enabled the identification of approximately a thousand peptides. In order to correlate abundance of the alkaloids with specific genes and polypeptides, the amount of reticuline, sanguinarine and protopine at different time intervals has been measured. Elicitor treatment resulted in a significant increase in sanguinarine as much as 40 folds after 100 hours of application. Several genes and enzymes potentially involved in BIA metabolism were identified as genetic and biochemical output of the study (Desgagné-Penix et al., 2010). Similarly, a more recent study of the same group reports deep transcriptome sequences for 8 different opium poppy cultivars. LC–MS/MS was used to determine the alkaloid levels of each one. In conclusion, the latter studies have shown that the combination of transcriptome data obtained by NGS sequencing and metabolite profiling through LC–MS/MS analysis could pave the way for understanding of BIA metabolism and regulation in plants (Desgagné-Penix et al., 2012).

### 1.3.8. Regulation of gene expression

Following integrated approach of transcriptomics and metabolite analysis, another group has identified novel ESTs with one encoding WRKY under wounding stress. Gene expression analysis of *PsWRKY* was done with qualitative real time PCR (qRT-PCR) and gel retardation assay showed specific binding of the protein to W-box element indicating a regulatory role (Mishra et al., 2013). Only a limited number of regulatory elements were found in opium poppy. (Kawano et al., 2012; Mishra et al., 2013). In a recent study, transcriptome datasets from previous studies have been collected to identify transcription factor (TF) gene families. Hence, transcriptional regulation at specific BIA pathways could be investigated. WRKY was found to be the most abundant TF family in all varieties. According to the comparative transcriptomics analysis, one C3H-type and two WRKY-type TFs were found to be candidate regulators in papaverine and thebaine pathways in opium poppy (Agarwal et al., 2015).

Unver and his group (2010) identified a number of conserved microRNAs (miRNAs) for the first time in opium poppy (Unver et al., 2010). Bioinformatics approach has been used to elucidate miRNA-mediated regulation of gene expression. The expression of opium poppy miRNAs and their targets have been experimentally validated through qRT-PCR. In addition to that, the same group has studied miRNA regulation of BIA

metabolism in a tissue-specific manner. Eleven novel miRNAs have been identified and their targets were determined in opium poppy. In conclusion, pso-miR13, pso-miR2161 and pso-miR408 were predicted to be involved in BIA biosynthesis regulation (Boke et al., 2015)

Until quite recently, nearly ninety opium poppy specific genic markers have been developed by using EST database and some of them were validated for forensic use and genetic diversity analysis (Lee et al., 2011; Şelale et al., 2013). In 2014, Celik and coworkers reported a comprehensive study on genomic SSR marker development utilizing DNA pyrosequencing (Celik et al., 2014).

### 1.3.9. Metabolic engineering in opium poppy

Morphinan alkaloids including morphine, thebaine, codeine and oripavine are medicinally important metabolites used in production of narcotic analgesics and medicine for treatment of opiate addiction. Opium poppy remains one and only source producing these metabolites. Today, there are a limited number of countries which can legally produce opiates. In this sense, higher alkaloid content becomes an important issue in terms of financial efficiencies: land needed for cultivation, efforts in harvesting and transportation. Hence, plant breeders desire to grow high alkaloid producing opium poppy.

In this context, genetic engineering offers the opportunity to manipulate specific metabolic activity. More specifically, plant metabolic engineering can be used to alter the level of intermediates either by over-expressing the gene-of-interest for increased production or by silencing it through antisense and RNA interference (RNAi) in order to explore the nearly-exact function in morphinan alkaloid biosynthesis (Hughes and Shanks, 2002).

There are several metabolic engineering studies conducted in opium poppy using these methods. Frick et al was the first to alter the alkaloid content in latex. In this study, an industrial elite line has been transformed by cDNA encoding BBE enzyme which is isolated from opium poppy, PCR, northern and southern hybridization methods have been employed to evaluate transformation of the plant while the alkaloid type and abundance in latex have been measured by HPLC and LC-MS. The concentration of

several pathway intermediates such as laudanine, reticuline, salutaridine and (S)-scoulerine have been shown to have an increasing pattern. Transformation caused an altered ratio of morphinan and tetrahydrobenzylisoquinoline alkaloids in latex but did not affect benzophenanthridine alkaloids in roots. This study is known to be the first metabolic engineering in opium poppy which succeeded in the transformation of a BIA pathway gene and further analyzed the changes caused by transgene in the offsprings (Frick et al., 2004).

In another research, COR enzyme which is responsible for the reduction of codeinone to codeine was reported to be silenced through RNAi. Following gene silencing, the transgenic opium poppy has shown to accumulate a 7 step upstream codeninone precursor (*S*)-reticuline. The authors evaluated this result as a feedback mechanism which inhibits morphinan alkaloid biosynthesis but does not affect the other BIA branches. Besides, this is the first gene silencing experience in transgenic opium poppy (Allen et al., 2004).

Moreover, opium poppy was transformed with NMCH encoding sense and antisense copies of *cyp80b3*. In the latex, overexpression of the cDNA increased the alkaloid content up to 450% and antisense-suppression decreased the levels to 84%. *Cyp80b3* was found to have a regulatory role in morphine biosynthesis (Frick et al., 2007).

Concurrently, Larkin and his colleagues reported the first transgenic opium poppy with higher morphinan alkaloid synthesis. Upon transformation with constitutively expressed cDNA of codeninone reductase (*cor1*), the overexpressor plant showed up to 30% (dry-weight) increase in capsule alkaloid content in both glasshouse and field while other organs remained unaffected. This increase was concordant with the increase in morphine, thebaine and codeine content. The thebaine level was not expected to increase, since COR is a downstream enzyme. However, the researchers explained this situation in two ways: one was the feedback mechanism and the other was the inhibition of thebaine transport into vesicles. Compared to non-transgenic control parent, a high-morphine genotype has been obtained in the offsprings (Larkin et al., 2007).

In another metabolic engineering study, a morphinan branch enzyme SalAT cDNA was both overexpressed and suppressed by RNAi in opium poppy. SalAT overexpressor plant demonstrated an increased morphine, codeine and thebaine levels, while the

suppressed plant resulted in unexpected accumulation of an upstream intermediate salutaridine suggesting a channeling between thebaine in reversible manner in which both SalAT and SalR enzymes involved (Allen et al., 2008). In the meanwhile, RNAi-mediated suppression of SalAT was performed by another research group and similarly salutaridine acumulation was observed too. The interaction of two enzymes (SalAT and SalR) was determined by yeast two-hybrid and co-immunoprecipitation analyses supporting the previous explanation. Furthermore, not morphine but codeine and thebaine levels were found to be reduced in the latex (Kempe et al., 2009).

Recently, VIGS method has become a widely used functional genomics tool to investigate BIA metabolism in opium poppy (Hileman et al., 2005; Hagel and Facchini, 2010; Lee and Facchini, 2011; Dang and Facchini, 2012; Hagel et al., 2012; Wijekoon and Facchini, 2012; Winzer et al., 2012; Farrow and Facchini, 2013). Hileman and coworkers described the method in opium poppy for the first time and silenced an endogenous gene *phytoene desaturase* (*PDS*) with a tobacco rattle virus (TRV) based vector. Consequently, *PDS* expression was reduced and photo-bleached appearance was observed. This easy loss-of-function method suggests a rapid and alternative strategy for functional genomics studies on BIA pathway genes and enzymes (Hileman et al., 2005).

Hagel and Facchini used functional genomics to isolate T6ODM and CODM involved in morphine biosynthesis and explained their physiological roles by knocking the genes down using VIGS. Silencing of T6ODM and CODM caused the inhibition of thebaine metabolism and codeinone accumulation, respectively. No oripavine was detected in CODM-specific vector treated plants, since the conversion of thebaine to oripavine was blocked (Hagel and Facchini, 2010). Recently a further VIGS study has been conducted in opium poppy in which T6ODM and CODM genes are silenced revealing extensive roles in protopine, sanguinarine and rhoeadine production in addition to morphine biosynthesis (Farrow and Facchini, 2013).

Isolation and characterization of TyrAT, an enzyme in upstream precursor production of BIAs have been achieved using 454 pyrosequencing. VIGS based functional genomics are employed to investigate the role of TyrAT in alkaloid metabolism. Gene silencing reduced the level of *TyrAT* transcripts up to 80% and showed a decrease in total alkaloid content suggesting a role in the formation of BIA precursors (Lee and Facchini, 2011).

In another study, the cDNA libraries from eight cultivars have been constructed for sequencing and further metabolite measurements have been undertaken. *SOMT1*, *SOMT2* and *SOMT3* gene transcripts are found to be correlated with noscapine accumulation. Their metabolic functions have been explored using VIGS. *SOMT* silencing resulted in a significant reduction in noscapine and papaverine production (Dang and Facchini, 2012).

Wijekoon and coworkers performed systematic and individual knockdown of six enzymes involved in morphine biosynthesis in order to investigate the regulation of the metabolism (Wijekoon and Facchini, 2012). After gene silencing by VIGS, *SalSyn*, *SalR*, *T6ODM*, and *CODM* knockout plants showed reduced morphine levels and increased s-reticuline, salutaridine, thebaine and codeine levels respectively. On the other hand, *COR* and *SalAT* knockouts accumulated reticuline and salutaridine respectively.

In another metabolic engineering study, a ten gene cluster encoding enzymes of five different classes expressed in high noscapine opium poppy cultivar were identified with transcriptome analysis. Furthermore, VIGS-based silencing of individual genes was performed and subsequent accumulation of intermediates was evaluated. Consequently, six out of ten genes were found to be involved in noscapine pathway (Winzer et al., 2012).

Hagel and his colleagues reported the identification, characterization and silencing of a cDNA encoding DBOX which catalyze the last reaction in sanguinarine formation. VIGS caused reduced sanguinarine levels but papaverine accumulation in knockout plants suggesting the enzyme a multifunctional role in BIA metabolism (Hagel et al., 2012).

In a very recent study, a cytochrome P450/aldo-keto reductase (AKR) fusion which catalyzes (*S*)-reticuline to (*R*)-reticuline isomerization is proposed. This reticuline epimerase fusion (REPI) has been detected in opium poppy. VIGS-based suppression of REPI mRNAs caused a significant decrease in (*R*)-reticuline level in accordance with the decrease in morphinan alkaloids. In contrast, (*S*)-reticuline levels increased indicating a role in stereochemical epimerization from S to R with the activity of REPI (Farrow et al., 2015).

Concordantly, Winzer and coworkers identified a genetic locus (s to r: STORR) responsible for S to R epimerization that express P450 and aldo/keto reductase together. *Storr* mutants were subjected to metabolite analysis. Results confirmed that P450 converts (S)-reticuline to 1,2-dehydroreticuline and aldo/keto reductase family enzyme oxidoreductase catalyzes the further step to (R)-reticuline. Consequently, following a proteomics approach, a fusion protein is found to be responsible for 2-step isomerization (Winzer et al., 2015).

### 1.3.10. Compartmentalization and trafficking of BIA metabolism

#### 1.3.10.1. Cellular localization and transport

In plants, alkaloid biosynthesis and accumulation is largely based on their cytotoxicity and role in defense mechanism (Facchini and De Luca, 2008). Different plants can accumulate BIAs in different cells and tissues. For instance, Madagascar periwinkle and opium poppy store alkaloids in laticifers, whereas meadow rue does in root epidermis and stem cortex (St-Pierre et al., 1999; Bird et al., 2003; Samanani et al., 2005).

For a long time, laticifers were known as the cells where both the biosynthesis of BIAs and their accumulation takes place (Fairbairn and Wassel, 1964; Fairbairn et al., 1968; Wilson and Coscia, 1975; Roberts et al., 1983). However, the localization of transcripts of BIA pathway genes and corresponding enzymes were shown in different cells via *in situ* hybridization and immunofluorescent detection assays (Bird et al., 2003; Weid et al., 2004). Early studies demonstrated that *TYDC* transcripts were localized in phloem instead of laticifers (Facchini and De Luca, 1995). Also, SalSyn (Cyp719B1) and SalR enzymes which are belong to morphinan pathway were not found to accumulate in latex (Gerardy and Zenk, 1992; Gerardy and Zenk, 1993). Further detections with immunoflourescence labeling using polyclonal antibodies revealed two additional cell-types involved in BIA biosynthesis and localization: sieve elements and companion cells of phloem. In this study, morphine and sanguinarine pathway enzymes CYP80B1, BBE, and COR were found to be localized in sieve elements while their corresponding gene transcripts were detected in companion cells using *in situ* hybridization technique (Bird et al., 2003). A similar study demonstrated the localization of four more BIA enzymes (CNMT, 6OMT, 4OMT, SalAT) and their gene transcripts in sieve elements

and companion cells, respectively (Samanani et al., 2006). The distinct localization of BIA metabolism indicated a diverse and complicated mechanism in plants involving intercellular transport of pathway intermediates and alkaloids (Facchini and De Luca, 2008).

According to translocation model; biosynthetic genes are expressed in companion cells, their corresponding enzymes are transported to sieve elements where biosynthesis of alkaloids take place. Eventually alkaloids are translocated to adjacent laticifers to be accumulated in the latex vesicles (Bird et al., 2003; Samanani et al., 2006; Facchini and De Luca, 2008). Contradicting to this model, gene transcripts of SalAT and COR were found in cytoplasmic content of laticifers (Decker et al., 2000; Larkin et al., 2007; Allen et al., 2008).

Another opium poppy enzyme localization study based on immunofluorescence labeling demonstrated that 4OMT, 7OMT, SalAT and COR enzymes were synthesized in phloem paranchyma and laticifer cells (Weid et al., 2004). Thus, an alternative model has been developed suggesting that initial stages of biosynthesis starts at parenchymatic cells in phloem and upcoming intermediates are transferred to laticifer cells to complete biosynthesis and accumulate.

To study cell-specific expression of BIA biosynthesis genes and site-specific accumulation of alkaloids is necessary for metabolite engineering of opium poppy since proper transcriptional promoters could be selected in this way. In spite of considerable efforts in cell-type specific localization of enzymes and related products in morphine synthesis, it remained controversial. A further research has been carried out to bring a solution to disparate results obtained before (Onoyovwe et al., 2013). In this study, subsequent to identification and localization of morphinan-branch enzymes through immunofluorescence labeling with polyclonal antibodies, shotgun proteomics approach has been used to measure relative abundance of them. Three thebaine-to-morphine enzymes were extensively found in latex while all six enzymes were detected in sieve elements too. Additionally, the intermediates from (*S*)-norcoclaurine to salutaridine were found to occur largely in sieve elements indicating a requirement for the translocation of pathway intermediates between cells (Onoyovwe et al., 2013).

In opium poppy a translocation mechanism based on plasmodesmata structures between sieve elements and laticifers has been suggested in which both symplastic and apoplastic transport could be used (Facchini and De Luca, 2008). Although it is not demonstrated in opium poppy yet, ATP binding cassette (ABC) transporter is found to be involved in berberine efflux and uptake and in *C. japonica*. Figure 1.8 depicts the possible translocation points as well as gene expression pattern and enzyme localization in opium poppy.

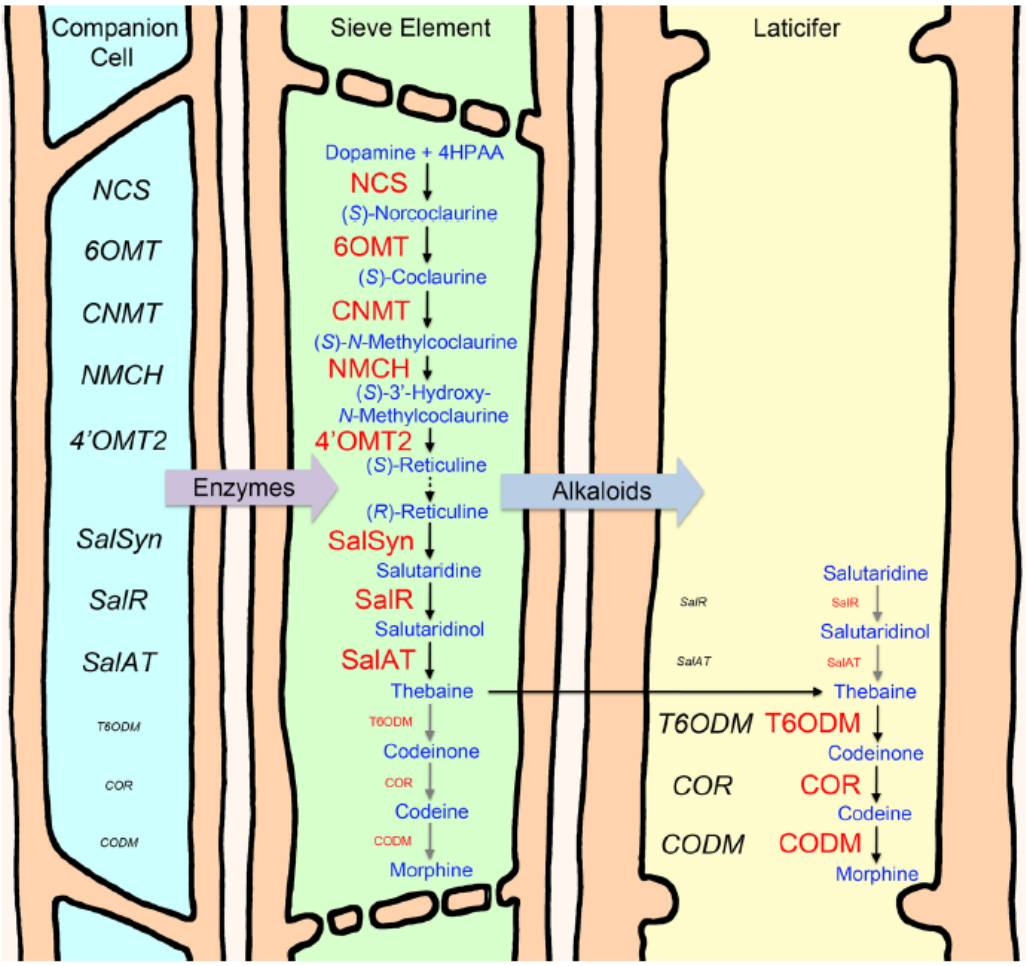



**Figure 1.7.** Cellular localization of morphine biosynthetic enzymes (Beaudoin et al., 2014)

### 1.3.10.2. Subcellular compartmentalization of BIA enzymes

Some of the enzymes involved in biosynthesis of specific BIAs linked with a subcellular organelle. For example, sanguinarine branch enzymes have been localized in opium cell cultures at subcellular level with different methods. Initially BBE, CheSyn, StySyn, MSH, and P6H were shown to associate with microsomes rather than relatively heavy endoplasmic reticulum (ER) by density gradient fractionation method (Amann et al., 1986; Ruffer and Zenk, 1987; Tanahashi and Zenk, 1990; Bauer and Zenk, 1991). Enzymes other than BBE were all P450 dependent endomembrane proteins and thus BBE was shown to possess a 25-aa *N*-terminal signal peptide directing the enzyme to the ER lumen and further targeting to vacuole. It was also understood that the enzyme is

functional until translocation to the vacuole where the interior pH is under 8.8 which is required for BBE activity (pH:8.8) (Bird and Facchini, 2001).

Another enzyme NCS was also reported to have an ER-targeting signal peptide suggesting a subcellular translocation network of sanguinarine pathway intermediates (Samanani et al., 2004; Lee and Facchini, 2010). It is then postulated that the final transformations of sanguinarine pathway intermediates such as dihydrosanguinarine, could probably occur in specialized alkaloid synthesizing vesicles (Amann et al., 1986). Contradicting to that, later studies reported that these vesicles were not distinct organelles, instead ER dilations which leads to vacuole without involving Golgi apparatus suggesting BIA metabolism to be entirely associated with the ER (Alcantara et al., 2005).

Further localization studies based on immunofluorescence labeling enlightened subcellular compartmentalization of BIA biosynthetic enzymes and products to some extent. In a recent study opium poppy cell cultures has been investigated using particle bombardment mediated transient expression and subsequent GFP imaging approach to localize sanguinarine pathway enzymes systematically (Hagel and Facchini, 2012). As depicted in the Figure 1. 9, NCS is found to be present in ER lumen which necessitates the inflow of dopamine and 4HPAA. Newly produced (*S*)-norcoclaurine is expected to be subsequently transported outside to react with a cytosolic enzyme 6OMT. Next enzyme CNMT is also located in cytosole. (*S*)-reticuline biosynthesis is thought to involve a metabolic channel on the cytosolic face of the ER. There are five enzymes (NMCH, CheSyn, StySyn, MSH and P6H) found to be P450-dependent endomembrane proteins whose catalytic sites supposed to be on the cytosolic site of ER membrane. Detection of BBE enzyme inside ER indicates the influx of (*S*)-reticuline into ER to form (*S*)-scoulerine which subsequently transported into the cytosole from a second membrane transit point. A third point of translocation between ER and cytosole is predicted to be present in order to import intermediates formed after five consecutive steps, since the next pathway enzyme DBOX has a putative N-terminal transit peptide (Hagel et al., 2012) and its related enzymes, STOX, found to be ER-associated (Amann et al., 1988). After its biosynthesis, sangunarine is finally transported to vacuole within the vesicles (Alcantara et al., 2005). These studies shows that there is a complex transfer of intermediates between subcellular compartments in consistent with the related

enzymes which is important in the regulation of subcellular localization (Beaudoin and Facchini, 2014).

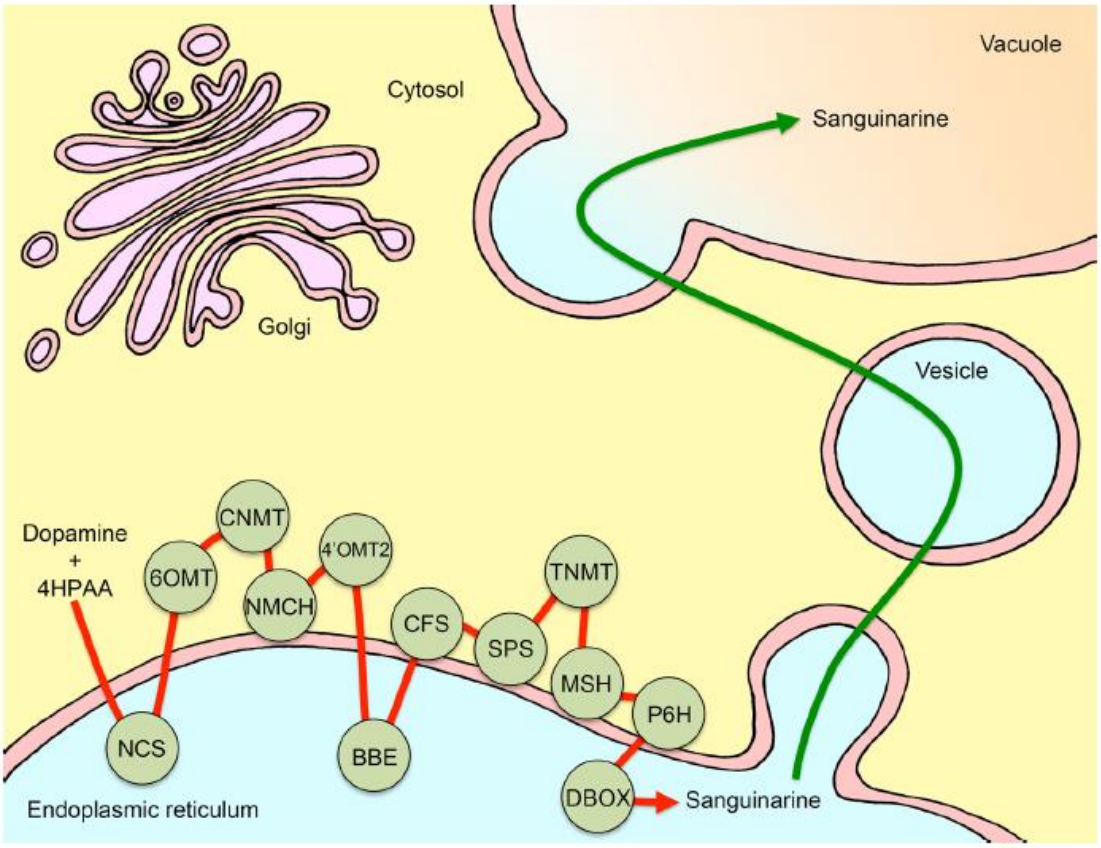



**Figure 1.8.** Subcellular localization of BIA pathway enzymes in sanguinarine biosynthesis (Beaudoin et al., 2014)

### *1.3.10.2.1.* Metabolic channeling

Generally, enzymes in the same pathway locate in close proximity to cooperate more efficiently (Winkel, 2004). Such enzyme clustering functions as a metabolic channel which provides direct transfer of the intermediates. Hence, dilution based decrease in concentration of pathway intermediates in cytosol could be prevented. In addition, metabolic channeling also circumvents other competitive enzymes in favor of intermediate stability (Ziegler and Facchini, 2008).

Metabolic channeling of some pathway enzymes have been investigated and found to be related with both primary and secondary metabolism (Burbulis and Winkel-Shirley, 1999; Panicot et al., 2002). Furthermore, *COR* genes in morphinan pathway have been silenced with RNA interference to investigate a possible enzyme complex. According to

the results, instead of COR substrates codeinone and morphinone, only (*S*)-reticuline has been accumulated. Absence of downstream intermediates of (*S*)-reticuline in morphine biosynthesis pathway suggested to break the metabolic channeling in this branch (Allen et al., 2004). (*S*)-reticuline accumulation could also be explained by the possible co-suppression of *COR* and *DRR* due to their homology (Facchini and De Luca, 2008).

## 1.4. State-of-the-art Sequencing Technologies Used in Plant Omics Research

As next generation sequencing (NGS) technologies emerged and developed, genomic research based on genome sequencing have been revolutionized (Türktaş et al., 2014). Standard genome and RNA sequencing applications of NGS have extensively contributed to comprehensive DNA and transcriptome analysis (Morozova and Marra, 2008). Before describing the NGS systems in detail, it is necessary to review the history of sequencing technology development for a better understanding of the current status.

### 1.4.1. Sanger sequencing

Almost two decades later than the double helix structure of DNA was revealed by Watson and Crick, plus and minus sequencing method was introduced for the first time (Sanger and Coulson, 1975) and dideoxy sequencing method (also known as chain-terminator sequencing) has been developed in the following years (Sanger et al., 1977). The latter one has been the most extensively used sequencing method, also employed in human genome project. The idea behind this method is that the DNAPoly-dependent DNA synthesis occurs in the presence of deoxynucleotide-triphosphate (dNTPs) but terminates if a dideoxynucleotide (ddNTP) is incorporated into the stand which happens in a random manner (Sanger et al., 1977). Each time a ddNTP is bound to the 3’ end of the growing chain, the polymerization is disturbed leaving oligonucleotides in different lengths. These oligonucleotides are then separated using PAGE resulting in a pattern in which the DNA sequence could be determined.

Initially, four individual reactions, each time with a different ddNTP, are needed to complete sequencing. However, automation of DNA sequencing along with the advances in fluorescence detection has led to combine four reactions in a single one via labelling with different fluorescent dye colors (Smith et al., 1986; Prober et al., 1987).

Moreover, slab gels have been replaced by capillary electrophoresis to increase separation capacity (Cohen et al., 1988). Further improvements in sequencing have led the birth of state-of-the art Sanger sequencers with low-error rate and long read length. Applied Biosystems introduced ABI 370A, ABI Prism 373, ABI Prism 377, ABI Prism 310, ABI Prism 3730 and ABI Prism 3730xl to the market respectively. With the recent one, read capacity has increased up to 900bp with 96 capillary (Morozova and Marra, 2008). However, Sanger sequencing is still time-consuming and not cost-effective for large sequencing projects (Morozova and Marra, 2008).

### 1.4.2. Next generation sequencing

To circumvent the limitations of Sanger sequencing, second generation technologies have emerged. Development of novel instruments provides high throughput and speed due to the massively parallel nature which enables generation of huge data in a single experiment compared to Sanger method. In spite of assembly difficulties resulting from shorter read lengths, accuracy could still be higher because of high level of sequence coverage. Additionally, these new technologies greatly reduce the cost and manpower. Currently, there are three main widely used NGS platforms: 454 Pyrosequencing, ABI SOLiD and Illumina.

#### 1.4.2.1. Roche 454 sequencing

In 2005, 454 Life Sciences introduced the first NGS system "Genome Sequencer (GS)" to the market. This method is based on the principle of sequencing by synthesis using pyrosequencing technology. Different from Sanger sequencing, detection of released pyrophosphate is employed instead of ddNTP-based chain termination. After Roche purchased the company in 2007, GS-FLX Titanium as a newer version of the platform has been developed.

454 sequencing has 3 main steps involving DNA library construction, emulsion PCR (emPCR) and sequencing. Amplification of the sample DNA is imaged on a solid surface where it is attached (Thudi et al., 2012). Sequencing libraries can be created from any kind of dsDNA (including cDNA) of any size. Following nebulization of the library (fragmentation under high pressure), 454-specific adaptors are ligated to the dsDNA. Small fragments are then exposed to size selection using magnetic beads and

direct quantification of the library is carried out by fluorescent tag on the adaptors. Prior to emPCR, library quality is checked utilizing Bioanalyzer High Sensitivity DNA chip.

Next, emulsion-based clonal amplification of the sample library is performed. This second step involves preparation of the live and mock amplification mixes followed by DNA capture. DNA capture beads and PCR ingredients are emulsified in thermostable water-in-oil microreactors where further amplification takes place. After emPCR, microreactors are broken to enrich DNA-bound beads. All reactions in each microreactor processed in parallel. The beads are then counted in a cell counter. Afterwards, sequencing primers are annealed to the amplified DNA sequences attached on the beads. Then, sequencing has been performed on a metal coated picotiter plate (PTP) which has 3.400.000 wells each 29 micrometer (μm) in diameter. The size of ssDNA amplified during emPCR is about 20 micrometers that only one DNA-carrying bead can fit into a well. Layers of packing, enzyme and $PP_i$ase beads are deposited on PTP before the plate is located into GS-FLX for sequencing.

In this final step following reactions takes place: Four nucleotides of dNTPs (ATCG) are flowed on the PTP for 200 cycles. Whenever a complementary base is added to the DNA strand by DNA polymerase, one pyrophosphate ($PP_i$) is released and converted to ATP by sulfurylase in the presence of APS. This ATP is used by luciferase to convert luciferin to oxyluciferin thus created a chemiluminescent signal which is captured by a CCD camera inside the instrument (Figure 1.10).

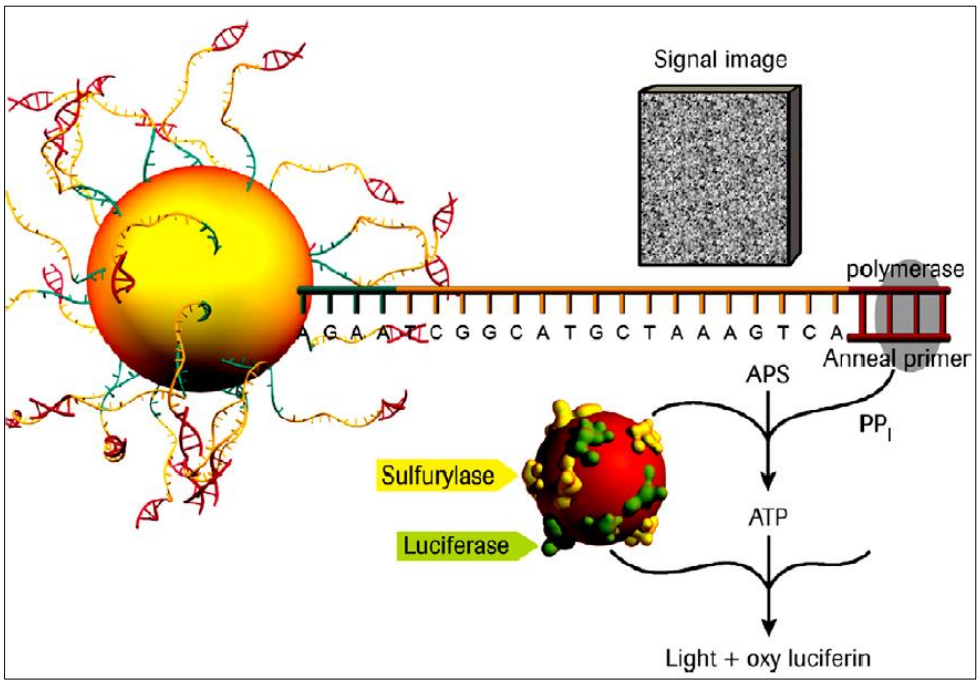



**Figure 1.9.** Sequencing-by-synthesis in 454 GS-FLX Platform (extracted from Roche 454 GS-FLX Catalogue)

A pyrogram is recorded. This raw data is then processed into signals and the sequence of the DNA is determined. Further mapping and assembly can be performed by GS-FLX softwares. The signal strength is proportional to the number of PPi released, so that the length of homopolymers could be misinterpreted which makes the pyrosequencing approach error-prone. However, 454 technology is still one of the most widely used and published NGS platforms (Liu et al., 2012).

#### 1.4.2.2. ABI/Solid sequencing

Applied Biosystems has acquired SOLID (Sequencing by Oligo Ligation Detection) in 2006 and launched their system in the following year. The technology relies on massively parallel sequencing by hybridization-ligation. Library preparation and clonal amplification of the DNA by emPCR are similar to that of 454 platforms in principle.

After fragmentation of the DNA sample into smaller sizes (400-850 bps), P1 and P2 adaptors are ligated to the both ends of the library DNA. DNA fragments are immobilized onto P1-coated magnetic beads. So that, the starting sequence of every fragment will be known and also identical. emPCR is then carried out in microreactors.

The resulting amplicon-carrying beads are covalently attached to a glass slide. The open slides could be divided up to eight segments to increase the number of sequenced samples contributing higher levels of throughput (Valouev et al., 2008).

In ligation-based sequencing, first, sequencing primers are hybridized to P1 adaptors on the beads. A collection of all possible fluorescently labeled complementary di-base probes are introduced to ligate to the primers. Multiple cycles of hybridization, ligation, dephosphoylation, detection and cleavage are repeated until the read length has been determined. A fluorescent signal is generated for each ligation and fades away by the cleavage of fluorophores. After 5 rounds of sequencing 35 base-pair (bp) read is obtained. And a second round of ligation series is performed subsequent to the removal of extended sequences. Hence sequencing accuracy can increase up to 99.85% (Liu et al., 2012). After several upgrades in years the state of the art SOLID technology has higher read length (up to 85 bp) and accuracy (99.99%). SOLID is employed in whole genome resequencing, targeted resequencing, transcriptome sequencing and epigenome studies (Liu et al., 2012).

#### 1.4.2.3. Illumina/Solexa platform

"Genome analyzer (GA)" was first introduced to the market shortly before the acquisition of Solexa (the developer) by Illumina in 2007 and rapidly improved in the following years. The instrument adopts the "sequencing-by-synthesis" (SBS) technology which is based on immobilizing the DNA/cDNA fragments on a solid surface for amplification and sequencing using fluorescent reversible terminator dNTPs (Bentley, 2006; Liu et al., 2012; Mutz et al., 2013) (Figure 1.11).

Sequencing library preparation involves fragmentation of sample DNA/cDNA and also ligation of specific adapters to the both ends of the fragments. 3' and 5' overhangs resulting from fragmentation are converted into blunt ends by end repair enzymes. Subsequent to that, only one Adenine base is added to the 3' ends of each fragment for proper ligation to specific adapter sequences containing a single T at the 3' ends. Addition of adapter sequences enables library amplification with end-specific primers. The amplification takes place after adapter-ligated fragments are attached to the inner surface of a flow cell (or single-molecule array) which confers ultra-dense layer of

primers with complementary sequences to the adapters (Koboldt et al., 2012). The flow cell device has eight channels that allow the amplification of separate libraries at a time in a massively parallel manner. A great number of immobilized ds-DNA copies are produced per cluster which represents one fragment. Each colony containing nearly 1.000.000 copies becomes sufficient for signal production for detection in sequencing. To achieve amplification on a solid surface with immobilized primers using isothermal conditions, bridge-PCR is performed instead of typical PCR. Single-stranded DNA (ssDNA) fragments hybridize to immobilized primers resulting in a molecular bridge structure which serves as a template for extension step. Repeated reagent flush cycles of denaturation, annealing, extension and wash, result in ssDNA clusters on the surface which is then hybridized to sequencing primers (Mardis, 2008; Myllykangas et al., 2012; Türktaş et al., 2014).

The sequencing step involves the addition of reversible terminator dNTPs each labeled with base-specific fluorophores. Each time an appropriate base is incorporated into the growing chain by DNA polymerase, a characteristic fluorescent signal is emitted which is detected by the CCD camera. Following each incorporation and resulting signal detection, 3' blocking unit is removed from the terminator in order to proceed to the next cycle. This is why the dNTPs called as "reversible terminators". The number of cycles determines the length of the read, also can be defined by the user of the instrument to establish a discrete and uniform read length. Besides, the emission wavelength with the signal intensity determines the base call. Each cluster is sequenced simultaneously (Mardis, 2008; Myllykangas et al., 2012; Türktaş et al., 2014).

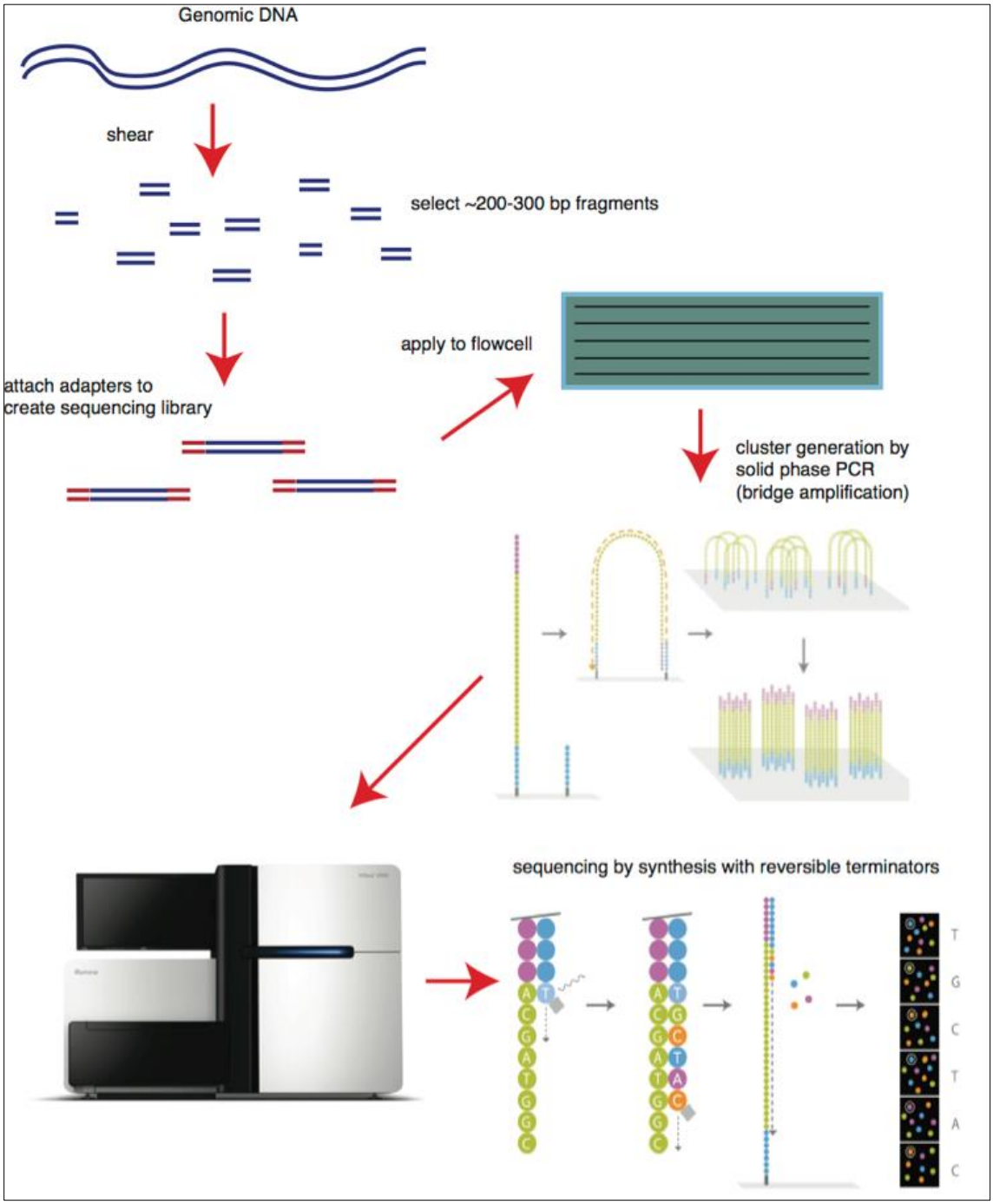



**Figure 1.10.** Overview of Illumina sequencing (adapted from http://bitesizebio.com/13546/sequencing-by-synthesis-explaining-the-illumina-sequencing-technology/, June 2016)

Illumina is one of the most extensively used high-throughput sequencing platform. In 2010, Illumina launched HiSeq, utilizing the same principle with GA, which produces the highest output (600 Gb per run, max 100bp read length for both ends) among the

NGS systems (Liu et al., 2012). BGI has adopted 'Illumina HiSeq' to serve globally. Today, Illumina is responsible for a large fraction of DNA sequence information produced worldwide in terms of number of total bases. There is also a cheaper and faster bench-top sequencer called "Illumina MiSeq". However, it is capable of processing only one sample per run and can produce less sequencing data (200 Mb to 2 GB). MiSeq is convenient for amplicon and bacterial DNA sequencing whereas HiSeq can be utilized in *de novo* sequencing, whole-genome sequencing, DNA sequencing, targeted resequencing, RNA sequencing (such as small RNA and transcriptome analysis) (Liu et al., 2012).

#### *1.4.2.3.1.* Comparison of Illumina HiSeq and 454 GS-FLX platforms

At the first glance, there are a few important differences between the two widely used technologies. To start with; the read length is significantly longer in 454 compared to Illumina. 454 can produce up to 700-800 bp while Illumina HiSeq can achieve 100 bp at most. Adversely, Illumina generates significantly more reads than 454. A typical HiSeq can end up with 200 million reads; however, 454 can only produce 1 million reads on an entire picotiter plate. Thirdly, the cost of sequencing greatly differs among all NGS platforms. Comparisons in terms of 'per-base' prize can be helpful but sometimes misleading; as the accuracy increases it may be worth. Illumina seems more advantageous than 454 with a drastic prize difference per base. In contrast, 454 can be the technology of choice for specific goals where long reads are critical such as *de novo* assembly of a genome (Shendure and Ji, 2008).

From a sequencing mechanism perspective, 454 differs with an advanced method of pyrosequencing in comparison to llumina which employs flour-labeled reversible terminators. This chemistry difference leads the users to completely different error modes: 454 mostly prone to have indels, especially at homopolymer sites, since the signal detection weakens after continuous 6 bases. Illumina is more accurate regarding homopolymers, but because of producing shorter sequence reads, it is not good at resolving short sequence repeats (SSRs) effectively and additionally tends to have substitutions owing to the use of reversible terminator and modified DNA polymerases (Bentley, 2006; Hutchison, 2007; Morozova and Marra, 2008).

**Table 1.1.** Comparison of 2nd generation sequencing platforms

| Sequencer | 454 GS FLX | HiSeq 2000 | SOLiDv4 |
|---|---|---|---|
| **Sequencing mechanism** | Pyrosequencing | Sequencing-by-synthesis | Bridge amplification |
| **Read length** | 700-800 bp | 50SE, 50PE, 100PE, 150PE | 50 + 35 bp, 50 + 50 bp |
| **Reads** | 1M | 3G | 1200-1400 M |
| **Output data/run** | 700 Mb | 600 Gb | 120 Gb |
| **Run time** | 23h | 8.5 days | 2 weeks |
| **Advantage** | Read length, fast | High throughput | Accuracy |
| **Disadvantage** | High cost, error rate with polybase >6 | Short read assembly | Short read assembly |

### 1.4.3. Third-Generation Sequencing

In addition to NGS technologies, third-generation sequencing with mmore advanced technology has been emerged recently. Single Molecule Real-Time (SMRT) sequencing by Pacific Biosystems (PacBio) (Eid et al., 2009), and nanopore sequencing by Oxford Nanopore Techniques (ONT) (Branton et al., 2008) has been developed to enable single molecule sequencing with longer reads. Different from NGS, 3rd generation sequencing does not involve amplification during library preparation step.

**PacBio SMRT sequencer** was first introduced to the market in 2010 (Lee et al., 2016). The technique is mainly based on "sequencing-by-synthesis". As each fluorescently labelled dNTP is incorporated into the newly synthesized strand, the DNA polymerase activity is detected real-time. The reactions take place on a surface where the polymerase molecules are attached (SMRT-cells). In this way, the process becomes faster and produces longer reads than NGS sequencers (over 10 kilo bp) (Bleidorn, 2016). However, high error rate and relatively small output per cell can be considered as the weaknesses of the technique.

More recently, **Oxford nanopore sequencer, MinION**, has been released in 2014 (Lee et al., 2016). MinION technique is based on biological nanopores. Small channels

within a phospholipid bilayer are formed by specific proteins. The ionic gradient which is formed by electrodes in a salty buffer, forces DNA to pass the nanopore due to its negative charge. As it moves down, the ion current is detected by a sensor. During library preparation step, a hairpin adapter and a motor protein are ligated to the ends of each DNA strand. Motor protein pulls the single stranded DNA strand through the nanopore, while the hairpin adapter keeps the strands from separating and this way, the second strand is sequenced immediately after the first one. Compared to PacBio, read lengths are similar. But MinION is less accurate and used mostly in amplicon or small genome sequencing because of low throughput (Bleidorn, 2016).

### 1.4.4. Transcriptome sequencing for gene expression profiling in plants

Almost every cell in a multicellular organism contains the same genetic material but specific genes are only transcribed into RNA and processed through proteins at specific times. According to phenotypic response to the condition in which they exist, cells may differ in gene expression profiles in a specific manner giving rise to physical, biochemical, and developmental variations.

A transcriptome represents the entire set of transcripts of the cells for a specific developmental phase or a physiological condition. It is important to understand the transcriptome to reveal the functional units of the genome and also to gain insight on development and disease. Thus, researchers collect and compare transcriptomes from different plant cells or tissues and investigate the molecular key players behind differential expression. Factors such as RNA editing, varying transcription initiation/termination sites and alternative splicing cause transcriptome to exhibit a complex structure.

By considering transcriptome all kinds of transcripts can be catalogued (mRNA, ncRNA, sRNA, miRNAs etc.). Besides, transcriptome studies also reveal a comprehensive frame of active and silent genes as well as transcriptional structures of the genes and patterns of RNA splicing under various conditions. Furthermore, dynamic levels of gene expression in different circumstances can be measured at the whole transcriptome level (Wang et al., 2009; Kogenaru et al., 2012). For this purpose, several hybridization and sequence-based methods have been developed in time.

#### 1.4.4.1. Microarray technology

Along with the invention of microarrays in 1990s, it has been widely used in large-scale transcriptome profiling studies (Zhao et al., 2014). Microarray experiments are based on the hybridization of target oligonucleotides with immobilized probe sequences on a solid surface. Fluorescent dyes in different colors are used to label the transcripts of different conditions. Subsequent to hybridization of the transcripts to the probes on the array, the intensity of fluorescence is measured to quantify each transcript (Hegde et al., 2000; Xiang and Chen, 2000). Agilent and Affimetrix are the most prominent array platforms (Tan et al., 2003; Baginsky et al., 2010). Currently, microarrays could be easily afforded by many researchers as being an inexpensive (except specialized tilling arrays for large genomes) and high-throughput approach. However, limitations of array technology such as the need for previous DNA sequence knowledge and background signals due to cross hybridization (Wang et al., 2009; Zhao et al., 2014). These limitations resulted in a more sensitive technology demand. In this respect, novel technologies such as transcriptome sequencing (RNAseq) have emerged in following years. Compared to microarrays, transcription levels are measured more quantitatively by using sequence-based methods.

#### 1.4.4.2. cDNA libraries, ESTs, SAGE and RNAseq

In contrast to array-based methods, sequence-based approaches involve direct sequencing of cDNAs to determine gene expression levels. Quantity of cDNAs reflects corresponding gene expression level. A cDNA library represents actively expressed genes in a specific organ, tissue or a developmental stage. Initially total RNA is isolated and mRNAs are then selected for reverse transcription with their poly(A) tails which distinguish mRNAs from other types of transcribed RNAs. cDNA synthesis is carried out with the use of complementary to poly(T) primer and reverse transcriptase. After the construction of cDNA libraries, Sanger sequencing is performed and the sequences are used to design shorter cDNA regions called expressed sequence tags (ESTs). ESTs can be used as probes to find similar transcripts in other samples. As new ESTs are identified, they are collected in publicly available databases.

In addition to Sanger sequencing of cDNAs or ESTs, tag-based approaches have been developed for gene expression profiling. Among them serial analysis of gene expression (SAGE) was the first high-throughput sequencing method used in transcript analysis followed by massively parallel signature sequencing (MPSS) (Velculescu et al., 1995; Brenner et al., 2000). In SAGE, tags representing different expressed sequences are used for sequencing. Although low abundance transcripts can be quantified by SAGE, only a limited number of expression profiles could be obtained at a time due to the technical restrictions and amount of RNA samples required (Eom et al., 2006). Thus, it is still not being utilized largely as microarrays.

Along with the breakthrough in next generation sequencing (NGS) technologies, **RNA sequencing (RNAseq)** has emerged as a valuable tool for accurate transcriptome profiling with powerful advantages (such as high level of sensitivity and throughput) over other existing methods discussed above. This method is also known as whole transcriptome shotgun sequencing and mainly used for characterization of RNA content and composition of a given sample. Taking the advantage of being derived directly from mRNA constituent, mostly the coding elements, RNAseq provides a good source for functional genomics studies. RNAseq analysis is extensively used for gene expression profiling. It enables researchers to determine the differentially expressed genes under different stress treatments or different developmental stages. Small RNA sequencing (miRNA, tRNA, rRNA) is also implicated by RNAseq (Morozova and Marra, 2008). Furthermore, transcriptome characterization and annotation can be achieved using this method. Transcription start sites, exon-intron boundaries, splice variants, regulatory elements can all be determined. Besides, gene-associated molecular marker development is another field where RNAseq can be utilized.

In principle, RNAseq approach involves three steps involving library preparation, sequencing and analysis.

First the RNA samples collected from a tissue or cell type are reverse transcribed into cDNAs and adapters are added to the ends of the fragments. Then these cDNA libraries are sequenced using one of the high throughput sequencing technologies (Mortazavi et al., 2008). Lastly, the reads are mapped on a reference genome (if the organism has a

sequenced genome) or subject to *de novo* assembly (for species without a reference genome) for the further data analysis.

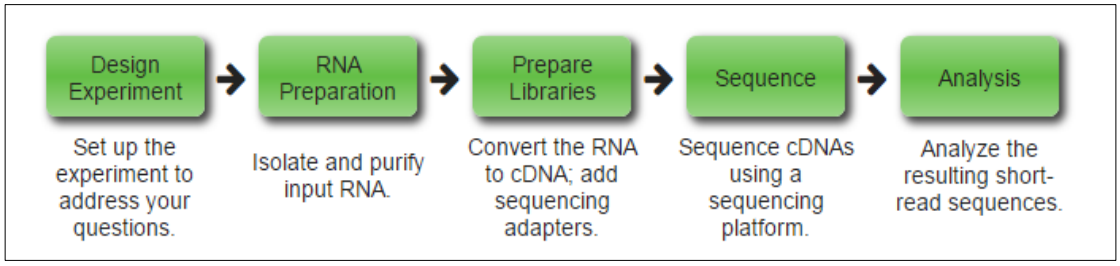



**Figure 1.11.** Steps of a typical RNAseq experiment

Practically, 454 GS-FLX pyrosequencing, Illumina HiSeq sequencing-by-synthesis (SBS) and ABI SOLID sequencing-by-ligation technologies can be implemented for RNAseq (Costa et al., 2010). The resulting read length, throughput and cost may vary according to the method followed. 454 GS-FLX system surpass the other NGS technologies by means of long read length (up to 700 bp). Long reads enable more efficient transcriptome assembly or more accurate mapping on a reference. However, the number of reads produced by Illumina and SOLID is undisputedly high when compared to 454 GS-FLX, which also affects the cost per run. More reads could also contribute to more accurate expression measurements. In addition to that, rare transcripts are more achievable with high sequence depth obtained by short-read sequencing methods (Nagalakshmi et al., 2008). There is also a hybrid strategy which combines three methods employing short-reads to obtain high coverage and long-reads to scaffold the relatively short contigs (Martin and Wang, 2011; Saroglia and Liu, 2012).

Illumina, in particular, is compared to microarray and shown to be greater in terms of catching novel transcripts or transcripts with low expression levels (Marioni et al., 2008).

#### 1.4.4.3. Advantages of RNAseq and comparison with other methods

Initially, microarray technology has been the most widely used tool in transcriptome-wide gene expression profiling studies. However, RNAseq has started to supersede microarray with many advantages. Recent studies have compared microarray data to

that of RNAseq from different aspects such as technical reproducibility, expression analysis, differentially expressed genes or isoforms, etc (Marioni et al., 2008; Fu et al., 2009; Bradford et al., 2010; Giorgi et al., 2013; Wang et al., 2014) (Table 1.2).

Unlike microarray, RNAseq does not require prior genomic sequence knowledge. Furthermore, RNAseq is remarkably high-throughput which will increase the sequence depth so that weakly expressed transcripts would be also detected (Wang et al., 2014). In addition to that, exon-intron boundaries could be revealed precisely allowing to analysis of complex transcriptomes (Wang et al., 2009). Likewise, the sequence variations such as SNPs are also detectable by this method (Wang et al., 2009).

Another advantage over microarray is that the background signals in RNAseq are very low and cross-hybridization problem has been surpassed. RNAseq also provides alternative splice variants which could not be detected by microarray easily (Wolf, 2013). Moreover, a higher dynamic range of expression levels is obtained as long as increasing number of reads are produced (Wang et al., 2009; van Vliet, 2010).

According to the comparison studies, RNAseq was found to be highly reproducible and compared to microarray, lesser amount of RNA sample is sufficient for testing (Marioni et al., 2008; Kogenaru et al., 2012)

All in all, RNAseq is a high-throughput powerful method with highly quantitative and accurate digital expression measurements in whole transcriptome analysis. Nonetheless, microarray could still be the method of choice for analysis of multiple samples. So that, microarray and RNAseq can be considered as complementary tools utilized for different purposes (Saroglia and Liu, 2012; Wolf, 2013).

**Table 1.2.** Advantages of RNASeq compared with other transcriptomics methods (extracted from Wang et al., 2009)

| **Technology** | **Tiling microarray** | **cDNA or EST sequencing** | **RNA-Seq** |
|---|---|---|---|
| Principle | Hybridization | Sanger sequencing | High-throughput sequencing |
| Resolution | From several to 100 bp | Single base | Single base |
| Throughput | High | Low | High |
| Reliance on genomic sequence | Yes | No | In some cases |
| Background noise | High | Low | Low |
| Dynamic range to quantify gene expression level | Up to a few-hundredfold | Not practical | >8000-fold |
| Ability to distinguish different isoforms | Limited | Yes | Yes |
| Required amount of RNA | High | High | Low |
| Cost for mapping transcriptomes of large genomes | High | High | Relatively low |

#### 1.4.4.4. Challenges and limitations of RNAseq

Despite of many advantages, RNAseq may end up with excess amount of rRNA sequences, short reads, differing read density throughout the fragment which makes the technique challenging (Kogenaru et al., 2012). Another challenge comes up when larger RNA molecules are tried to be analyzed since further manipulations are required during cDNA library construction step. RNA molecules that are large in length must be subject to fragmentation prior to deep sequencing, while small RNAs like siRNAs and miRNAs can be directly sequenced subsequent to adapter ligation. Fragmentation methods may cause bias on the transcripts. Moreover, identical short reads generated during PCR may be confusing since they may reflect either the abundance of the RNAs or the PCR artefacts (Wang et al., 2009).

Another key issue is the rate of mRNA stability/turnover which is generally ignored during transcriptome profiling. So, it is important to consider that the expression level

of a specific gene is not a direct measure of corresponding protein abundance (Vogel et al., 2010).

Further precautions should be taken while deciphering the expression patterns. Although the tissue is collected carefully, gene expression patterns may still represent different cell populations within the same tissue.

Lastly, strand-specific libraries could be prepared to have the transcript orientation knowledge which is especially useful in gene annotation. Overlapping genes located on opposite strands could be distinguished in this way. Nevertheless, construction of strand specific library is troublesome with additional steps. It is important to keep in mind that antisense strands would easily be mixed with reverse transcription artefacts (Wang et al., 2009; Wolf, 2013).

#### 1.4.4.5. RNAseq data analysis

Regardless of the NGS platform used, a single sequencing run generates millions of transcripts to be analyzed. High throughput sequencing systems brought about the need of advanced computational algorithms to handle the enormous raw data. A typical RNAseq analysis pipeline involves four major steps. For the beginning, the fluorescent signals produced by NGS platform needs to be converted into read sequences which are subsequently subjected to quality control. Low quality reads are eliminated due to the quality score of each base. Meanwhile, adapters should be removed. Secondly, the short reads are aligned or mapped to the reference genome/transcriptome if there is. Otherwise *de novo* assembly is required. Read mapping is a preliminary but the most crucial step for accurate quantification of the reads. Problems such as sequencing errors, splice variants, single nucleotide polymorphisms (SNPs), repeats and reads corresponding to exon junctions makes this step complicated. However, computational methods like TopHat, SOAP and Bowtie have been developed with different algorithms for better alignment results. Thirdly, the read counts for each gene or transcript are determined to estimate the transcript abundance for gene expression analysis since the number of assigned reads for a particular gene is accepted as a relative measure of transcript abundance. Reads Per Kilobase of transcript per Million mapped reads. Reads per kilo base per million of mapped reads (RPKM) is a widely used method in

differential gene expression of samples in different conditions. Normalization of read counts is required for proper comparison of transcripts in varying lengths. It is obvious that a long transcript will be covered by more reads than a shorter one. Finally, statistical analysis is performed to identify the genes/transcripts differentially expressed among samples and treatments. Several bioinformatics tools currently available for differential gene expression: Cufflinks, edgeR, DEGseq, DESeq and etc each with unique features which make the method preferable regarding the dataset (Jain, 2012; Mutz et al., 2013; Wolf, 2013).

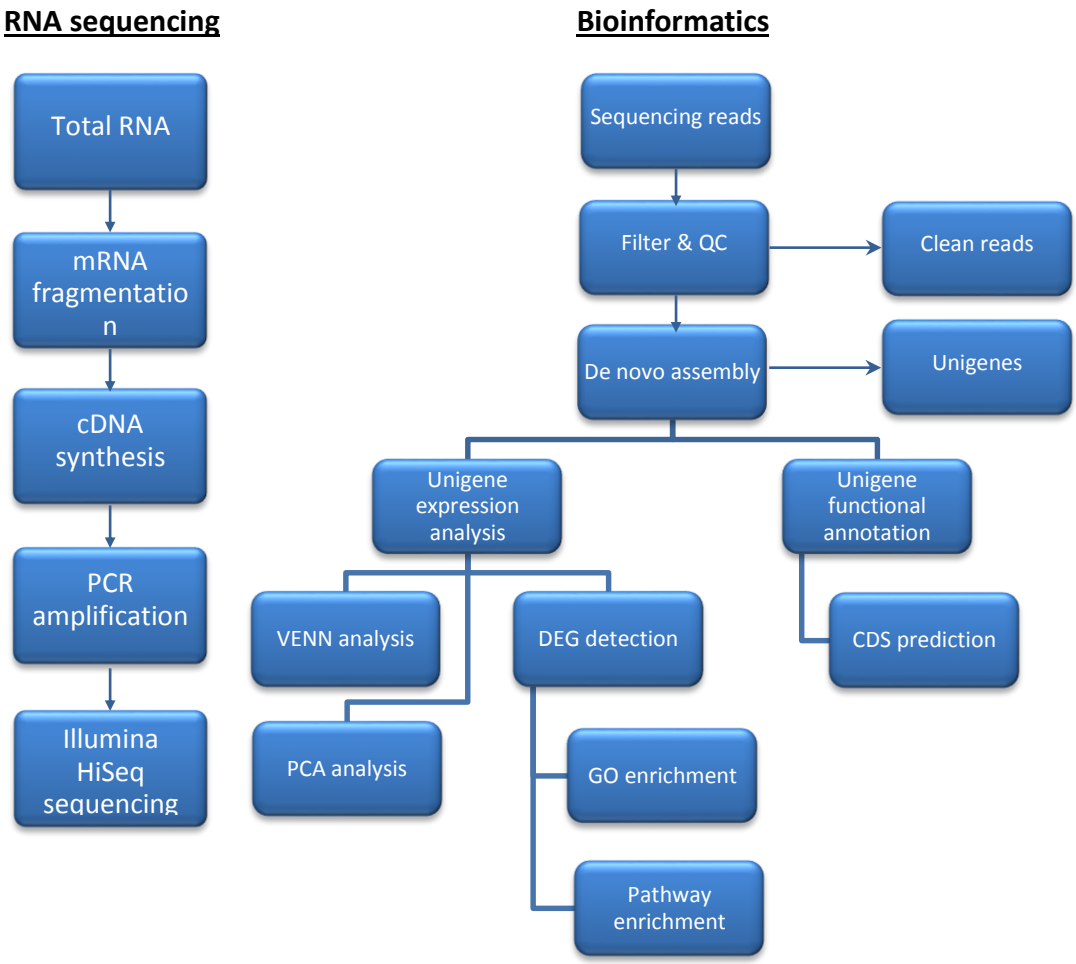



**Figure 1.12.** *De novo* transcriptome analysis flow chart

In this thesis study we aimed to investigate opium poppy transcriptome profiles in MeJA treated stem and capsule tissues of two different cultivars with a focus on morphine biosynthesis pathway. RNAseq approach has been used to dissect the transcriptome profiles. Methods and results are given and discussed in the following sections.

# 2. MATERIAL AND METHODS

## 2.1. Plant Material and Growth Conditions

Two opium poppy (*Papaver somniferum* L.) cultivars used in this study ((morphine-rich) Ofis8 cv. and (morphine-poor) Ofis NM cv.) were obtained from Turkish Grain Board (TMO). The poppy seeds were first sown in plastic viols and they were transferred into pots after germination. Plants were grown in greenhouse conditions (20°C/18°C; 16h light and 8h dark cycles with 60% humidity). Stem and capsule tissues were harvested after eight months. Samples were designated as follows throughout the thesis:

**Table 2.1.** Control and MeJA-treated tissue samples

| Tissues | Control | MeJA Applied |
|---|---|---|
| Mophine Rich Capsule | MRCC | MRCT |
| Mophine Rich Stem | MRSC | MRST |
| Mophine Poor Capsule | MPCC | MPCT |
| Mophine Poor Stem | MPSC | MPST |



**Figure 2.1.** Opium poppy grown in green house

## 2.2. MeJA Treatment

Methyl jasmonate (MeJA; Sigma, Germany) is used as a chemical elicitor to induce morphine biosynthesis in opium poppy. Prior to MeJA treatment, capsule and stem tissues were wounded using a sterile needle without causing a major injury on the plant (Mishra et al., 2013; Gurkok et al., 2015).

In order to prepare the treatment solution, first, 229.2 µl MeJA was dissolved in 10 ml ethanol (100mM stock). Then, 100µl of the stock solution was mixed with sterile water to make 100 ml. Lastly, 100 µl of triton X-100 (0.1 %) was added inside the solution. The treatment solution for control plants contained 100 µl of ethanol instead of MeJA.

Two days after the petals fell; plants similar in length and capsule size were selected and MeJA solution was sprayed over the capsule and stem tissues. Plants were treated with MeJA-free solution were taken as control. Following the treatment, the tissues were collected at a time interval of 3, 6, 9 hours for metabolite and transcriptome analysis. Each group were set as triplets. Tissues were frozen in liquid nitrogen and kept at −80 °C until metabolite analysis and RNA isolation was performed.

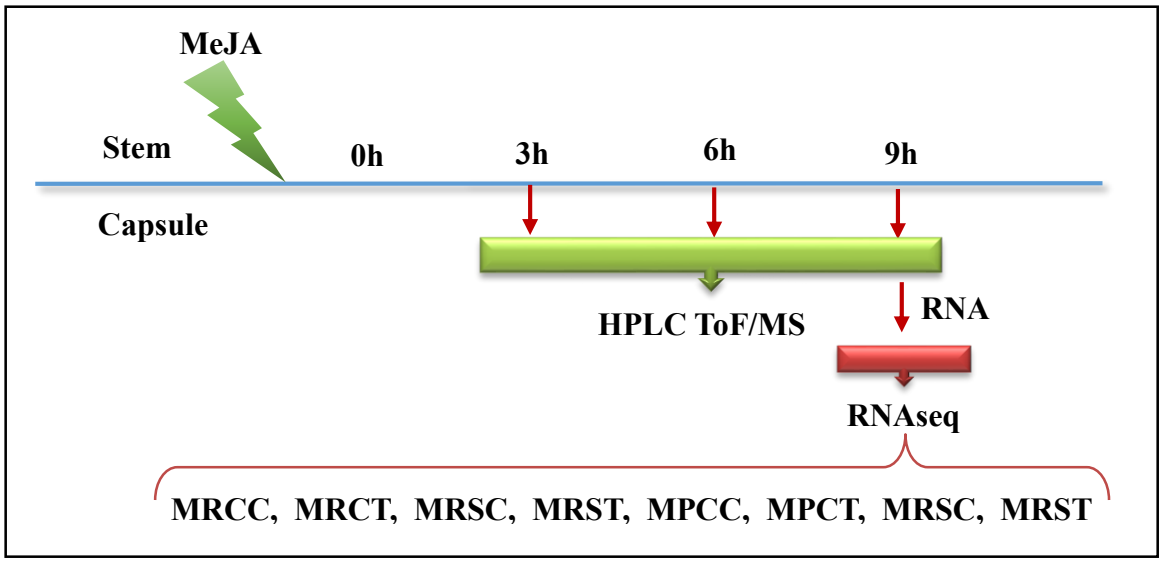



**Figure 2.2.** MeJA treatment and further analysis

## 2.3. Analysis of Alkaloid Contents by ToF LC/MS

ToF LC/MS analysis was performed in order to show the induction of BIA biosynthesis upon MeJA treatment. Alkaloid contents were measured using Agilent platform (http://www.chem.agilent.com). Representing each time point, three control and three MeJA treated plants were selected for both MR and MP cultivars. Stem and capsule samples were left at 28 °C for two days. Dried plant materials were weighed and 0.1 g of each sample was dissolved in 2.5 ml of methanol (HPLC grade, Merck) in a shaker for 24 h at room temperature (RT). Liquid extracts were filtered and resolved in methanol at 2000 ppm concentration. All extracts were then diluted with a ratio of 1:200 to reach a 10 ppm final concentration and were filtered by 0.45 μm membranes to be ready for the ToF LC/MS analysis.

To detect quantitative amounts of the alkaloids; morphine, noscapine, codeine, papaverine, thebaine, and s-reticuline alkaloid solutions were prepared at 10, 25, 50, 100, 250, 500 and 1000 ppb concentrations and loaded into the instrument to be used as standards. Retention time of the standards were identified utilizing mass of the standard alkaloids and their relative ion abundance versus concentration graph (calibration curve) was plotted. Alkaloid concentrations of the samples were then calculated in ppb by employing the plot drawn. Taking into account the concentration and the amount of the samples, alkaloid abundance in the plants were obtained.



**Figure 2.3.** Agilent 1260 infinity LC and 6210 ToF-LC/MS Sytems (www.agilent.com)

## 2.4. RNA Isolation and Characterization

According to the results of ToF LC/MS analysis, 9h samples were found to be induced for morphine biosynthesis. Three biological replicates were set for each control and treatment samples. Total RNA was isolated from capsule and stem tissues by TRIzol® Reagent (Life Technologies, Rockville, MD) following the instructions of the manufacturer as described below:

Each tissue sample (100 mg) was grinded in liquid nitrogen with a pre-chilled, sterile mortar and pestle. Powdered samples were transferred into 2 ml microcentrifuge tubes and 1 ml of TRIzol® Reagent was added on each tube then vortexed. After 5 min incubation of the tubes at room temperature, 0.2 ml of chloroform was added and the caps was closed tightly. The tubes were shaken vigorously by hand for 15 s then left for 2-3 min at room temperature. Following the centrifugation at 15000 rpm for 15 min at 4°C, the aqueous phase formed in the upper side of each tube was transferred to a new microcentrifuge tube. Subsequently, 0.5 ml of pre-chilled isopropyl alcohol (half of the volume of TRIzol used initially) was added. The tubes were left for 10 min incubation at room temperature followed by centrifugation at 15,000 rpm for 15 min at 4°C. The supernatant was discarded and the pellet was washed with 1 ml of 75% ethanol (the

same amount of TRIzol used initially). Sample tubes were vortexed briefly and centrifuged at 10,000 rpm for 5 min at 4°C. The RNA pellet was air-dried for 10 min on bench and then resuspended in 30 µl of DEPC treated water by pipetting up and down. After incubating for 10 min at 55-60ºC, RNA samples were stored at -80ºC until use (Unver et al., 2008; Gurkok, 2013)

Quality and quantity of the isolated RNA samples were evaluated based on Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and NanoDrop 2000c spectrophotometer (Thermo Scientific, Wilmington, DE) measurements at 230, 260 and 280 wavelengths. The ratio of A260/A230 and A260/A280 falling in 2.0–2.2 and 1.8–2.1 range indicates the purity of RNA samples. The RNA integrity was checked by separating the major rRNA bands in 2% agarose gels for 1 h at 100 V.

DNase treatment of 1 ug of total RNA was performed in a 10 µl reaction mixture containing 1 U of DNase I dioxyribonuclease I (Fermentas) and 1x reaction buffer. Reaction took place at 37°C for 30 min and terminated by incubation with adding 1 µl of 25 mM EDTA at 65°C for 10 min. RNA was stored at −20°C (Kantar et al., 2010; Kurtoglu et al., 2014).

### 2.5. Library Construction for Illumina Sequencing

Prior to sequencing library preparation, isolated RNA samples were purified using RNeasy Plant Mini Kit (Qiagen, CA) and quantified by NanoDrop 2000c spectrophotometer (Thermo Scientific, Wilmington, DE). Samples were further analyzed by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) for RNA purity and integrity check. The samples with a RIN greater than 6.5 were used. RIN is a qualitative unit representing RNA integrity. A total of 8 paired-end stranded sequencing libraries (MRCT, MRCC, MRST, MRSC, MPCT, MPCC, MPST and MPSC) were prepared using “TruSeq® Total RNA Library Prep Kit” following the instructions of the manufacturer (Illumina). Library preparation procedure involves; 1) purification and fragmentation of mRNA; 2) first-strand cDNA synthesis; 3) second-strand cDNA synthesis; 4) end repair and adenylation of 3’ ends; 5) adapter ligation; 6) enrichment of DNA fragments; 7) validation of the library and 8) normalization steps (Figure 2.3).

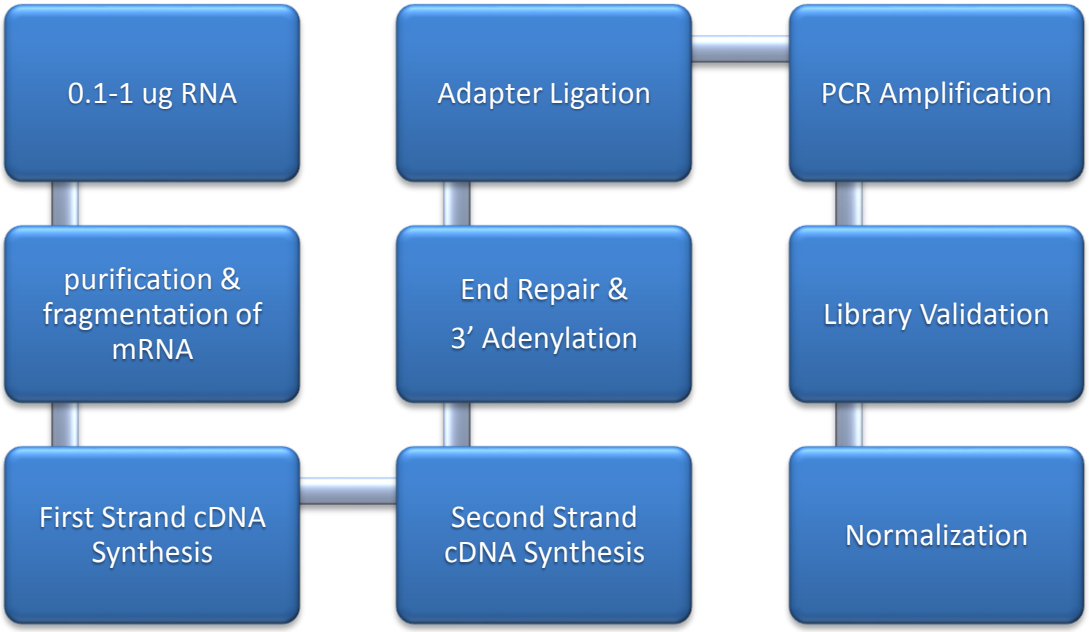



**Figure 2.4.** TruSeq Total RNA sequencing library preparation workflow

Briefly, the polyA containing mRNA molecules was purified with oligo-dT attached magnetic beads and fragmented. Following mRNA fragmentation, cleaved RNA fragments were reverse transcribed into first-strand cDNA by using random primers and reverse transcriptase. Then, second-strand cDNA was synthesized. Afterwards, the overhangs resulting from fragmentation were converted to blunt-ends for proper adapter ligation. the 3' overhangs were removed by 3' to 5' exonuclease and the gaps were filled with the polymerase. In the following step, a single 'Adenine' (A) nucleotide is added to the 3' ends of the blunt ends to prevent them from ligating to each other during the adapter ligation reaction. On the other hand, 'Tymine' (T) nucleotide on the 3' end of the adapter allows ligating the adapter to the fragment by providing a complementary overhang. In this sense, chimera (concatenated template) formation was prevented. In the next step, multiple indexing adapters were ligated to the ends of the ds cDNA, preparing them for hybridization onto an Illumina flow cell. The flow cell has complementary oligonucleotides and allows hybridization for "sequencing by synthesis" process. Adapter-ligated cDNA fragments on both ends were then enriched by PCR with limited cycles (only 15 cycles) to avoid redundant amplification. Since, an optimum density of DNA is required for cluster generation and sequencing steps, amplicons were quality checked by 2100 Bioanalyzer (Agilent Technologies) and

quantified by qRT-PCR. Indexed libraries were diluted to 10 nM and pooled in the cluster template plates.

### 2.6. Cluster Generation and Transcriptome Sequencing

After the tissue and application specific libraries were constructed, they should be enriched through bridge-amplification on the flow cell so that, each amplicon would be specifically amplified. In bridge-amplification process, single strand short oligonucleotides attached to the flow cell were hybridized to adapters ligated to amplicons. Then clusters were generated through repeated denaturation and extension reactions with the use of "TruSeq Cluster Kit v5" (Illumina).

Subsequently, the flow cell was loaded into "Illumina HiSeq 4000 System". Dye-tagged deoxyribonucleotides were bound to bridge-amplified amplicons and sequencing reaction was achieved. Resulting raw image files were processed with "Illumina Base calling software V1.7" and converted to raw nucleotide sequences in ".fastq" format.

The construction of the RNA libraries, cluster generation and transcriptome sequencing were performed by Beijing Genomics Institute (BGI) (Hong Kong). Illumina Hi-seq 4000 platform was used for RNA sequencing (San Diego, CA).

### 2.7. Raw Data Processing

Raw data should be pre-processed to eliminate sequencing errors. Quality control (QC) process includes filtering steps such as;

i) removal of adaptor sequences (adaptor trimming);
ii) removal of low quality reads with phred score <20;
iii) removal of first 10 bases misread due to the library preparation with random hexamers;
iv) removal of sequences shorter than 50 bases;
v) removal of reads in which ambiguous (N) bases are more than 5%.

Quality parameters such as nucleotide base quality, sequence quality score, GC% content and length distribution were determined and visualized by fastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Additional scripts in fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) were used for further filtering steps

Adapter sequences, first 10 unreliable bases and low-quality reads were filtered by FASTQ/A Clipper, FASTQ/A Trimmer and FASTQ Quality Trimmer, respectively. Clean and high quality (HQ) sequences were obtained for *de novo* transcriptome assembly.

Command lines:

```
fastx_clipper -a adapter_sequence -l 20 -c -k -i input.fastq
-o output.fastq
fastx_trimmer -f 10 -v -i input_file.fastq -o output_file
fastq_quality_filter -v -p 80 -q 20 -i input_file.fastq -o
clean_HQ_output_file.fastq
```

### 2.8. *De novo* Transcriptome Assembly

Short read assembly program 'Trinity' was utilized to perform *de novo* assembly with clean reads that PCR duplication removed (in order to improve the efficiency), then TIGR Gene Indices clustering tools (TGICL) was used to cluster transcripts to Unigenes.

First, Trinity combines overlapping reads to form longer fragments which are coined as contigs. The sequencing reads are than mapped to these contigs. Afterwards, Trinity connects them to end up with transcripts. Large volume of reads are processed with the Trinity submodules: Inchworm, Chrysalis, and Butterfly (Grabherr et al., 2011).

The resulting sequences of Trinity are called transcripts. Then perform gene family clustering with TGICL to get final Unigenes (if more than one sample, we would execute TGICL again with each sample's Unigene to get final Unigene for downstream analyses), the Unigenes will be classified as 1) clusters, with a CL prefix followed by cluster id (clusters contain more than one Unigenes with more than 70% similarity), 2) singletons, named with Unigene as the prefix. Schematic overview of the process is shown in Figure 2.4.

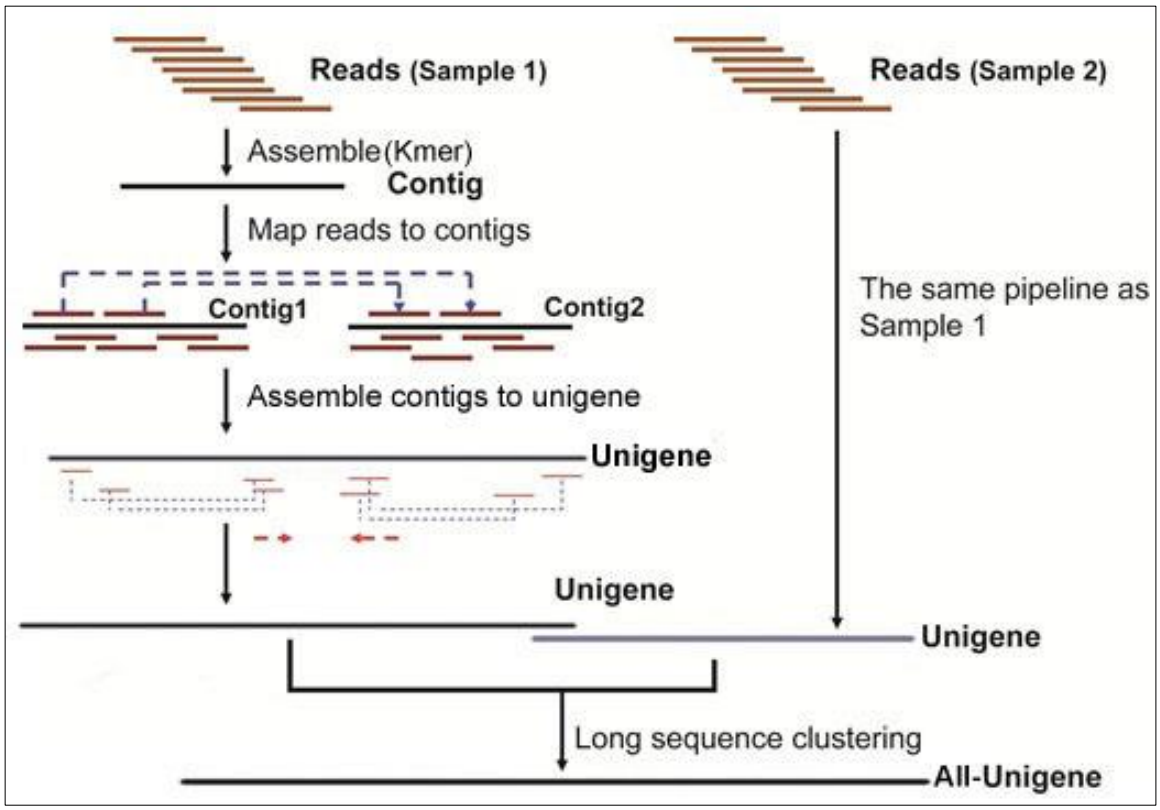



**Figure 2.5.** Schematic overview of the assembly process.

- Software information:

**Trinity:**

version: v2.0.6

parameters: --CPU 8 --min_kmer_cov 3 --min_glue 3 --min_contig_length 150 --bfly_opts '-V 5 --edge-thr=0.1 --stderr'

**Tgicl:**

version: v2.0.6

parameters: -v 25 -c 10 -l 40 -O '-repeat_stringency 0.95 -minmatch 35 -minscore 35'

## 2.9. Reference Transcriptome Dataset

After *de novo* assembly analysis, the resulting fasta file contains a large number of redundant sequences, misassembled sequences and chimeric transkripts. In this sense, it is important to create a concensus reference transcript dataset for further analysis. First, the same transcript fragments in differing lengths were removed from the dataset by CD-HIT-EST (Huang et al., 2010) and iAssembler (Zheng et al., 2011). Hence, an appropriate reference transcript library was created for annotation and gene expression analysis.

## 2.10. Transcriptome Annotation

Total number of transcripts, total number of bases, average transcript length, min and max length of the sequences, average GC% content, N25, N50, N75 values were calculated by Trinity.

### 2.10.1. Functional annotation of transcripts: GO and KEGG analysis

Gene annotation of the contigs was carried out using NCBI BLAST: Basic Local Alignment Search Tool (e-value <1e-5) in NCBI database (Altschul et al., 1990). Unigenes were blasted (BLASTx) against NCBI non-redundant protein (Nr) database, and Swiss-Prot protein database, for further annotation (e-value <1e-5). Duplicates of annotated unigenes found in both databases were removed. By using NR annotation, GO annotation based on biological process (P), molecular function (F) and cellular component (C) was performed by Blast2GO software (Conesa et al., 2005). Additionally, metabolic pathways were assigned according to Kyoto Encyclopedia of Genes and Genomes pathway database (KEGG) by BLASTx, e-value <1e-5 (Kanehisa and Goto, 2000).

- Software information:

```
Blast:
version: v2.2.23
parameters: default
website: http://blast.ncbi.nlm.nih.gov/Blast.cgi
```

**Blast2GO:**

version: v2.5.0

parameters: default

website: https://www.blast2go.com

- Database information:

**NT:**

nucleotide sequence database, with GenBank, EMBL, and DDBJ entries

Not non-redundant.

website: ftp://ftp.ncbi.nlm.nih.gov/blast/db

**NR:**

non-redundant protein squence database with entries from GenPept, Swissprot, PIR, PDF, PDB and NCBI RefSeq

website: ftp://ftp.ncbi.nlm.nih.gov/blast/db

**GO:**

Gene Ontology (GO) is used to represent biological functions of the genes at the molecular, cellular and tissue levels.

website: http://geneontology.org

**KEGG:**

KEGG (Kyoto Encyclopedia of Genes and Genomes) is composed of databases of biological pathways genomes, drugs, diseases, and chemicals. KEGG is used in bioinformatics of data analysis especially in omics studies.

website: http://www.genome.jp/kegg

**SwissProt:**

Swiss-Prot is a section under UniProtKB which is manually annotated.

It is a high quality non-redundant protein database, which combines experimental results, computed features and scientific conclusions.

website: http://ftp.ebi.ac.uk/pub/databases/swissprot

### 2.10.2. Unigene CDS prediction

With functional annotation, the segment of unigene that best mapped to functional databases was selected in a priority order of NR, SwissProt, KEGG as its CDS, and display from 5' to 3' in FASTA format. Coding regions of the unigenes that unaligned to any database mentioned above are predicted by ESTScan software with Blast-predicted CDS as model (Iseli et al., 1999).

- Software information:

**ESTScan:**

version: v3.0.2

parameters: default

website: http://sourceforge.net/projects/estscan

### 2.10.3. Detection of gene expression levels

Raw reads were matched to reference transcript dataset and the number of mapped transcripts were calculated and normalized. For this purpose; all clean reads were mapped to unigenes using ***Bowtie2*** (Langmead and Salzberg, 2012) and gene expression levels were calculated with ***RSEM*** (Li and Dewey, 2011). Then, principal component analysis (PCA) with all samples were performed by princomp, a function of R.

- Software information:

***Bowtie2*** **:**

version: v2.2.5

parameters: -q --phred64 --sensitive --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1 -I 1 -X 1000 --no-mixed --no-discordant -p 1 -k 200

website:

http://bowtie-bio.sourceforge.net/ ***Bowtie2*** /index.shtml

***RSEM*** **:**

version: v1.2.12

parameters: default

website: http://deweylab.biostat.wisc.edu/ ***RSEM***

### 2.10.4. Detection of differentially expressed genes (DEGs)

DEGs were detected with DEseq2 (Love et al., 2014) and PossionDis. DEseq2 is based on the negative binomial distribution while PossionDis is based on the poisson distribution and peformed as described at (Audic and Claverie, 1997).

- Software information:

**DEseq2:**

parameters: Fold Change >= 2.00 and *Adjusted Pvalue* <= 0.05

**PossionDis:**

parameters: Fold Change >= 2.00 and ***FDR*** <= 0.001

### 2.10.5. Gene ontology analysis of DEGs

With the GO annotation result, DEGs were classified according to official classification, and GO functional enrichment was also performed using phyper, a function of R. False discovery rate (*FDR*) was calculated for each pvalue. The terms with *FDR* smaller than 0.001 were defined as significant.

### 2.10.6. Pathway analysis of DEGs

With the KEGG annotation result, we classify DEGs according to official classification, and we also perform pathway functional enrichment using phyper, a function of R. False discovery rate (*FDR*) was calculated for each pvalue with the above formula. The terms with *FDR* smaller than 0.001 were defined as significant.

# 3. RESULTS AND DISCUSSION

## 3.1. Alkaloid Content Analysis of Capsule and Stem Upon MeJA Treatment

Alkaloid content may differ according to the tissue of interest, cultivar, growing conditions and external elicitor usage. In this sense, it is necessary to consider alkaloid content in transcriptome studies (Ziegler et al., 2006)

MeJA is an important regulator involved in diverse set of biological processes such as development and also activates plant defense mechanisms in response to wounding, pathogen attack and several abiotic stresses as well (Cheong and Choi, 2003). It has been used as an elicitor for the induction of secondary metabolite biosynthesis. (Pauwels et al., 2008; Tang et al., 2012; Mishra et al., 2013). This hormonal elicitor is used to enhance secondary metabolites in *in vitro* cultures of many medicinal plants such as *Scrophularia kakudensis* (Manivannan et al., 2016) *Hypericum perforatum* (Wang et al., 2015) and *Salvia miltiorrhiza* (Hao et al., 2015). MeJA is also used in opium poppy cell culture and seedlings to trigger alkaloid biosynthesis (Huang and Kutchan, 2000). In a more recent study, it has been clearly shown that the morphine and noscapine production have been increased in opium poppy capsule tissue upon MeJA treatment (Gurkok et al., 2015). In this research two opium poppy cultivars (morphine-rich, morphine-poor) were dissected for their metabolite and transcriptome profiles.

MeJA treated opium poppy capsule and stem tissues in both morphine-rich and morphine-poor cultivars were subjected to alkaloid content analysis by ToF-LS/MS.

Morphine level was observed to reach at its highest level in 9 hours after treatment in capsule tissues of both cultivars (Figure 3.1). So that the tissue samples of this time point is selected for transcriptome analysis. Besides, morphine content was decreased in stem tissues of both cultivars after MeJA treatment. This result was in concordance with the differentially expressed morphine related transcript profiles of the same libraries.

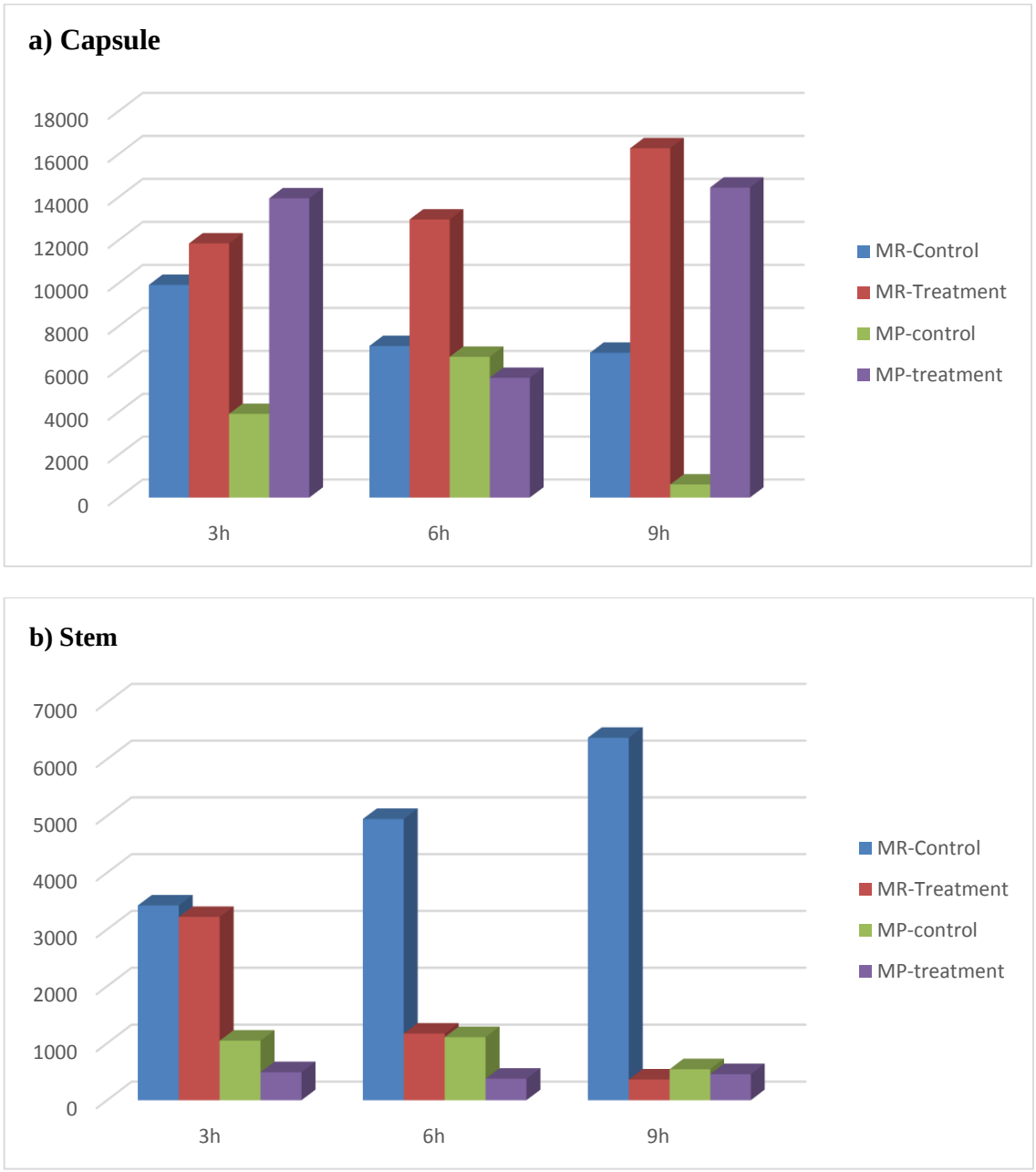



**Figure 3.1.** Morphine contents of a) capsule and b) stem in different time points.

Relative abundance values of morphine, *S*-reticuline, noscapine, codeine, thebaine and papaverine alkaloids at 9-h time point were given in Figure 3.2 and Figure 3.3.

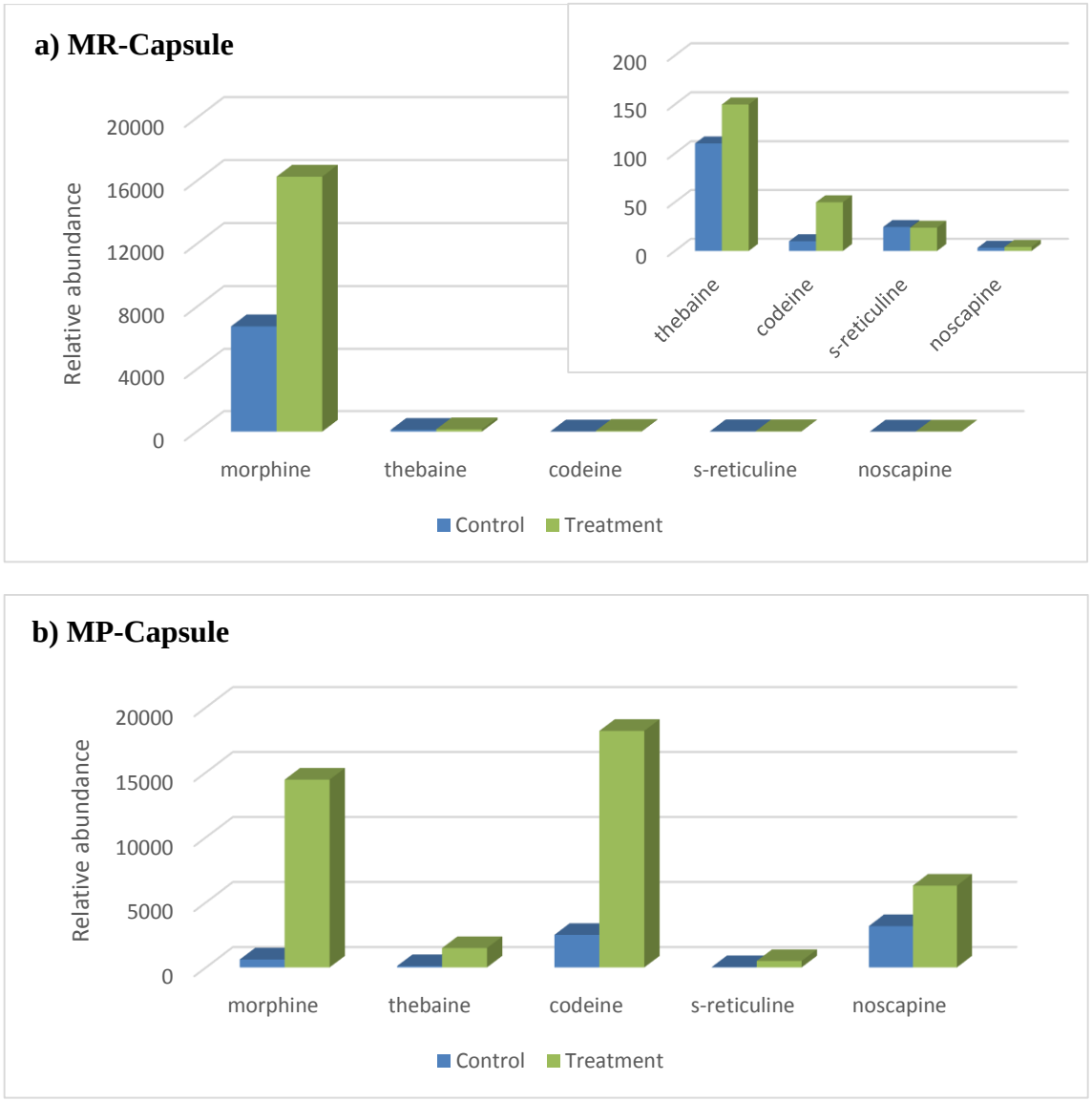



**Figure 3.2.** Comparative alkaloid contents in capsule tissues of a) MR and b) MP

*The amount of morphine dominated and the amounts of other alkaloids could not be differentiated. So that a small explanatory graph is added to the upper graph.

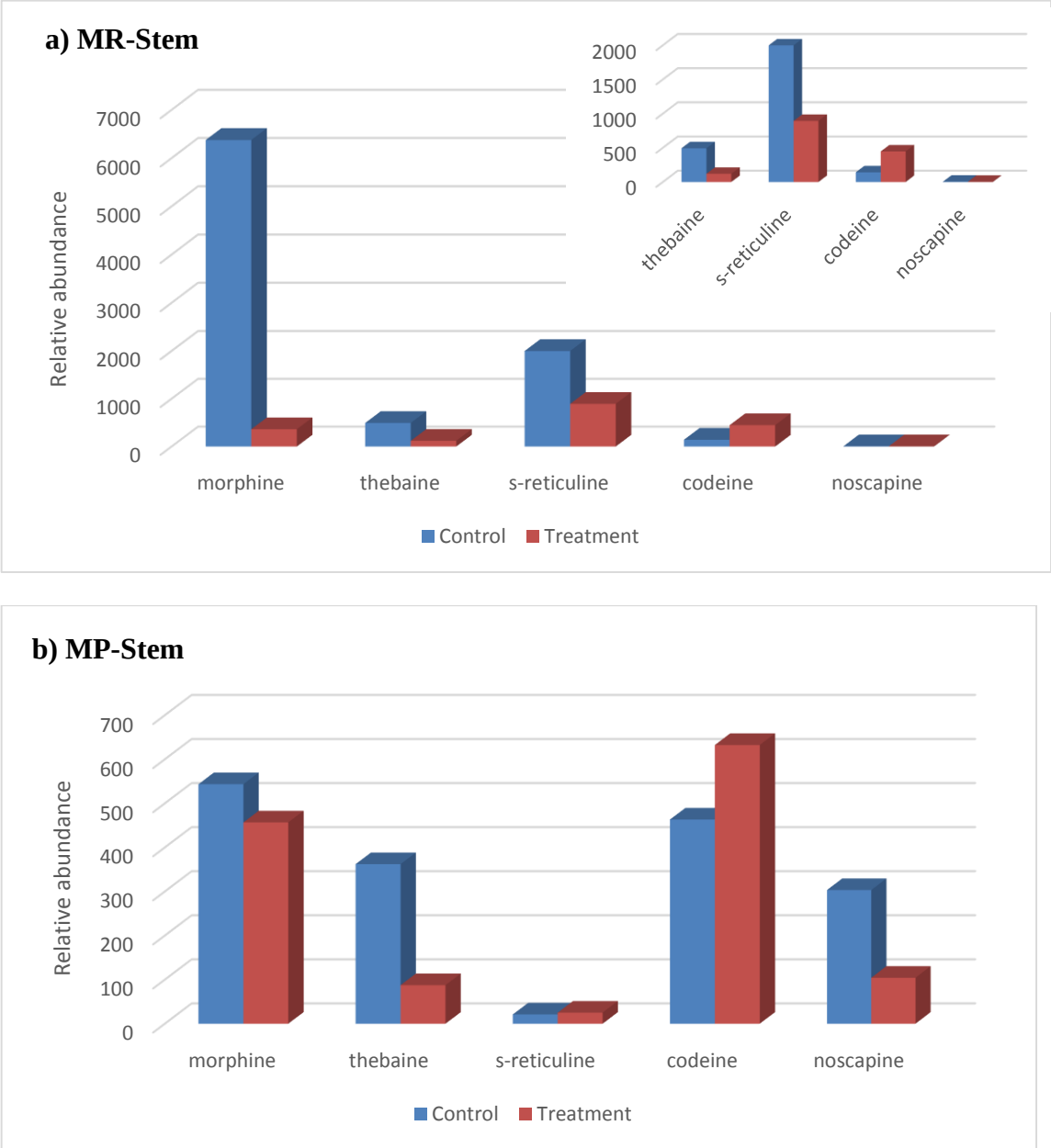



**Figure 3.3.** Comparative alkaloid contents in stem tissues of a) MR and b) MP

*The amount of morphine dominated and the amounts of other alkaloids could not be differentiated. So that a small explanatory graph is added to the upper graph.

According to the results, morphine was found to be the prominent alkaloid in control and MeJA treated capsule tissues of morphine-rich cultivar. Noscapine was the most abundant in morphine-poor control tissues, so does codeine in MeJA treated capsules. Upon elicitor application, a considerable increase has been observed for morphine, noscapine, codeine, thebaine and *S*-reticuline in capsule tissues as expected. On the

other hand, the alkaloid content has demonstrated a very different pattern in stem tissue. Except codeine and *S*-reticuline, all alkaloids were decreased after MeJA treatment as shown Figure 3.2. Papaverine could not be detected in both cultivars. Consistent with previous studies, this alkaloid is synthesized at very low levels in capsule tissue so that could not be detected during the ToF-LC/MS analysis (Pathak et al., 2013; Singh et al., 2014). In another study, no papaverine was detected in capsule tissue et al (Shukla and Singh, 2001).

### 3.2. RNA Quality Assessment and Quantitation

Transcriptome sequencing with next generation technologies is dependent on the quality and quantity of the isolated RNA. Based on the alkaloid analysis, MeJA-treated 9-hat (9 hours after treatment) and control samples were selected and RNA isolation was performed from capsule and stem tissues. RNA quantity, purity and integrity were determined by NanoDrop 2000c (Thermo Scientific, DE) and Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). RNA samples with absorption ratio of (A260/A280) 1.8-2.1 and RIN $\geq$ 6.5 were selected for Illumina transcriptome sequencing. The results are given in Table 3.1. Bioanalyzer electropherograms are depicted in Figure 3.4. All RNA samples were found to be convenient for further analysis.

**Table 3.1.** RNA concentration and quality check results

| | **Concentration (ng/µL)** | **Total Mass (µg)** | **OD 260/280** | **OD 260/230** | **RIN** |
|---|---|---|---|---|---|
| MRSC | 90 | 2.43 | 2.01 | 1.44 | 7.4 |
| MRST | 117 | 3.16 | 2.05 | 1.53 | 6.7 |
| MRCC | 175 | 4.73 | 2.08 | 1.46 | 7.5 |
| MRCT | 203 | 5.48 | 2.11 | 1.23 | 6.5 |
| MPSC | 104 | 2.81 | 2.03 | 1.39 | 8.4 |
| MPST | 113 | 3.05 | 2.03 | 0.18 | 7.4 |
| MPCC | 133 | 3.59 | 2.04 | 0.97 | 6.7 |
| MPCT | 71 | 1.99 | 1.97 | 0.66 | 7.4 |

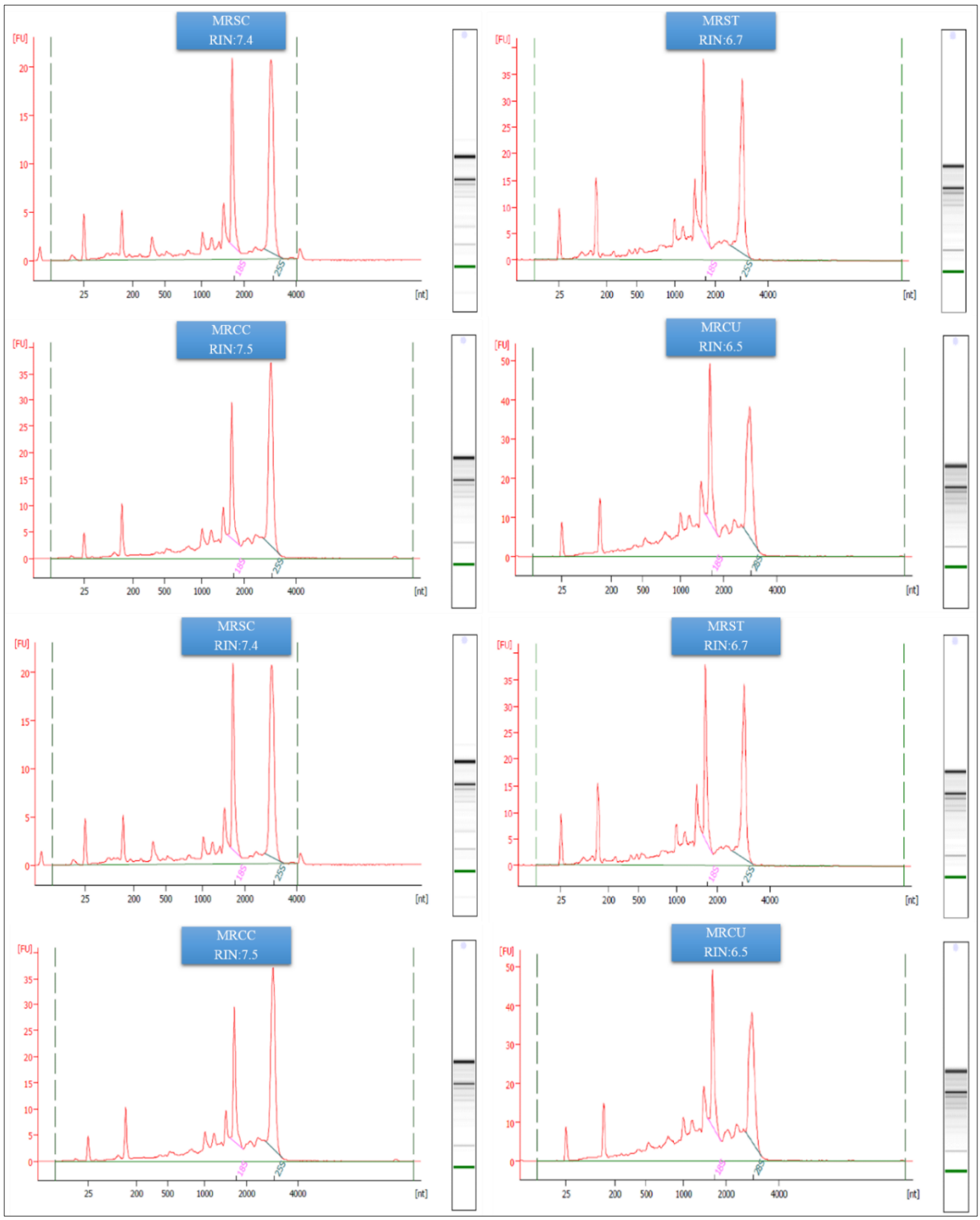



**Figure 3.4.** Bioanalyzer electropherograms for all RNA samples

## 3.3. Transcriptome Sequencing & Analysis

### 3.3.1. Illumina HiSeq RNA-Seq results

A total of 36,16 Gbp and 364.18 million raw sequencing reads from 8 distinct libraries were generated by Illumina paired-end sequencing. With a 45-fold coverage, this is the largest opium poppy transcriptomic data obtained so far. The base calling software assigns a quality score (phred score) for each nucleotide. This score is determined in 20-60 value range and shows how accurate the nucleotide is read. Q20 indicates that the nucleotide 90% accuracy, while Q30 gives 99.9% accuracy. Raw reads were subjected to quality control and the sequencing reads which contains low-quality, adaptor-polluted and high content of unknown base (N) reads were removed with FastQC before downstream analyses. After filtering, reads quality metrics are shown as Table 3.2. The distribution of base content and quality are shown as Figure 3.5 and Figure 3.6 respectively.

**Table 3.2.** Summary of sequencing after filtering

| | **Total Raw Reads(M)** | **Total Clean Reads(M)** | **Total Clean Bases(Gb)** | **Clean Reads Q20(%)** | **Clean Reads Q30(%)** | **Clean Reads Ratio(%)** |
|---|---|---|---|---|---|---|
| MRCC | 45.25 | 44.96 | 4.5 | 97 | 92.67 | 99.36 |
| MRCT | 45.92 | 45.52 | 4.55 | 97.81 | 94.71 | 99.13 |
| MRSC | 45.25 | 45.05 | 4.51 | 95.31 | 88.78 | 99.56 |
| MRST | 46.09 | 45.83 | 4.58 | 97.9 | 94.8 | 99.43 |
| MPCC | 45.25 | 44.98 | 4.5 | 96.95 | 92.64 | 99.39 |
| MPCT | 45.92 | 45.42 | 4.54 | 97.94 | 94.91 | 98.93 |
| MPSC | 45.25 | 44.53 | 4.45 | 96.5 | 91.96 | 98.42 |
| MPST | 45.25 | 45 | 4.5 | 96.74 | 92.23 | 99.45 |

*M: million; Gb: Giga base pair*

Quality of the nucleotides is crucial in *de novo* transcriptome analysis, since these nucleotides are the building blocks of the transcripts build by the assembly. So that, quality control filtering parameters were kept stringent. After filtering, each tissue sample with approximately 45 Mb high quality clean reads was ready for transcriptome assembly.

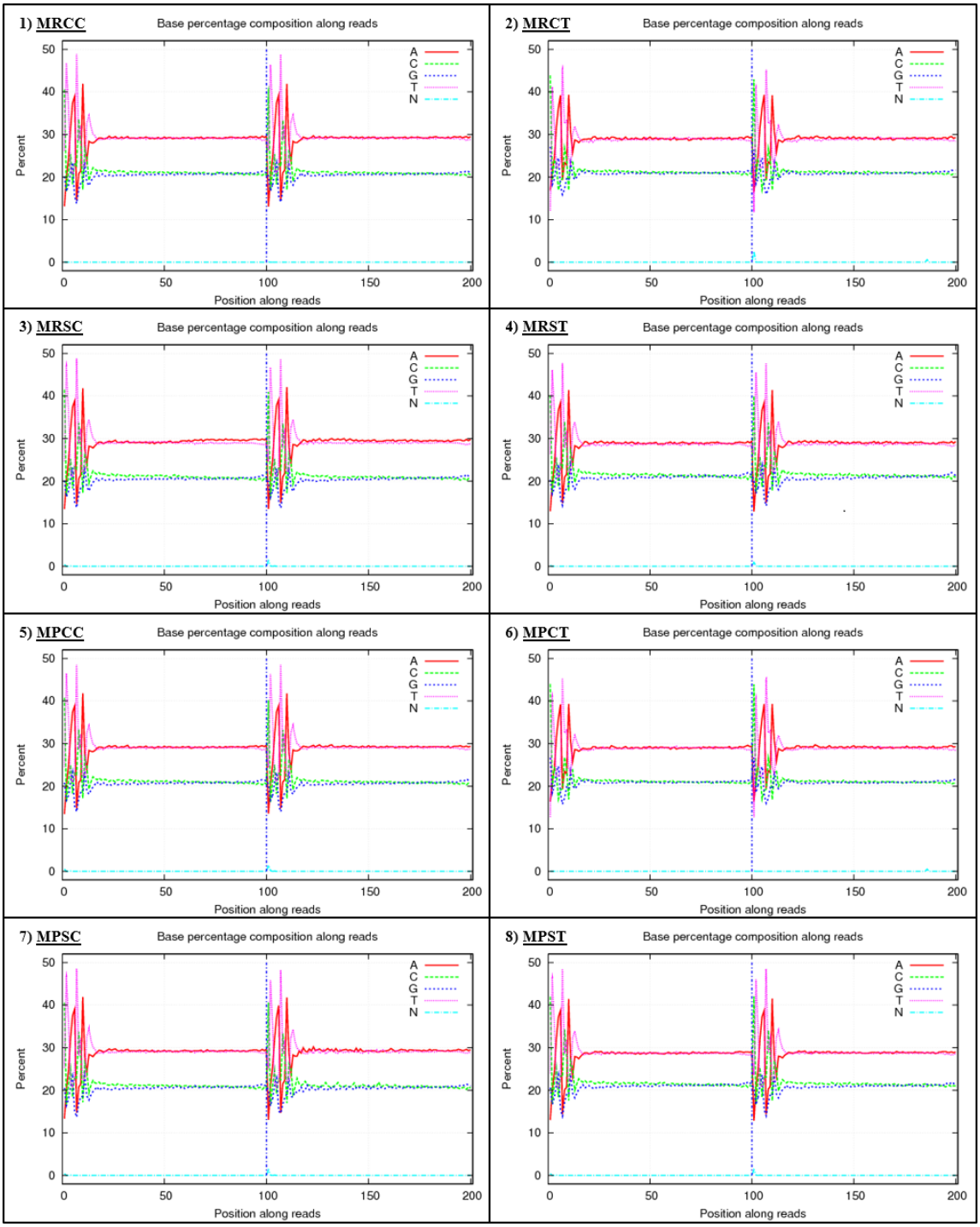



**Figure 3.5.** Distribution of base composition on clean reads.

X-axis: base position on reads, Y-axis: base content percentage. In high quality reads, A and T curves should be strictly overlapped as well as G and C curves. Fluctuations in the first six bases is caused by random-hexamer primers in PCR.

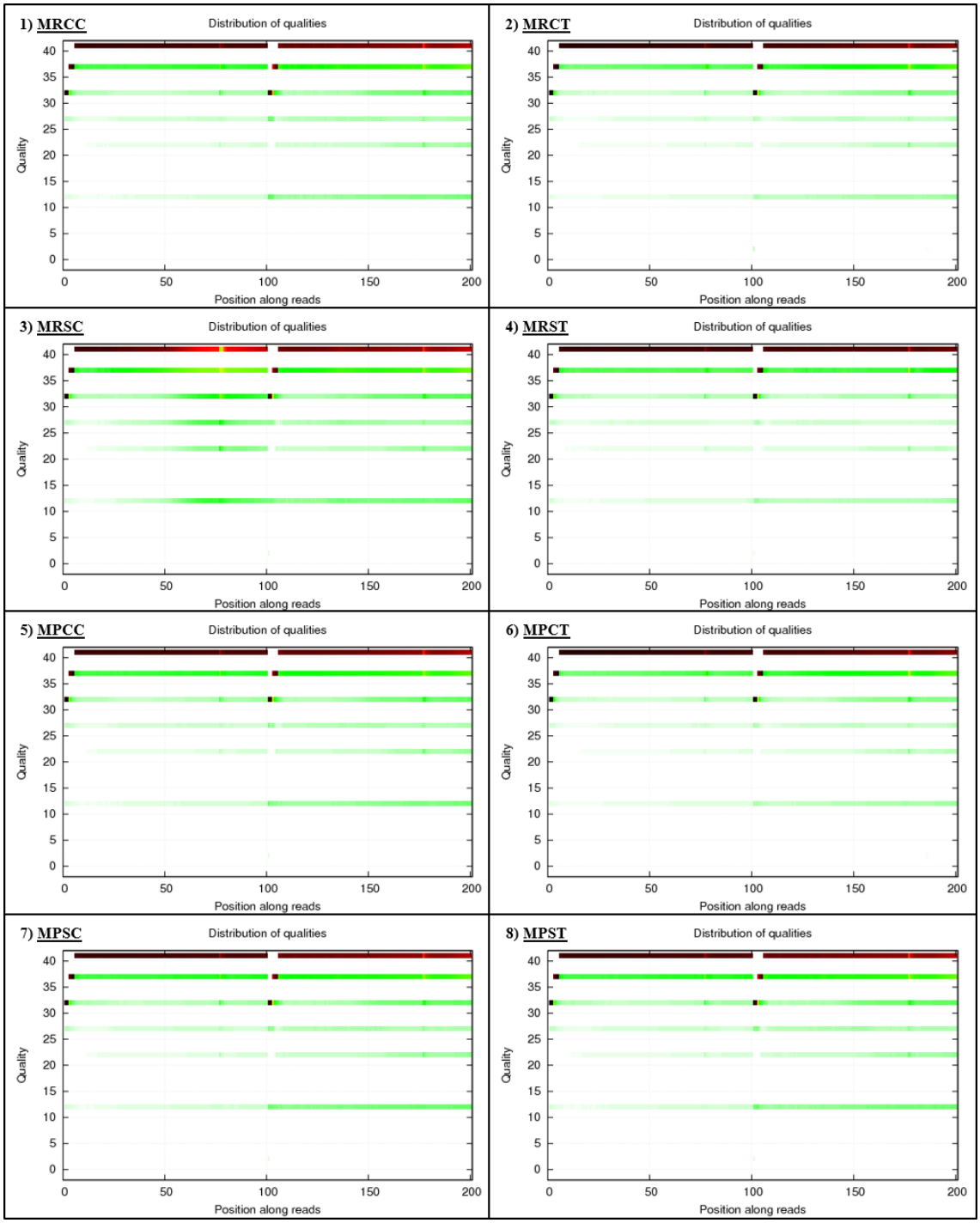



**Figure 3.6.** Distribution of base quality on clean reads.

X-axis: base positions on reads, Y-axis: base quality value. Each dot in the image is a representative of the number of total bases with certain quality value of the corresponding base along reads. Darker dot color means greater bases number. If the percentage of the bases with low quality (< 20) is very high, then the sequencing quality of this lane is bad.

### 3.3.2. *De novo* transcriptome assembly

*De novo* transcriptome assembly is a strategy which is used to generate structural and functional transcripts from Illumina-derived millions of short nucleotide sequences. Trinity was used to perform the transcriptome assembly with filtered 'clean' reads. Three independent software modules are combined in Trinity to divide the reads into *de Brujin* graphs which represent gene/locus variations which will be processed in parallel to report full-length transcripts for alternatively spliced isoforms, and separate transcripts of different genes.

The longer sequences produced by Trinity are defined as 'transcripts'. The assembly quality metrics are shown in Table 3.3. In the next step, TGICL is used to cluster transcripts into 'Unigenes' (gene family clustering). The clustering quality metrics are shown in Table 3.4, and the Unigene length distribution is shown in Figure 3.7. A total of 915,917 transcripts from all samples were produced by Trinity and combined into 570,498 Unigenes. Morphine rich capsule tissue (control) has the highest number of transcripts/contigs while stem tissue (treatment) of the same cultivar has the lowest number for both counts.

Along with the numbers of the transcripts and unigenes, their lengths are also important. N50 is a statistics used in bioinformatics as a measure of the success of genome or transcriptome assembly. For a given set of contigs with different lengths, all contigs are first sorted from longest to shortest. Then the lengths are summed starting from the longest contig. N50 is the length represents contigs of that length or longer contains at least half of the sum of the lengths of all contigs in the assembly. In short, "N50 value is the size of the smallest contig that 50% of the genome is contained in contigs of size N50 or larger" (Salzberg et al., 2012).

In the primary transcriptome assembly, different forms of fragments in different lengths that correspond to the same transcript take place. This redundancy should be removed before annotation. So that, transcripts with high similarity were clustered together and filtered. Only one transcript fragment was left to represent that transcript. Hence, a consensus transcriptome set has been created.

**Table 3.3.** Quality metrics of transcripts (N50: a weighted median statistic that 50% of the Total Length is contained in transcripts great than or equal to this value)

| Sample | Total Number | Total Length | Mean Length | N50 | N70 | N90 | GC (%) |
|---|---|---|---|---|---|---|---|
| MRCC | 129,077 | 90,004,798 | 697 | 1,168 | 648 | 272 | 39.91 |
| MRCT | 116,422 | 74,747,772 | 642 | 1,038 | 577 | 259 | 40.01 |
| MRSC | 116,431 | 80,686,237 | 692 | 1,152 | 651 | 272 | 39.94 |
| MRST | 89,612 | 57,138,514 | 637 | 999 | 594 | 263 | 40.27 |
| MPCC | 114,999 | 75,437,020 | 655 | 1,075 | 600 | 263 | 40.06 |
| MPCT | 120,592 | 81,165,929 | 673 | 1,126 | 612 | 263 | 40.02 |
| MPSC | 113,154 | 78,924,599 | 697 | 1,160 | 660 | 274 | 39.91 |
| MPST | 115,630 | 80,620,571 | 697 | 1,168 | 649 | 273 | 40.04 |

**Table 3.4.** Quality metrics of unigenes (N50: a weighted median statistic that 50% of the Total Length is contained in Unigenes great than or equal to this value)

| Sample | Total Number | Total Length | Mean Length | N50 | N70 | N90 | GC (%) |
|---|---|---|---|---|---|---|---|
| MRCC | 78,023 | 66,623,732 | 853 | 1,370 | 850 | 344 | 39.82 |
| MRCT | 71,074 | 55,755,630 | 784 | 1,222 | 762 | 323 | 39.93 |
| MRSC | 72,620 | 60,343,425 | 830 | 1,314 | 819 | 340 | 39.89 |
| MRST | 59,146 | 44,000,861 | 743 | 1,113 | 713 | 318 | 40.24 |
| MPCC | 73,133 | 57,105,465 | 780 | 1,227 | 762 | 320 | 39.99 |
| MPCT | 73,460 | 60,077,220 | 817 | 1,318 | 801 | 326 | 39.93 |
| MPSC | 71,620 | 59,547,743 | 831 | 1,321 | 822 | 339 | 39.84 |
| MPST | 71,422 | 59,990,714 | 839 | 1,342 | 830 | 341 | 39.97 |
| All-Unigene | 126,747 | 142,290,032 | 1,122 | 1,680 | 1,190 | 544 | 39.63 |

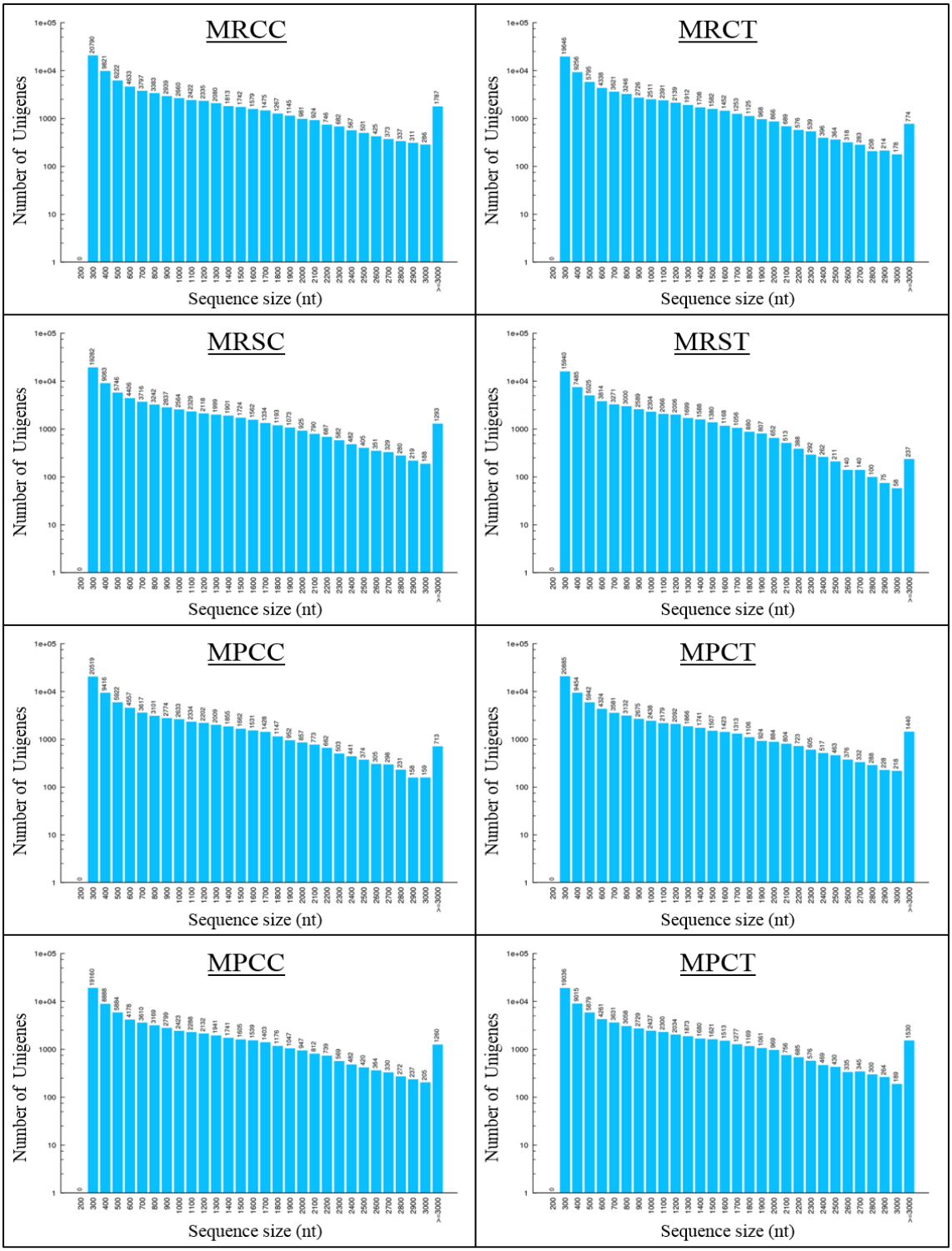
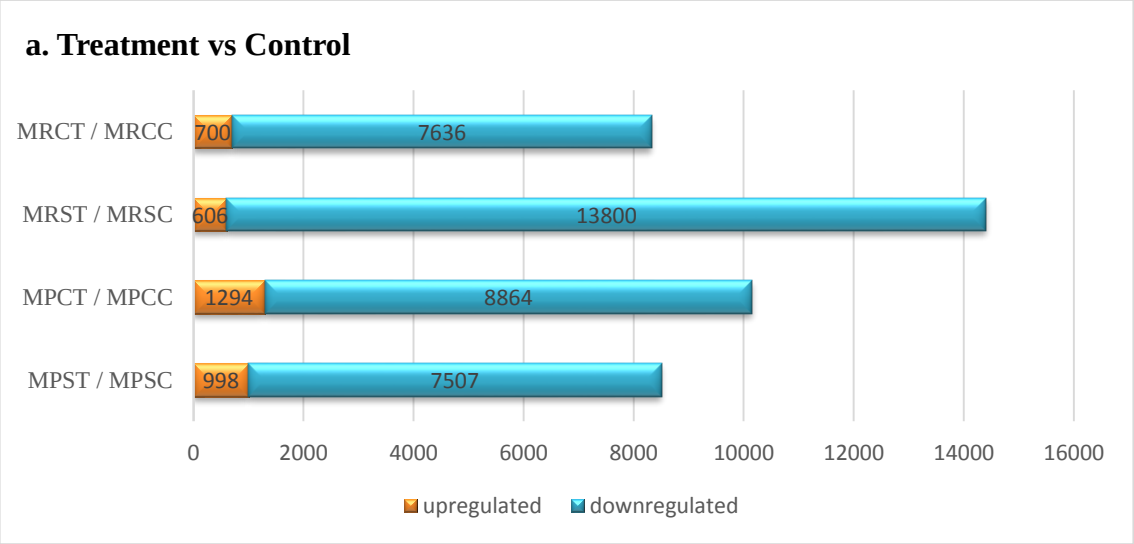


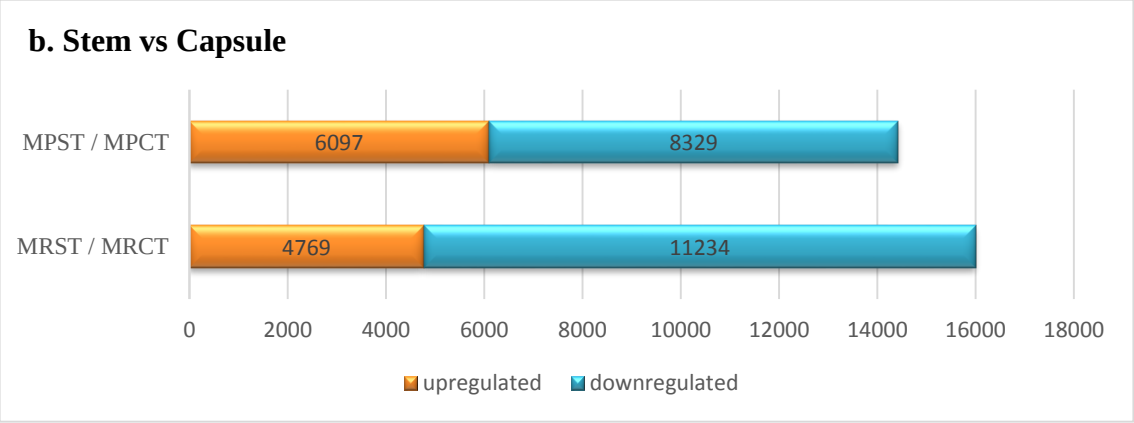

**Figure 3.7.** Unigene length distribution.

(X-axis: the length of Unigenes. Y-axis: the number of Unigenes)

### 3.4. Transcriptome Annotation

According to the annotation results, correlation of the change in the number of transcripts in relation with MeJA-treatment, tissue and cultivar was assessed. Following groups are evaluated own between in terms of differentially expressed transcript levels: i) Control vs MeJA-treated samples of each cultivar and tissue (MRCT/MRCC, MRST/MRSC, MPCT/MPCC, and MPST/MPSC), ii) comparison of different tissues in morphine rich and poor cultivars respectively (MPST/MPCT and MRST/MRCT), iii) comparison of the same tissues in different cultivars (MPCT/MRCT and MPST/MRST). In total 8 distinct transcriptome profiles have been generated. Changes in transcript numbers are given for each group in Figure 3.8.



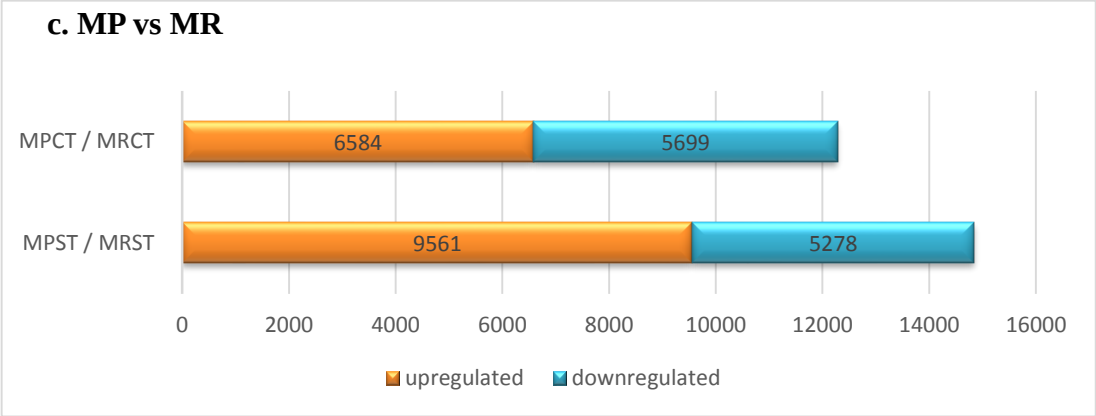



**Figure 3.8.** Changes in differentially expressed transcript numbers

Digital gene expression analysis of eight libraries has been performed. For each library the number of transcripts before and after MeJA treatment are given in Figure 3.9a. Upon MeJA treatment, the number of transcripts are decreased in all samples.

Transcript numbers in terms of capsule and stem tissue comparisons in both MR and MP cultivars are given in 3.9b, and cultivar comparisons at tissue level are given in 3.9c.

Unique transcripts were abundant in capsules of both MR and MP cultivars compared to stem tissues. Moreover, MP cultivar was rich in unique transcripts compared to MR for both capsule and stem tissues.

Venn diagrams representing these groups were plotted using Venny 2.1 tool (Oliveros, 2014).

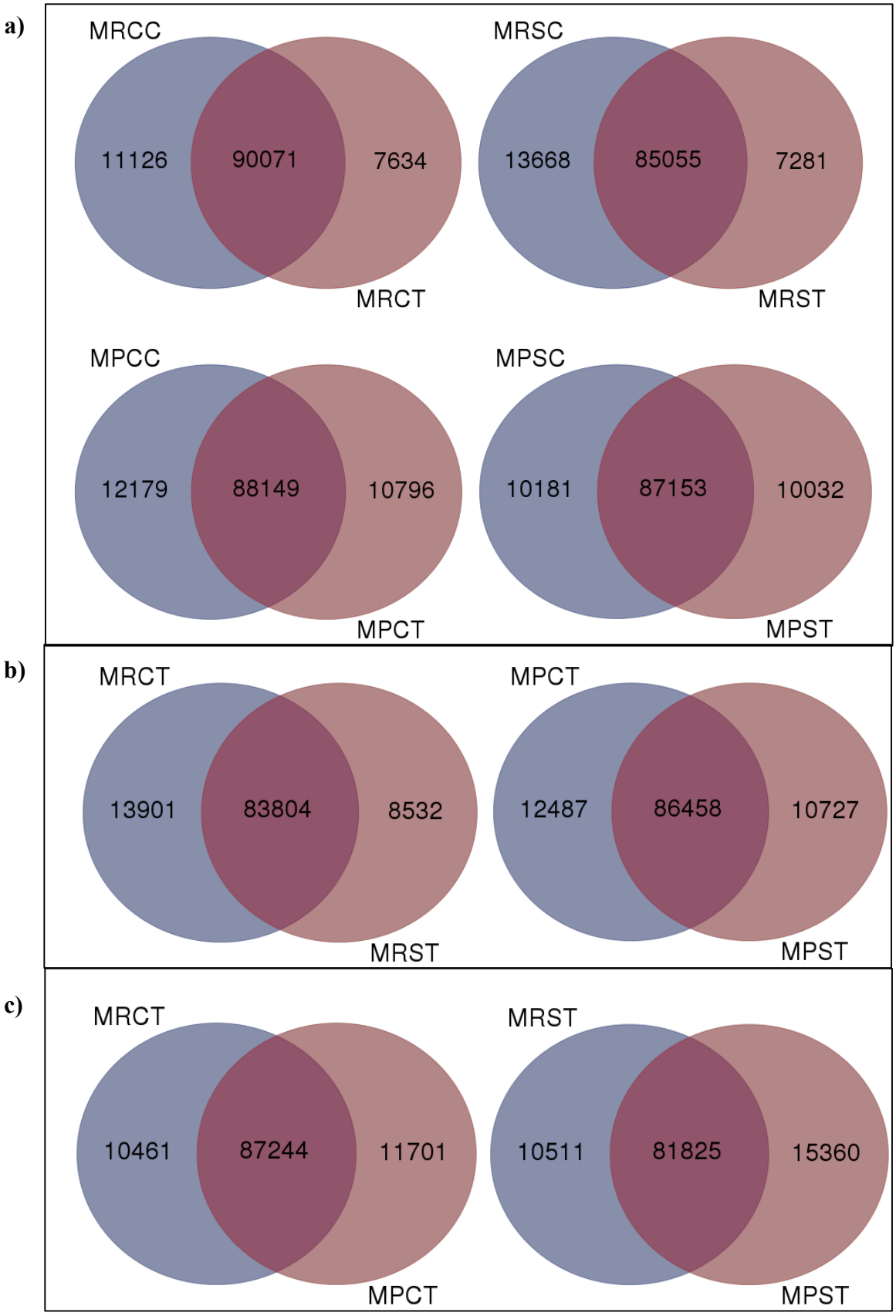



**Figure 3.9.** Number of the transcripts expressed in all control and treatment groups

### 3.4.1. Comparative GO analysis of transcriptional profiles

GO and KEGG analysis were performed on differentially expressed genes (DEGs). GO analysis provide standard terms which represent gene product properties in terms of three aspects: **Biological Process** represent defined sets of molecular events relevant to their functions. The main difference between a biological process and a molecular function is that a process must involve have more than one distinct steps. Besides, a biological process is not equivalent to a pathway. GO does not meet the requirements for fully described pathway. **Molecular Function** describes the activities of a gene product at the molecular level. These activities can be performed by single or assembled complex gene products. Molecular function terms are mostly represented as "activity" in order not to confuse with a gene product name. **Cellular Component** represents the parts of a cell or a gene product group. Using Blast2GO statistically significant ontology terms that are differentially represented in different experimental group, tissues and cultivars were identified.

In opium poppy capsule and stem tissues, nine hours after MeJA treatment, the majority of the genes assigned to biological process were grouped as metabolic process, response to stimulus and transport. Alkaloid metabolic processes, isoquinoline biosynthesis, response to ABA, SA and JA, regulation of jasmonic acid mediated signaling pathway and defense responses were main subgroups altered by elicitor treatment. Related genes were discussed in Section 3.6.1 and detailed GO term lists are given in Supplementary Table 1-8.

In the following subsections a summary of the corresponding GO results is described for each comparison group. Differential expression in the first 4 groups are based on MeJA treatment (T/C), while the following 2 are tissue (S/C) and the last 2 are cultivar-dependent (MP/MR).

#### 3.4.1.1. MRCT/MRCC comparison

Gene expression levels in control and MeJA-treated capsule tissues of morphine-rich cultivar were compared. The $\log_2$ expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

Upon MeJA treatment, a total of 8336 transcripts were found to be differentially expressed in two tissues, of which 700 and 7636 transcripts were upregulated and downregulated in treatment tissue, respectively (Figure 3.8a). Using blastn homology search against NCBI gene database (e-value <1e-5), 4029 transcripts (48.3%) were annotated to a gene.

According to GO results, in total, 10428 gene ontology terms were assigned to differentially expressed transcripts, of which 4890 were determined to be belonged to biological processes, 3314 to molecular functions and 2224 to cellular components.

In terms of biological processes, the amount of transcripts that are belong to metabolic process, response to stimulus and transport are found to be higher compared to other processes (Figure 3.11). In treatment group, 490 transcripts were upregulated while 4398 were downregulated.

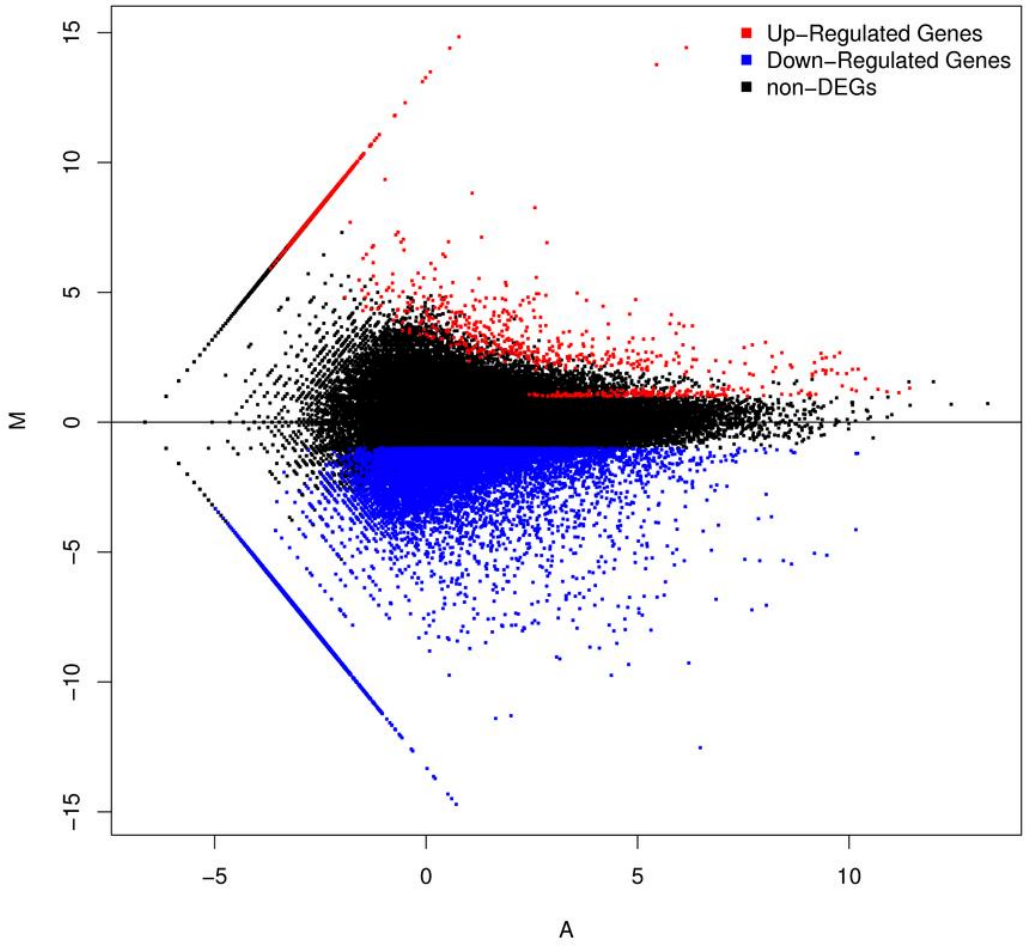


**Figure 3.10.** MA plot of MRCT/MRCC comparison library

Transcripts with functions in **metabolic process** were shown to have their 91% downregulated. Phosphorylation and methylation processes were prominent ontologies of this category. Morphine biosynthetic process (*Codeine O-demethylase, Salutaridine synthase*) and isoquinoline alkaloid biosynthetic process were also detected. 89% of the transcripts categorized in **response to stimulus** process were downregulated. Response to wounding, heat, cold, salt and biotic stresses are major processes of this category. Out of 495 transcripts involved in **transport** process, 87% was downregulated and the major group was transmembrane transport with 87 transcripts. Out of 319 transcripts having role in **development** process, 91% was downregulated. Almost 20% of these transcripts were related to pollen, flower, root development and pollen tube growth. Out of 316 transcripts in **protein modification** category, 91% were downregulated, of which 134 were belonged to protein phosphorylation, 31 to protein folding and 23 to protein ubiquitination. Another major group **oxidation-reduction** process has 252 transcripts in total and 84% were downregulated. In this group, *(S)-stylopine synthase, Salutaridine reductase, Reticuline oxidase, Salutaridine synthase* genes were detected to be relevant to alkaloid biosynthesis. Under **gene expression** process category with 90% downregulated transcripts, most of them were belonged to mRNA splicing, transcription and its regulation. A total of 220 transcripts were detected in **carbohydrate metabolism** and 85% were downregulated. Out of 205 transcripts in **hormone** process category, 188 (92%) were downregulated. Particularly, response to jasmonic acid and jasmonic acid biosynthesis genes were downregulated. Besides, the expression of transcripts involved in response to abscisic acid, cytokinin, auxin, gibberellin and ethylene were also decreased. Under the category of **signal transduction**, almost all of the transcripts (97%) were downregulated. MAPK cascade was the most abundant transcript showing differential expression. Almost half of the transcripts (42%) in **protein metabolism** category were belonged to proteolysis process. Out of 111 transcripts belonged to **lipid and fatty acid metabolism** category, 89% were downregulated.

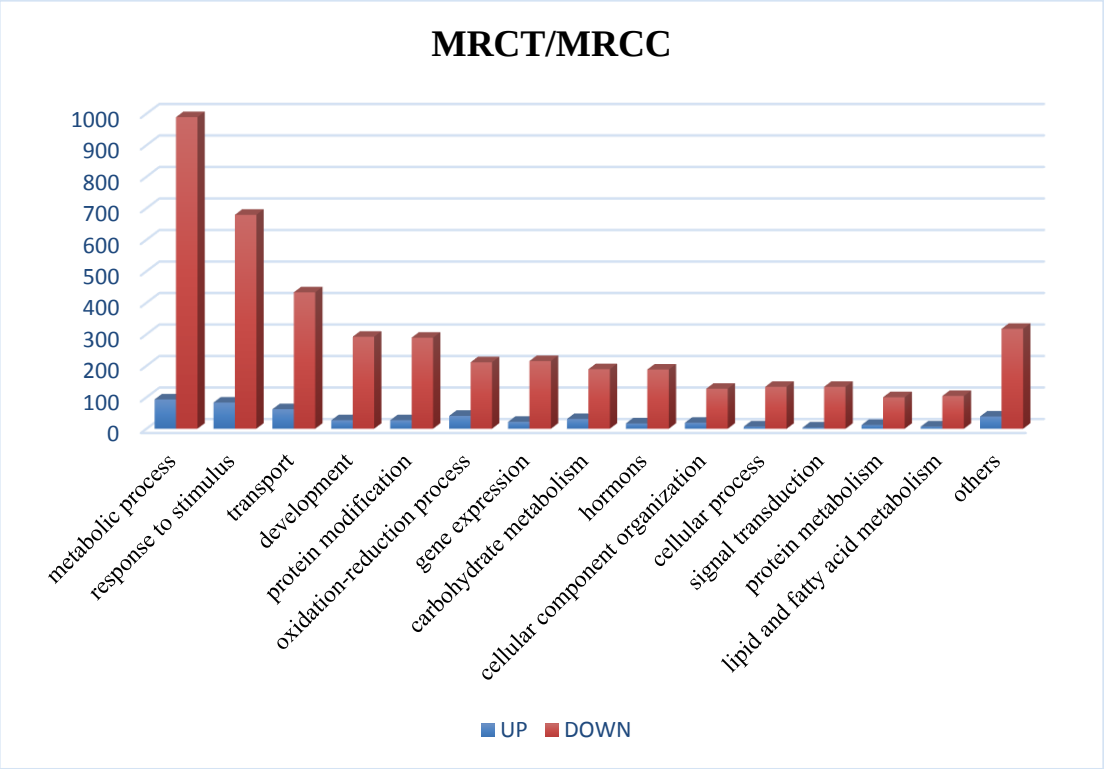



**Figure 3.11.** Most abundant biological processes in MRCT/MRCC comparison

### 3.4.1.2. MRST/MRSC comparison

Gene expression levels in control and MeJA-treated stem tissues of morphine-rich cultivar were compared. The $\log_2$ expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

After MeJA treatment, a total of 14406 transcripts were found to be differentially expressed in two tissues, of which 606 and 13800 transcripts were upregulated and downregulated, respectively (Figure 3.8a). Using blastn homology search against NCBI gene database (e-value <1e-5), 6682 transcripts (46.3%) were annotated to a gene.

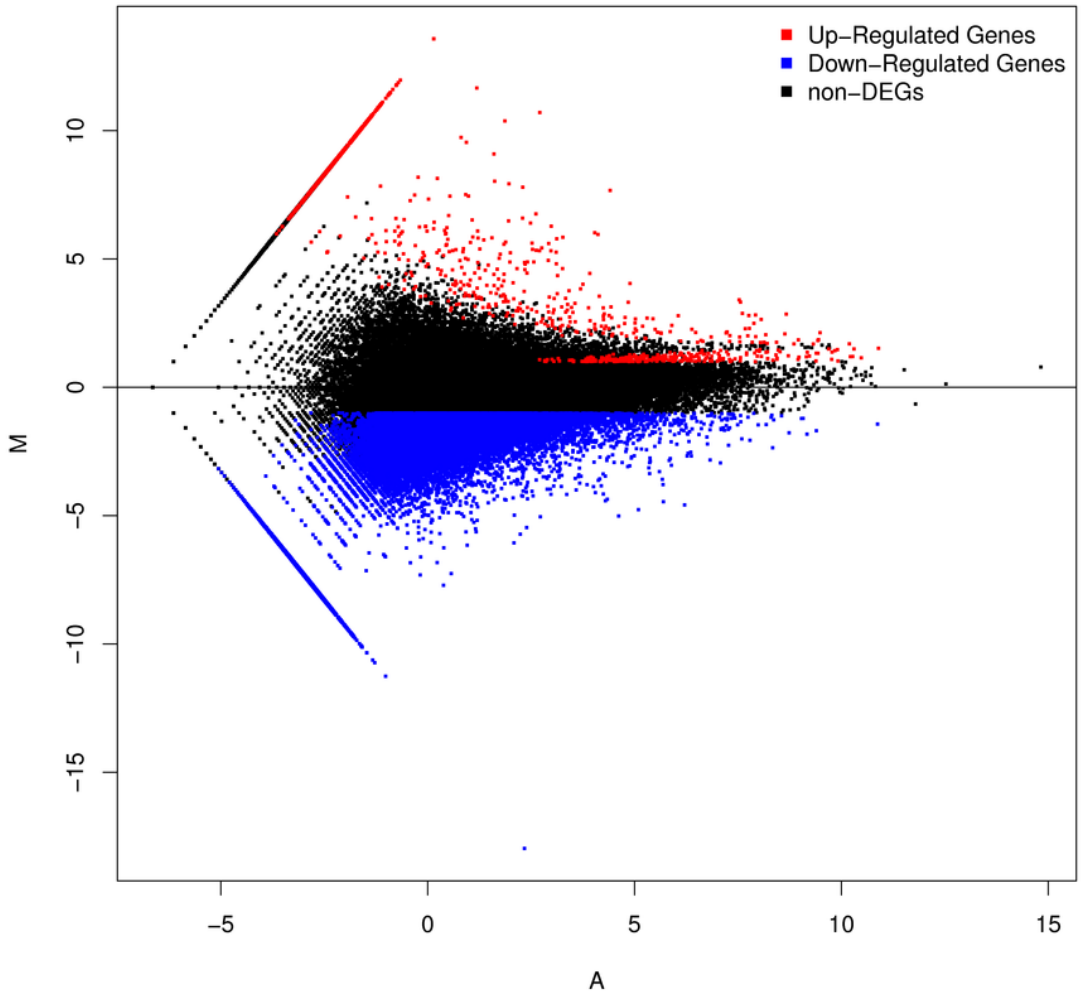



**Figure 3.12.** MA plot of MRST/MRSC comparison library

Based on GO analysis, out of 14496 transcripts with assigned gene ontology terms, it is determined that 6915 belong to biological processes, 4380 to molecular functions and 3201 to cellular components.

In terms of biological processes, the amount of transcripts that are belong to metabolic process, response to stimulus and transport are found to be higher compared to other processes (Figure 3.13). In treatment group, 477 transcripts were upregulated while 6438 were downregulated.

Out of 1442 transcripts in **metabolic process,** 1367 (95%) were downregulated. In this category, phosphorylation (114), ATP catabolic process (53) and methylation (29) are the prominent terms. Besides, differentially expressed BIA related transcripts found in this category (*tyrosine/DOPA decarboxylase, codeine O-demethylase, (RS)-*

*norcoclaurine 6-O-methyltransferase, (R,S)-reticuline 7-O-methyltransferase, (RS)-norcoclaurine 6-O-methyltransferase, (S)-stylopine synthase, codeine O-demethylase*) were all downregulated. 93% of the transcripts with functions in **response to stimulus process** were upregulated. Response to cadmium ion (71), high light intensity (38), heat (36), cold (31), salt (59), defense response to bacterium (35) and wounding (27) were prominent processes in this category. Transcripts involved in **transport process** were found to have their 91% downregulated, most of which (109) were involved in transmembrane transport. Out of 582 transcripts categorized in **development process,** 93% were downregulated. Mostly, regulation of flower development (46), xylem and phloem pattern formation (20) and pollen development (24) subcategories were recorded. A total of 502 transcripts were detected in **gene expression** category, 83% were downregulated, most of which were assigned to translation (95) and regulation of transcription (98) processes. Under **protein modification**, 97% out of 436 were found to be upregulated. Most of the transcripts in this category were involved in protein phosphorylation (141), protein folding (78), protein ubiquitination (38). Out of 278 transcripts in **hormone** process category, 252 (91%) were downregulated. Response to abscisic acid (53) and auxin (46) were the major groups in hormone response. **Oxidation-reduction process** has 276 transcripts, of which 93% were downregulated. BIA pathway related transcripts were also detected in this group (*salutaridine reductase, (S)-stylopine synthase, codeine O-demethylase*). Out of 270 transcripts in **carbohydrate metabolism**, 94% were upregulated. Under **lipid and fatty acid metabolism** category, 97% of 159 transcripts were downregulated. **Signal transduction** process was found to have 135 transcripts in total and almost all of them were downregulated. MAPK cascade (22) was the most frequent subgroup that the transcripts exhibited differential expression. Under the category of **protein metabolism**, 98% of the transcripts were downregulated, 61 of them were belonged to proteolysis and 35 to ubiquitin-dependent protein catabolic process.

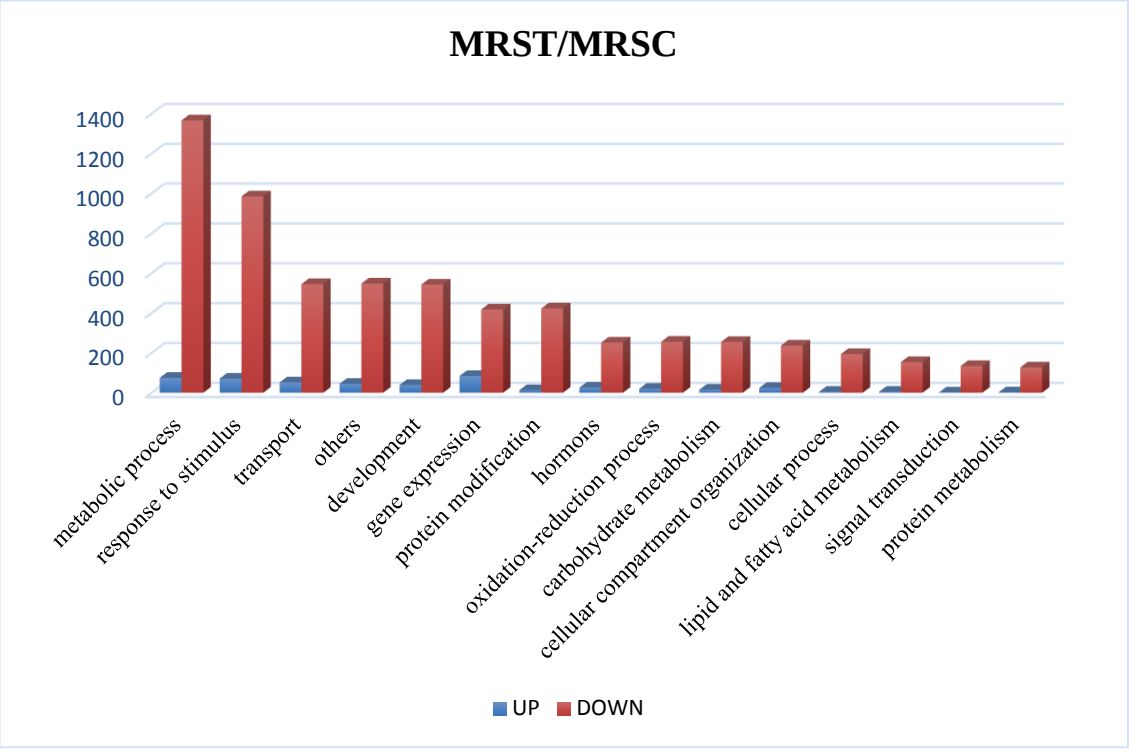



**Figure 3.13.** Most abundant biological processes in MRST/MRSC comparison

### 3.4.1.3. MPCT/MPCC comparison

Gene expression levels in control and MeJA-treated capsule tissues of morphine-poor cultivar were compared. The $\log_2$ expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

Upon MeJA treatment, a total of 10158 transcripts were found to be differentially expressed in two tissues, of which 1294 transcripts were upregulated and 8864 were downregulated in capsule treatment tissue, respectively (Figure 3.8a). By NCBI blastn search, 4657 of the transcripts (45.8%) were annotated to a gene in NCBI gene database (e-value <1e-5).

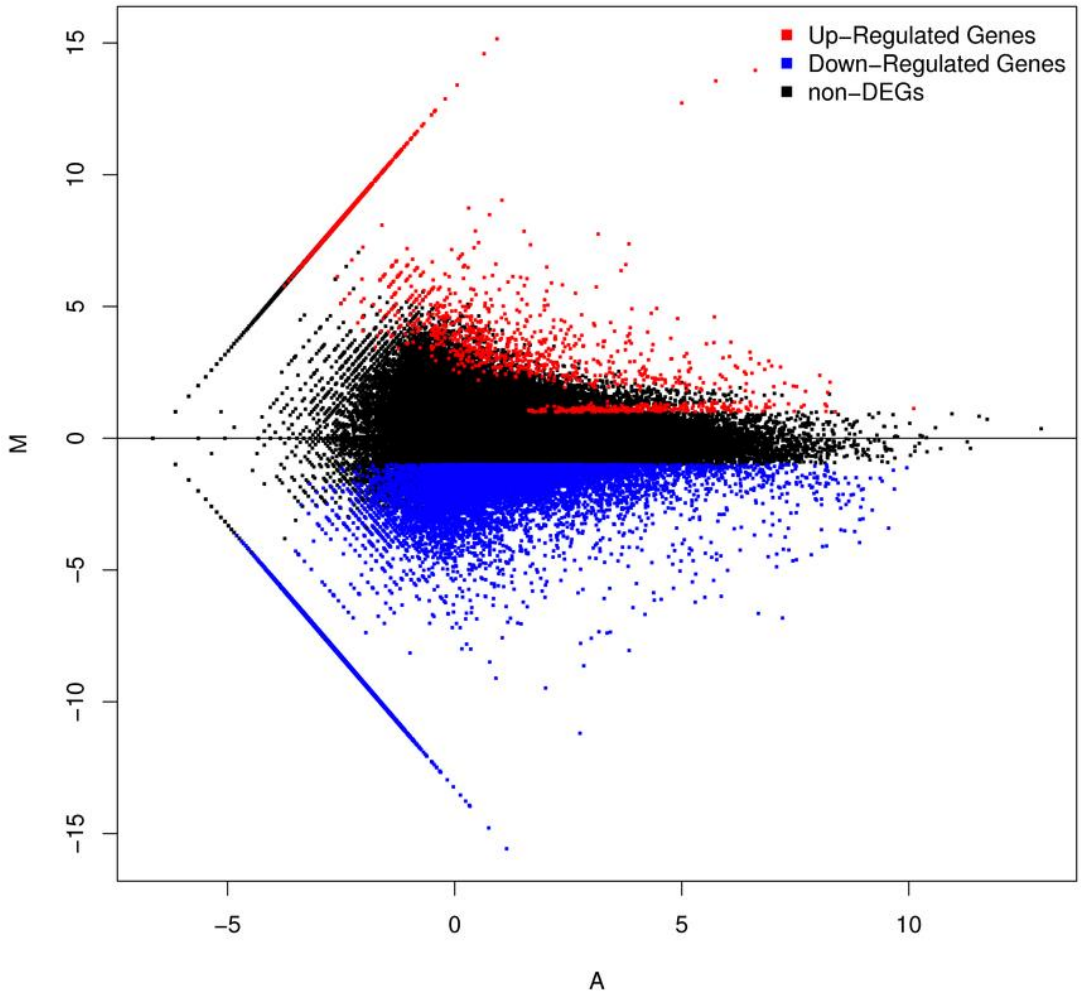



**Figure 3.14.** MA plot of MPCT/MPCC comparison

According to Blast2GO results, in total, 12271 gene ontology terms were assigned to differentially expressed transcripts, of which 5826 were determined to be belonged to biological processes, 3353 to molecular functions and 3092 to cellular components.

In the assessment of biological processes, metabolic process, response to stimulus and transport are found to be the major categories demonstrating differential expression (Figure 3.15). In treatment group, 616 transcripts were upregulated while 5210 were downregulated in the capsule tissue of morphine poor opium poppy cultivar.

Out of 1238 transcripts in **metabolic process,** 1064 (86%) were downregulated. In this category phosphorylation, methylation, DNA replication and regulation were frequent ontology groups. Besides, alkaloid metabolic process (*reticuline oxidase*), isoquinoline alkaloid biosynthetic process (*(S)-stylopine synthase*) and morphine biosynthetic

process (*cheilanthifoline synthase* and *salutaridine synthase*) transcripts were all found to be downregulated after treatment. 91% of the transcripts with functions in **response to stimulus** process were downregulated. Response to osmotic stress, heat, salt, wounding and biotic stresses were prominent processes in this category. Transcripts involved in **transport process** were found to have their 89% downregulated, most of which were transmembrane transport. Out of 472 transcripts categorized in **development process,** 91% were downregulated. Most of them were belonged to leaf morphogenesis (14), regulation of meristem growth (23), flower (25) and pollen (22) development. Under **gene expression** category, 94% out of 366 transcripts were found to be downregulated. Most of them were translation (94) and regulation of transcription processes (65). **Oxidation-reduction** process has 294 transcripts, of which 88% were downregulated. BIA pathway related transcripts were all downregulated (*salutaridine reductase, (S)-stylopine synthase, cheilanthifoline synthase, salutaridine synthase, reticuline oxidase, salutaridine reductase* and *thebaine 6-O-demethylase*). Out of 262 transcripts in **carbohydrate metabolism,** 85% were downregulated. A total of 254 transcripts were detected in **protein modification category,** 86% were downregulated most of which were assigned to protein phosphorylation (82), protein folding (38) and ubiquitination (19). Out of 218 transcripts in **hormone** process category, 204 (94%) were downregulated. Particularly, response to abscisic acid (46), cytokinin (5), auxin (26), gibberellin (4) and ethylene (9) were downregulated. Besides, the expression of transcripts in response to jasmonic acid and jasmonic acid biosynthesis genes were decreased. Out of 184 transcripts belonged to **lipid and fatty acid metabolism** category, 91% were downregulated. Under the category of **protein metabolism**, 85% of the transcripts were downregulated. 53 of them were belonged to proteolysis and 21 to ubiquitin-dependent protein catabolic process. Lastly, **signal transduction** process was found to have 103 transcripts in total and 95% with downregulation. Small GTPase mediated signal transduction (22) and activation of MAPKK activity (16) were the most frequent ontology that the transcripts exhibited differential expression.

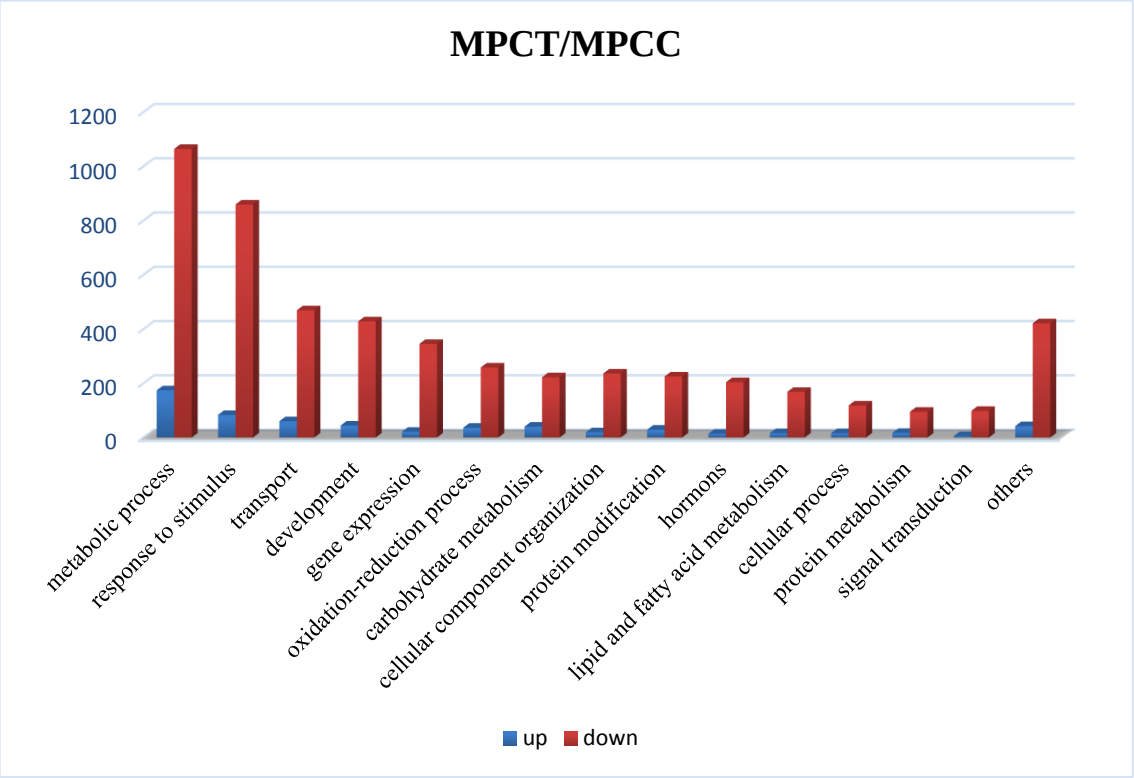



**Figure 3.15.** Most abundant biological processes in MPCT/MPCC comparison

### 3.4.1.4. MPST/MPSC comparison

Gene expression levels in control and MeJA-treated stem tissues of morphine-poor cultivar were compared. The $\log_2$ expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

After MeJA treatment, a total of 8505 transcripts were found to be differentially expressed in two tissues, of which 998 and 7507 transcripts were upregulated and downregulated in stem treatment tissue, respectively (Figure 3.8a). Using blastn homology search against NCBI gene database (e-value <1e-5), 3824 transcripts (44.9%) were annotated to a gene.

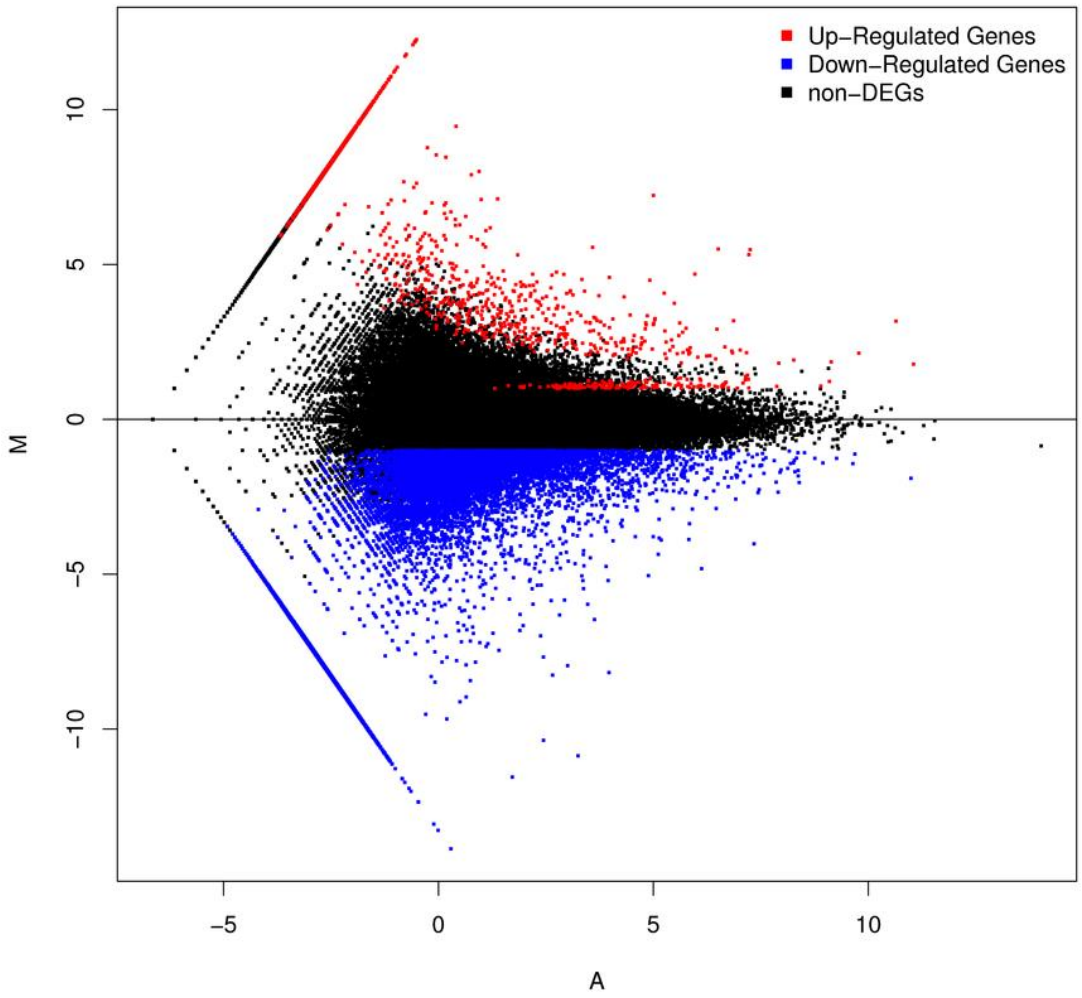



**Figure 3.16.** MA plot of MPST/MPSC comparison

Based on GO analysis, out of 9136 transcripts with assigned gene ontology terms, it is determined that 4225 belong to biological processes, 2722 to molecular functions and 2189 to cellular components.

In terms of biological processes, the amount of transcripts that are belong to metabolic process, response to stimulus and transport are found to be higher compared to other processes (Figure 3.17). In treatment group, 611 transcripts were upregulated while 3614 were downregulated.

Transcripts with functions in **metabolic process** were shown to have their 86% downregulated. Phosphorylation (71) and methylation (27) processes were prominent ontologies of this category. Morphine biosynthetic process (*Cheilanthifoline synthase*), alkaloid metabolic process (NADPH-dependent *codeinone reductase*, *salutaridine*

*reductase*) and flavonoid biosynthetic process (*Codeine O-demethylase* and *Thebaine 6-O-demethylase*) were also detected. Out of 647 transcripts categorized in **response to stimulus** process, 535 (83%) were downregulated. Response to wounding, salt and biotic stresses are major processes of this category. 89% of the transcripts involved in **transport** process was downregulated and the major group was transmembrane transport with 79 transcripts. Out of 316 transcripts having role in **development** process, 91% was downregulated. Regulation of meristem growth (21), flower development (20), pollen development (19) were the major ontologies. Out of 244 transcripts categorized in **gene expression** process, 91% were downregulated. Most of them were belonged to translation and regulation of transcription. Under **protein modification** category, 91% of the transcripts were downregulated, of which 70 were belonged to protein phosphorylation, 31 to protein folding and 20 to histone H3-K9 methylation. Another major group **oxidation-reduction** process has 219 transcripts in total and 81% of them were downregulated. In this group, downregulated *cheilanthifoline synthase, salutaridine reductase, reticuline oxidase, thebaine 6-O-demethylase,* NADPH-dependent *codeinone reductase* and *upregulated salutaridine reductase* were detected to be relevant to BIA biosynthesis. A total of 191 transcripts were detected in **carbohydrate metabolism** and 82% were downregulated. Out of 135 transcripts in **hormone** process category, 121 (90%) were downregulated. The expression of transcripts involved in response to abscisic acid (25), auxin (16), jasmonic acid (12) and ethylene (8) were also decreased.

Out of 97 transcripts belonged to **lipid and fatty acid metabolism** category, 76% were downregulated. Under the category of **protein metabolism**, 79% of the transcripts were downregulated. 49 of them were belonged to proteolysis and 17 to ubiquitin-dependent protein catabolic process. Lastly, **signal transduction** process was found to have 103 transcripts in total and 95% with downregulation. MAPK cascade (11) and small GTPase mediated signal transduction (8) were the most frequent ontology that the transcripts exhibited differential expression.

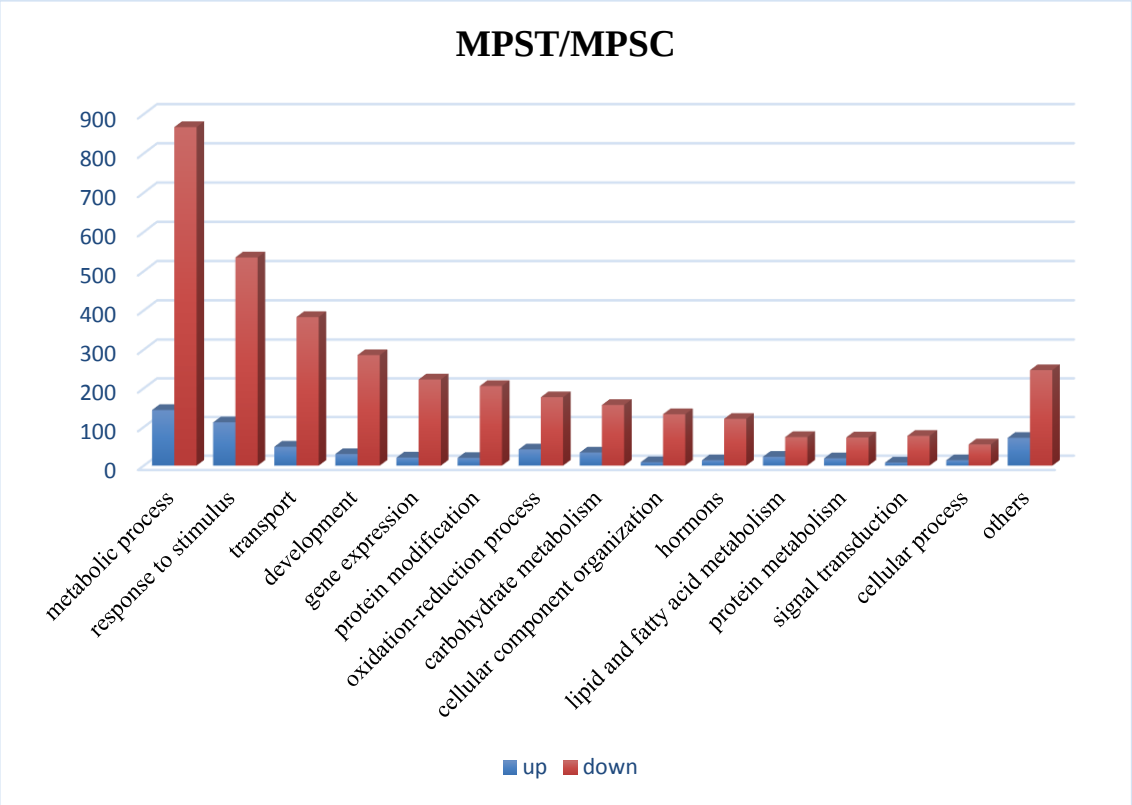



**Figure 3.17.** Most abundant biological processes in MPST/MPSC comparison

### 3.4.1.5. MRST/MRCT comparison

Gene expression levels in MeJA-treated capsule and stem tissues of morphine-rich cultivar were compared. The $\log_2$ expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

A total of 16003 transcripts were found to be differentially expressed, of which 4769 and 11234 transcripts were upregulated and downregulated respectively in the stem tissue of MP cultivar (Figure 3.8b). Using NCBI blast search, 8377 of the transcripts (52.3%) were annotated to a gene in NCBI gene database (e-value <1e-5).

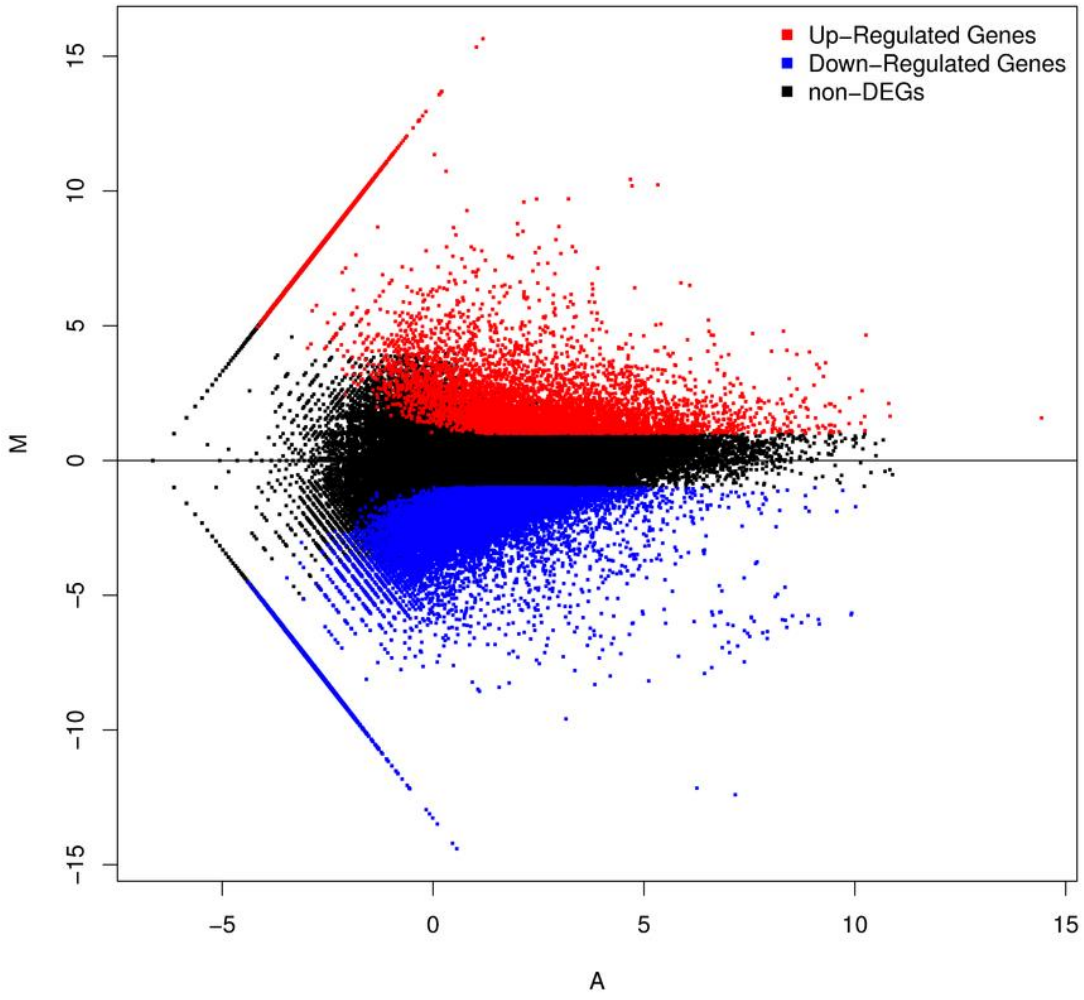



**Figure 3.18.** MA plot of MRST/MRCT comparison

With GO analysis, in total, 19328 gene ontology terms were assigned to differentially expressed transcripts, of which 9199 were determined to be belonged to biological processes, 5658 to molecular functions and 4471 to cellular components.

In terms of biological processes, the amount of transcripts that are belong to metabolic process, response to stimulus and transport are found to be higher compared to other processes (Figure 3.19). In stem tissue, 3723 transcripts were upregulated while 5476 were downregulated compared to capsule.

Transcripts with functions in **metabolic process** were shown to have their 61% downregulated. Phosphorylation and methylation processes were prominent ontologies

of this category. More than 80% of the transcripts related to morphine, isoquinoline alkaloid metabolic and biosynthetic processes found to be upregulated in stem tissue: (*salutaridine synthase, salutaridine reductase, Reticuline oxidase, (R,S)-norcoclaurine 6-O-methyltransferase, (R,S)-reticuline 7-O-methyltransferase, (S)-stylopine synthase* and *Codeine O-demethylase*). 54% of the transcripts categorized in **response to stimulus** process were downregulated. Response to response to cadmium ion (93), salt (76), cold (64), defense response to bacterium (59), water deprivation (44), wounding (44) and heat (41) are major processes detected in this category. Out of 907 transcripts involved in **transport** process, 51% was downregulated and the major group was transmembrane transport with 165 transcripts. Out of 768 transcripts having role in **development** process, 71% was downregulated. Regulation of flower development (44), root hair elongation (35) and pollen development (21) were the prominent processes. G**ene expression** category with 57% downregulated transcripts, most of them were belonged to translation, transcription and its regulation. A total of 454 transcripts were detected in **carbohydrate metabolism** and 64% were downregulated. Out of 451 transcripts in **protein modification** category, 66% were downregulated, of which 134 were belonged to protein phosphorylation (139), protein folding (71) and protein ubiquitination (40). Out of 431 transcripts in **hormone** process category, 241 (56%) were downregulated. Particularly, response to jasmonic acid and jasmonic acid biosynthesis genes were mostly upregulated. Besides, the expression of transcripts involved in response to abscisic acid, cytokinin, auxin, gibberellin, ethylene and karrikin were also decreased.

Another major group **oxidation-reduction process** has 426 transcripts in total and 51% were downregulated. In this group, transcripts found to be related to alkaloid biosynthesis were all upregulated: *thebaine 6-O-demethylase,* NADPH-dependent *codeinone reductase, (S)-stylopine synthase, salutaridine synthase, salutaridine reductase, reticuline oxidase, salutaridine reductase,* NADPH-dependent *codeinone reductase*. Out of 258 transcripts belonged to **lipid and fatty acid metabolism** category, 66% were downregulated. Out of 182 transcripts in **protein metabolism** category, 67% were downregulated. Lastly, **signal transduction process**, (59%) of the transcripts were downregulated. Activation of MAPKK activity was the most abundant transcript showing differential expression.

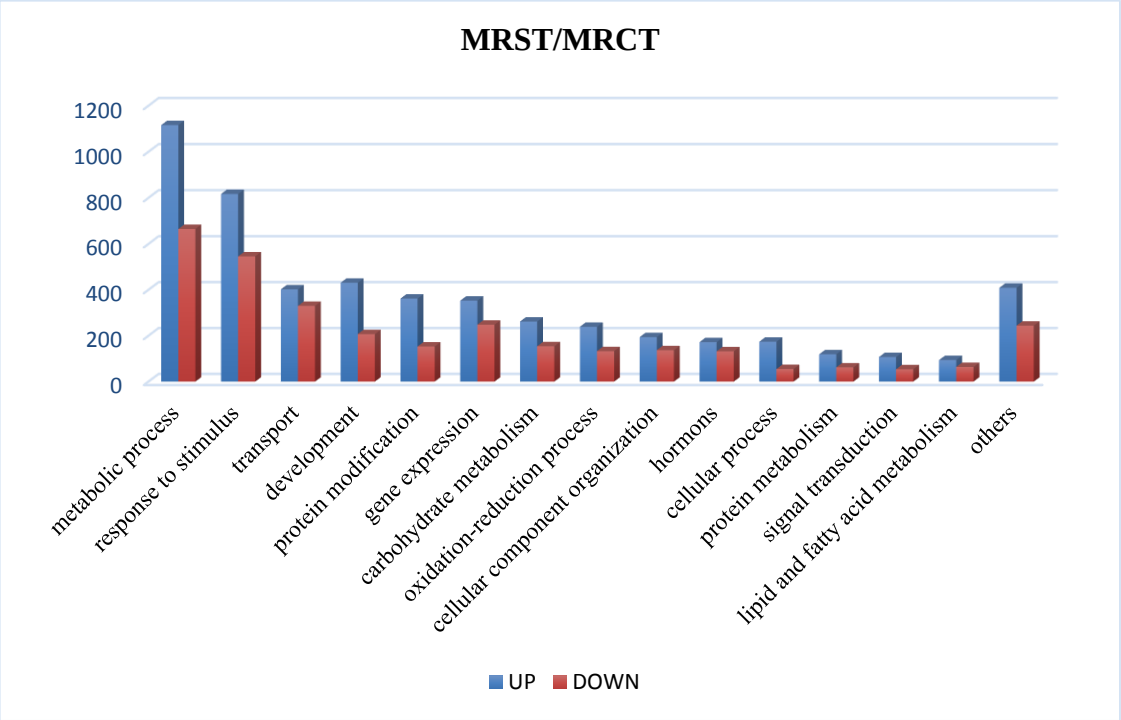



**Figure 3.19.** Most abundant biological processes in MRST/MRCT comparison

### 3.4.1.6. MPST/MPCT comparison

Gene expression levels in MeJA-treated capsule and stem tissues of morphine-poor cultivar were compared. The $\log_2$ expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

A total of 14426 transcripts were found to be differentially expressed, of which 6097 and 8329 transcripts were upregulated and downregulated respectively in the stem tissue of MP cultivar (Figure 3.8b). By NCBI blastn search, 8186 of the transcripts (56.7%) were annotated to a gene in NCBI gene database (e-value <1e-5).

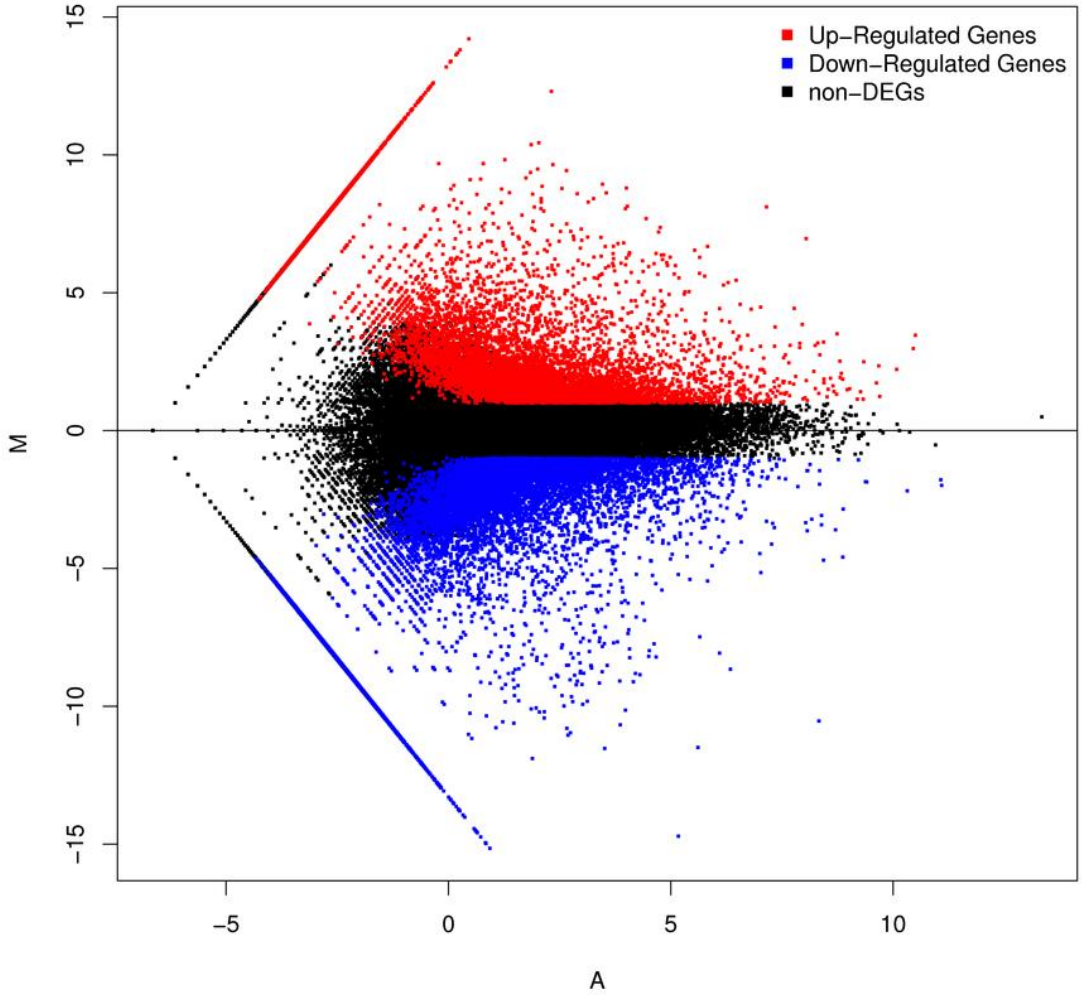



**Figure 3.20.** MA plot of MPST/MPCT comparison

Based on GO analysis, out of 19729 transcripts with assigned gene ontology terms, it is determined that 9383 belong to biological processes, 5780 to molecular functions and 4566 to cellular components.

In the assessment of biological processes, metabolic process, response to stimulus and transport are found to be the major categories demonstrating differential expression (Figure 3.21). In total, 4621 transcripts were upregulated while 4762 were downregulated in the morphine-poor stem tissue compared to capsule tissue.

Out of 1849 transcripts in **metabolic process,** 995 (54%) were downregulated. In this category phosphorylation (143), DNA replication and regulation (66), and methylation (55) were frequent ontology groups. Besides, *tyrosine/DOPA decarboxylase*,

*salutaridine reductase, codeine O-demethylase, (R,S)-reticuline 7-O-methyltransferase* were found to be downregulated while *(R,S)-norcoclaurine 6-O-methyltransferase, (S)-stylopine synthase* were found to be upregulated in stem tissue. Half of the transcripts assigned to **response to stimulus** process were downregulated. Response to cadmium ion (102), salt (97), cold (65), water deprivation (52), wounding (51) and bacterium (60) were prominent processes in this category. Transcripts involved in **transport process** were found to have their 51% upregulated, most of which were belonged to transmembrane transport process. Out of 719 transcripts categorized in **development process,** 52% were downregulated. Most of them were belonged to flower development and regulation (65), pollen development (35), and pollen tube growth (22), regulation of meristem growth (21). Under **gene expression** category, 53% out of 515 transcripts were found to be upregulated. Most of them were assigned to transcription (57) and regulation of transcription processes (117), translation (70) processes. **Oxidation-reduction** process has 509 transcripts, of which 53% were upregulated. BIA pathway related transcripts were detected in this category (*salutaridine reductase, (S)-stylopine synthase,* NADPH-dependent *codeinone reductase, Codeine O-demethylase*). Out of 436 transcripts in **carbohydrate metabolism,** 52% were downregulated. A total of 500 transcripts were detected in **protein modification category,** half of them were upregulated most of which were assigned to protein phosphorylation (187), protein folding (65) and ubiquitination (39). Out of 408 transcripts in **hormone** process category, 226 (55%) were downregulated. Particularly, transcripts belonged to response to abscisic acid (88), auxin (44), response to karrikin (38), ethylene (22), jasmonic acid (22) were found to be differentially expressed.

Out of 259 transcripts belonged to **lipid and fatty acid metabolism** category, 62% were downregulated. **Signal transduction** process was found to have 238 transcripts in total, of which 55% was upregulated. Most of them were assigned to following processes; MAPK cascade (33), small GTPase mediated signal transduction (27) and activation of MAPKK activity (26). Lastly, under the category of **protein metabolism**, 54% of the transcripts were downregulated; 93 of them were belonged to proteolysis and 28 to ubiquitin-dependent protein catabolic process.

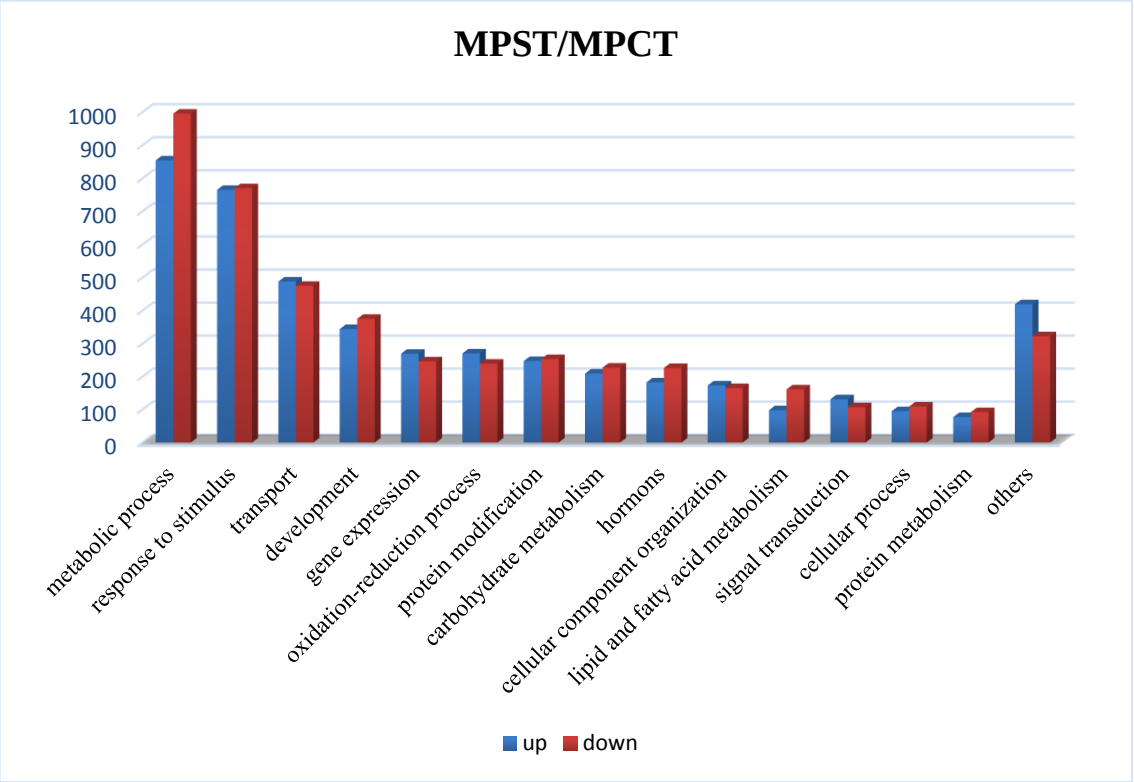



**Figure 3.21.** Most abundant biological processes in MPST/MPCT comparison

### 3.4.1.7. MPCT/MRCT comparison

Gene expression levels in MeJA-treated capsule tissues of morphine-rich and morphine-poor cultivar were compared. The log2 expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

A total of 12283 transcripts were found to be differentially expressed, of which 6584 and 5699 transcripts were upregulated and downregulated respectively in the capsule tissue of morphine-poor cultivar (Figure 3.8c). Using blastn homology search against NCBI gene database (e-value <1e-5), 6563 transcripts (53.4%) were annotated to a gene.

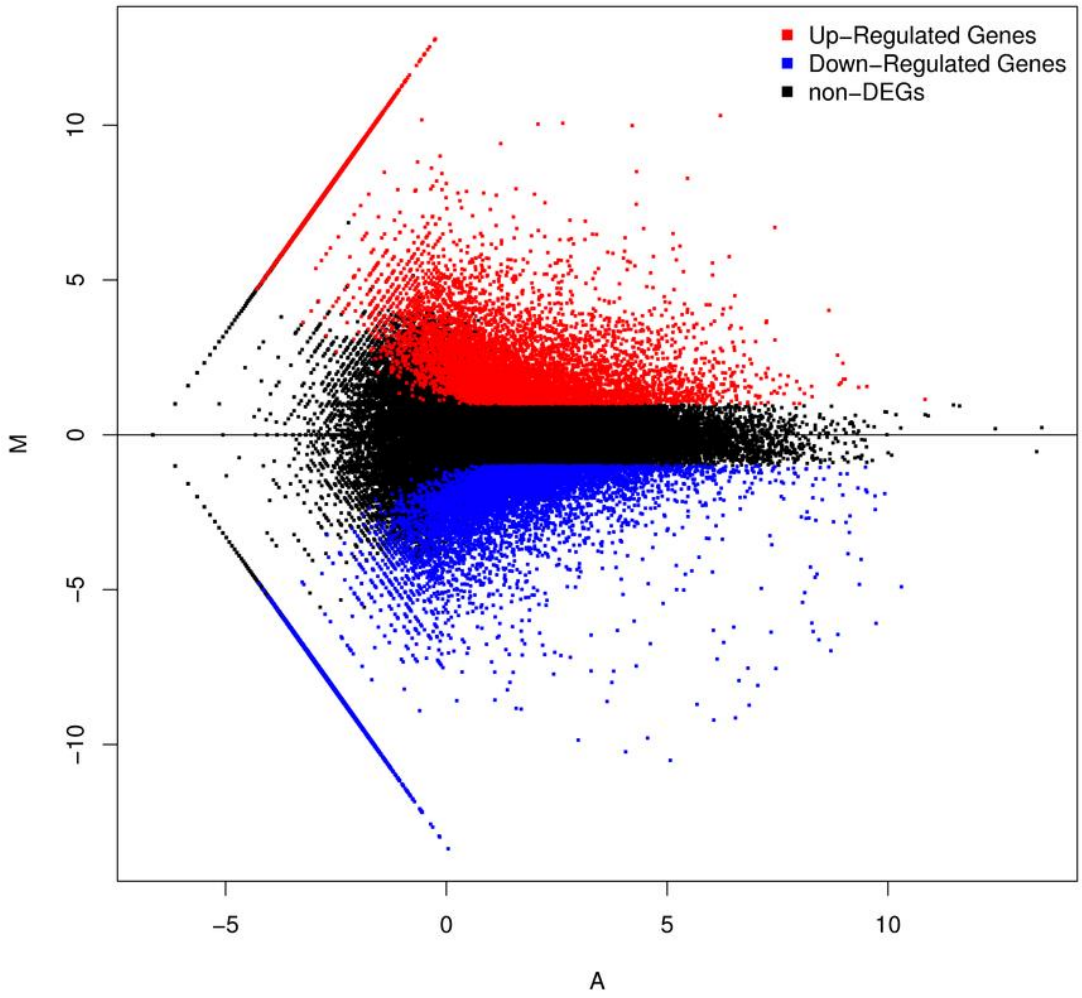



**Figure 3.22.** MA plot of MPCT/MRCT comparison

With GO analysis, out of 16177 transcripts with assigned gene ontology terms, it is determined that 7775 belong to biological processes, 4697 to molecular functions and 3705 to cellular components.

In the assessment of biological processes, metabolic process, response to stimulus and transport are found to be the major categories demonstrating differential expression (Figure 3.23). In total, 4271 transcripts were upregulated while 3504 were downregulated in the morphine-poor capsule tissue compared to morphine-rich capsule tissue.

Out of 1604 transcripts in **metabolic process,** 856 (54%) were upregulated. In this category phosphorylation (120), ATP catabolic process (41) and methylation (36) were frequent ontology groups. Besides, *tyrosine/DOPA decarboxylase, salutaridine reductase,* NADPH-dependent *codeinone reductase, reticuline oxidase, codeine O-demethylase, (R,S)-reticuline 7-O-methyltransferase, (RS)-norcoclaurine 6-O-methyltransferase, (S)-stylopine synthase, (R,S)-reticuline 7-O-methyltransferase* were found to be related to alkaloid metabolic processes. 62% of the transcripts assigned to **response to stimulus** process were upregulated. Response to cadmium ion (72), salt (50), cold (58), heat (56), high light intensity (53), salt (50) and defense response to bacterium (47) were the major processes found in this category. Transcripts involved in **transport process** were found to have their 55% upregulated, most of which were belonged to transmembrane transport process. Out of 577 transcripts categorized in **development process,** 52% were downregulated. Most of them were belonged to flower development/regulation (43), and pollen development (28). A total of 453 transcripts were detected in **protein modification** category and 62% were upregulated most of which were assigned to protein phosphorylation (130), protein folding (100) and ubiquitination (55). **Oxidation-reduction** process has 449 transcripts, of which 56% were upregulated. BIA pathway related transcripts were detected in this category (*salutaridine reductase, (S)-stylopine synthase, NADPH-dependent codeinone reductase, cheilanthifoline synthase, salutaridine synthase, reticuline oxidase*). Under **gene expression** category, half of the transcripts (210) were found to be upregulated. Most of them were assigned to regulation of transcription processes (118) and translation (45) processes. Out of 371 transcripts in **carbohydrate metabolism,** 51% were upregulated. Out of 338 transcripts in **hormone** process category, 195 (58%) were upregulated. Particularly, transcripts belonged to response to abscisic acid (82), auxin (27), response to karrikin (32), ethylene (25), jasmonic acid (21) were found to be differentially expressed. Out of 230 transcripts belonged to **lipid and fatty acid metabolism** category, 54% were upregulated. **Signal transduction** process was found to have 188 transcripts in total, the half was upregulated. Most of them were assigned to following processes; small GTPase mediated signal transduction (22), MAPK cascade (19) and activation of MAPKK activity (22). Lastly, under the category of **protein metabolism**, 53% of the transcripts were upregulated. proteolysis (62) and ubiquitin-

dependent protein catabolic process (42) were the major groups that the differentially expressed transcripts were belonged to.

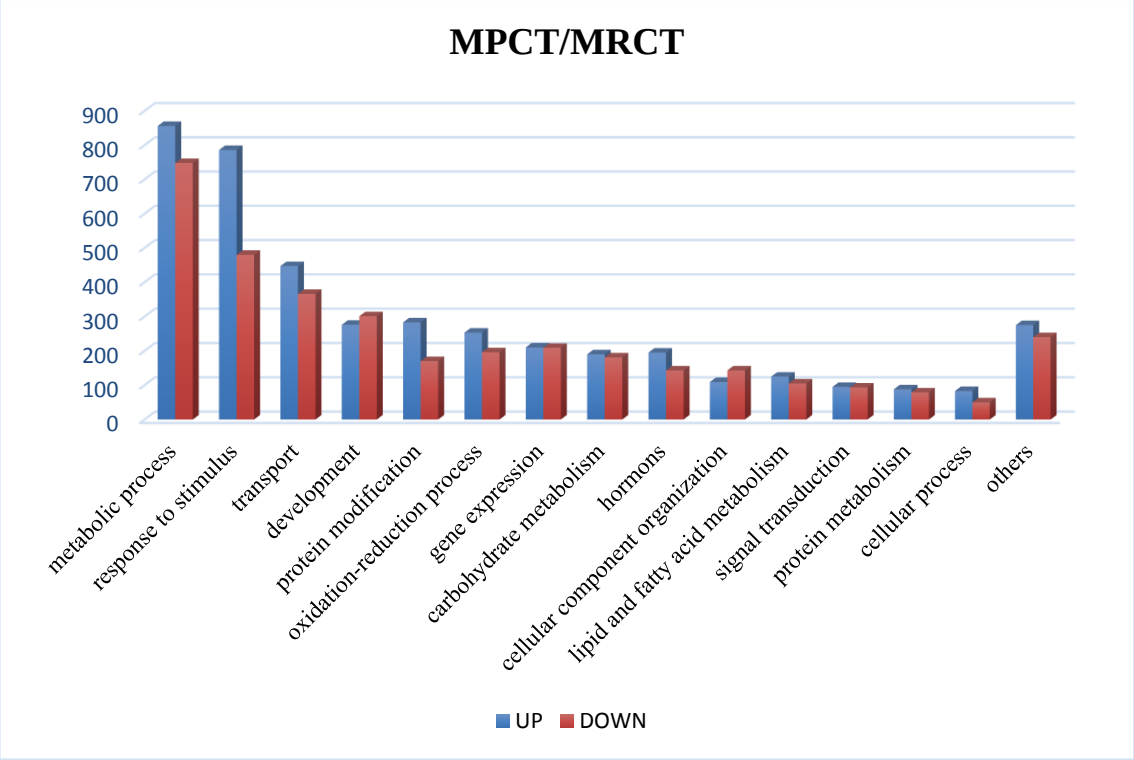



**Figure 3.23.** Most abundant biological processes in MPCT/MRCT comparison

### 3.4.1.8. MPST/MRST comparison

Gene expression levels in MeJA-treated stem tissues of morphine-rich and morphine-poor were compared. The log2 expression ratio cutoff was set to ± 1.5 fold and student's t-test p-value to ≤0.05.

A total of 14839 transcripts were found to be differentially expressed, of which 9561 and 5278 transcripts were upregulated and downregulated respectively in the stem tissue of morphine-poor cultivar (Figure 3.8c). By NCBI blastn search, 7740 of the transcripts (52.1%) were annotated to a gene in NCBI gene database (e-value <1e-5).

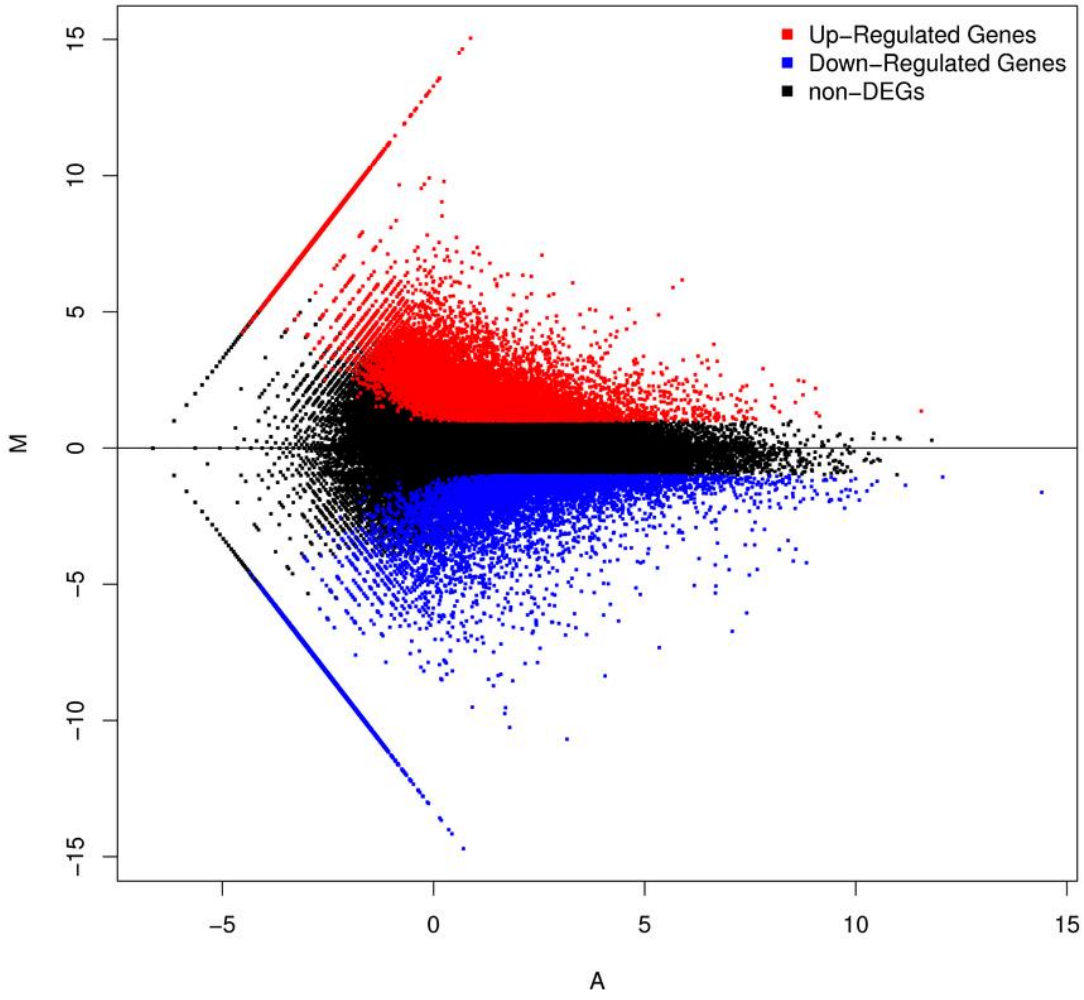



**Figure 3.24.** MA plot of MPST/MRST comparison

According to GO results, in total, 17995 gene ontology terms were assigned to differentially expressed transcripts, of which 8435 were determined to be belonged to biological processes, 5298 to molecular functions and 4262 to cellular components.

In the assessment of biological processes, metabolic process, response to stimulus and transport are again found to be the major categories representing differentially expressed transcripts (Figure 3.25). In the comparison of MeJA treated stem tissues 5251 transcripts were detected to be upregulated while 3184 were downregulated in morphine poor opium poppy cultivar.

Out of 1781 transcripts in **metabolic process,** 1116 (62%) were upregulated. In this category, phosphorylation (132), ATP catabolic process (65) and methylation (37) are the prominent terms. Besides, *tyrosine/DOPA decarboxylase, salutaridine reductase, codeine O-demethylase, reticuline oxidase,* NADPH-dependent *codeinone reductase, (R,S)-norcoclaurine 6-O-methyltransferase, (R,S)-reticuline 7-O-methyltransferase* were detected under metabolic processes. 60% of the transcripts with functions in **response to stimulus** process were upregulated. Response to cadmium ion (95), high light intensity (50), hydrogen peroxide (38), heat (50), cold (53), salt (72), defense response to bacterium (62) and wounding (34) were prominent processes in this category. Transcripts involved in **transport process** were found to have their 55% upregulated, most of which (131) were involved in transmembrane transport. Out of 638 transcripts categorized in **development process,** 68% were upregulated. Most of them were belonged to regulation of flower development (48), root development (25), leaf morphogenesis (21), and pollen development (21). Under **protein modification**, 70% out of 515 were found to be upregulated. Most of the transcripts in this category were involved in protein phosphorylation (142), protein folding (101), protein ubiquitination (45). A total of 501 transcripts were detected in **gene expression** category, 70% were upregulated, most of which were assigned to translation (108) and regulation of transcription (105) processes. **Oxidation-reduction** process has 372 transcripts, of which 64% were upregulated. BIA pathway related transcripts were also detected in this group (*salutaridine reductase, reticuline oxidase,* NADPH-dependent *codeinone reductase*). Out of 416 transcripts in **carbohydrate metabolism,** 63% were upregulated. Out of 304 transcripts in **hormone** process category, 172 (57%) were upregulated. The transcripts in response to abscisic acid (75), cytokinin (5), auxin (19), gibberellin (3), ethylene (16) and response to karrikin (30) were detected. Out of 158 transcripts belonged to **lipid and fatty acid metabolism** category, 60% were downregulated. Under the category of **protein metabolism**, 66% of the transcripts were upregulated, 90 of them were belonged to proteolysis and 26 to ubiquitin-dependent protein catabolic process. Lastly, **signal transduction** process was found to have 161 transcripts in total and 66% were upregulated. Small GTPase mediated signal transduction (21), and MAPK cascade (27) were the most frequent ontology terms that the transcripts exhibited differential expression.

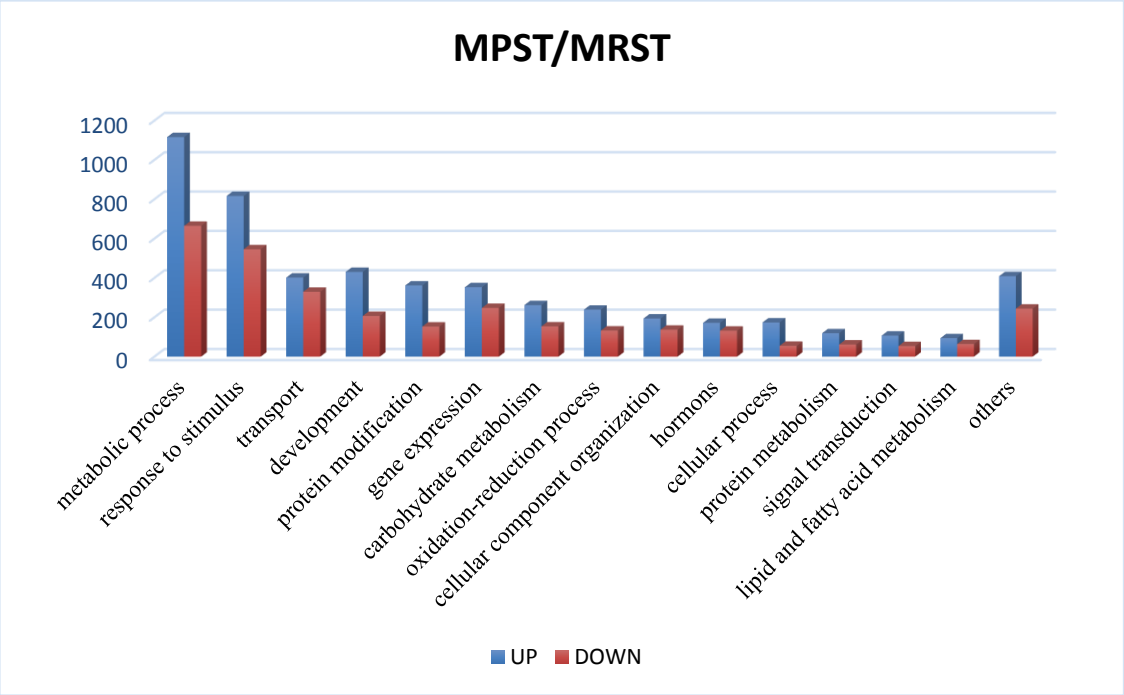



**Figure 3.25.** Most abundant biological processes in MPST/MRST comparison

### 3.5. KEGG Pathway Analysis

KEGG pathway analysis is important for a better understanding of the function of the genes. In this sense, the metabolic processes in which the opium poppy differentially expressed genes are involved is determined. As the most comprehensive metabolic pathway database, KEGG (Kyoto Encyclopedia of Genes and Genomes) database is used for pathway annotation. Hence, we identified 7094 (MRCT/MRCC), 11718 (MRST/MRSC), 7801 (MPCT/MRCC), 6336 (MPST/MPSC), 12413 (MRST/MRCT), 11876 (MPST/MPCT), 9554 (MPCT/MRCT), 12045 (MPST/MRST) DEG genes with pathway annotation.

KEGG-enrichment analyses resulted that the most significant pathway groups were related to metabolic pathway and biosynthesis of secondary metabolites and were common in all libraries. The third group of pathway represented by most genes, was plant-pathogen interaction and endocytosis especially in capsule, while splicesome and ribosome in stem. MeJA altered the expression of plant-pathogen related genes in capsule tissue as expected. Splicesome genes were higher in stem tissues indicating alternative-splicing of the transcripts which are affected by MeJA-induction. Besides, highly annotated another group of genes were active in plant hormone signal transduction.

Additionally, the isoquinoline biosynthesis metabolism was another prominent pathway in both capsule and stem tissues, as it leads to morphine and other BIA alkaloid production. Several studies showed that biosynthesis of defensive secondary metabolites is associated with pathogen response (Zhao et al., 2005). Figure 3.26 illustrates isoquinoline pathway, with the transcripts detected in this study.

A representative KEGG pathway annotation list is given in Table 3.5 which is derived for morphine rich capsule tissue. Besides, functional KEGG pathway enrichment for each comparison group is given as dot plot graphs. (Figure 3.27 - Figure 3.34)

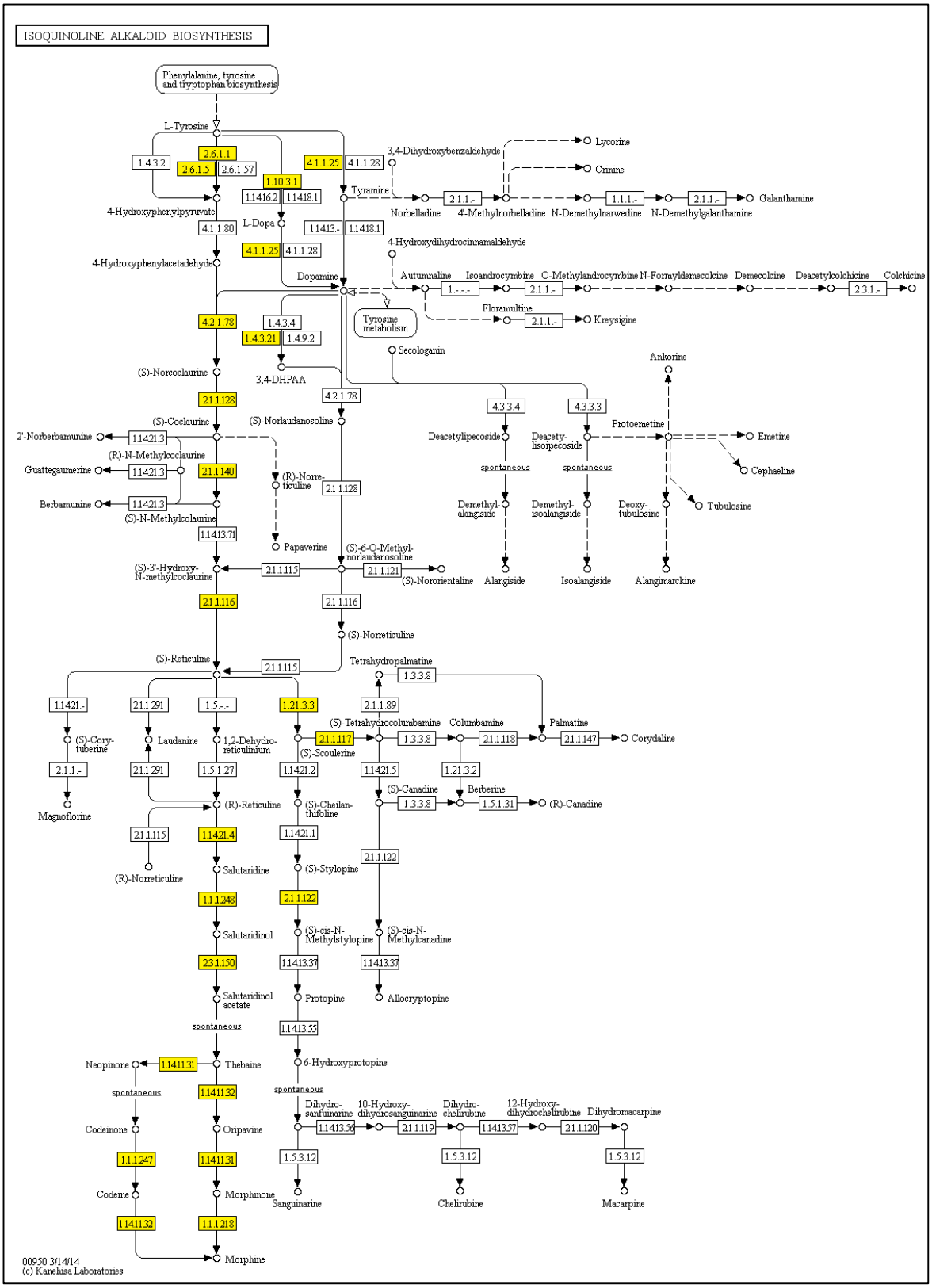



**Figure 3.26.** A representative BIA pathway KEGG map (DEGs are shown in yellow)

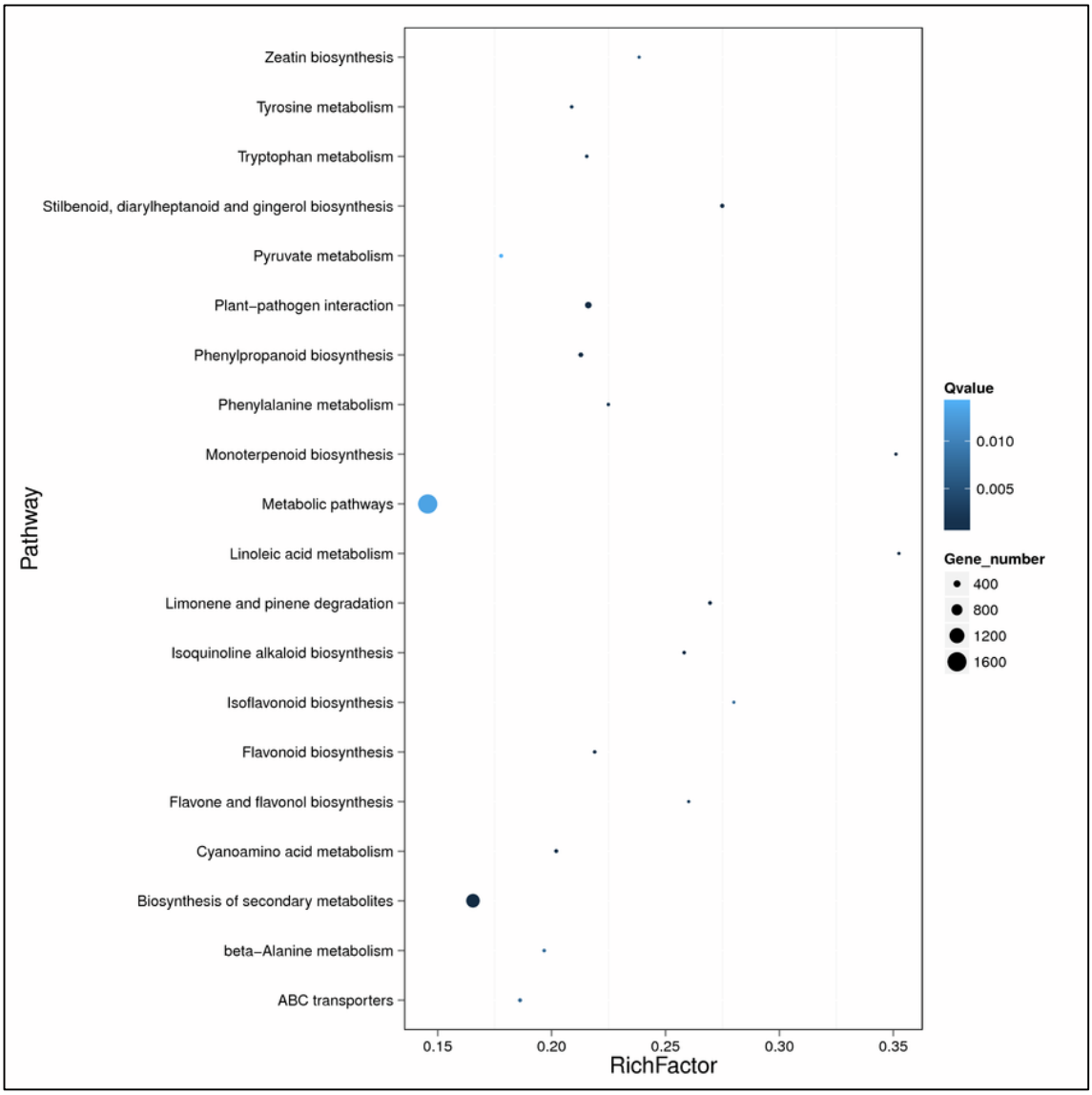



**Figure 3.27.** Pathway functional enrichment of DEGs for MRCT/MRCC*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

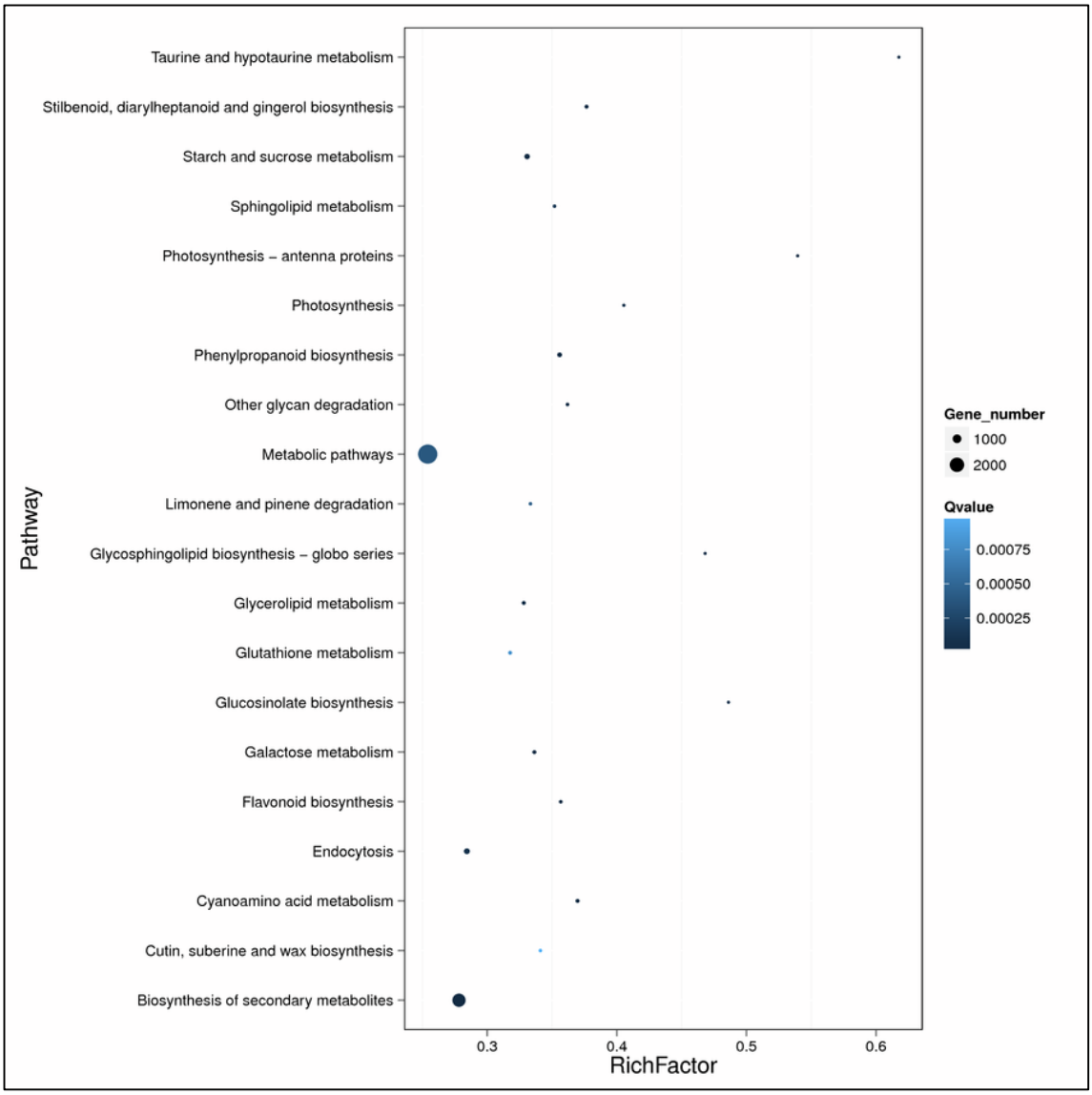



**Figure 3.28.** Pathway functional enrichment of DEGs for MRST/MRCT*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

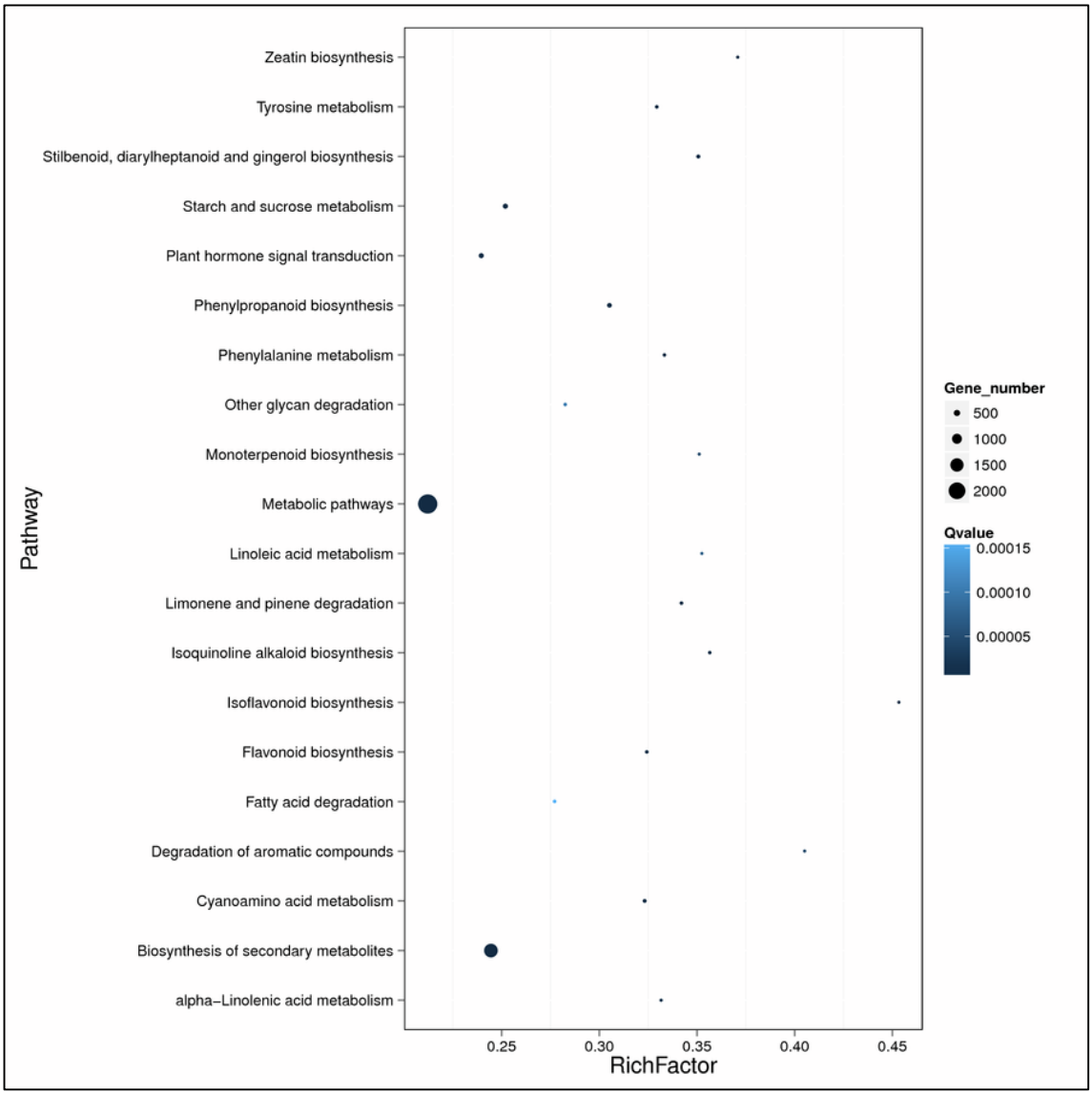



**Figure 3.29.** Pathway functional enrichment of DEGs for MPCT/MRCT*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

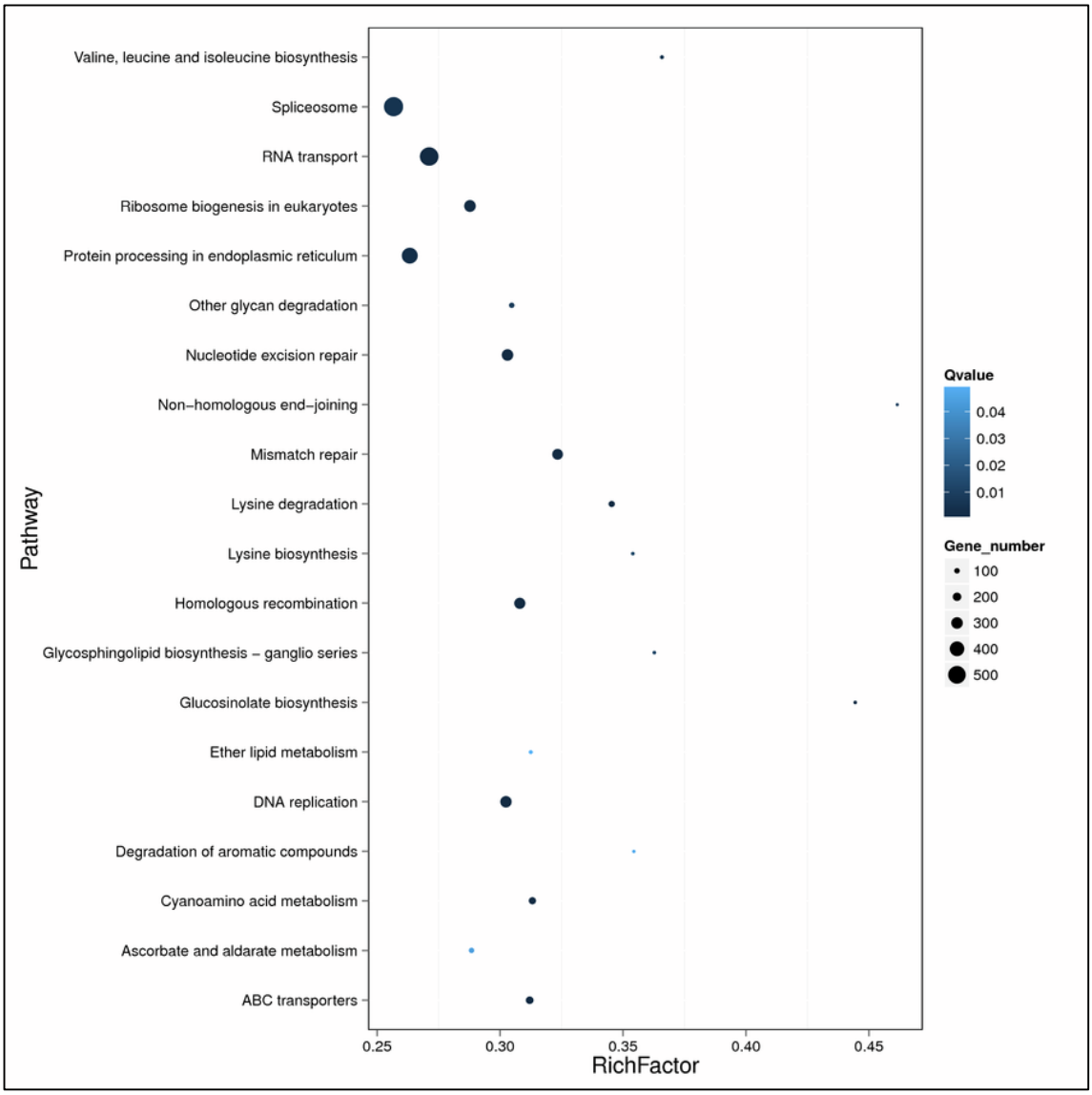



**Figure 3.30.** Pathway functional enrichment of DEGs for MRST/MRSC*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

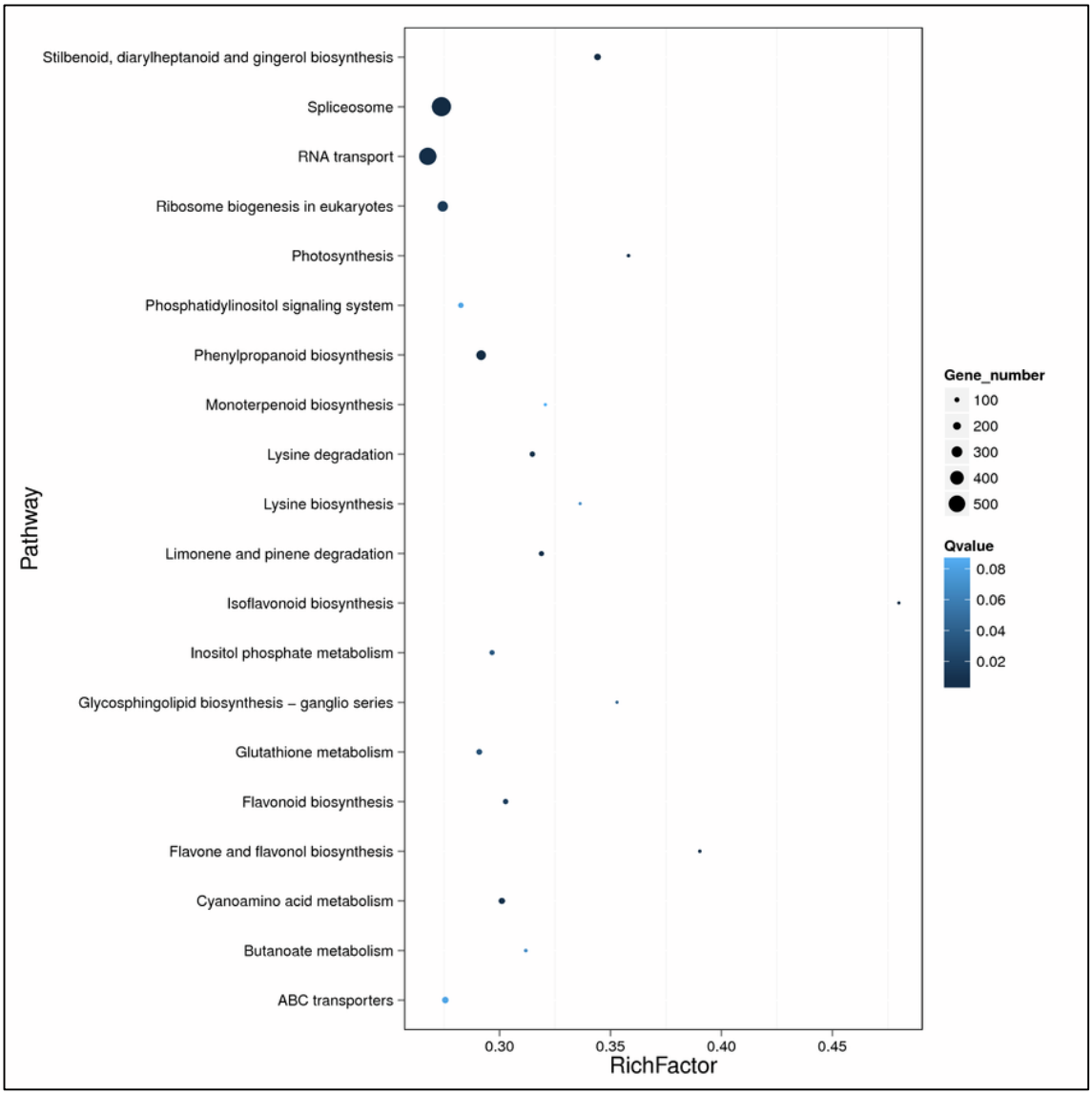



**Figure 3.31.** Pathway functional enrichment of DEGs for MPST/MRST*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

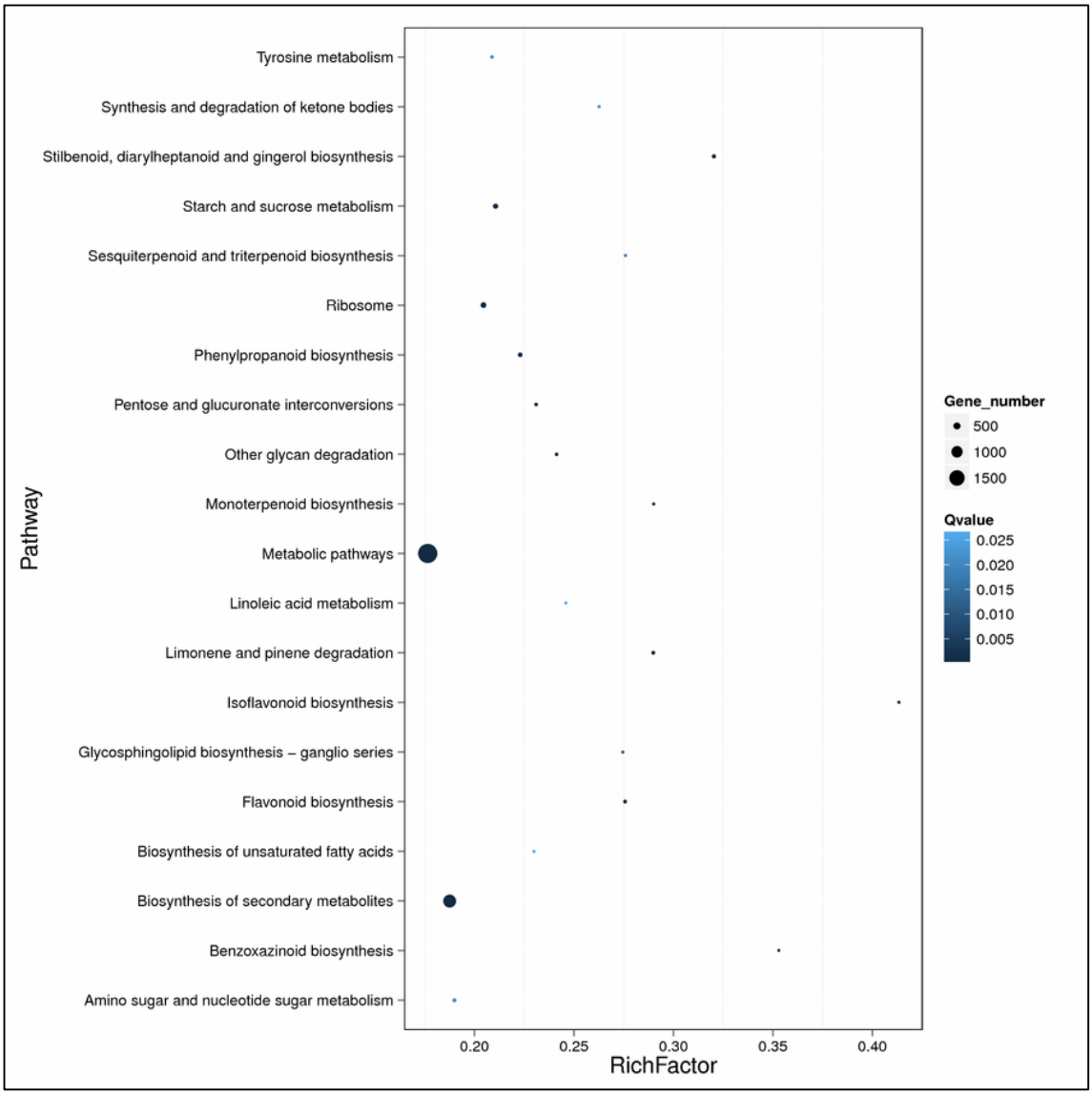



**Figure 3.32.** Pathway functional enrichment of DEGs for MPCT/MPCC*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

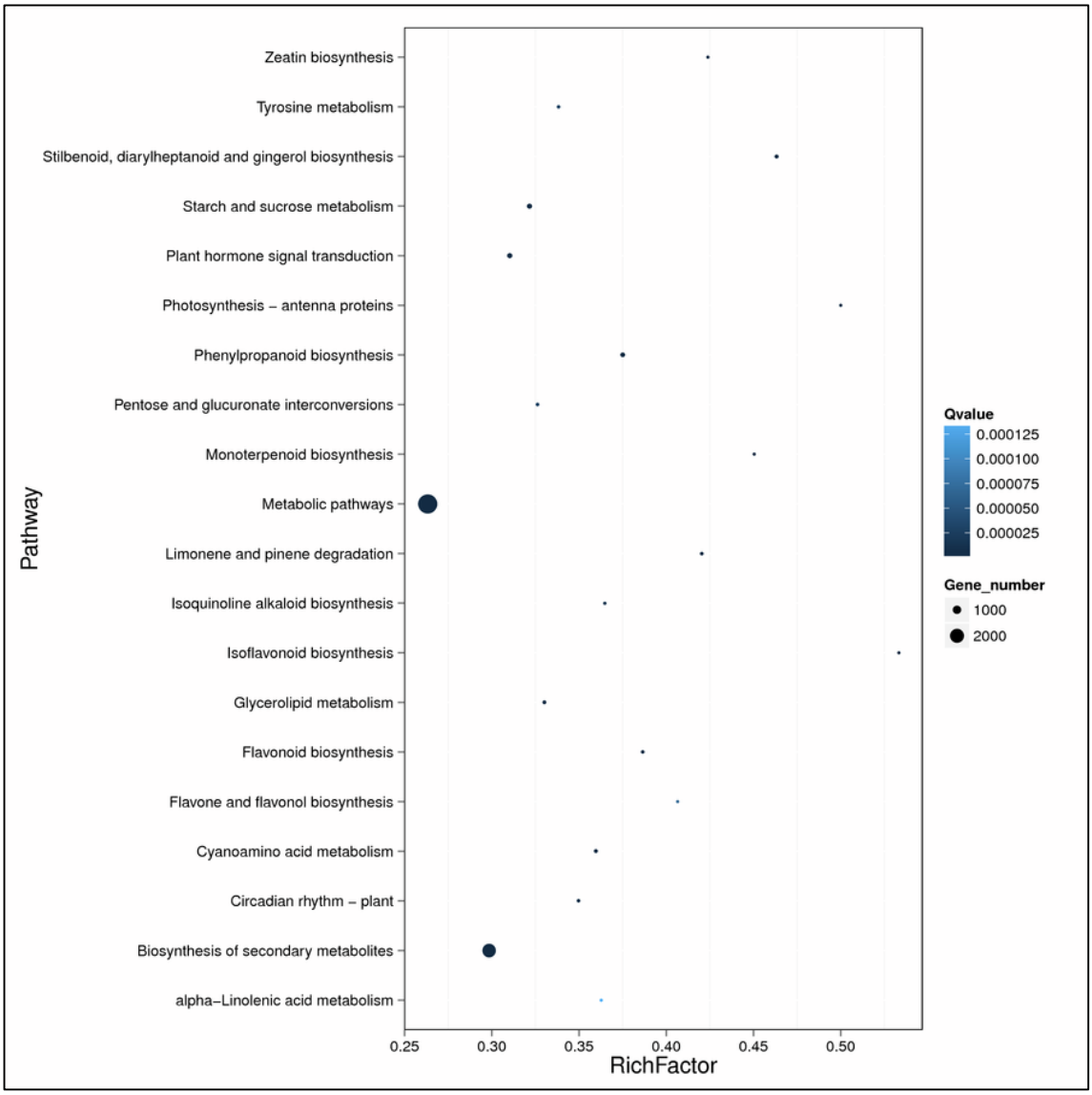



**Figure 3.33.** Pathway functional enrichment of DEGs for MPST/MPCT*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

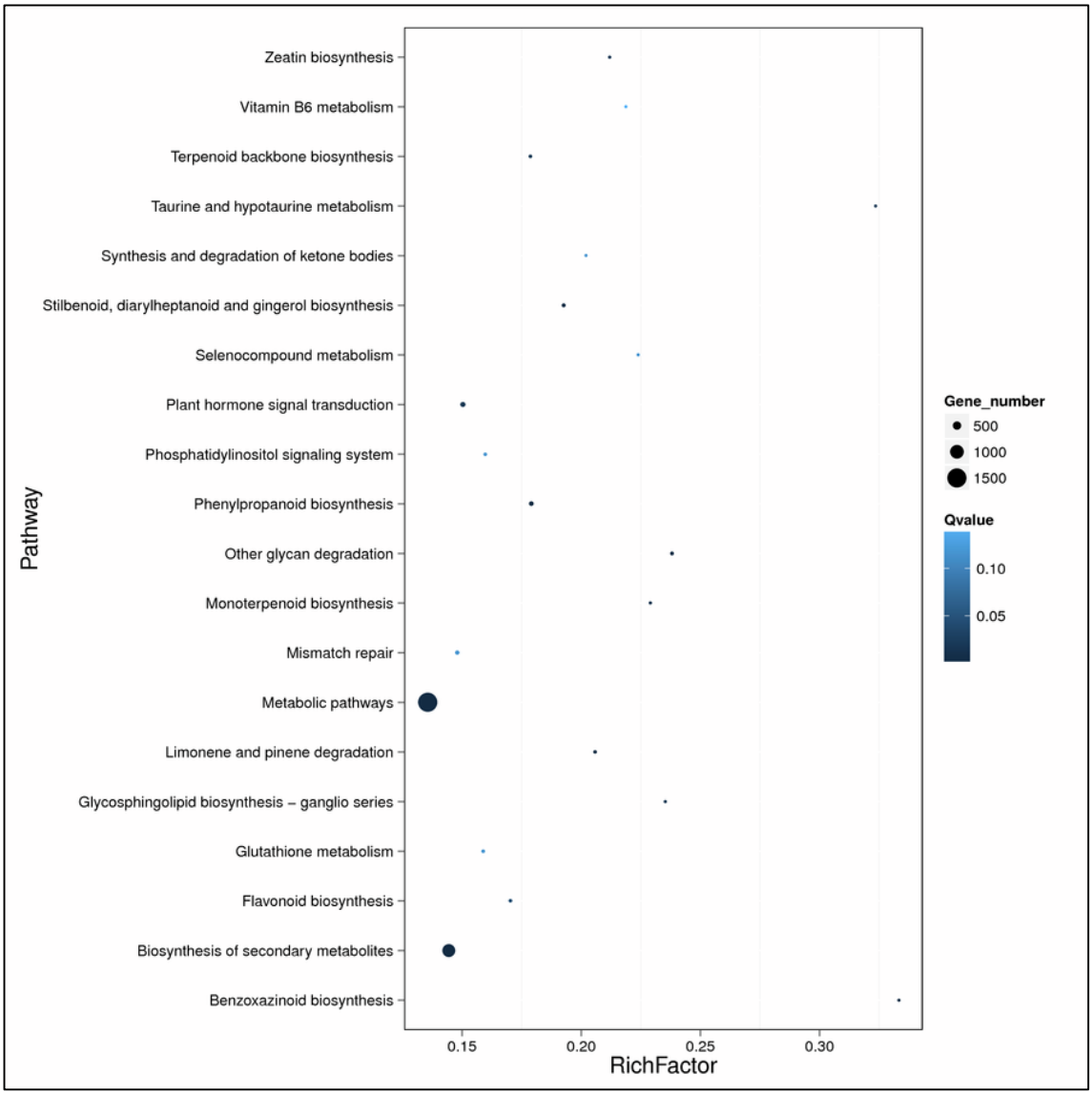



**Figure 3.34.** Pathway functional enrichment of DEGs for MPST/MPSC*

***X-axis represents enrichment factor. Y-axis represents pathway name. Coloring indicate Qvalue (high: white, low: blue), the lower q value indicates the more significant enriched. Point size indicate number of DEGs (more: big, less: small).

**Table 3.5.** KEGG Pathway enrichment analysis in MR capsule tissues Pathways with P-value <0.05 and Q value <0.05 indicate significantly enriched in differentially expressed genes (DEGs).

| # | Pathway | DEGs genes with pathway annotation (7094) | All genes with pathway annotation (51729) | P value | Q value | Pathway ID |
|---|---|---|---|---|---|---|
| **1** | Metabolic pathways | 1623 (22.88%) | 11145 (21.54%) | 0.00181 | 0.01275 | ko01100 |
| **2** | Biosynthesis of secondary metabolites | 1053 (14.84%) | 6363 (12.3%) | 3.8E-12 | 1.7E-10 | ko01110 |
| **3** | Plant-pathogen interaction | 368 (5.19%) | 1703 (3.29%) | 1.1E-19 | 1.5E-17 | ko04626 |
| **4** | Spliceosome | 258 (3.64%) | 2166 (4.19%) | 0.99482 | 1 | ko03040 |
| **5** | Endocytosis | 247 (3.48%) | 1801 (3.48%) | 0.51036 | 0.9779 | ko04144 |
| **6** | Carbon metabolism | 211 (2.97%) | 1527 (2.95%) | 0.46379 | 0.94163 | ko01200 |
| **7** | Protein processing in endoplasmic reticulum | 210 (2.96%) | 1694 (3.27%) | 0.9508 | 1 | ko04141 |
| **8** | Starch and sucrose metabolism | 207 (2.92%) | 1306 (2.52%) | 0.01399 | 0.06942 | ko00500 |
| **9** | RNA transport | 203 (2.86%) | 1973 (3.81%) | 1 | 1 | ko03013 |
| **10** | Biosynthesis of amino acids | 202 (2.85%) | 1618 (3.13%) | 0.93423 | 1 | ko01230 |
| **11** | Phenylpropanoid biosynthesis | 189 (2.66%) | 888 (1.72%) | 3.3E-10 | 8.9E-09 | ko00940 |
| **12** | Plant hormone signal transduction | 183 (2.58%) | 1344 (2.6%) | 0.55434 | 0.99042 | ko04075 |
| **13** | Ribosome | 157 (2.21%) | 1584 (3.06%) | 1 | 1 | ko03010 |
| **14** | mRNA surveillance pathway | 144 (2.03%) | 1150 (2.22%) | 0.89202 | 1 | ko03015 |
| **15** | Purine metabolism | 138 (1.95%) | 1197 (2.31%) | 0.98956 | 1 | ko00230 |
| **16** | Glycolysis / Gluconeogenesis | 131 (1.85%) | 900 (1.74%) | 0.24269 | 0.6254 | ko00010 |
| **17** | Nucleotide excision repair | 131 (1.85%) | 990 (1.91%) | 0.68545 | 1 | ko03420 |
| **18** | Stilbenoid, diarylheptanoid and gingerol biosynthesis | 127 (1.79%) | 462 (0.89%) | 4.5E-15 | 3E-13 | ko00945 |
| **19** | Homologous recombination | 121 (1.71%) | 922 (1.78%) | 0.71441 | 1 | ko03440 |
| **20** | DNA replication | 121 (1.71%) | 969 (1.87%) | 0.87942 | 1 | ko03030 |
| **21** | Ubiquitin mediated proteolysis | 120 (1.69%) | 902 (1.74%) | 0.65557 | 1 | ko04120 |
| **22** | Pyrimidine metabolism | 117 (1.65%) | 976 (1.89%) | 0.95044 | 1 | ko00240 |
| **23** | Pyruvate metabolism | 114 (1.61%) | 641 (1.24%) | 0.00209 | 0.01401 | ko00620 |
| **24** | Mismatch repair | 114 (1.61%) | 838 (1.62%) | 0.55261 | 0.99042 | ko03430 |
| **25** | Ribosome biogenesis in eukaryotes | 114 (1.61%) | 1046 (2.02%) | 0.9974 | 1 | ko03008 |
| **26** | Amino sugar and nucleotide sugar metabolism | 109 (1.54%) | 711 (1.37%) | 0.11474 | 0.375 | ko00520 |
| **27** | RNA degradation | 103 (1.45%) | 921 (1.78%) | 0.99072 | 1 | ko03018 |

| # | Pathway | DEGs genes with pathway annotation (7094) | All genes with pathway annotation (51729) | P value | Q value | Pathway ID |
|---|---|---|---|---|---|---|
| **28** | ABC transporters | 102 (1.44%) | 548 (1.06%) | 0.00078 | 0.00651 | ko02010 |
| **29** | Cyanoamino acid metabolism | 100 (1.41%) | 495 (0.96%) | 4.1E-05 | 0.00054 | ko00460 |
| **30** | Cysteine and methionine metabolism | 95 (1.34%) | 653 (1.26%) | 0.28241 | 0.67929 | ko00270 |
| **31** | Limonene and pinene degradation | 93 (1.31%) | 345 (0.67%) | 6.1E-11 | 2E-09 | ko00903 |
| **32** | Flavonoid biosynthesis | 81 (1.14%) | 370 (0.72%) | 1.1E-05 | 0.00017 | ko00941 |
| **33** | Galactose metabolism | 76 (1.07%) | 553 (1.07%) | 0.51084 | 0.9779 | ko00052 |
| **34** | Glutathione metabolism | 74 (1.04%) | 447 (0.86%) | 0.0487 | 0.18646 | ko00480 |
| **35** | Aminoacyl-tRNA biosynthesis | 74 (1.04%) | 643 (1.24%) | 0.95742 | 1 | ko00970 |
| **36** | beta-Alanine metabolism | 72 (1.01%) | 366 (0.71%) | 0.00096 | 0.00712 | ko00410 |
| **37** | Tyrosine metabolism | 71 (1%) | 340 (0.66%) | 0.00018 | 0.00182 | ko00350 |
| **38** | Glycerophospholipid metabolism | 71 (1%) | 567 (1.1%) | 0.81274 | 1 | ko00564 |
| **39** | Glyoxylate and dicarboxylate metabolism | 70 (0.99%) | 618 (1.19%) | 0.96625 | 1 | ko00630 |
| **40** | Alanine, aspartate and glutamate metabolism | 69 (0.97%) | 493 (0.95%) | 0.44725 | 0.92203 | ko00250 |
| **41** | Oxidative phosphorylation | 68 (0.96%) | 628 (1.21%) | 0.98716 | 1 | ko00190 |
| **42** | Tryptophan metabolism | 67 (0.94%) | 311 (0.6%) | 0.0001 | 0.00125 | ko00380 |
| **43** | Pentose and glucuronate interconversions | 65 (0.92%) | 420 (0.81%) | 0.16255 | 0.45379 | ko00040 |
| **44** | Fatty acid metabolism | 64 (0.9%) | 495 (0.96%) | 0.71394 | 1 | ko01212 |
| **45** | Isoquinoline alkaloid biosynthesis | 63 (0.89%) | 244 (0.47%) | 3.6E-07 | 6.1E-06 | ko00950 |
| **46** | Glycine, serine and threonine metabolism | 61 (0.86%) | 389 (0.75%) | 0.14529 | 0.43265 | ko00260 |
| **47** | Peroxisome | 61 (0.86%) | 473 (0.91%) | 0.71759 | 1 | ko04146 |
| **48** | Valine, leucine and isoleucine degradation | 60 (0.85%) | 373 (0.72%) | 0.10534 | 0.35288 | ko00280 |
| **49** | Glycerolipid metabolism | 60 (0.85%) | 521 (1.01%) | 0.93968 | 1 | ko00561 |
| **50** | Insulin resistance | 59 (0.83%) | 376 (0.73%) | 0.14855 | 0.43274 | ko04931 |
| **51** | Terpenoid backbone biosynthesis | 58 (0.82%) | 375 (0.72%) | 0.17924 | 0.49016 | ko00900 |
| **52** | Phagosome | 58 (0.82%) | 525 (1.01%) | 0.9706 | 1 | ko04145 |
| **53** | Regulation of autophagy | 57 (0.8%) | 292 (0.56%) | 0.00358 | 0.02085 | ko04140 |
| **54** | Other glycan degradation | 57 (0.8%) | 315 (0.61%) | 0.01696 | 0.08116 | ko00511 |
| **55** | Phenylalanine metabolism | 56 (0.79%) | 249 (0.48%) | 0.00011 | 0.00125 | ko00360 |
| **56** | Carbon fixation in photosynthetic organisms | 56 (0.79%) | 387 (0.75%) | 0.35369 | 0.8033 | ko00710 |
| **57** | Diterpenoid biosynthesis | 55 (0.78%) | 279 (0.54%) | 0.00335 | 0.02042 | ko00904 |
| **58** | Arginine and proline metabolism | 54 (0.76%) | 317 (0.61%) | 0.05338 | 0.19869 | ko00330 |

| # | Pathway | DEGs genes with pathway annotation (7094) | All genes with pathway annotation (51729) | P value | Q value | Pathway ID |
|---|---|---|---|---|---|---|
| **59** | Ascorbate and aldarate metabolism | 54 (0.76%) | 319 (0.62%) | 0.05869 | 0.21255 | ko00053 |
| **60** | Nitrogen metabolism | 53 (0.75%) | 395 (0.76%) | 0.59038 | 1 | ko00910 |
| **61** | Fructose and mannose metabolism | 52 (0.73%) | 363 (0.7%) | 0.38969 | 0.82887 | ko00051 |
| **62** | Arginine biosynthesis | 52 (0.73%) | 449 (0.87%) | 0.92002 | 1 | ko00220 |
| **63** | Inositol phosphate metabolism | 51 (0.72%) | 354 (0.68%) | 0.37466 | 0.81731 | ko00562 |
| **64** | Phosphatidylinositol signaling system | 51 (0.72%) | 407 (0.79%) | 0.77693 | 1 | ko04070 |
| **65** | Proteasome | 49 (0.69%) | 447 (0.86%) | 0.96462 | 1 | ko03050 |
| **66** | Cutin, suberine and wax biosynthesis | 48 (0.68%) | 261 (0.5%) | 0.0203 | 0.08499 | ko00073 |
| **67** | Fatty acid degradation | 48 (0.68%) | 332 (0.64%) | 0.36977 | 0.81731 | ko00071 |
| **68** | Porphyrin and chlorophyll metabolism | 48 (0.68%) | 333 (0.64%) | 0.37816 | 0.81731 | ko00860 |
| **69** | Monoterpenoid biosynthesis | 46 (0.65%) | 131 (0.25%) | 5.2E-10 | 1.2E-08 | ko00902 |
| **70** | Propanoate metabolism | 45 (0.63%) | 282 (0.55%) | 0.15584 | 0.44431 | ko00640 |
| **71** | Basal transcription factors | 45 (0.63%) | 365 (0.71%) | 0.80048 | 1 | ko03022 |
| **72** | Tropane, piperidine and pyridine alkaloid biosynthesis | 44 (0.62%) | 238 (0.46%) | 0.02334 | 0.09477 | ko00960 |
| **73** | Sphingolipid metabolism | 44 (0.62%) | 287 (0.55%) | 0.2349 | 0.61719 | ko00600 |
| **74** | 2-Oxocarboxylic acid metabolism | 44 (0.62%) | 357 (0.69%) | 0.79884 | 1 | ko01210 |
| **75** | Lysine degradation | 44 (0.62%) | 359 (0.69%) | 0.81023 | 1 | ko00310 |
| **76** | Linoleic acid metabolism | 43 (0.61%) | 122 (0.24%) | 1.7E-09 | 3.2E-08 | ko00591 |
| **77** | Carotenoid biosynthesis | 43 (0.61%) | 229 (0.44%) | 0.0194 | 0.08499 | ko00906 |
| **78** | RNA polymerase | 43 (0.61%) | 421 (0.81%) | 0.98744 | 1 | ko03020 |
| **79** | Base excision repair | 42 (0.59%) | 359 (0.69%) | 0.88481 | 1 | ko03410 |
| **80** | Pentose phosphate pathway | 42 (0.59%) | 360 (0.7%) | 0.88872 | 1 | ko00030 |
| **81** | Butanoate metabolism | 40 (0.56%) | 186 (0.36%) | 0.00236 | 0.01507 | ko00650 |
| **82** | Circadian rhythm - plant | 40 (0.56%) | 326 (0.63%) | 0.79828 | 1 | ko04712 |
| **83** | Zeatin biosynthesis | 36 (0.51%) | 151 (0.29%) | 0.00056 | 0.00497 | ko00908 |
| **84** | SNARE interactions in vesicular transport | 35 (0.49%) | 220 (0.43%) | 0.19565 | 0.52434 | ko04130 |
| **85** | Nicotinate and nicotinamide metabolism | 34 (0.48%) | 156 (0.3%) | 0.00382 | 0.02136 | ko00760 |
| **86** | Biosynthesis of unsaturated fatty acids | 34 (0.48%) | 174 (0.34%) | 0.02019 | 0.08499 | ko01040 |
| **87** | Flavone and flavonol biosynthesis | 32 (0.45%) | 123 (0.24%) | 0.00021 | 0.00203 | ko00944 |
| **88** | Citrate cycle (TCA cycle) | 32 (0.45%) | 282 (0.55%) | 0.89588 | 1 | ko00020 |

| # | Pathway | DEGs genes with pathway annotation (7094) | All genes with pathway annotation (51729) | P value | Q value | Pathway ID |
|---|---|---|---|---|---|---|
| **89** | alpha-Linolenic acid metabolism | 30 (0.42%) | 193 (0.37%) | 0.25763 | 0.65136 | ko00592 |
| **90** | Phenylalanine, tyrosine and tryptophan biosynthesis | 29 (0.41%) | 235 (0.45%) | 0.75709 | 1 | ko00400 |
| **91** | Ubiquinone and other terpenoid-quinone biosynthesis | 28 (0.39%) | 242 (0.47%) | 0.85755 | 1 | ko00130 |
| **92** | Glycosaminoglycan degradation | 27 (0.38%) | 141 (0.27%) | 0.04409 | 0.17376 | ko00531 |
| **93** | Sulfur metabolism | 27 (0.38%) | 180 (0.35%) | 0.33876 | 0.79639 | ko00920 |
| **94** | N-Glycan biosynthesis | 27 (0.38%) | 267 (0.52%) | 0.96862 | 1 | ko00510 |
| **95** | Glycosphingolipid biosynthesis - ganglio series | 24 (0.34%) | 102 (0.2%) | 0.00511 | 0.02738 | ko00604 |
| **96** | Fatty acid elongation | 24 (0.34%) | 154 (0.3%) | 0.2817 | 0.67929 | ko00062 |
| **97** | Pantothenate and CoA biosynthesis | 24 (0.34%) | 201 (0.39%) | 0.7961 | 1 | ko00770 |
| **98** | Fatty acid biosynthesis | 24 (0.34%) | 229 (0.44%) | 0.94033 | 1 | ko00061 |
| **99** | Synthesis and degradation of ketone bodies | 23 (0.32%) | 99 (0.19%) | 0.00709 | 0.03653 | ko00072 |
| **100** | Isoflavonoid biosynthesis | 21 (0.3%) | 75 (0.14%) | 0.00088 | 0.00695 | ko00943 |
| **101** | Protein export | 21 (0.3%) | 207 (0.4%) | 0.94967 | 1 | ko03060 |
| **102** | Degradation of aromatic compounds | 18 (0.25%) | 79 (0.15%) | 0.01932 | 0.08499 | ko01220 |
| **103** | Lysine biosynthesis | 18 (0.25%) | 113 (0.22%) | 0.28388 | 0.67929 | ko00300 |
| **104** | Riboflavin metabolism | 18 (0.25%) | 118 (0.23%) | 0.35158 | 0.8033 | ko00740 |
| **105** | Ether lipid metabolism | 18 (0.25%) | 160 (0.31%) | 0.84732 | 1 | ko00565 |
| **106** | Valine, leucine and isoleucine biosynthesis | 16 (0.23%) | 123 (0.24%) | 0.62957 | 1 | ko00290 |
| **107** | Photosynthesis - antenna proteins | 15 (0.21%) | 76 (0.15%) | 0.09106 | 0.31288 | ko00196 |
| **108** | Histidine metabolism | 15 (0.21%) | 140 (0.27%) | 0.87878 | 1 | ko00340 |
| **109** | Brassinosteroid biosynthesis | 14 (0.2%) | 74 (0.14%) | 0.13009 | 0.40539 | ko00905 |
| **110** | Thiamine metabolism | 14 (0.2%) | 107 (0.21%) | 0.61759 | 1 | ko00730 |
| **111** | Sesquiterpenoid and triterpenoid biosynthesis | 13 (0.18%) | 87 (0.17%) | 0.41531 | 0.86955 | ko00909 |
| **112** | Glycosphingolipid biosynthesis - globo series | 13 (0.18%) | 94 (0.18%) | 0.53229 | 0.99042 | ko00603 |
| **113** | Benzoxazinoid biosynthesis | 11 (0.16%) | 51 (0.1%) | 0.08225 | 0.29005 | ko00402 |
| **114** | Monobactam biosynthesis | 11 (0.16%) | 78 (0.15%) | 0.5098 | 0.9779 | ko00261 |
| **115** | Folate biosynthesis | 11 (0.16%) | 87 (0.17%) | 0.66096 | 1 | ko00790 |
| **116** | Steroid biosynthesis | 11 (0.16%) | 135 (0.26%) | 0.98325 | 1 | ko00100 |
| **117** | Glucosinolate biosynthesis | 10 (0.14%) | 72 (0.14%) | 0.5345 | 0.99042 | ko00966 |
| **118** | Glycosylphosphatidylinositol(GPI)-anchor biosynthesis | 10 (0.14%) | 115 (0.22%) | 0.96257 | 1 | ko00563 |

| # | Pathway | DEGs genes with pathway annotation (7094) | All genes with pathway annotation (51729) | P value | Q value | Pathway ID |
|---|---|---|---|---|---|---|
| **119** | Caffeine metabolism | 9 (0.13%) | 44 (0.09%) | 0.14057 | 0.42811 | ko00232 |
| **120** | Other types of O-glycan biosynthesis | 9 (0.13%) | 65 (0.13%) | 0.54207 | 0.99042 | ko00514 |
| **121** | Photosynthesis | 9 (0.13%) | 148 (0.29%) | 0.99906 | 1 | ko00195 |
| **122** | Arachidonic acid metabolism | 8 (0.11%) | 94 (0.18%) | 0.95492 | 1 | ko00590 |
| **123** | Selenocompound metabolism | 7 (0.1%) | 67 (0.13%) | 0.8298 | 1 | ko00450 |
| **124** | One carbon pool by folate | 7 (0.1%) | 89 (0.17%) | 0.96886 | 1 | ko00670 |
| **125** | C5-Branched dibasic acid metabolism | 6 (0.08%) | 52 (0.1%) | 0.7352 | 1 | ko00660 |
| **126** | Sulfur relay system | 6 (0.08%) | 66 (0.13%) | 0.90454 | 1 | ko04122 |
| **127** | Biotin metabolism | 6 (0.08%) | 138 (0.27%) | 0.99993 | 1 | ko00780 |
| **128** | Anthocyanin biosynthesis | 5 (0.07%) | 20 (0.04%) | 0.12871 | 0.40539 | ko00942 |
| **129** | Taurine and hypotaurine metabolism | 5 (0.07%) | 34 (0.07%) | 0.5076 | 0.9779 | ko00430 |
| **130** | Vitamin B6 metabolism | 5 (0.07%) | 64 (0.12%) | 0.94965 | 1 | ko00750 |
| **131** | Indole alkaloid biosynthesis | 4 (0.06%) | 31 (0.06%) | 0.63102 | 1 | ko00901 |
| **132** | Non-homologous end-joining | 4 (0.06%) | 39 (0.08%) | 0.80134 | 1 | ko03450 |
| **133** | Vancomycin resistance | 1 (0.01%) | 14 (0.03%) | 0.87322 | 1 | ko01502 |
| **134** | Lipoic acid metabolism | 1 (0.01%) | 22 (0.04%) | 0.96106 | 1 | ko00785 |

### 3.6. Integrative Correlation of Metabolic and RNA-Seq Transcriptomic Data

In response to elicitor treatment, a defensive mechanism has been induced by the accumulation of alkaloids in different organs of opium poppy as reported in several studies (Huang and Kutchan, 2000; Holkova et al., 2010; Mishra et al., 2013; Gurkok et al., 2015). With respect to the metabolic profiles of capsule and stem tissues belonged to MR and MP cultivars, it can be obviously inferred from increasing alkaloid content that MeJA acts as a morphine-elicitor especially in capsule tissue at almost all time points in both cultivars. On the contrary, a completely opposite pattern has been determined in stem tissues of both MR and MP. This result supported the hypothesis of that the morphine synthesized in stem immediately transported into the capsule in order to be accumulated. Morphine production occurs in both tissues as will be explained by the transcriptome results.

### 3.6.1. Digital gene expression analysis

Although each cell of an organism has the same set of genes, different genes are activated in different conditions. Digital gene expression analysis sheds light on the transcriptional behavior of the organisms. It is mainly used to identify genes differentially expressed among tissues, developmental stages or given conditions.

The mapped reads of the RNAseq data is summed in order to be converted into a quantitative measure of gene expression. Normalization is employed to eliminate the bias caused by the variation in transcript lengths and number of fragments. In our case, fragments per kilobase of transcript per million mapped (FPKM) method of normalization is preferred since paired-end sequencing data is produced. FPKM normalization enabled us to directly compare relative gene expression levels in different tissues/cultivars.

Using normalized RNAseq data, a heatmap describing the distance between samples is plotted considering 50-mostly expressed transcripts (Figure 3.35). The expression levels of specific transcripts such as storage proteins are explicitly differed in capsule and stem tissues. 11S globulin seed storage protein, 2S albumin, legumin B and glycine-rich RNA-binding proteins are known to play essential roles in plant survival along with plant defense mechanism with their antimicrobial activity (de Souza Cândido et al., 2011). As MeJA treatment elicits pathogen response in the plant, these proteins are expected to accumulate in different plant tissues.

Gene expression levels of transcripts of thaumatin-like protein, pathogen related protein 1, antimicrobial peptide, chitinase 6, non-specific lipid-transfer protein, polyubiquitin, defensine-like protein are evenly distributed among capsule and stem tissues in both cultivars. These proteins have a common point that they are all pathogen-related and expressed in response to pathogen attack (Stotz et al., 2009; Ebrahim et al., 2011; Misra et al., 2016).

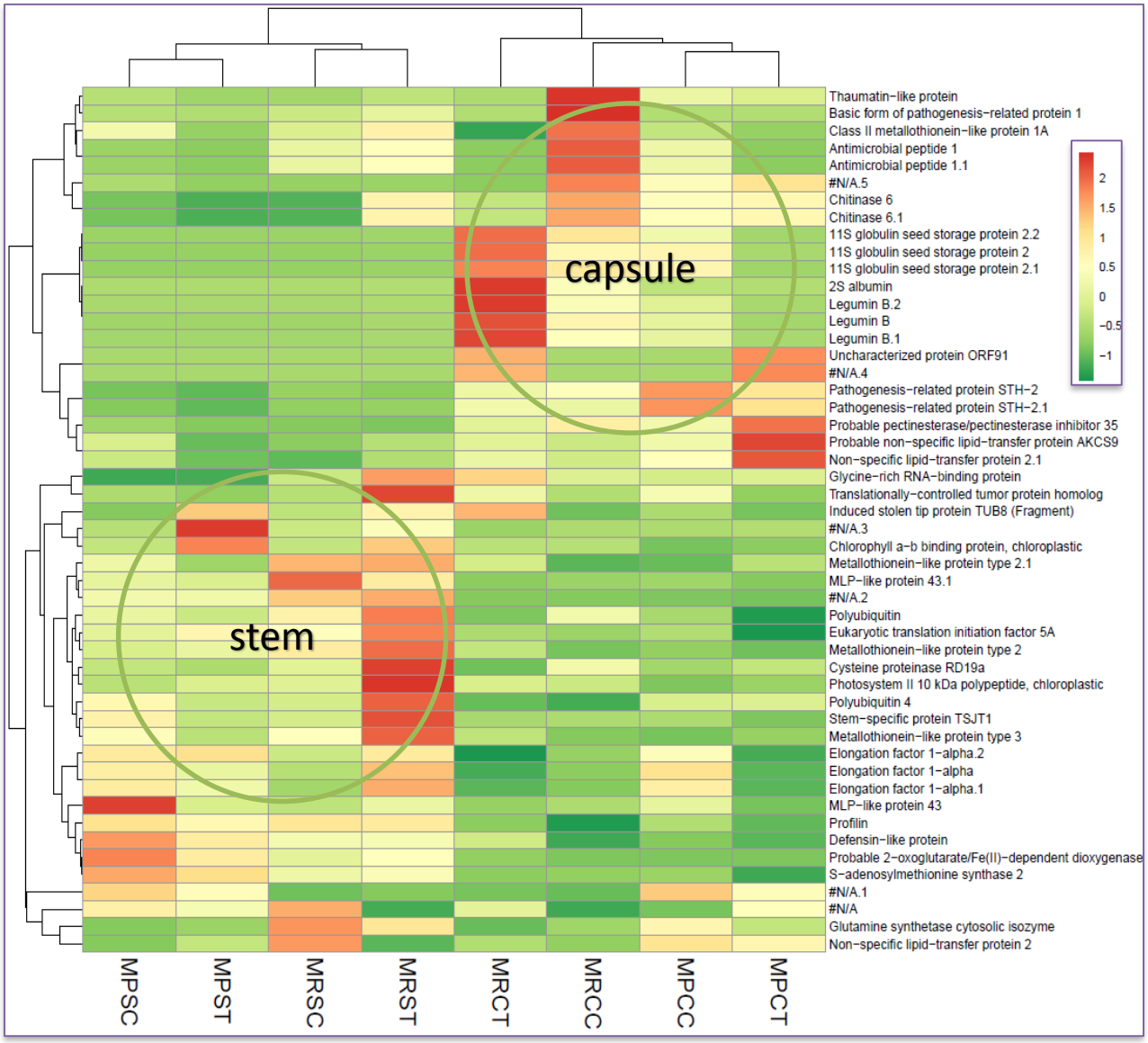
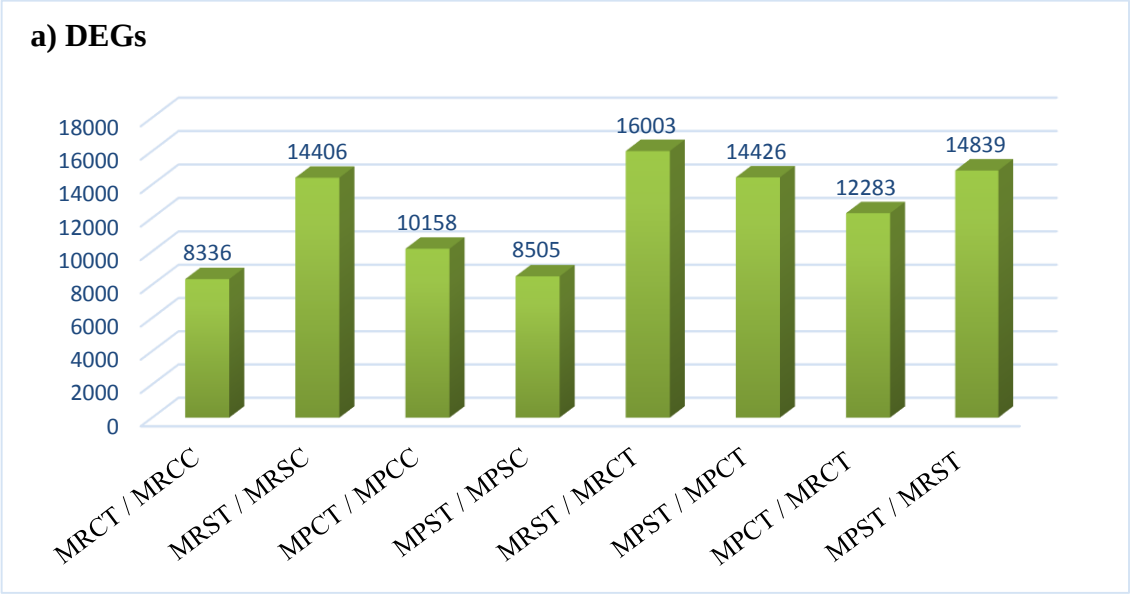



**Figure 3.35.** Heatmap of top 50-mostly expressed transcripts among libraries

### 3.6.2. DEGs in MeJA applied opium poppy cultivars and tissues

A large number of differentially expressed genes were found in MeJA treated stem and capsule tissues in morphine rich opium poppy, indicating tissue and cultivar-specific expression (Figure 3.36).



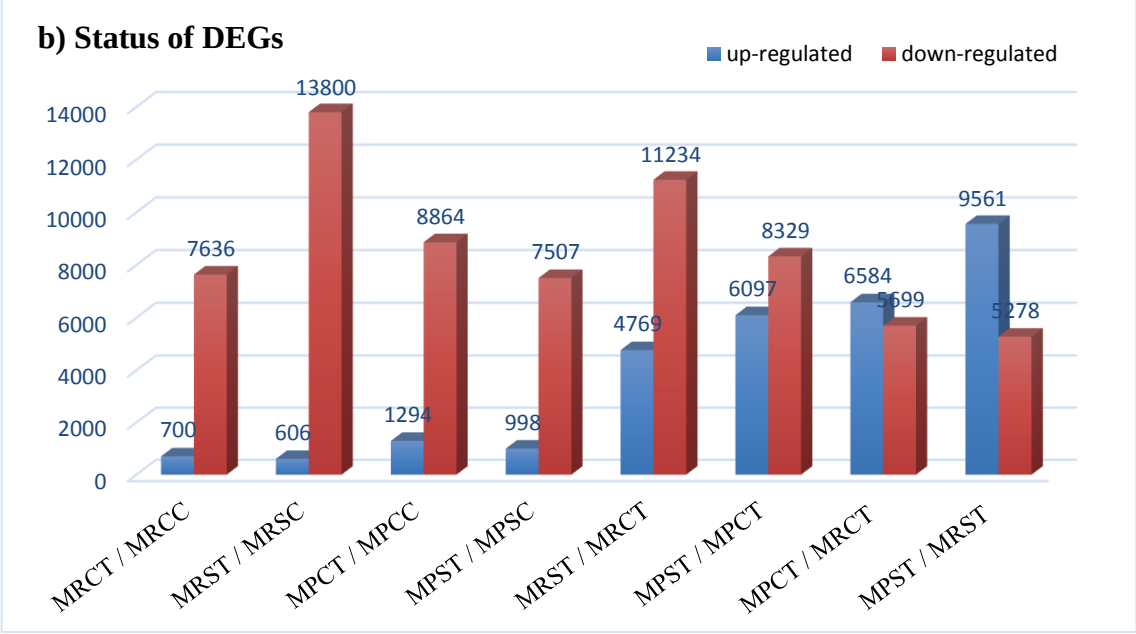




**Figure 3.36.** Statistics of differentially expressed transcripts

Top-50 differentially expressed genes in response to exogenous MeJA application were listed for all treatment/tissue/cultivar combinations (Supplementary Table 9-16). Besides, some of the remarkable genes are discussed here. (Table 3.6)

**In morphine rich capsule tissue,** in response to MeJA stress, the highest levels of differential expression are detected in transcripts annotated as Aquaporin TIP4, 2-oxoglutarate-dependent dioxygenase, F-box/LRR protein, Transcription factor bHLH113, ABC transporter G family member, and Chaperone/DNA-J proteins. Some

of them are known to be related to defense and stress mechanisms as well as morphine biosynthesis:

- Aquaporins, are composed of a highly conserved super family of intrinsic membrane proteins responsible for water transport and stress response (Vera-Estrella and Bohnert, 2011). They have been shown to express differential responses in several transcriptome studies of different plant tissues (Afzal et al., 2016). A previous research conducted in opium poppy laticifers demonstrated a significant correlation between variation in the latex water content and that of the alkaloids morphine and codeine. The study explained the alkaloid content fluctuation as the result of water transport between the laticifers and the surrounding vascular tissue through water channels and aquaporins (Itenov et al., 1999). In this study, aquaporins seem to regulate osmotic pressure in latex upon elicitor treatment. Yet, another *aquaporin (NIP2-2)* gene was found to be highly upregulated in capsule compared to stem tissue.
- 2-oxoglutarate-dependent dioxygenases (2-ODDs) are known to function in both primary and secondary plant metabolisms (Farrow and Facchini, 2014). In BIA biosynthesis pathway, norcoclaurine synthase (NCS) catalyzes the first commited step to form (S)-Norcoclaurine which is then converted to (S)-reticuline, a key branch point to all BIA alkaloids. Lee and his colleagues has reported that 2-ODD-like enzyme exhibited NCS activity (Lee and Facchini, 2010). In a DNA microarray study, transcriptomes of high-morphine and morphine-free opium poppy chemotypes were compared and 2-ODD encoding gene was identified (Hagel and Facchini, 2010). The same group has further isolated dioxygenases, thebaine 6-O-demethylase (T6ODM) and codeine O-demethylase (CODM), with a functional genomics approach and confirmed their functions in morphine biosynthesis (Hagel and Facchini, 2010). Our study also supports the role of 2-ODD in morphine biosynthesis pathway.
- F-box/LRR was previously found to be encoded by *coi1* in *A. thaliana* with a function in JA signaling pathway for pollen development and pathogen/insect response (Xie et al., 1998). F-box motifs are the receptors responsible for recalling regulatory proteins for degradation through ubiquitination, thus activates JA signaling genes (Devoto et al., 2002).

- ABC transporters were shown to be involved in berberine transport; uptake to the plasma membrane and efflux from the cytosole in *Coptus japonica* (Sakai et al., 2002). The presence of the transcripts at high levels in capsule tissue of MR cultivar may suggest a probable role in morphine transport in opium poppy.

On the other hand, transcripts coding for Formin-like protein (related to defense mechanism), Elicitor-responsive protein, Scoulerine demethylase, (S)-tetrahydroprotoberberine N-methyltransferase (TNMT) and Calmodulin were found to be highly up or downregulated in morphine poor capsule after MeJA application.

- Calmodulin is a ubiquitous $Ca^{+2}$ binding protein which modulates stress-induced interior $Ca^{+2}$ changes and functions a regulator of several proteins in plant response mechanism (Peng et al., 2014). An expression study in tomato fruit showed that *calmodulin* genes were highly responsive to MeJA and upregulated activating defense pathways (Peng et al., 2014). Another study on opium poppy cell culture showed that calmodulin levels were increased in response to elicitor treatment (Zulak et al., 2009). Ca+2 flux is important for elicitor-induced accumulation of secondary metabolites (Zhao et al., 2005). In another study preformed in opium poppy cell culture, *calmodulin* gene expression levels were suppressed 20 hours later the elicitor treatment. On the contrary, *calmodulin* is downregulated in morphine poor capsule upon MeJA treatment in our study. The transcriptional induction of capsule *calmodulin* in this cultivar might occur too rapidly for detection given the selected time point (9 h after MeJA application). Our result was similar to another expression study performed in soybean cell cultures, which were immediately downregulated to basal level after 3 hours of elicitor-induction (Heo et al., 1999).
- Transcripts corresponding to (S)-tetrahydroprotoberberine N-methyltransferase (TNMT) enzyme are found to be upregulated in both morphine poor capsule and stem tissues. (S)-tetrahydroprotoberberine N-methyltransferase is a BIA biosynthesis enzyme which converts both (S)-canadine to (S)-N-methylcanadine in the noscapine path, and (S)-Stylopine to (S)-*cis*-*N*-Methyl stylotine in the sanguinarine path. This cultivar-specific upregulation of *TNMT* is probably

result in higher level of protoberberine alkaloid biosynthesis (noscapine and sanguinarine) than morphine.

**In morphine rich stem tissue,** transcripts encoding Major latex protein (MLP) 146, defensin-like, polyubiquitine and (+)-neomenthol dehydrogenase were found to be highly induced by MeJA treatment.

- Major latex protein 146 (MLP) coding genes specifically expressed in stem tissues of both cultivars. In response to stimulus, MLPs are known to be associated with pathogen defense response (Lytle et al., 2009). The proteins are further characterized as a pathogen related Bet v 1 protein superfamily through sequence similarity (Osmark et al., 1998). Based on additional sequence comparisons Bet v 1 superfamily is expanded to include the (S)-norcoclaurine synthases and intracellular pathogenesis-related class 10 (PR-10) families (Liscombe et al., 2005; Radauer and Breiteneder, 2007). MLP induction in stem tissue proves that a high level of MeJA response is developed there.
- (+)-neomenthol dehydrogenase is a short chain dehydrogenase/reductase (SDR) family member, known to be involved in plant defense. In a previous study, (+)-neomenthol dehydrogenase was shown to have a homology with salutaridine reductase (SalR) enzyme which is specific to morphine production (Ziegler et al., 2009). In our study, the morphine biosynthesis may be induced in the stem tissue by the upregulation of this transcript in MR cultivar.

**In MP stem tissue** one of the most differentially expressed gene was *pheophytinase* which is known to function in cyclopentenone response MeJA is a cyclopentenone compound and as the stress perceived as a pathogen attack, plant start to degrade chlorophyll by *pheophytinase* as a response (Eckardt, 2009).

Additionally, a transcript coding for Q3 NRT1/PTR FAMILY protein demonstrates high level of differential expression in MR stem tissue compared to MP stem. NRT1/PTR FAMILY proteins were initially known as nitrate or di/tri-peptide transporters. However recent studies revealed that stress responsive plant hormones auxin (indole-3-acetic acid), abscisic acid (ABA), and gibberellin (GA), as well as secondary metabolites (glucosinolates) are also transported by this transporter family (Chiba et al., 2015).

Furthermore, Cytochrome P450 71A9, Transcription factor MYB39 and Indole-3-acetic acid-amido synthetase are found to be upregulated in morphine poor capsule compared to morphine rich capsule, hence corresponding transcripts are found to be highly expressed in MP cultivar.

- Cytochrome P450 71A9 is an elicitor induced protein shown to have a role in secondary metabolite biosynthetic process (Schopfer and Ebel, 1998), thus may lead to accumulation of not morphine but other alkaloids.
- The transcript encoding Indole-3-acetic acid-amido synthetase is highly upregulated in the capsule tissue of morphine poor cultivar compared to morphine rich cultivar. As GO results indicate, this gene functions in response to cyclopentenone, like *pheophytinase*, biological process which cause a higher response to MeJA explaining the drastic increase in morphine content in MP capsule (Figure 3.1b).

**Table 3.6.** FPKM values of **s**elected transcripts remarkable in DEG analysis

| Gene ID | Description | MRCC | MRCT | MRSC | MRST | MPCC | MPCT | MPSC | MPST |
|---|---|---|---|---|---|---|---|---|---|
| *CL7922.Contig2_All* | Aquaporin TIP4-1 | 0.01 | 12.28 | NA | NA | NA | NA | NA | NA |
| *Unigene49484_All* | 2-oxoglutarate-dependent dioxygenase AOP2 | 0.01 | 9.43 | NA | NA | NA | NA | NA | NA |
| *CL120.Contig5_All* | Transcription factor bHLH113 | 0.01 | 6.44 | NA | NA | NA | NA | NA | NA |
| *CL2230.Contig10_All* | ABC transporter G family member 11 | 0.01 | 6.4 | NA | NA | NA | NA | NA | NA |
| *Unigene23170_All* | F-box/LRR-repeat protein 15 | 0.01 | 6.46 | NA | NA | NA | NA | NA | NA |
| *CL2230.Contig10_All* | ABC transporter G family member 11 | 0.01 | 6.4 | NA | NA | NA | NA | NA | NA |
| *CL7285.Contig6_All* | Chaperone protein dnaJ 15 | 0.01 | 5.7 | NA | NA | NA | NA | NA | NA |
| *Unigene1201_All* | Aquaporin NIP2-2 | NA | 13.04 | NA | 0.01 | NA | NA | NA | NA |
| *Unigene30047_All* | Major latex protein 146 | NA | NA | NA | NA | NA | NA | 0.01 | 38.74 |
| *Unigene35911_All* | Major latex protein 146 | NA | NA | 0.01 | 132.8 | NA | NA | NA | NA |
| *CL3489.Contig1_All* | Defensin-like protein | NA | NA | 0.01 | 9.74 | NA | NA | NA | NA |
| *Unigene44551_All* | Polyubiquitin | NA | NA | 0.01 | 10.86 | NA | NA | NA | NA |
| *CL15179.Contig28_All* | (+)-neomenthol dehydrogenase | NA | NA | 0.01 | 8.67 | NA | NA | NA | NA |
| *CL11400.Contig2_All* | Elicitor-responsive protein 1 | NA | NA | NA | NA | 0.01 | 8.17 | NA | NA |
| *Unigene23985_All* | scoulerine O-demethylase (DIOX2) mRNA | NA | NA | NA | NA | 0.01 | 11.62 | NA | NA |
| *CL15259.Contig1_All* | Calmodulin | NA | NA | NA | NA | 40.85 | 0.01 | NA | NA |
| *Unigene16736_All* | Formin-like protein 3 | NA | NA | NA | NA | 0.01 | 9.26 | NA | NA |
| *CL1828.Contig3_All* | Pheophytinase, chloroplastic | NA | NA | NA | NA | NA | NA | 0.01 | 7.75 |
| *CL2921.Contig4_All* | Cytochrome P450 71A9 | NA | 0.01 | NA | NA | NA | 10.8 | NA | NA |
| *CL3887.Contig2_All* | Transcription factor MYB39 | NA | 0.01 | NA | NA | NA | 11.69 | NA | NA |
| *CL13183.Contig2_All* | Indole-3-acetic acid-amido synthetase GH3.6 | NA | 0.01 | NA | NA | NA | 12.19 | NA | NA |
| *CL10453.Contig4_All* | Q3Protein NRT1/ PTR FAMILY 7.1 | NA | NA | NA | 14.96 | NA | NA | NA | 0.01 |
| *CL570.Contig17_All* | Salutaridine reductase | NA | NA | NA | 0.01 | NA | NA | NA | 22.44 |
| *Unigene17049_All* | (S)-tetrahydroprotoberberine N-methyltransferase | NA | NA | NA | 0.01 | NA | NA | NA | 28.96 |
| *Unigene646_All* | (S)-tetrahydroprotoberberine N-methyltransferase 1 | NA | NA | NA | NA | 0.01 | 29.28 | NA | 0.01 |

### 3.6.3. Differential expression of morphine biosynthesis related genes

RNAseq analysis shows that morphine biosynthesis specific *SalSyn, SalAT, SalR, T6ODM, COR, CODM* transcripts are differentially expressed in MR capsule transcriptome. In particular, all of these transcripts except *CODM* are found to be upregulated in the treatment group which indicates that the morphine pathway is triggered upon MeJA application. On the other hand, *CODM* is interestingly downregulated which may result in activation of the codeinone route to morphine and cause an accumulation of codeine before morphine until it is required by the organism again. Since the data is collected at 9 hours later than the treatment, morphine biosynthesis could be reached to a maximum level before this time point and slowed down. *COR* is found to be expressed independent from MeJA treatment in different studies (Huang and Kutchan, 2000; Facchini and Park, 2003; Alcantara et al., 2005), however in our data it is again upregulated by MeJA.

In MR cultivar, transcripts involved in morphine biosynthesis are mostly upregulated in stem tissue indicating that the main morphine production takes place in stems. However, in MP cultivar it is more likely to happen in capsule tissue at this time point.

Comparison of morphine related transcripts in capsule and stem tissues of both cultivars are shown in Table 3.7 and 3.8. Figure 3.36 illustrates status of all differentially expressed BIA-related transcripts.

**Table 3.7.** Morphine biosynthesis specific genes compared at tissue level

| | **MR** | | **MP** | |
|---|---|---|---|---|
| | **Capsule** | **Stem** | **Capsule** | **Stem** |
| ***SalSyn*** | ↓ | ↑ | ↑ | - |
| ***SalR*** | ↑↓ | ↑↓ | ↑↓ | ↑↓ |
| ***SalAT*** | ↑↓ | ↑↓ | ↑↓ | ↑↓ |
| ***T6ODM*** | ↑↓ | ↑↓ | ↑ | ↓ |
| ***COR*** | ↓ | ↑ | ↑↓ | ↑↓ |
| ***CODM*** | ↓ | ↑ | ↑ | ↓ |

**Table 3.8.** Morphine biosynthesis specific genes compared at cultivar level

| | Capsule | | Stem | |
|---|---|---|---|---|
| | **MR** | **MP** | **MR** | **MP** |
| ***SalSyn*** | ↑ | ↓ | - | - |
| ***SalR*** | ↑↓ | ↑↓ | ↑↓ | ↑↓ |
| ***SalAT*** | ↑↓ | ↑↓ | ↑↓ | ↑↓ |
| ***T6ODM*** | ↑ | ↓ | ↑ | ↓ |
| ***COR*** | ↑ | ↓ | ↑↓ | ↑↓ |
| ***CODM*** | ↓ | ↑ | ↑↓ | ↑↓ |

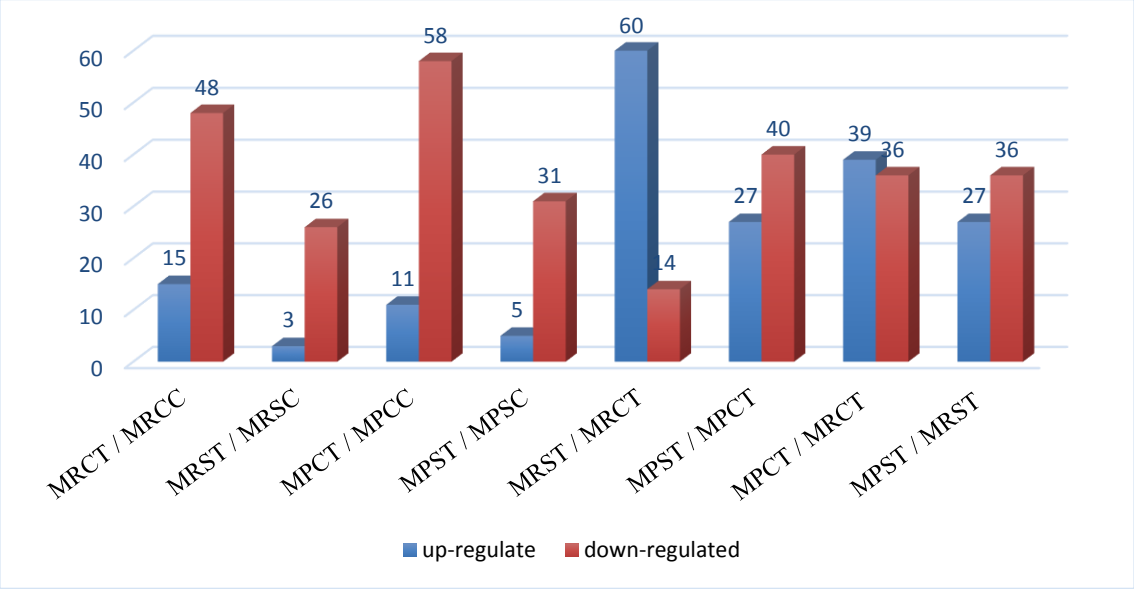



**Figure 3.37.** Statistics of BIA-related transcripts

# 4. CONCLUSION

This study is conducted on MeJA elicitor applied opium poppy capsule and stem tissues of morphine rich and poor cultivars. A genome-wide transcriptome is generated by an RNA-Seq approach using Illumina HiSeq platform. Alkaloid contents are investigated before and after the stress treatment and an integrative transcriptomic correlation has been presented. Accordingly, a significant increase in morphine content of capsule tissue indicated a successful induction by MeJA. On the contrary, it was decreased in stem tissues, together with the transcriptomic data, suggesting a possible production and transfer of morphine throughout the capsule tissue to be accumulated there.

Following *de novo* assembly, sequence-based transcriptome analyses has been performed. As gene annotation, GO and KEGG pathway analysis revealed, a large number of genes encoding secondary metabolite biosynthesis enzymes and stress response proteins have been identified. Mechanisms altered by MeJA in different tissues of opium poppy cultivars are evaluated by *in silico* comparative approach.

MeJA has been shown to activate stress response mechanisms thus increasing morphine production in opium poppy. In this sense, differentially expressed genes are identified in all comparison libraries.

The novelty of this thesis lies on that we produce the largest opium poppy transcriptome data (45x coverage) by next generation sequencing up to date. Moreover, this is the first study that compares morphine producing and accumulating tissues; the capsule and the stem. In addition to that, cultivars rich and poor in producing morphine are compared at genome wide transcriptome level.

As a further study, genes found to be highly expressed could be transferred to local opium poppy varieties and employing metabolic engineering, plants with higher morphine producing capacity could be achieved either by transgenic or non-transgenic approaches. In short future, the value of the data obtained in this thesis will be highly increased after the genome sequence is launched. With the use of functional genomics techniques, molecular background of morphine and other BIA biosynthesis mechanisms, along with their regulation and functions of the related genes could be investigated.

Additionally, overexpression and silencing of prominent SSRs, transcription factors (TFs), long noncoding RNAs (lncRNAs), microRNAs, and endogenous target mimic (eTMs) modules in opium poppy a more comprehensive frame of morphine mechanism will be revealed.

# REFERENCES


Afzal, Z., Howton, T., Sun, Y. and Mukhtar, M. S. (2016) The roles of aquaporins in plant stress responses. Journal of Developmental Biology, 4(1), 9.

Agarwal, P., Pathak, S., Lakhwani, D., Gupta, P., Asif, M. H. and Trivedi, P. K. (2015) Comparative analysis of transcription factor gene families from Papaver somniferum: identification of regulatory factors involved in benzylisoquinoline alkaloid biosynthesis. Protoplasma, 1-15.

Alcantara, J., Bird, D. A., Franceschi, V. R. and Facchini, P. J. (2005) Sanguinarine biosynthesis is associated with the endoplasmic reticulum in cultured opium poppy cells after elicitor treatment. Plant Physiology, 138(1), 173-183.

Allen, R. S., Miller, J. A., Chitty, J. A., Fist, A. J., Gerlach, W. L. and Larkin, P. J. (2008) Metabolic engineering of morphinan alkaloids by over-expression and RNAi suppression of salutaridinol 7-O-acetyltransferase in opium poppy. Plant Biotechnology Journal, 6(1), 22-30.

Allen, R. S., Millgate, A. G., Chitty, J. A., Thisleton, J., Miller, J. A., Fist, A. J., Gerlach, W. L. and Larkin, P. J. (2004) RNAi-mediated replacement of morphine with the nonnarcotic alkaloid reticuline in opium poppy. Nature Biotechnology, 22(12), 1559-1566.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403-410.

Amann, M., Nagakura, N. and Zenk, M. H. (1988) Purification and properties of (S)-tetrahydroprotoberberine oxidase from suspension-cultured cells of Berberis wilsoniae. European Journal of Biochemistry, 175(1), 17-25.

Amann, M., Wanner, G. and Zenk, M. H. (1986) Intracellular compartmentation of two enzymes of berberine biosynthesis in plant cell cultures. Planta, 167(3), 310-320.

Audic, S. and Claverie, J. M. (1997) The significance of digital gene expression profiles. Genome Research, 7(10), 986-995.

Badri, D. V., Loyola-Vargas, V. M., Du, J., Stermitz, F. R., Broeckling, C. D., Iglesias-Andreu, L. and Vivanco, J. M. (2008) Transcriptome analysis of Arabidopsis roots treated with signaling compounds: a focus on signal transduction, metabolic regulation and secretion. New Phytologist, 179(1), 209-223.

Baginsky, S., Hennig, L., Zimmermann, P. and Gruissem, W. (2010) Gene expression analysis, proteomics, and network discovery. Plant Physiology, 152(2), 402-410.

Barken, I., Geller, J. and Rogosnitzky, M. (2008) Noscapine inhibits human prostate cancer progression and metastasis in a mouse model. Anticancer Research, 28(6A), 3701-3704.

Battersby, A. R., Staunton, J., Wiltshire, H. R., Francis, R. J. and Southgate, R. (1975) Biosynthesis. Part XXII. The origin of chelidonine and of other alkaloids derived from the tetrahydroprotoberberine skeleton. Journal of the Chemical Society, Perkin Transactions 1(12), 1147-1156.

Bauer, W. and Zenk, M. (1989) Formation of both methylenedioxy groups in the alkaloid (S)-stylopine is catalyzed by cytochrome P-450 enzymes. Tetrahedron Letters, 30(39), 5257-5260.

Bauer, W. and Zenk, M. H. (1991) Two methylenedioxy bridge forming cytochrome P-450 dependent enzymes are involved in (S)-stylopine biosynthesis. Phytochemistry, 30(9), 2953-2961.

Beaudoin, G. A. and Facchini, P. J. (2014) Benzylisoquinoline alkaloid biosynthesis in opium poppy. Planta, 240(1), 19-32.

Bentley, D. R. (2006) Whole-genome re-sequencing. Current opinion in genetics & development, 16(6), 545-552.

Bird, D. A. and Facchini, P. J. (2001) Berberine bridge enzyme, a key branch-point enzyme in benzylisoquinoline alkaloid biosynthesis, contains a vacuolar sorting determinant. Planta, 213(6), 888-897.

Bird, D. A., Franceschi, V. R. and Facchini, P. J. (2003) A tale of three cell types: alkaloid biosynthesis is localized to sieve elements in opium poppy. The Plant Cell, 15(11), 2626-2635.

Bleidorn, C. (2016) Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. Systematics and Biodiversity, 14(1), 1-8.

Blume, B., Nurnberger, T., Nass, N. and Scheel, D. (2000) Receptor-mediated increase in cytoplasmic free calcium required for activation of pathogen defense in parsley. Plant Cell, 12(8), 1425-1440.

Boke, H., Ozhuner, E., Turktas, M., Parmaksiz, I., Ozcan, S. and Unver, T. (2015) Regulation of the alkaloid biosynthesis by miRNA in opium poppy. Plant Biotechnology Journal, 13(3), 409-420.

Bradford, J. R., Hey, Y., Yates, T., Li, Y., Pepper, S. D. and Miller, C. J. (2010) A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. BMC Genomics, 11(1), 282.

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A. and Huang, X. (2008) The potential and challenges of nanopore sequencing. Nature Biotechnology, 26(10), 1146-1153.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., DuBridge, R. B., Kirchner, J., Fearon, K., Mao, J. and Corcoran, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature Biotechnology, 18(6), 630-634.

Brochmann-Hanssen, E. (1984) A second pathway for the terminal steps in the biosynthesis of morphine. Planta Medica, 50(4), 343-345.

Brownstein, M. J. (1993) A brief history of opiates, opioid peptides, and opioid receptors. Proceedings of the National Academy of Sciences, 90(12), 5391-5393.

Bryant, R. (1988) The manufacture of medicinal alkaloids from the opium poppy-a review of a traditional biotechnology. Chemistry and Industry, 5, 146-153.

Burbulis, I. E. and Winkel-Shirley, B. (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proceedings of the National Academy of Sciences, 96(22), 12929-12934.

Celik, I., Gultekin, V., Allmer, J., Doganlar, S. and Frary, A. (2014) Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing. Molecular Breeding, 34(2), 323-334.

Chen, X. and Facchini, P. J. (2014) Short-chain dehydrogenase/reductase catalyzing the final step of noscapine biosynthesis is localized to laticifers in opium poppy. The Plant Journal, 77(2), 173-184.

Cheong, J. J. and Choi, Y. D. (2003) Methyl jasmonate as a vital substance in plants. Trends in Genetics, 19(7), 409-413.

Chiba, Y., Shimizu, T., Miyakawa, S., Kanno, Y., Koshiba, T., Kamiya, Y. and Seo, M. (2015) Identification of Arabidopsis thaliana NRT1/PTR FAMILY (NPF) proteins capable of transporting plant hormones. Journal of Plant Research, 128(4), 679-686.

Choi, K.-B., Morishige, T., Shitan, N., Yazaki, K. and Sato, F. (2002) Molecular cloning and characterization of coclaurinen-methyltransferase from cultured cells of Coptis japonica. Journal of Biological Chemistry, 277(1), 830-835.

Cohen, A., Najarian, D., Paulus, A., Guttman, A., Smith, J. A. and Karger, B. (1988) Rapid separation and purification of oligonucleotides by high-performance capillary gel electrophoresis. Proceedings of the National Academy of Sciences, 85(24), 9660-9663.

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

Costa, V., Angelini, C., De Feis, I. and Ciccodicola, A. (2010) Uncovering the complexity of transcriptomes with RNA-Seq. Biomed Research International, 2010.

Dang, T.-T. T. and Facchini, P. J. (2012) Characterization of three O-methyltransferases involved in noscapine biosynthesis in opium poppy. Plant Physiology, 159(2), 618-631.

Dang, T.-T. T. and Facchini, P. J. (2014) Cloning and characterization of canadine synthase involved in noscapine biosynthesis in opium poppy. FEBS Letters, 588(1), 198-204.

Dang, T.-T. T. and Facchini, P. J. (2014) CYP82Y1 is n-methylcanadine 1-hydroxylase, a key noscapine biosynthetic enzyme in opium poppy. Journal of Biological Chemistry, 289(4), 2013-2026.

Dang, T. T. T., Onoyovwi, A., Farrow, S. C. and Facchini, P. J. (2012) Biochemical genomics for gene discovery in benzylisoquinoline alkaloid biosynthesis in opium poppy and related species. Methods in enzymology, 515, 231.

De-Eknamkul, W. and Zenk, M. H. (1992) Purification and properties of 1, 2-dehydroreticuline reductase from Papaver somniferum seedlings. Phytochemistry, 31(3), 813-821.

de Souza Cândido, E., Pinto, M. F. S., Pelegrini, P. B., Lima, T. B., Silva, O. N., Pogue, R., Grossi-de-Sá, M. F. and Franco, O. L. (2011) Plant storage proteins with antimicrobial activity: novel insights into plant defense mechanisms. The FASEB Journal.

Decker, G., Wanner, G., Zenk, M. H. and Lottspeich, F. (2000) Characterization of proteins in latex of the opium poppy (Papaver somniferum) using two-dimensional gel electrophoresis and microsequencing. Electrophoresis, 21(16), 3500-3516.

Desgagné-Penix, I., Farrow, S. C., Cram, D., Nowak, J. and Facchini, P. J. (2012) Integration of deep transcript and targeted metabolite profiles for eight cultivars of opium poppy. Plant molecular biology, 79(3), 295-313.

Desgagné-Penix, I., Khan, M. F., Schriemer, D. C., Cram, D., Nowak, J. and Facchini, P. J. (2010) Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. BMC Plant Biology, 10(1), 252.

Devoto, A., Nieto-Rostro, M., Xie, D., Ellis, C., Harmston, R., Patrick, E., Davis, J., Sherratt, L., Coleman, M. and Turner, J. G. (2002) COI1 links jasmonate signalling and fertility to the SCF ubiquitin-ligase complex in Arabidopsis. Plant Journal, 32(4), 457-466.

Dittrich, H. and Kutchan, T. M. (1991) Molecular cloning, expression, and induction of berberine bridge enzyme, an enzyme essential to the formation of benzophenanthridine alkaloids in the response of plants to pathogenic attack. Proceedings of the National Academy of Sciences, 88(22), 9969-9973.

Ebrahim, S., Usha, K. and Singh, B. (2011) Pathogenesis related (PR) proteins in plant defense mechanism. Science Against Microbial Pathogens: Communicating Current Research and Technological Advances, 2, 1043-1054.

Eckardt, N. A. (2009) A new chlorophyll degradation pathway. The Plant Cell, 21(3), 700.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P. and Bettman, B. (2009) Real-time DNA sequencing from single polymerase molecules. Science, 323(5910), 133-138.

Eom, E., Lee, J., Park, H., Byun, Y., Ha-Lee, Y. and Lee, D. (2006) Comparison between SAGE and cDNA microarray for quantitative accuracy in transcript profiling analyses. Journal of Plant Biology, 49(6), 498-506.

Facchini, P. J. (2006) Regulation of alkaloid biosynthesis in plants. The Alkaloids: Chemistry and Biology, 63, 1-44.

Facchini, P. J. and De Luca, V. (1994) Differential and tissue-specific expression of a gene family for tyrosine/dopa decarboxylase in opium poppy. Journal of Biological Chemistry, 269(43), 26684-26690.

Facchini, P. J. and De Luca, V. (1995) Phloem-specific expression of tyrosine/dopa decarboxylase genes and the biosynthesis of isoquinoline alkaloids in opium poppy. Plant Cell, 7(11), 1811-1821.

Facchini, P. J. and De Luca, V. (2008) Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. Plant Journal, 54(4), 763-784.

Facchini, P. J., Hagel, J. M., Liscombe, D. K., Loukanina, N., MacLeod, B. P., Samanani, N. and Zulak, K. G. (2007) Opium poppy: blueprint for an alkaloid factory. Phytochemistry Reviews, 6(1), 97-124.

Facchini, P. J. and Park, S.-U. (2003) Developmental and inducible accumulation of gene transcripts involved in alkaloid biosynthesis in opium poppy. Phytochemistry, 64(1), 177-186.

Fairbairn, J., Djote, M. and Paterson, A. (1968) The alkaloids of Papaver somniferum L.—VII: Biosynthetic activity of the isolated latex. Phytochemistry, 7(12), 2111-2116.

Fairbairn, J. and Wassel, G. (1964) The alkaloids of Papaver somniferum L.—III.: Biosynthesis in the isolated latex. Phytochemistry, 3(5), 583-585.

Farrow, S. C. and Facchini, P. J. (2013) Dioxygenases catalyze O-demethylation and O, O-demethylenation with widespread roles in benzylisoquinoline alkaloid metabolism in opium poppy. Journal of Biological Chemistry, 288(40), 28997-29012.

Farrow, S. C. and Facchini, P. J. (2014) Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. Frontiers in Plant Science, 5, 524.

Farrow, S. C., Hagel, J. M., Beaudoin, G. A., Burns, D. C. and Facchini, P. J. (2015) Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. Nature Chemical Biology, 11(9), 728-732.

Fisinger, U., Grobe, N. and Zenk, M. H. (2007) Thebaine synthase: a new enzyme in the morphine pathway in Papaver somniferum. Natural Product Communications, 2(3), 249-253.

Frick, S., Chitty, J. A., Kramell, R., Schmidt, J., Allen, R. S., Larkin, P. J. and Kutchan, T. M. (2004) Transformation of opium poppy (Papaver somniferum L.) with antisense berberine bridge enzyme gene (anti-bbe) via somatic embryogenesis results in an altered ratio of alkaloids in latex but not in roots. Transgenic research, 13(6), 607-613.

Frick, S., Kramell, R. and Kutchan, T. M. (2007) Metabolic engineering with a morphine biosynthetic P450 in opium poppy surpasses breeding. Metabolic engineering, 9(2), 169-176.

Frick, S., Kramell, R., Schmidt, J., Fist, A. J. and Kutchan, T. M. (2005) Comparative qualitative and quantitative determination of alkaloids in narcotic and condiment papaver s omniferum cultivars. Journal of natural products, 68(5), 666-673.

Fu, X., Fu, N., Guo, S., Yan, Z., Xu, Y., Hu, H., Menzel, C., Chen, W., Li, Y. and Zeng, R. (2009) Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics, 10(1), 161.

Gerardy, R. and Zenk, M. H. (1992) Formation of salutaridine from (R)-reticuline by a membrane-bound cytochrome P-450 enzyme from Papaver somniferum. Phytochemistry, 32(1), 79-86.

Gerardy, R. and Zenk, M. H. (1993) Purification and characterization of salutaridine: NADPH 7-oxidoreductase from Papaver somniferum. Phytochemistry, 34(1), 125-132.

Gesell, A., Rolf, M., Ziegler, J., Chávez, M. L. D., Huang, F.-C. and Kutchan, T. M. (2009) CYP719B1 is salutaridine synthase, the CC phenol-coupling enzyme of morphine biosynthesis in opium poppy. Journal of Biological Chemistry, 284(36), 24432-24442.

Giorgi, F. M., Del Fabbro, C. and Licausi, F. (2013) Comparative study of RNA-seq- and Microarray-derived coexpression networks in Arabidopsis thaliana. Bioinformatics, 29(6), 717-724.

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

Grothe, T., Lenz, R. and Kutchan, T. M. (2001) Molecular characterization of the salutaridinol 7-O-acetyltransferase involved in morphine biosynthesis in opium poppy Papaver somniferum. Journal of Biological Chemistry, 276(33), 30717-30723.

Gurkok, T. (2013) Morfin içeriği indüklenmiş haşhaş (Papaver somniferum L.) bitkisinde genom düzeyinde transcriptome analizi. Gaziosmanpasa University. Graduate School of Natural and Applied Sciences, Tokat, Turkey. PhD Thesis, 340.

Gurkok, T., Parmaksiz, I., Boztepe, G. and Kaymak, E. (2010) Alkaloid biosynthesis mechanism in opium poppy (Papaver somniferum L.). Electronic Journal of Biotechnology, 1(2), 31-45.

Gurkok, T., Turktas, M., Parmaksiz, I. and Unver, T. (2015) Transcriptome profiling of alkaloid biosynthesis in elicitor induced opium poppy. Plant Molecular Biology Reporter, 33(3), 673-688.

Hagel, J. M., Beaudoin, G. A., Fossati, E., Ekins, A., Martin, V. J. and Facchini, P. J. (2012) Characterization of a flavoprotein oxidase from opium poppy catalyzing the final steps in sanguinarine and papaverine biosynthesis. Journal of Biological Chemistry, 287(51), 42972-42983.

Hagel, J. M. and Facchini, P. J. (2008) Plant metabolomics: analytical platforms and integration with functional genomics. Phytochemistry Reviews, 7(3), 479-497.

Hagel, J. M. and Facchini, P. J. (2010) Biochemistry and occurrence of o-demethylation in plant metabolism. Frontiers in Physiology, 1, 14.

Hagel, J. M. and Facchini, P. J. (2010) Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. Nature Chemical Biology, 6(4), 273-275.

Hagel, J. M. and Facchini, P. J. (2012) Subcellular localization of sanguinarine biosynthetic enzymes in cultured opium poppy cells. In Vitro Cellular & Developmental Biology-Plant, 48(2), 233-240.

Hagel, J. M. and Facchini, P. J. (2013) Benzylisoquinoline alkaloid metabolism–a century of discovery and a brave new world. Plant and Cell Physiology, pct020.

Hagel, J. M., Weljie, A. M., Vogel, H. J. and Facchini, P. J. (2008) Quantitative 1H nuclear magnetic resonance metabolite profiling as a functional genomics platform to investigate alkaloid biosynthesis in opium poppy. Plant Physiology, 147(4), 1805-1821.

Hamilton, G. and Baskett, T. (2000) In the arms of morpheus: the development of morphine for postoperative pain relief. Canadian Journal of Anesthesia, 47(4), 367-374.

Han, X., Lamshöft, M., Grobe, N., Ren, X., Fist, A. J., Kutchan, T. M., Spiteller, M. and Zenk, M. H. (2010) The biosynthesis of papaverine proceeds via (S)-reticuline. Phytochemistry, 71(11), 1305-1312.

Hao, X., Shi, M., Cui, L., Xu, C., Zhang, Y. and Kai, G. (2015) Effects of methyl jasmonate and salicylic acid on tanshinone production and biosynthetic gene expression in transgenic Salvia miltiorrhiza hairy roots. Biotechnology and Applied Biochemistry, 62(1), 24-31.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J. E., Snesrud, E., Lee, N. and Quackenbush, J. (2000) A concise guide to cDNA microarray analysis. Biotechniques, 29(3), 548-550, 552-544, 556 passim.

Heo, W. D., Lee, S. H., Kim, M. C., Kim, J. C., Chung, W. S., Chun, H. J., Lee, K. J., Park, C. Y., Park, H. C., Choi, J. Y. and Cho, M. J. (1999) Involvement of specific calmodulin isoforms in salicylic acid-independent activation of plant disease resistance responses. Proceedings of the National Academy of Sciences U S A, 96(2), 766-771.

Hileman, L. C., Drea, S., Martino, G., Litt, A. and Irish, V. F. (2005) Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species Papaver somniferum (opium poppy). The Plant Journal, 44(2), 334-341.

Hirata, K., Poeaknapo, C., Schmidt, J. and Zenk, M. H. (2004) 1, 2-Dehydroreticuline synthase, the branch point enzyme opening the morphinan biosynthetic pathway. Phytochemistry, 65(8), 1039-1046.

Holkova, I., Bezakova, L., Bilka, F., Balazova, A., Vanko, M. and Blanarikova, V. (2010) Involvement of lipoxygenase in elicitor-stimulated sanguinarine accumulation in Papaver somniferum suspension cultures. Plant Physiology and Biochemistry, 48(10-11), 887-892.

Huang, F.-C. and Kutchan, T. M. (2000) Distribution of morphinan and benzo [c] phenanthridine alkaloid gene transcript accumulation in Papaver somniferum. Phytochemistry, 53(5), 555-564.

Huang, Y., Niu, B., Gao, Y., Fu, L. and Li, W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics, 26(5), 680-682.

Hughes, E. H. and Shanks, J. V. (2002) Metabolic engineering of plants for alkaloid production. Metabolic Engineering, 4(1), 41-48.

Hutchison, C. A. (2007) DNA sequencing: bench to bedside and beyond. Nucleic Acids Research, 35(18), 6227-6237.

Ikezawa, N., Iwasa, K. and Sato, F. (2007) Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylopine biosynthesis in Eschscholzia californica. FEBS Journal, 274(4), 1019-1035.

Inal, B. (2015) Haşhaş (Papaver somniferum L.)'da Tebain Üretim Mekanizmasının Yeni Nesil Dizileme Sistemi ile Transkriptom Düzeyinde İncelenmesi. Ankara University. Biotechnology Institute, Ankara, Turkey. PhD Thesis, 169.

Iseli, C., Jongeneel, C. V. and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. ISMB.

Itenov, K., Mølgaard, P. and Nyman, U. (1999) Diurnal fluctuations of the alkaloid concentration in latex of poppy Papaver somniferum is due to day–night fluctuations of the latex water content. Phytochemistry, 52(7), 1229-1234.

Jain, M. (2012) Next-generation sequencing technologies for gene expression profiling in plants. Briefings in Functional Genomics, 11(1), 63-70.

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

Kantar, M., Unver, T. and Budak, H. (2010) Regulation of barley miRNAs upon dehydration stress correlated with target gene expression. Functional and Integrative Genomics, 10(4), 493-507.

Kawano, N., Kiuchi, F., Kawahara, N. and Yoshimatsu, K. (2012) Genetic and phenotypic analyses of a Papaver somniferum T-DNA insertional mutant with altered alkaloid composition. Pharmaceuticals, 5(2), 133-154.

Kempe, K., Higashi, Y., Frick, S., Sabarna, K. and Kutchan, T. M. (2009) RNAi suppression of the morphine biosynthetic gene salAT and evidence of association of pathway enzymes. Phytochemistry, 70(5), 579-589.

Koboldt, D. C., Larson, D. E., Chen, K., Ding, L. and Wilson, R. K. (2012). Massively parallel sequencing approaches for characterization of structural variation. Genomic Structural Variants, Springer**,** 369-384.

Kogenaru, S., Yan, Q., Guo, Y. and Wang, N. (2012) RNA-seq and microarray complement each other in transcriptome profiling. BMC Genomics, 13(1), 629.

Kurtoglu, K. Y., Kantar, M. and Budak, H. (2014) New wheat microRNA using whole-genome sequence. Functional Integrative Genomics. 14, 363-379.

Kutchan, T. M., Frick, S. and Weid, M. (2008) Engineering plant alkaloid biosynthetic pathways: progress and prospects. Advances in Plant Biochemistry and Molecular Biology, 1, 283-310.

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

Larkin, P. J., Miller, J. A., Allen, R. S., Chitty, J. A., Gerlach, W. L., Frick, S., Kutchan, T. M. and Fist, A. J. (2007) Increasing morphinan alkaloid production by over-expressing codeinone reductase in transgenic Papaver somniferum. Plant Biotechnology Journal, 5(1), 26-37.

Lee, E.-J. and Facchini, P. (2010) Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. The Plant Cell, 22(10), 3489-3503.

Lee, E.-J. and Facchini, P. J. (2011) Tyrosine aminotransferase contributes to benzylisoquinoline alkaloid biosynthesis in opium poppy. Plant Physiology, 157(3), 1067-1078.

Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., McCombie, W. R. and Schatz, M. (2016) Third-generation sequencing and the future of genomics. bioRxiv: p048603

Lee, S., Park, Y., Han, E., Choi, H., Chung, H., Oh, S. M. and Chung, K. H. (2011) Thebaine in hair as a marker for chronic use of illegal opium poppy substances. Forensic Science International, 204(1), 115-118.

Lenz, R. and Zenk, M. H. (1995) Acetyl coenzyme A:salutaridinol-7-O-acetyltransferase from papaver somniferum plant cell cultures. The enzyme catalyzing the formation of thebaine in morphine biosynthesis. The Journal of Biological Chemistry, 270(52), 31091-31096.

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

Liscombe, D. K. and Facchini, P. J. (2007) Molecular cloning and characterization of tetrahydroprotoberberine cis-N-methyltransferase, an enzyme involved in alkaloid biosynthesis in opium poppy. The Journal of Biological Chemistry, 282(20), 14741-14751.

Liscombe, D. K., MacLeod, B. P., Loukanina, N., Nandi, O. I. and Facchini, P. J. (2005) Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. Phytochemistry, 66(11), 1374-1393.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012) Comparison of next-generation sequencing systems. BioMed Research International, vol. 2012, Article ID 251364

Lorenzo, O., Piqueras, R., Sanchez-Serrano, J. J. and Solano, R. (2003) ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. Plant Cell, 15(1), 165-178.

Love, M. I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 1-21.

Lytle, B. L., Song, J., de la Cruz, N. B., Peterson, F. C., Johnson, K. A., Bingman, C. A., Phillips, G. N., Jr. and Volkman, B. F. (2009) Structures of two Arabidopsis thaliana major latex proteins represent novel helix-grip folds. Proteins, 76(1), 237-243.

Manivannan, A., Soundararajan, P., Park, Y. G. and Jeong, B. R. (2016) Chemical elicitor-induced modulation of antioxidant metabolism and enhancement of secondary metabolite accumulation in cell suspension cultures of Scrophularia kakudensis Franch. International Journal of Molecular Sciences, 17(3), 399.

Mardis, E. R. (2008) Next-generation DNA sequencing methods. Annual Review of Genomics Human Genetics, 9, 387-402.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. and Gilad, Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Research, 18(9), 1509-1517.

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

Memelink, J., Kijne, J. W., van der Heijden, R. and Verpoorte, R. (2001) Genetic modification of plant secondary metabolite pathways using transcriptional regulators. Advances in Biochemical Engineering/Biotechnology, 72, 103-125.

Millgate, A. G., Pogson, B. J., Wilson, I. W., Kutchan, T. M., Zenk, M. H., Gerlach, W. L., Fist, A. J. and Larkin, P. J. (2004) Analgesia: morphine-pathway block in top1 poppies. Nature, 431(7007), 413-414.

Mishra, S., Triptahi, V., Singh, S., Phukan, U. J., Gupta, M., Shanker, K. and Shukla, R. K. (2013) Wound induced tanscriptional regulation of benzylisoquinoline pathway and characterization of wound inducible PsWRKY transcription factor from Papaver somniferum. PLoS One, 8(1), e52784.

Misra, R. C., Sandeep, M. K., Kumar, S. and Ghosh, S. (2016) A thaumatin-like protein of Ocimum basilicum confers tolerance to fungal pathogen and abiotic stress in transgenic Arabidopsis. Scientific Reports, 6.

Morishige, T., Dubouzet, E., Choi, K. B., Yazaki, K. and Sato, F. (2002) Molecular cloning of columbamine O-methyltransferase from cultured Coptis japonica cells. European Journal of Biochemistry, 269(22), 5659-5667.

Morishige, T., Tsujita, T., Yamada, Y. and Sato, F. (2000) Molecular characterization of thes-adenosyl-l-methionine: 3′-hydroxy-n-methylcoclaurine 4′-o-methyltransferase involved in isoquinoline alkaloid biosynthesis in Coptis japonica. Journal of Biological Chemistry, 275(30), 23398-23405.

Morozova, O. and Marra, M. A. (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics, 92(5), 255-264.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods, 5(7), 621-628.

Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G. and Stahl, F. (2013) Transcriptome analysis using next-generation sequencing. Current Opinion in Biotechnology, 24(1), 22-30.

Myllykangas, S., Buenrostro, J. and Ji, H. P. (2012). Overview of sequencing technology platforms. Bioinformatics for high throughput sequencing, Springer**,** 11-25.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, 320(5881), 1344-1349.

Oliveros, J. (2014) VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.

Onoyovwe, A., Hagel, J. M., Chen, X., Khan, M. F., Schriemer, D. C. and Facchini, P. J. (2013) Morphine biosynthesis in opium poppy involves two cell types: sieve elements and laticifers. The Plant Cell, 25(10), 4110-4122.

Osmark, P., Boyle, B. and Brisson, N. (1998) Sequential and structural homology between intracellular pathogenesis-related proteins and a group of latex proteins. Plant Molecular Biology, 38(6), 1243-1246.

Ounaroon, A., Decker, G., Schmidt, J., Lottspeich, F. and Kutchan, T. M. (2003) (R, S)-Reticuline 7-O-methyltransferase and (R, S)-norcoclaurine 6-O-methyltransferase of Papaver somniferum–cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. The Plant Journal, 36(6), 808-819.

Panicot, M., Minguet, E. G., Ferrando, A., Alcázar, R., Blázquez, M. A., Carbonell, J., Altabella, T., Koncz, C. and Tiburcio, A. F. (2002) A polyamine metabolon involving aminopropyl transferase complexes in Arabidopsis. The Plant Cell, 14(10), 2539-2551.

Pathak, S., Lakhwani, D., Gupta, P., Mishra, B. K., Shukla, S., Asif, M. H. and Trivedi, P. K. (2013) Comparative transcriptome analysis using high papaverine mutant of Papaver somniferum reveals pathway and uncharacterized steps of papaverine biosynthesis. PloS one 8.5 (2013): e65622.

Pauli, H. H. and Kutchan, T. M. (1998) Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3′-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. The Plant Journal, 13(6), 793-801.

Pauwels, L., Morreel, K., De Witte, E., Lammertyn, F., Van Montagu, M., Boerjan, W., Inzé, D. and Goossens, A. (2008) Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured Arabidopsis cells. Proceedings of the National Academy of Sciences, 105(4), 1380-1385.

Peng, H., Yang, T. and Ii, W. M. (2014) Calmodulin Gene Expression in Response to Mechanical Wounding and Botrytis cinerea Infection in Tomato Fruit. Plants (Basel), 3(3), 427-441.

Pienkny, S., Brandt, W., Schmidt, J., Kramell, R. and Ziegler, J. (2009) Functional characterization of a novel benzylisoquinoline O-methyltransferase suggests its involvement in papaverine biosynthesis in opium poppy (Papaver somniferum L). The Plant Journal, 60(1), 56-67.

Pilatzke-Wunderlich, I. and Nessler, C. L. (2001) Expression and activity of cell-wall-degrading enzymes in the latex of opium poppy, Papaver somniferum L. Plant molecular biology, 45(5), 567-576.

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science, 238(4825), 336-341.

Radauer, C. and Breiteneder, H. (2007) Evolutionary biology of plant food allergens. Journal of Allergy and Clinical Immunology, 120(3), 518-525.

Reymond, P. and Farmer, E. E. (1998) Jasmonate and salicylate as global signals for defense gene expression. Current Opinion Plant Biology, 1(5), 404-411.

Roberts, M. F., McCarthy, D., Kutchan, T. M. and Coscia, C. J. (1983) Localization of enzymes and alkaloidal metabolites in Papaver latex. Archives of Biochemistry and Biophysics, 222(2), 599-609.

Ruffer, M. and Zenk, M. H. (1987) Enzymatic formation of protopines by a microsomal cytochrome P-450 system of Corydalisvaginans. Tetrahedron Letters, 28(44), 5307-5310.

Ruiz-May, E., Galaz-Avalos, R. M. and Loyola-Vargas, V. M. (2009) Differential secretion and accumulation of terpene indole alkaloids in hairy roots of Catharanthus roseus treated with methyl jasmonate. Molecular Biotechnology, 41(3), 278-285.

Sakai, K., Shitan, N., Sato, F., Ueda, K. and Yazaki, K. (2002) Characterization of berberine transport into Coptis japonica cells and the involvement of ABC protein. Journal of Experimental Botany, 53(376), 1879-1886.

Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., Delcher, A. L., Roberts, M., Marçais, G., Pop, M. and Yorke, J. A. (2012) GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome research, 22(3), 557-567.

Samanani, N., Alcantara, J., Bourgault, R., Zulak, K. G. and Facchini, P. J. (2006) The role of phloem sieve elements and laticifers in the biosynthesis and accumulation of alkaloids in opium poppy. Plant Journal, 47(4), 547-563.

Samanani, N., Liscombe, D. K. and Facchini, P. J. (2004) Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. Plant Journal, 40(2), 302-313.

Samanani, N., Park, S.-U. and Facchini, P. J. (2005) Cell type–specific localization of transcripts encoding nine consecutive enzymes involved in protoberberine alkaloid biosynthesis. The Plant Cell, 17(3), 915-926.

Sanger, F. and Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology, 94(3), 441-448.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, 74(12), 5463-5467.

Saroglia, M. and Liu, Z. (2012). Functional Genomics in Aquaculture. Wiley-Blackwell, pp. 397-403

Schopfer, R. C. and Ebel, J. (1998) Identification of elicitor-induced cytochrome P450s of soybean (Glycine max L.) using differential display of mRNA. Molecular and General Genetics, 258(4), 315-322.

Şelale, H., Çelik, I., Gültekin, V., Allmer, J., Doğanlar, S. and Frary, A. (2013) Development of EST-SSR markers for diversity and breeding studies in opium poppy. Plant Breeding, 132(3), 344-351.

Shendure, J. and Ji, H. (2008) Next-generation DNA sequencing. Nature Biotechnology, 26(10), 1135-1145.

Shukla, S. and Singh, S. (2001) Alkaloid profile in relation to different developmental stages of Papaver somniferum L. PHYTON, 41(1), 87-96.

Singh, S. K., Shukla, A. K., Dhawan, O. P. and Shasany, A. K. (2014) Recessive loci Pps-1 and OM differentially regulate PISTILLATA-1 and APETALA3-1 expression for sepal and petal development in Papaver somniferum. PLoS One, 9(6), e101272.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. and Hood, L. E. (1986) Fluorescence detection in automated DNA sequence analysis. Nature, 321(6071), 674-679.

St-Pierre, B., Vazquez-Flota, F. A. and De Luca, V. (1999) Multicellular compartmentation of catharanthus roseus alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. Plant Cell, 11(5), 887-900.

Stotz, H. U., Thomson, J. G. and Wang, Y. (2009) Plant defensins: Defense, development and application. Plant signaling & behaviour, 4(11), 1010-1012.

Takao, N., Kamigauchi, M. and Okada, M. (1983) Biosynthesis of benzo [c] phenanthridine alkaloids sanguinarine, chelirubine and macarpine. Helvetica Chimica Acta, 66(2), 473-484.

Takemura, T., Ikezawa, N., Iwasa, K. and Sato, F. (2013) Molecular cloning and characterization of a cytochrome P450 in sanguinarine biosynthesis from Eschscholzia californica cells. Phytochemistry, 91, 100-108.

Takeshita, N., Fujiwara, H., Mimura, H., Fitchen, J. H., Yamada, Y. and Sato, F. (1995) Molecular cloning and characterization of S-adenosyl-L-methionine: scoulerine-9-O-methyltransferase from cultured cells of Coptis japonica. Plant and cell physiology, 36(1), 29-36.

Tan, P. K., Downey, T. J., Spitznagel, E. L., Jr., Xu, P., Fu, D., Dimitrov, D. S., Lempicki, R. A., Raaka, B. M. and Cam, M. C. (2003) Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Research, 31(19), 5676-5684.

Tanahashi, T. and Zenk, M. H. (1990) Elicitor induction and characterization of microsomal protopine-6-hydroxylase, the central enzyme in benzophenanthridine alkaloid biosynthesis. Phytochemistry, 29(4), 1113-1122.

Tang, S., Wang, Y., Li, Z., Gui, Y., Xiao, B., Xie, J., Zhu, Q.-H. and Fan, L. (2012) Identification of wounding and topping responsive small RNAs in tobacco (Nicotiana tabacum). BMC Plant Biology, 12(1), 1.

Thudi, M., Li, Y., Jackson, S. A., May, G. D. and Varshney, R. K. (2012) Current state-of-art of sequencing technologies for plant genomics research. Briefings in Functional Genomics, 11(1), 3-11.

Türktaş, M., Kurtoğlu, K. Y., Dorado, G., Zhang, B., Hernandez, P. and Unver, T. (2014) Sequencing of plant genomes–a review. Turkish Journal of Agriculture and Forestry, 38.

Türktaş, M., Kurtoğlu, K. Y., Dorado, G., Zhang, B., Hernandez, P. and ÜNVER, T. (2014) Sequencing of plant genomes–a review. Turkish Journal of Agriculture and Forestry, 38.

Tytgat, T. O., Verhoeven, K. J., Jansen, J. J., Raaijmakers, C. E., Bakx-Schotman, T., McIntyre, L. M., van der Putten, W. H., Biere, A. and van Dam, N. M. (2013) Plants know where it hurts: root and shoot jasmonic acid induction elicit differential responses in Brassica oleracea. PLoS One, 8(6), e65502.

Unterlinner, B., Lenz, R. and Kutchan, T. M. (1999) Molecular cloning and functional expression of codeinone reductase: the penultimate enzyme in morphine biosynthesis in the opium poppy Papaver somniferum. The Plant Journal, 18(5), 465-475.

Unver, T., Bozkurt, O. and Akkaya, M. S. (2008) Identification of differentially expressed transcripts from leaves of the boron tolerant plant Gypsophila perfoliata L. Plant Cell Reports, 27(8), 1411-1422.

Unver, T., Parmaksız, İ. and Dündar, E. (2010) Identification of conserved micro-RNAs and their target transcripts in opium poppy (Papaver somniferum L.). Plant Cell Reports, 29(7), 757-769.

Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R. M. and Sidow, A. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nature Methods, 5(9), 829-834.

van Vliet, A. H. (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiology Letters, 302(1), 1-7.

Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995) Serial analysis of gene expression. Science, 270(5235), 484-487.

Vera-Estrella, R. and Bohnert, H. J. (2011). Physiological roles for the PIP family of plant aquaporins. The Plant Plasma Membrane. S. A. Murphy, B. Schulz and W. Peer. Berlin, Heidelberg, Springer Berlin Heidelberg**,** 193-222.

Vogel, C., de Sousa Abreu, R., Ko, D., Le, S. Y., Shapiro, B. A., Burns, S. C., Sandhu, D., Boutz, D. R., Marcotte, E. M. and Penalva, L. O. (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Molecular Systems Biology, 6(1), 400.

Vogel, M., Lawson, M., Sippl, W., Conrad, U. and Roos, W. (2010) Structure and mechanism of sanguinarine reductase, an enzyme of alkaloid detoxification. The Journal of Biological Chemistry, 285(24), 18397-18406.

Wang, C., Gong, B., Bushel, P. R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., Fang, H., Hong, H., Shen, J. and Su, Z. (2014) The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nature Biotechnology, 32(9), 926-932.

Wang, J., Qian, J., Yao, L. and Lu, Y. (2015) Enhanced production of flavonoids by methyl jasmonate elicitation in cell suspension culture of Hypericum perforatum. Bioresources and Bioprocessing, 2(1), 1-9.

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

Wasternack, C. and Hause, B. (2002) Jasmonates and octadecanoids: signals in plant stress responses and development. Progress in Nucleic Acid Research and Molecular Biology, 72, 165-221.

Weid, M., Ziegler, J. and Kutchan, T. M. (2004) The roles of latex and the vascular bundle in morphine biosynthesis in the opium poppy, Papaver somniferum. Proceedings of the National Academy of Sciences, 101(38), 13957-13962.

Weiss, D., Baumert, A., Vogel, M. and Roos, W. (2006) Sanguinarine reductase, a key enzyme of benzophenanthridine detoxification. Plant, Cell & Environment, 29(2), 291-302.

Wesley, S. V., Liu, Q., Wielopolska, A., Ellacott, G., Smith, N., Singh, S. and Helliwell, C. (2003). Custom knock-outs with hairpin RNA-mediated gene silencing. Plant Functional Genomics, Springer**,** 273-286.

Wijekoon, C. P. and Facchini, P. J. (2012) Systematic knockdown of morphine pathway enzymes in opium poppy using virus-induced gene silencing. The Plant Journal, 69(6), 1052-1063.

Wilson, M. L. and Coscia, C. J. (1975) Studies on the early stages of Papaver alkaloid biogenesis. Journal of the American Chemical Society, 97(2), 431-432.

Winkel, B. S. (2004) Metabolic channeling in plants. Annual Review of Plant Biology., 55, 85-107.

Winkler, A., Łyskowski, A., Riedl, S., Puhl, M., Kutchan, T. M., Macheroux, P. and Gruber, K. (2008) A concerted mechanism for berberine bridge enzyme. Nature Chemical Biology, 4(12), 739-741.

Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T. R., Li, Y., Meade, F., Teodor, R., Vaistij, F. E., Walker, C., Bowser, T. A. and Graham, I. A. (2012) A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science, 336(6089), 1704-1708.

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., Li, Y., Dowle, A. A., Cartwright, J. and Bates, R. (2015) Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. Science, 349(6245), 309-312.

Wolf, J. B. (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Molecular Ecology Resources, 13(4), 559-572.

Wu, S. and Chappell, J. (2008) Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. Current Opinion in Biotechnology, 19(2), 145-152.

Xiang, C. C. and Chen, Y. (2000) cDNA microarray technology and its applications. Biotechnology Advances, 18(1), 35-46.

Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M. and Turner, J. G. (1998) COI1: an Arabidopsis gene required for jasmonate-regulated defense and fertility. Science, 280(5366), 1091-1094.

Ye, K., Ke, Y., Keshava, N., Shanks, J., Kapp, J. A., Tekmal, R. R., Petros, J. and Joshi, H. C. (1998) Opium alkaloid noscapine is an antitumor agent that arrests metaphase and induces apoptosis in dividing cells. Proceedings of the National Academy of Sciences, 95(4), 1601-1606.

Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pre, M., Gantet, P. and Memelink, J. (2011) The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in Catharanthus roseus. The Plant Journal, 67(1), 61-71.

Zhao, J., Davis, L. C. and Verpoorte, R. (2005) Elicitor signal transduction leading to production of plant secondary metabolites. Biotechnology advances, 23(4), 283-333.

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. and Liu, X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One, 9(1), e78644.

Zheng, Y., Zhao, L., Gao, J. and Fei, Z. (2011) iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. BMC Bioinformatics, 12(1), 1.

Ziegler, J., Diaz-Chávez, M. L., Kramell, R., Ammer, C. and Kutchan, T. M. (2005) Comparative macroarray analysis of morphine containing Papaver somniferum and eight morphine free Papaver species identifies an O-methyltransferase involved in benzylisoquinoline biosynthesis. Planta, 222(3), 458-471.

Ziegler, J. and Facchini, P. J. (2008) Alkaloid biosynthesis: metabolism and trafficking. Annual Review of Plant Biology, 59, 735-769.

Ziegler, J., Facchini, P. J., Geissler, R., Schmidt, J., Ammer, C., Kramell, R., Voigtlander, S., Gesell, A., Pienkny, S. and Brandt, W. (2009) Evolution of morphine biosynthesis in opium poppy. Phytochemistry, 70(15-16), 1696-1707.

Ziegler, J., Voigtländer, S., Schmidt, J., Kramell, R., Miersch, O., Ammer, C., Gesell, A. and Kutchan, T. M. (2006) Comparative transcript and alkaloid profiling in Papaver species identifies a short chain dehydrogenase/reductase involved in morphine biosynthesis. The Plant Journal, 48(2), 177-192.

Zulak, K. G., Cornish, A., Daskalchuk, T. E., Deyholos, M. K., Goodenowe, D. B., Gordon, P. M., Klassen, D., Pelcher, L. E., Sensen, C. W. and Facchini, P. J. (2007) Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism. Planta, 225(5), 1085-1106.

Zulak, K. G., Khan, M. F., Alcantara, J., Schriemer, D. C. and Facchini, P. J. (2009) Plant defense responses in opium poppy cell cultures revealed by liquid chromatography-tandem mass spectrometry proteomics. Molecular & Cellular Proteomics, 8(1), 86-98.

# SUPPLEMENTARY

**Supplementary Table 1**. Top 50 Significantly enriched gene ontology (GO) terms in MR capsule tissue exposed MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (3652)** | **DEGs (27508)** | |
| metabolic process | 2676 genes, 73.3% | 20203 genes, 73.4% | 1 |
| cellular process | 2523 genes, 69.1% | 19544 genes, 71.0% | 1 |
| single-organism process | 2206 genes, 60.4% | 16566 genes, 60.2% | 1 |
| cellular metabolic process | 1967 genes, 53.9% | 15535 genes, 56.5% | 1 |
| organic substance metabolic process | 1963 genes, 53.8% | 15731 genes, 57.2% | 1 |
| primary metabolic process | 1788 genes, 49.0% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 1586 genes, 43.4% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 1431 genes, 39.2% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 1144 genes, 31.3% | 10232 genes, 37.2% | 1 |
| response to stimulus | 1091 genes, 29.9% | 7575 genes, 27.5% | 0.84573 |
| cellular macromolecule metabolic process | 1047 genes, 28.7% | 9374 genes, 34.1% | 1 |
| nitrogen compound metabolic process | 876 genes, 24.0% | 7163 genes, 26.0% | 1 |
| biosynthetic process | 857 genes, 23.5% | 7290 genes, 26.5% | 1 |
| organic substance biosynthetic process | 809 genes, 22.2% | 6884 genes, 25.0% | 1 |
| cellular biosynthetic process | 784 genes, 21.5% | 6657 genes, 24.2% | 1 |
| organic cyclic compound metabolic process | 769 genes, 21.1% | 6656 genes, 24.2% | 1 |
| protein metabolic process | 760 genes, 20.8% | 6661 genes, 24.2% | 1 |
| phosphorus metabolic process | 759 genes, 20.8% | 5135 genes, 18.7% | 0.56344 |
| cellular aromatic compound metabolic process | 749 genes, 20.5% | 6393 genes, 23.2% | 1 |
| phosphate-containing compound metabolic process | 747 genes, 20.5% | 5007 genes, 18.2% | 0.21003 |
| biological regulation | 717 genes, 19.6% | 5742 genes, 20.9% | 1 |
| localization | 715 genes, 19.6% | 5319 genes, 19.3% | 1 |
| cellular nitrogen compound metabolic process | 702 genes, 19.2% | 6195 genes, 22.5% | 1 |
| establishment of localization | 694 genes, 19.0% | 5161 genes, 18.8% | 1 |
| cellular protein metabolic process | 689 genes, 18.9% | 5927 genes, 21.5% | 1 |
| heterocycle metabolic process | 680 genes, 18.6% | 6146 genes, 22.3% | 1 |
| transport | 676 genes, 18.5% | 5009 genes, 18.2% | 1 |
| small molecule metabolic process | 638 genes, 17.5% | 4814 genes, 17.5% | 1 |
| regulation of biological process | 635 genes, 17.4% | 5202 genes, 18.9% | 1 |
| response to stress | 617 genes, 16.9% | 4333 genes, 15.8% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (3652)** | **DEGs (27508)** | |
| nucleobase-containing compound metabolic process | 598 genes, 16.4% | 5461 genes, 19.9% | 1 |
| single-organism transport | 567 genes, 15.5% | 3911 genes, 14.2% | 1 |
| single-organism biosynthetic process | 562 genes, 15.4% | 4177 genes, 15.2% | 1 |
| regulation of cellular process | 542 genes, 14.8% | 4283 genes, 15.6% | 1 |
| response to chemical | 511 genes, 14.0% | 3785 genes, 13.8% | 1 |
| oxidation-reduction process | 504 genes, 13.8% | 3471 genes, 12.6% | 1 |
| organonitrogen compound metabolic process | 504 genes, 13.8% | 3673 genes, 13.4% | 1 |
| macromolecule modification | 501 genes, 13.7% | 3777 genes, 13.7% | 1 |
| phosphorylation | 463 genes, 12.7% | 2673 genes, 9.7% | 4.47E-07 |
| cellular protein modification process | 454 genes, 12.4% | 3263 genes, 11.9% | 1 |
| protein modification process | 454 genes, 12.4% | 3263 genes, 11.9% | 1 |
| multicellular organismal process | 443 genes, 12.1% | 3926 genes, 14.3% | 1 |
| developmental process | 435 genes, 11.9% | 3987 genes, 14.5% | 1 |
| single-organism developmental process | 431 genes, 11.8% | 3914 genes, 14.2% | 1 |
| cellular component organization or biogenesis | 426 genes, 11.7% | 4379 genes, 15.9% | 1 |
| single-multicellular organism process | 423 genes, 11.6% | 3776 genes, 13.7% | 1 |
| response to abiotic stimulus | 408 genes, 11.2% | 3221 genes, 11.7% | 1 |
| anatomical structure development | 401 genes, 11.0% | 3567 genes, 13.0% | 1 |
| multicellular organismal development | 395 genes, 10.8% | 3581 genes, 13.0% | 1 |
| cellular response to stimulus | 392 genes, 10.7% | 2758 genes, 10.0% | 1 |

**Supplementary Table 2.** Top 50 Significantly enriched gene ontology (GO) terms in MR stem tissue exposed MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (5320)** | **DEGs (27508)** | |
| metabolic process | 3887 genes, 73.1% | 20203 genes, 73.4% | 1 |
| cellular process | 3750 genes, 70.5% | 19544 genes, 71.0% | 1 |
| single-organism process | 3207 genes, 60.3% | 16566 genes, 60.2% | 1 |
| organic substance metabolic process | 3016 genes, 56.7% | 15731 genes, 57.2% | 1 |
| cellular metabolic process | 2999 genes, 56.4% | 15535 genes, 56.5% | 1 |
| primary metabolic process | 2817 genes, 53.0% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 2316 genes, 43.5% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 2001 genes, 37.6% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 1966 genes, 37.0% | 10232 genes, 37.2% | 1 |
| cellular macromolecule metabolic process | 1821 genes, 34.2% | 9374 genes, 34.1% | 1 |
| response to stimulus | 1530 genes, 28.8% | 7575 genes, 27.5% | 1 |
| nitrogen compound metabolic process | 1379 genes, 25.9% | 7163 genes, 26.0% | 1 |
| biosynthetic process | 1308 genes, 24.6% | 7290 genes, 26.5% | 1 |
| organic cyclic compound metabolic process | 1276 genes, 24.0% | 6656 genes, 24.2% | 1 |
| cellular aromatic compound metabolic process | 1241 genes, 23.3% | 6393 genes, 23.2% | 1 |
| organic substance biosynthetic process | 1228 genes, 23.1% | 6884 genes, 25.0% | 1 |
| protein metabolic process | 1226 genes, 23.0% | 6661 genes, 24.2% | 1 |
| cellular biosynthetic process | 1194 genes, 22.4% | 6657 genes, 24.2% | 1 |
| cellular nitrogen compound metabolic process | 1193 genes, 22.4% | 6195 genes, 22.5% | 1 |
| heterocycle metabolic process | 1178 genes, 22.1% | 6146 genes, 22.3% | 1 |
| biological regulation | 1169 genes, 22.0% | 5742 genes, 20.9% | 1 |
| cellular protein metabolic process | 1110 genes, 20.9% | 5927 genes, 21.5% | 1 |
| nucleobase-containing compound metabolic process | 1065 genes, 20.0% | 5461 genes, 19.9% | 1 |
| regulation of biological process | 1052 genes, 19.8% | 5202 genes, 18.9% | 1 |
| localization | 1011 genes, 19.0% | 5319 genes, 19.3% | 1 |
| phosphorus metabolic process | 990 genes, 18.6% | 5135 genes, 18.7% | 1 |
| establishment of localization | 966 genes, 18.2% | 5161 genes, 18.8% | 1 |
| phosphate-containing compound metabolic process | 963 genes, 18.1% | 5007 genes, 18.2% | 1 |
| transport | 942 genes, 17.7% | 5009 genes, 18.2% | 1 |
| cellular component organization or biogenesis | 887 genes, 16.7% | 4379 genes, 15.9% | 1 |
| response to stress | 882 genes, 16.6% | 4333 genes, 15.8% | 1 |
| regulation of cellular process | 880 genes, 16.5% | 4283 genes, 15.6% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (5320)** | **DEGs (27508)** | |
| small molecule metabolic process | 851 genes, 16.0% | 4814 genes, 17.5% | 1 |
| multicellular organismal process | 831 genes, 15.6% | 3926 genes, 14.3% | 1 |
| developmental process | 812 genes, 15.3% | 3987 genes, 14.5% | 1 |
| single-organism developmental process | 801 genes, 15.1% | 3914 genes, 14.2% | 1 |
| single-multicellular organism process | 788 genes, 14.8% | 3776 genes, 13.7% | 1 |
| cellular component organization | 787 genes, 14.8% | 3922 genes, 14.3% | 1 |
| nucleic acid metabolic process | 776 genes, 14.6% | 3851 genes, 14.0% | 1 |
| single-organism transport | 756 genes, 14.2% | 3911 genes, 14.2% | 1 |
| single-organism biosynthetic process | 756 genes, 14.2% | 4177 genes, 15.2% | 1 |
| macromolecule modification | 746 genes, 14.0% | 3777 genes, 13.7% | 1 |
| response to chemical | 746 genes, 14.0% | 3785 genes, 13.8% | 1 |
| multicellular organismal development | 738 genes, 13.9% | 3581 genes, 13.0% | 1 |
| anatomical structure development | 733 genes, 13.8% | 3567 genes, 13.0% | 1 |
| response to abiotic stimulus | 681 genes, 12.8% | 3221 genes, 11.7% | 1 |
| macromolecule biosynthetic process | 676 genes, 12.7% | 3656 genes, 13.3% | 1 |
| cellular protein modification process | 662 genes, 12.4% | 3263 genes, 11.9% | 1 |
| protein modification process | 662 genes, 12.4% | 3263 genes, 11.9% | 1 |

**Supplementary Table 3.** Top 50 Significantly enriched gene ontology (GO) terms in MP capsule tissue exposed MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (4202)** | **DEGs (27508)** | |
| metabolic process | 3025 genes, 72.0% | 20203 genes, 73.4% | 1 |
| cellular process | 2780 genes, 66.2% | 19544 genes, 71.0% | 1 |
| single-organism process | 2603 genes, 61.9% | 16566 genes, 60.2% | 1 |
| organic substance metabolic process | 2269 genes, 54.0% | 15731 genes, 57.2% | 1 |
| cellular metabolic process | 2156 genes, 51.3% | 15535 genes, 56.5% | 1 |
| primary metabolic process | 2098 genes, 49.9% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 1797 genes, 42.8% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 1632 genes, 38.8% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 1322 genes, 31.5% | 10232 genes, 37.2% | 1 |
| response to stimulus | 1199 genes, 28.5% | 7575 genes, 27.5% | 1 |
| cellular macromolecule metabolic process | 1197 genes, 28.5% | 9374 genes, 34.1% | 1 |
| biosynthetic process | 1114 genes, 26.5% | 7290 genes, 26.5% | 1 |
| organic substance biosynthetic process | 1060 genes, 25.2% | 6884 genes, 25.0% | 1 |
| cellular biosynthetic process | 1000 genes, 23.8% | 6657 genes, 24.2% | 1 |
| nitrogen compound metabolic process | 996 genes, 23.7% | 7163 genes, 26.0% | 1 |
| organic cyclic compound metabolic process | 904 genes, 21.5% | 6656 genes, 24.2% | 1 |
| cellular aromatic compound metabolic process | 848 genes, 20.2% | 6393 genes, 23.2% | 1 |
| protein metabolic process | 839 genes, 20.0% | 6661 genes, 24.2% | 1 |
| biological regulation | 823 genes, 19.6% | 5742 genes, 20.9% | 1 |
| cellular nitrogen compound metabolic process | 812 genes, 19.3% | 6195 genes, 22.5% | 1 |
| heterocycle metabolic process | 799 genes, 19.0% | 6146 genes, 22.3% | 1 |
| localization | 791 genes, 18.8% | 5319 genes, 19.3% | 1 |
| establishment of localization | 765 genes, 18.2% | 5161 genes, 18.8% | 1 |
| transport | 750 genes, 17.8% | 5009 genes, 18.2% | 1 |
| small molecule metabolic process | 747 genes, 17.8% | 4814 genes, 17.5% | 1 |
| cellular protein metabolic process | 740 genes, 17.6% | 5927 genes, 21.5% | 1 |
| regulation of biological process | 733 genes, 17.4% | 5202 genes, 18.9% | 1 |
| response to stress | 687 genes, 16.3% | 4333 genes, 15.8% | 1 |
| nucleobase-containing compound metabolic process | 686 genes, 16.3% | 5461 genes, 19.9% | 1 |
| phosphorus metabolic process | 669 genes, 15.9% | 5135 genes, 18.7% | 1 |
| single-organism biosynthetic process | 650 genes, 15.5% | 4177 genes, 15.2% | 1 |
| phosphate-containing compound metabolic process | 645 genes, 15.3% | 5007 genes, 18.2% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (4202)** | **DEGs (27508)** | |
| developmental process | 620 genes, 14.8% | 3987 genes, 14.5% | 1 |
| cellular component organization or biogenesis | 613 genes, 14.6% | 4379 genes, 15.9% | 1 |
| response to chemical | 605 genes, 14.4% | 3785 genes, 13.8% | 1 |
| single-organism developmental process | 605 genes, 14.4% | 3914 genes, 14.2% | 1 |
| regulation of cellular process | 605 genes, 14.4% | 4283 genes, 15.6% | 1 |
| single-organism transport | 598 genes, 14.2% | 3911 genes, 14.2% | 1 |
| oxidation-reduction process | 594 genes, 14.1% | 3471 genes, 12.6% | 1 |
| multicellular organismal process | 574 genes, 13.7% | 3926 genes, 14.3% | 1 |
| cellular component organization | 559 genes, 13.3% | 3922 genes, 14.3% | 1 |
| single-multicellular organism process | 550 genes, 13.1% | 3776 genes, 13.7% | 1 |
| anatomical structure development | 536 genes, 12.8% | 3567 genes, 13.0% | 1 |
| organonitrogen compound metabolic process | 536 genes, 12.8% | 3673 genes, 13.4% | 1 |
| multicellular organismal development | 532 genes, 12.7% | 3581 genes, 13.0% | 1 |
| macromolecule biosynthetic process | 519 genes, 12.4% | 3656 genes, 13.3% | 1 |
| response to abiotic stimulus | 516 genes, 12.3% | 3221 genes, 11.7% | 1 |
| cellular macromolecule biosynthetic process | 502 genes, 11.9% | 3571 genes, 13.0% | 1 |
| nucleic acid metabolic process | 485 genes, 11.5% | 3851 genes, 14.0% | 1 |

**Supplementary Table 4.** Top 50 Significantly enriched gene ontology (GO) terms in MP stem tissue exposed MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (3260)** | **DEGs (27508)** | |
| metabolic process | 2361 genes, 72.4% | 20203 genes, 73.4% | 1 |
| cellular process | 2180 genes, 66.9% | 19544 genes, 71.0% | 1 |
| single-organism process | 2018 genes, 61.9% | 16566 genes, 60.2% | 1 |
| organic substance metabolic process | 1731 genes, 53.1% | 15731 genes, 57.2% | 1 |
| cellular metabolic process | 1680 genes, 51.5% | 15535 genes, 56.5% | 1 |
| primary metabolic process | 1539 genes, 47.2% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 1437 genes, 44.1% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 1231 genes, 37.8% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 1023 genes, 31.4% | 10232 genes, 37.2% | 1 |
| response to stimulus | 925 genes, 28.4% | 7575 genes, 27.5% | 1 |
| cellular macromolecule metabolic process | 920 genes, 28.2% | 9374 genes, 34.1% | 1 |
| biosynthetic process | 771 genes, 23.7% | 7290 genes, 26.5% | 1 |
| nitrogen compound metabolic process | 769 genes, 23.6% | 7163 genes, 26.0% | 1 |
| organic substance biosynthetic process | 720 genes, 22.1% | 6884 genes, 25.0% | 1 |
| organic cyclic compound metabolic process | 701 genes, 21.5% | 6656 genes, 24.2% | 1 |
| cellular biosynthetic process | 690 genes, 21.2% | 6657 genes, 24.2% | 1 |
| cellular aromatic compound metabolic process | 677 genes, 20.8% | 6393 genes, 23.2% | 1 |
| localization | 663 genes, 20.3% | 5319 genes, 19.3% | 1 |
| establishment of localization | 651 genes, 20.0% | 5161 genes, 18.8% | 1 |
| protein metabolic process | 651 genes, 20.0% | 6661 genes, 24.2% | 1 |
| biological regulation | 641 genes, 19.7% | 5742 genes, 20.9% | 1 |
| transport | 636 genes, 19.5% | 5009 genes, 18.2% | 1 |
| cellular nitrogen compound metabolic process | 629 genes, 19.3% | 6195 genes, 22.5% | 1 |
| heterocycle metabolic process | 625 genes, 19.2% | 6146 genes, 22.3% | 1 |
| cellular protein metabolic process | 579 genes, 17.8% | 5927 genes, 21.5% | 1 |
| phosphorus metabolic process | 565 genes, 17.3% | 5135 genes, 18.7% | 1 |
| regulation of biological process | 565 genes, 17.3% | 5202 genes, 18.9% | 1 |
| phosphate-containing compound metabolic process | 545 genes, 16.7% | 5007 genes, 18.2% | 1 |
| response to stress | 540 genes, 16.6% | 4333 genes, 15.8% | 1 |
| small molecule metabolic process | 522 genes, 16.0% | 4814 genes, 17.5% | 1 |
| single-organism transport | 518 genes, 15.9% | 3911 genes, 14.2% | 1 |
| nucleobase-containing compound metabolic process | 517 genes, 15.9% | 5461 genes, 19.9% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (3260)** | **DEGs (27508)** | |
| single-organism biosynthetic process | 501 genes, 15.4% | 4177 genes, 15.2% | 1 |
| developmental process | 473 genes, 14.5% | 3987 genes, 14.5% | 1 |
| single-organism developmental process | 469 genes, 14.4% | 3914 genes, 14.2% | 1 |
| multicellular organismal process | 467 genes, 14.3% | 3926 genes, 14.3% | 1 |
| response to chemical | 466 genes, 14.3% | 3785 genes, 13.8% | 1 |
| regulation of cellular process | 458 genes, 14.0% | 4283 genes, 15.6% | 1 |
| cellular component organization or biogenesis | 454 genes, 13.9% | 4379 genes, 15.9% | 1 |
| single-multicellular organism process | 447 genes, 13.7% | 3776 genes, 13.7% | 1 |
| oxidation-reduction process | 433 genes, 13.3% | 3471 genes, 12.6% | 1 |
| multicellular organismal development | 432 genes, 13.3% | 3581 genes, 13.0% | 1 |
| organonitrogen compound metabolic process | 424 genes, 13.0% | 3673 genes, 13.4% | 1 |
| cellular component organization | 420 genes, 12.9% | 3922 genes, 14.3% | 1 |
| anatomical structure development | 411 genes, 12.6% | 3567 genes, 13.0% | 1 |
| macromolecule modification | 402 genes, 12.3% | 3777 genes, 13.7% | 1 |
| response to abiotic stimulus | 382 genes, 11.7% | 3221 genes, 11.7% | 1 |
| catabolic process | 364 genes, 11.2% | 3231 genes, 11.7% | 1 |
| cellular protein modification process | 364 genes, 11.2% | 3263 genes, 11.9% | 1 |

**Supplementary Table 5.** Top 50 Significantly enriched gene ontology (GO) terms in MR capsule and stem tissues exposed MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (6365)** | **DEGs (27508)** | |
| metabolic process | 4554 genes, 71.5% | 20203 genes, 73.4% | 1 |
| cellular process | 4354 genes, 68.4% | 19544 genes, 71.0% | 1 |
| single-organism process | 3954 genes, 62.1% | 16566 genes, 60.2% | 0.54114 |
| organic substance metabolic process | 3380 genes, 53.1% | 15731 genes, 57.2% | 1 |
| cellular metabolic process | 3333 genes, 52.4% | 15535 genes, 56.5% | 1 |
| primary metabolic process | 3099 genes, 48.7% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 2794 genes, 43.9% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 2446 genes, 38.4% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 2026 genes, 31.8% | 10232 genes, 37.2% | 1 |
| cellular macromolecule metabolic process | 1839 genes, 28.9% | 9374 genes, 34.1% | 1 |
| response to stimulus | 1824 genes, 28.7% | 7575 genes, 27.5% | 1 |
| biosynthetic process | 1556 genes, 24.4% | 7290 genes, 26.5% | 1 |
| nitrogen compound metabolic process | 1484 genes, 23.3% | 7163 genes, 26.0% | 1 |
| organic substance biosynthetic process | 1465 genes, 23.0% | 6884 genes, 25.0% | 1 |
| cellular biosynthetic process | 1402 genes, 22.0% | 6657 genes, 24.2% | 1 |
| organic cyclic compound metabolic process | 1383 genes, 21.7% | 6656 genes, 24.2% | 1 |
| biological regulation | 1344 genes, 21.1% | 5742 genes, 20.9% | 1 |
| cellular aromatic compound metabolic process | 1320 genes, 20.7% | 6393 genes, 23.2% | 1 |
| localization | 1261 genes, 19.8% | 5319 genes, 19.3% | 1 |
| protein metabolic process | 1248 genes, 19.6% | 6661 genes, 24.2% | 1 |
| establishment of localization | 1220 genes, 19.2% | 5161 genes, 18.8% | 1 |
| cellular nitrogen compound metabolic process | 1220 genes, 19.2% | 6195 genes, 22.5% | 1 |
| heterocycle metabolic process | 1208 genes, 19.0% | 6146 genes, 22.3% | 1 |
| regulation of biological process | 1191 genes, 18.7% | 5202 genes, 18.9% | 1 |
| transport | 1181 genes, 18.6% | 5009 genes, 18.2% | 1 |
| cellular protein metabolic process | 1099 genes, 17.3% | 5927 genes, 21.5% | 1 |
| phosphorus metabolic process | 1066 genes, 16.7% | 5135 genes, 18.7% | 1 |
| small molecule metabolic process | 1065 genes, 16.7% | 4814 genes, 17.5% | 1 |
| response to stress | 1044 genes, 16.4% | 4333 genes, 15.8% | 1 |
| phosphate-containing compound metabolic process | 1043 genes, 16.4% | 5007 genes, 18.2% | 1 |
| nucleobase-containing compound metabolic process | 1043 genes, 16.4% | 5461 genes, 19.9% | 1 |
| single-organism biosynthetic process | 1014 genes, 15.9% | 4177 genes, 15.2% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (6365)** | **DEGs (27508)** | |
| single-organism transport | 967 genes, 15.2% | 3911 genes, 14.2% | 1 |
| regulation of cellular process | 963 genes, 15.1% | 4283 genes, 15.6% | 1 |
| cellular component organization or biogenesis | 955 genes, 15.0% | 4379 genes, 15.9% | 1 |
| developmental process | 921 genes, 14.5% | 3987 genes, 14.5% | 1 |
| multicellular organismal process | 916 genes, 14.4% | 3926 genes, 14.3% | 1 |
| response to chemical | 910 genes, 14.3% | 3785 genes, 13.8% | 1 |
| single-organism developmental process | 906 genes, 14.2% | 3914 genes, 14.2% | 1 |
| single-multicellular organism process | 871 genes, 13.7% | 3776 genes, 13.7% | 1 |
| oxidation-reduction process | 863 genes, 13.6% | 3471 genes, 12.6% | 1 |
| cellular component organization | 863 genes, 13.6% | 3922 genes, 14.3% | 1 |
| anatomical structure development | 828 genes, 13.0% | 3567 genes, 13.0% | 1 |
| multicellular organismal development | 825 genes, 13.0% | 3581 genes, 13.0% | 1 |
| response to abiotic stimulus | 805 genes, 12.6% | 3221 genes, 11.7% | 1 |
| organonitrogen compound metabolic process | 802 genes, 12.6% | 3673 genes, 13.4% | 1 |
| macromolecule modification | 752 genes, 11.8% | 3777 genes, 13.7% | 1 |
| nucleic acid metabolic process | 739 genes, 11.6% | 3851 genes, 14.0% | 1 |
| macromolecule biosynthetic process | 686 genes, 10.8% | 3656 genes, 13.3% | 1 |

**Supplementary Table 6.** Top 50 Significantly enriched gene ontology (GO) terms in MP capsule and stem tissues exposed MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (6422)** | **DEGs (27508)** | |
| metabolic process | 4596 genes, 71.6% | 20203 genes, 73.4% | 1 |
| cellular process | 4323 genes, 67.3% | 19544 genes, 71.0% | 1 |
| single-organism process | 4012 genes, 62.5% | 16566 genes, 60.2% | 0.03286 |
| organic substance metabolic process | 3378 genes, 52.6% | 15731 genes, 57.2% | 1 |
| cellular metabolic process | 3321 genes, 51.7% | 15535 genes, 56.5% | 1 |
| primary metabolic process | 3102 genes, 48.3% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 2811 genes, 43.8% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 2520 genes, 39.2% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 1952 genes, 30.4% | 10232 genes, 37.2% | 1 |
| response to stimulus | 1843 genes, 28.7% | 7575 genes, 27.5% | 1 |
| cellular macromolecule metabolic process | 1795 genes, 28.0% | 9374 genes, 34.1% | 1 |
| biosynthetic process | 1538 genes, 23.9% | 7290 genes, 26.5% | 1 |
| nitrogen compound metabolic process | 1501 genes, 23.4% | 7163 genes, 26.0% | 1 |
| organic substance biosynthetic process | 1446 genes, 22.5% | 6884 genes, 25.0% | 1 |
| cellular biosynthetic process | 1386 genes, 21.6% | 6657 genes, 24.2% | 1 |
| organic cyclic compound metabolic process | 1343 genes, 20.9% | 6656 genes, 24.2% | 1 |
| biological regulation | 1301 genes, 20.3% | 5742 genes, 20.9% | 1 |
| localization | 1299 genes, 20.2% | 5319 genes, 19.3% | 1 |
| cellular aromatic compound metabolic process | 1296 genes, 20.2% | 6393 genes, 23.2% | 1 |
| establishment of localization | 1257 genes, 19.6% | 5161 genes, 18.8% | 1 |
| protein metabolic process | 1224 genes, 19.1% | 6661 genes, 24.2% | 1 |
| transport | 1213 genes, 18.9% | 5009 genes, 18.2% | 1 |
| cellular nitrogen compound metabolic process | 1191 genes, 18.5% | 6195 genes, 22.5% | 1 |
| heterocycle metabolic process | 1183 genes, 18.4% | 6146 genes, 22.3% | 1 |
| regulation of biological process | 1159 genes, 18.0% | 5202 genes, 18.9% | 1 |
| phosphorus metabolic process | 1130 genes, 17.6% | 5135 genes, 18.7% | 1 |
| small molecule metabolic process | 1121 genes, 17.5% | 4814 genes, 17.5% | 1 |
| phosphate-containing compound metabolic process | 1103 genes, 17.2% | 5007 genes, 18.2% | 1 |
| cellular protein metabolic process | 1101 genes, 17.1% | 5927 genes, 21.5% | 1 |
| response to stress | 1034 genes, 16.1% | 4333 genes, 15.8% | 1 |
| single-organism biosynthetic process | 1023 genes, 15.9% | 4177 genes, 15.2% | 1 |
| nucleobase-containing compound metabolic process | 1016 genes, 15.8% | 5461 genes, 19.9% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (6422)** | **DEGs (27508)** | |
| single-organism transport | 995 genes, 15.5% | 3911 genes, 14.2% | 1 |
| oxidation-reduction process | 969 genes, 15.1% | 3471 genes, 12.6% | 3.14E-08 |
| regulation of cellular process | 944 genes, 14.7% | 4283 genes, 15.6% | 1 |
| response to chemical | 927 genes, 14.4% | 3785 genes, 13.8% | 1 |
| developmental process | 907 genes, 14.1% | 3987 genes, 14.5% | 1 |
| cellular component organization or biogenesis | 892 genes, 13.9% | 4379 genes, 15.9% | 1 |
| single-organism developmental process | 890 genes, 13.9% | 3914 genes, 14.2% | 1 |
| multicellular organismal process | 886 genes, 13.8% | 3926 genes, 14.3% | 1 |
| organonitrogen compound metabolic process | 856 genes, 13.3% | 3673 genes, 13.4% | 1 |
| single-multicellular organism process | 847 genes, 13.2% | 3776 genes, 13.7% | 1 |
| response to abiotic stimulus | 812 genes, 12.6% | 3221 genes, 11.7% | 1 |
| multicellular organismal development | 810 genes, 12.6% | 3581 genes, 13.0% | 1 |
| cellular component organization | 804 genes, 12.5% | 3922 genes, 14.3% | 1 |
| anatomical structure development | 801 genes, 12.5% | 3567 genes, 13.0% | 1 |
| macromolecule modification | 771 genes, 12.0% | 3777 genes, 13.7% | 1 |
| organic acid metabolic process | 700 genes, 10.9% | 2723 genes, 9.9% | 1 |
| oxoacid metabolic process | 698 genes, 10.9% | 2716 genes, 9.9% | 1 |

**Supplementary Table 7.** Top 50 Significantly enriched gene ontology (GO) terms in MR and MP capsule tissues exposed to MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (5116)** | **DEGs (27508)** | |
| metabolic process | 3722 genes, 72.8% | 20203 genes, 73.4% | 1 |
| cellular process | 3460 genes, 67.6% | 19544 genes, 71.0% | 1 |
| single-organism process | 3147 genes, 61.5% | 16566 genes, 60.2% | 1 |
| organic substance metabolic process | 2737 genes, 53.5% | 15731 genes, 57.2% | 1 |
| cellular metabolic process | 2679 genes, 52.4% | 15535 genes, 56.5% | 1 |
| primary metabolic process | 2508 genes, 49.0% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 2237 genes, 43.7% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 2028 genes, 39.6% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 1573 genes, 30.7% | 10232 genes, 37.2% | 1 |
| response to stimulus | 1444 genes, 28.2% | 7575 genes, 27.5% | 1 |
| cellular macromolecule metabolic process | 1439 genes, 28.1% | 9374 genes, 34.1% | 1 |
| biosynthetic process | 1206 genes, 23.6% | 7290 genes, 26.5% | 1 |
| nitrogen compound metabolic process | 1178 genes, 23.0% | 7163 genes, 26.0% | 1 |
| organic substance biosynthetic process | 1135 genes, 22.2% | 6884 genes, 25.0% | 1 |
| organic cyclic compound metabolic process | 1075 genes, 21.0% | 6656 genes, 24.2% | 1 |
| cellular biosynthetic process | 1072 genes, 21.0% | 6657 genes, 24.2% | 1 |
| localization | 1050 genes, 20.5% | 5319 genes, 19.3% | 1 |
| cellular aromatic compound metabolic process | 1035 genes, 20.2% | 6393 genes, 23.2% | 1 |
| establishment of localization | 1013 genes, 19.8% | 5161 genes, 18.8% | 1 |
| protein metabolic process | 993 genes, 19.4% | 6661 genes, 24.2% | 1 |
| transport | 982 genes, 19.2% | 5009 genes, 18.2% | 1 |
| biological regulation | 974 genes, 19.0% | 5742 genes, 20.9% | 1 |
| cellular nitrogen compound metabolic process | 947 genes, 18.5% | 6195 genes, 22.5% | 1 |
| heterocycle metabolic process | 946 genes, 18.5% | 6146 genes, 22.3% | 1 |
| small molecule metabolic process | 918 genes, 17.9% | 4814 genes, 17.5% | 1 |
| cellular protein metabolic process | 886 genes, 17.3% | 5927 genes, 21.5% | 1 |
| phosphorus metabolic process | 874 genes, 17.1% | 5135 genes, 18.7% | 1 |
| phosphate-containing compound metabolic process | 859 genes, 16.8% | 5007 genes, 18.2% | 1 |
| regulation of biological process | 857 genes, 16.8% | 5202 genes, 18.9% | 1 |
| response to stress | 818 genes, 16.0% | 4333 genes, 15.8% | 1 |
| single-organism biosynthetic process | 816 genes, 15.9% | 4177 genes, 15.2% | 1 |
| single-organism transport | 815 genes, 15.9% | 3911 genes, 14.2% | 0.15262 |
| nucleobase-containing compound | 813 genes, 15.9% | 5461 genes, 19.9% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (5116)** | **DEGs (27508)** | |
| metabolic process | | | |
| oxidation-reduction process | 797 genes, 15.6% | 3471 genes, 12.6% | 6.65E-09 |
| response to chemical | 731 genes, 14.3% | 3785 genes, 13.8% | 1 |
| regulation of cellular process | 702 genes, 13.7% | 4283 genes, 15.6% | 1 |
| organonitrogen compound metabolic process | 669 genes, 13.1% | 3673 genes, 13.4% | 1 |
| cellular component organization or biogenesis | 652 genes, 12.7% | 4379 genes, 15.9% | 1 |
| developmental process | 634 genes, 12.4% | 3987 genes, 14.5% | 1 |
| response to abiotic stimulus | 628 genes, 12.3% | 3221 genes, 11.7% | 1 |
| single-organism developmental process | 624 genes, 12.2% | 3914 genes, 14.2% | 1 |
| multicellular organismal process | 615 genes, 12.0% | 3926 genes, 14.3% | 1 |
| macromolecule modification | 598 genes, 11.7% | 3777 genes, 13.7% | 1 |
| catabolic process | 584 genes, 11.4% | 3231 genes, 11.7% | 1 |
| single-multicellular organism process | 584 genes, 11.4% | 3776 genes, 13.7% | 1 |
| cellular component organization | 579 genes, 11.3% | 3922 genes, 14.3% | 1 |
| anatomical structure development | 562 genes, 11.0% | 3567 genes, 13.0% | 1 |
| organic substance catabolic process | 554 genes, 10.8% | 3033 genes, 11.0% | 1 |
| oxoacid metabolic process | 553 genes, 10.8% | 2716 genes, 9.9% | 1 |

**Supplementary Table 8.** Top 50 Significantly enriched gene ontology (GO) terms in MR and MP stem tissues exposed to MeJA

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (6045)** | **DEGs (27508)** | |
| metabolic process | 4424 genes, 73.2% | 20203 genes, 73.4% | 1 |
| cellular process | 4224 genes, 69.9% | 19544 genes, 71.0% | 1 |
| single-organism process | 3622 genes, 59.9% | 16566 genes, 60.2% | 1 |
| organic substance metabolic process | 3402 genes, 56.3% | 15731 genes, 57.2% | 1 |
| cellular metabolic process | 3359 genes, 55.6% | 15535 genes, 56.5% | 1 |
| primary metabolic process | 3162 genes, 52.3% | 14708 genes, 53.5% | 1 |
| single-organism cellular process | 2622 genes, 43.4% | 12028 genes, 43.7% | 1 |
| single-organism metabolic process | 2296 genes, 38.0% | 10409 genes, 37.8% | 1 |
| macromolecule metabolic process | 2195 genes, 36.3% | 10232 genes, 37.2% | 1 |
| cellular macromolecule metabolic process | 2018 genes, 33.4% | 9374 genes, 34.1% | 1 |
| response to stimulus | 1754 genes, 29.0% | 7575 genes, 27.5% | 1 |
| nitrogen compound metabolic process | 1561 genes, 25.8% | 7163 genes, 26.0% | 1 |
| biosynthetic process | 1492 genes, 24.7% | 7290 genes, 26.5% | 1 |
| organic cyclic compound metabolic process | 1453 genes, 24.0% | 6656 genes, 24.2% | 1 |
| cellular aromatic compound metabolic process | 1407 genes, 23.3% | 6393 genes, 23.2% | 1 |
| organic substance biosynthetic process | 1405 genes, 23.2% | 6884 genes, 25.0% | 1 |
| protein metabolic process | 1390 genes, 23.0% | 6661 genes, 24.2% | 1 |
| cellular biosynthetic process | 1360 genes, 22.5% | 6657 genes, 24.2% | 1 |
| cellular nitrogen compound metabolic process | 1349 genes, 22.3% | 6195 genes, 22.5% | 1 |
| heterocycle metabolic process | 1343 genes, 22.2% | 6146 genes, 22.3% | 1 |
| cellular protein metabolic process | 1246 genes, 20.6% | 5927 genes, 21.5% | 1 |
| nucleobase-containing compound metabolic process | 1201 genes, 19.9% | 5461 genes, 19.9% | 1 |
| biological regulation | 1199 genes, 19.8% | 5742 genes, 20.9% | 1 |
| localization | 1183 genes, 19.6% | 5319 genes, 19.3% | 1 |
| establishment of localization | 1144 genes, 18.9% | 5161 genes, 18.8% | 1 |
| phosphorus metabolic process | 1137 genes, 18.8% | 5135 genes, 18.7% | 1 |
| phosphate-containing compound metabolic process | 1115 genes, 18.4% | 5007 genes, 18.2% | 1 |
| transport | 1104 genes, 18.3% | 5009 genes, 18.2% | 1 |
| regulation of biological process | 1079 genes, 17.8% | 5202 genes, 18.9% | 1 |
| small molecule metabolic process | 1040 genes, 17.2% | 4814 genes, 17.5% | 1 |
| response to stress | 1015 genes, 16.8% | 4333 genes, 15.8% | 1 |
| cellular component organization or biogenesis | 963 genes, 15.9% | 4379 genes, 15.9% | 1 |

| Gene Ontology term | Cluster frequency | Genome frequency of use | Corrected P-value |
|---|---|---|---|
| **Biological Process** | **DEGs (6045)** | **DEGs (27508)** | |
| single-organism biosynthetic process | 920 genes, 15.2% | 4177 genes, 15.2% | 1 |
| regulation of cellular process | 901 genes, 14.9% | 4283 genes, 15.6% | 1 |
| response to chemical | 884 genes, 14.6% | 3785 genes, 13.8% | 1 |
| single-organism transport | 883 genes, 14.6% | 3911 genes, 14.2% | 1 |
| cellular component organization | 859 genes, 14.2% | 3922 genes, 14.3% | 1 |
| nucleic acid metabolic process | 846 genes, 14.0% | 3851 genes, 14.0% | 1 |
| developmental process | 843 genes, 13.9% | 3987 genes, 14.5% | 1 |
| single-organism developmental process | 831 genes, 13.7% | 3914 genes, 14.2% | 1 |
| multicellular organismal process | 830 genes, 13.7% | 3926 genes, 14.3% | 1 |
| organonitrogen compound metabolic process | 816 genes, 13.5% | 3673 genes, 13.4% | 1 |
| macromolecule modification | 814 genes, 13.5% | 3777 genes, 13.7% | 1 |
| oxidation-reduction process | 799 genes, 13.2% | 3471 genes, 12.6% | 1 |
| single-multicellular organism process | 799 genes, 13.2% | 3776 genes, 13.7% | 1 |
| response to abiotic stimulus | 783 genes, 13.0% | 3221 genes, 11.7% | 0.98265 |
| multicellular organismal development | 764 genes, 12.6% | 3581 genes, 13.0% | 1 |
| anatomical structure development | 755 genes, 12.5% | 3567 genes, 13.0% | 1 |
| macromolecule biosynthetic process | 748 genes, 12.4% | 3656 genes, 13.3% | 1 |

**Supplementary Table 9.** Top 50 most differentially expressed transcripts in MR capsule tissue exposed to MeJA.

| Gene ID | MRCC | MRCT | log2 Fold | Description |
|---|---|---|---|---|
| Unigene48528_All | 0.01 | 35.5 | 11.8 | Major latex protein 149 |
| Unigene2672_All | 0.01 | 21.57 | 11.1 | Non-legume hemoglobin |
| CL8382.Contig2_All | 0.01 | 18.44 | 10.8 | Protein transport protein Sec61 subunit alpha |
| CL8634.Contig2_All | 0.01 | 16.68 | 10.7 | Serine/ carboxypeptidase-like 18 |
| CL7922.Contig2_All | 0.01 | 12.28 | 10.3 | Aquaporin TIP4-1 |
| CL2962.Contig13_All | 0.01 | 11.92 | 10.2 | Probable metal-nicotianamine transporter YSL16 |
| CL5244.Contig4_All | 0.01 | 11.41 | 10.2 | 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase |
| Unigene49386_All | 0.01 | 11.42 | 10.2 | Proline-rich receptor-like protein kinase PERK8 |
| CL9340.Contig3_All | 0.01 | 10.26 | 10 | Protein TIC 20-I, chloroplastic |
| Unigene49484_All | 0.01 | 9.43 | 9.9 | 2-oxoglutarate-dependent dioxygenase AOP2 |
| Unigene4577_All | 0.01 | 8.24 | 9.7 | Luminal-binding protein |
| CL2174.Contig8_All | 0.01 | 7.9 | 9.6 | Potassium transporter 2 |
| CL15614.Contig8_All | 0.01 | 6.85 | 9.4 | Pyruvate decarboxylase 1 |
| CL2454.Contig2_All | 0.01 | 6.99 | 9.4 | Late embryogenesis abundant protein 2 |
| CL120.Contig5_All | 0.01 | 6.44 | 9.3 | Transcription factor bHLH113 |
| CL12639.Contig2_All | 0.01 | 6.49 | 9.3 | S-norcoclaurine synthase |
| CL15726.Contig32_All | 0.01 | 6.41 | 9.3 | Protein TPX2 |
| CL2230.Contig10_All | 0.01 | 6.4 | 9.3 | ABC transporter G family member 11 |
| Unigene23170_All | 0.01 | 6.46 | 9.3 | F-box/LRR-repeat protein 15 |
| CL12126.Contig1_All | 0.01 | 5.87 | 9.2 | Peroxidase 64 |
| CL7285.Contig6_All | 0.01 | 5.7 | 9.2 | Chaperone protein dnaJ 15 |
| CL2458.Contig1_All | 0.01 | 5.6 | 9.1 | Zinc finger A20 and AN1 domain-containing stress-associated protein 4 |
| Unigene16279_All | 0.01 | 5.66 | 9.1 | (-)-alpha-terpineol synthase |

| Gene ID | MRCC | MRCT | log2 Fold | Description |
|---|---|---|---|---|
| CL1755.Contig1_All | 0.01 | 4.99 | 9 | Limonoid UDP-glucosyltransferase |
| CL4571.Contig13_All | 0.01 | 4.83 | 8.9 | Ubiquitin-conjugating enzyme E2 4 |
| CL4577.Contig1_All | 6.99 | 0.01 | -9.4 | Peroxidase N1 |
| Unigene1663_All | 7.14 | 0.01 | -9.5 | Ethylene-responsive transcription factor 1B |
| Unigene1907_All | 7.13 | 0.01 | -9.5 | Pathogenesis-related protein PR-4B |
| Unigene51064_All | 7.1 | 0.01 | -9.5 | ATP synthase subunit a |
| Unigene23723_All | 7.97 | 0.01 | -9.6 | Proteinase inhibitor |
| Unigene44390_All | 7.5 | 0.01 | -9.6 | Extensin |
| Unigene53777_All | 7.87 | 0.01 | -9.6 | Cytochrome P450 76A2 |
| CL15726.Contig35_All | 8.06 | 0.01 | -9.7 | Protein TPX2 |
| Unigene50907_All | 8.34 | 0.01 | -9.7 | Cytochrome c oxidase subunit 1 |
| Unigene53552_All | 8.47 | 0.01 | -9.7 | Cytochrome c oxidase subunit 3 |
| Unigene50431_All | 9.01 | 0.01 | -9.8 | Cytochrome P450 71B10 |
| Unigene31231_All | 9.39 | 0.01 | -9.9 | Zeaxanthin epoxidase, chloroplastic |
| CL3964.Contig1_All | 10.02 | 0.01 | -10 | Major allergen Pru ar 1 |
| Unigene1906_All | 11.11 | 0.01 | -10.1 | S-norcoclaurine synthase |
| CL2803.Contig1_All | 11.61 | 0.01 | -10.2 | Cytochrome P450 71D8 |
| CL11200.Contig1_All | 13.29 | 0.01 | -10.4 | Flavonoid 3'-monooxygenase |
| CL1478.Contig6_All | 13.81 | 0.01 | -10.4 | Extensin |
| Unigene26202_All | 14.01 | 0.01 | -10.5 | Peroxidase 12 |
| Unigene31732_All | 14.86 | 0.01 | -10.5 | Calcium-binding protein CML39 |
| Unigene43472_All | 14.38 | 0.01 | -10.5 | Linoleate 9S-lipoxygenase 1 |
| CL3589.Contig2_All | 17.34 | 0.01 | -10.8 | V-type proton ATPase subunit F |
| Unigene42364_All | 17.31 | 0.01 | -10.8 | Dirigent protein 23 |
| CL6821.Contig2_All | 20.68 | 0.01 | -11 | Chitinase 6 |
| Unigene19545_All | 21.96 | 0.01 | -11.1 | Peroxidase 52 |
| CL8863.Contig2_All | 61.53 | 0.01 | -12.6 | Inhibitor of trypsin and hageman factor |

**Supplementary Table 10.** Top 50 most differentially expressed transcripts in MR stem tissue exposed to MeJA.

| Gene ID | MRSC | MRST | log2 Fold | Description |
|---|---|---|---|---|
| Unigene46282_All | 0.01 | 122.53 | 13.58 | Cytochrome c oxidase subunit 2 |
| Unigene46058_All | 0.01 | 121.98 | 13.57 | Cytochrome c oxidase subunit 3 |
| Unigene44464_All | 0.01 | 40.03 | 11.97 | Formin-like protein 20 |
| Unigene44998_All | 0.01 | 27.97 | 11.45 | Glutenin, high molecular weight subunit DX5 |
| CL1099.Contig4_All | 0.01 | 24.56 | 11.26 | 11S globulin seed storage protein 2 |
| Unigene46011_All | 0.01 | 20.82 | 11.02 | NADH-ubiquinone oxidoreductase chain 3 |
| CL2174.Contig8_All | 0.01 | 15.19 | 10.57 | Potassium transporter 2 |
| CL6917.Contig2_All | 0.01 | 14.76 | 10.53 | Golgi apparatus membrane protein-like protein ECHIDNA |
| CL4158.Contig8_All | 0.01 | 13.74 | 10.42 | 12S seed storage globulin 2 |
| Unigene35911_All | 0.1 | 132.8 | 10.38 | Major latex protein 146 |
| CL7572.Contig11_All | 0.01 | 12.94 | 10.34 | 5-methyltetrahydro--pteroyltriglutamate--homocysteine methyltransferase |
| CL610.Contig2_All | 0.01 | 12.18 | 10.25 | Aquaporin TIP3-1 |
| Unigene44551_All | 0.01 | 10.86 | 10.08 | Polyubiquitin |
| CL12395.Contig4_All | 0.01 | 10.52 | 10.04 | Non-specific lipid-transfer protein 1 |
| CL9340.Contig3_All | 0.01 | 10.23 | 10.00 | Protein TIC 20-I, chloroplastic |
| Unigene44820_All | 0.01 | 10.2 | 9.99 | Ferritin-1, chloroplastic |
| CL6111.Contig3_All | 0.01 | 9.97 | 9.96 | Non-specific lipid-transfer protein 2 |
| CL3489.Contig1_All | 0.01 | 9.74 | 9.93 | Defensin-like protein |
| CL3726.Contig5_All | 0.01 | 9.2 | 9.85 | Zinc transporter 1 |
| CL1788.Contig3_All | 0.01 | 9.19 | 9.84 | 2S albumin |
| CL8382.Contig2_All | 0.01 | 8.83 | 9.79 | Protein transport protein Sec61 subunit alpha |
| CL15179.Contig28_All | 0.01 | 8.67 | 9.76 | (+)-neomenthol dehydrogenase |
| CL9369.Contig2_All | 0.01 | 8.43 | 9.72 | Probable sulfate transporter 3.4 |
| CL1084.Contig4_All | 0.01 | 8.41 | 9.72 | Proteasome subunit alpha type-5 |
| Unigene26240_All | 4.26 | 0.01 | -8.73 | Phosphate transporter PHO1 homolog |

| Gene ID | MRSC | MRST | log2 Fold | Description |
|---|---|---|---|---|
| CL488.Contig13_All | 4.4 | 0.01 | -8.78 | Sucrose synthase 2 |
| CL1918.Contig7_All | 4.6 | 0.01 | -8.85 | Dihydroflavonol-4-reductase |
| CL11174.Contig1_All | 4.62 | 0.01 | -8.85 | Serine/threonine-protein kinase Nek2 |
| CL3889.Contig3_All | 4.67 | 0.01 | -8.87 | Leucine-rich repeat extensin-like protein 3 |
| CL13479.Contig47_All | 4.68 | 0.01 | -8.87 | Ankyrin repeat-containing protein At3g12360 |
| CL14880.Contig1_All | 4.99 | 0.01 | -8.96 | Hypersensitive-induced response protein 1 |
| CL13479.Contig31_All | 5.25 | 0.01 | -9.04 | Ankyrin repeat-containing protein At3g12360 |
| CL13479.Contig1_All | 5.35 | 0.01 | -9.06 | Ankyrin repeat-containing protein At3g12360 |
| CL6569.Contig4_All | 5.46 | 0.01 | -9.09 | Glycine-rich RNA-binding protein 3, mitochondrial |
| CL2217.Contig4_All | 5.51 | 0.01 | -9.11 | Vesicle-associated protein 2-2 |
| CL5185.Contig2_All | 5.67 | 0.01 | -9.15 | Aspartic proteinase PCS1 |
| CL15726.Contig32_All | 6.15 | 0.01 | -9.26 | Protein TPX2 |
| CL4586.Contig9_All | 6.32 | 0.01 | -9.30 | Serine carboxypeptidase-like 1 |
| CL2392.Contig4_All | 6.5 | 0.01 | -9.34 | Cytochrome b5 |
| CL8516.Contig4_All | 6.85 | 0.01 | -9.42 | Phototropin-2 |
| CL2468.Contig2_All | 7.01 | 0.01 | -9.45 | Monothiol glutaredoxin-S17 |
| CL1519.Contig11_All | 7.5 | 0.01 | -9.55 | Succinate-semialdehyde dehydrogenase, mitochondrial |
| Unigene27347_All | 7.63 | 0.01 | -9.58 | Mediator of RNA polymerase II transcription subunit 16 |
| CL4739.Contig5_All | 7.82 | 0.01 | -9.61 | Nascent polypeptide-associated complex subunit alpha-like protein 2 |
| CL1918.Contig12_All | 7.99 | 0.01 | -9.64 | Dihydroflavonol-4-reductase |
| CL4820.Contig1_All | 8.1 | 0.01 | -9.66 | Leucoanthocyanidin dioxygenase |
| Unigene246_All | 8.2 | 0.01 | -9.68 | Quinone-oxidoreductase homolog, chloroplastic |
| CL8982.Contig3_All | 8.43 | 0.01 | -9.72 | Proteasome subunit alpha type-3 |
| CL13537.Contig3_All | 8.79 | 0.01 | -9.78 | Ornithine aminotransferase, mitochondrial |
| CL8331.Contig10_All | 10.71 | 0.01 | -10.06 | Glycine-rich protein A3 |

**Supplementary Table 11.** Top 50 most differentially expressed transcripts in MP capsule tissue exposed to MeJA.

| Gene ID | MPCC | MPCT | log2 Fold | Description |
|---|---|---|---|---|
| CL8331.Contig6_All | 0.01 | 49.29 | 12.27 | Glycine-rich protein A3 |
| CL488.Contig48_All | 0.01 | 39.01 | 11.93 | Sucrose synthase 4 |
| Unigene39453_All | 0.01 | 31.99 | 11.64 | Extensin-1 |
| Unigene10188_All | 0.01 | 29.72 | 11.54 | Chitin-binding lectin 1 |
| Unigene646_All | 0.01 | 29.28 | 11.52 | (S)-tetrahydroprotoberberine N-methyltransferase 1 |
| CL15790.Contig16_All | 0.01 | 23.75 | 11.21 | Cytochrome P450 CYP72A219 |
| CL7783.Contig4_All | 0.01 | 21.23 | 11.05 | Aspartate aminotransferase 3, chloroplastic |
| Unigene40910_All | 0.01 | 17.93 | 10.81 | Chitin-binding lectin 1 |
| CL4586.Contig11_All | 0.01 | 17.67 | 10.79 | Serine carboxypeptidase-like 18 |
| CL8634.Contig2_All | 0.01 | 17.6 | 10.78 | Serine carboxypeptidase-like 18 |
| CL12541.Contig4_All | 0.01 | 17.4 | 10.76 | Wall-associated receptor kinase-like 1 |
| Unigene39056_All | 0.01 | 17.39 | 10.76 | Extensin-2 |
| CL5418.Contig2_All | 0.01 | 16.7 | 10.71 | Vacuolar protein sorting-associated protein 32 homolog 2 |
| CL535.Contig3_All | 0.01 | 15.13 | 10.56 | Caffeoylshikimate esterase |
| Unigene18778_All | 0.01 | 13.65 | 10.41 | UDP-glycosyltransferase 89B1 |
| Unigene39507_All | 0.01 | 12.18 | 10.25 | MLP-like protein 34 |
| Unigene23985_All | 0.01 | 11.62 | 10.18 | NA |
| CL11174.Contig4_All | 0.01 | 11.1 | 10.12 | Serine/threonine-protein kinase Nek2 |
| CL488.Contig13_All | 0.01 | 10.69 | 10.06 | Sucrose synthase 2 |
| CL13521.Contig5_All | 0.01 | 10.59 | 10.05 | Ninja-family protein 1 |
| CL414.Contig10_All | 0.01 | 9.74 | 9.93 | Rae1-like protein At1g80670 |
| CL2835.Contig4_All | 0.01 | 9.5 | 9.89 | Retrovirus-related Pol polyprotein from transposon TNT 1-94 |
| Unigene16736_All | 0.01 | 9.26 | 9.85 | Formin-like protein 3 |
| CL11400.Contig2_All | 0.01 | 8.17 | 9.67 | Elicitor-responsive protein 1 |
| CL15726.Contig30_All | 0.01 | 7.78 | 9.60 | Protein TPX2 |

| Gene ID | MPCC | MPCT | log2 Fold | Description |
|---|---|---|---|---|
| Unigene59582_All | 24.98 | 0.01 | -11.29 | Period clock protein |
| Unigene58818_All | 25.84 | 0.01 | -11.34 | Period clock protein |
| Unigene62165_All | 27.27 | 0.01 | -11.41 | Cytochrome c oxidase subunit 6b-1 |
| Unigene57982_All | 27.68 | 0.01 | -11.43 | Cytochrome c |
| Unigene58942_All | 28.22 | 0.01 | -11.46 | Kinesin-4 |
| Unigene61841_All | 31.02 | 0.01 | -11.60 | Cytochrome c oxidase subunit 3 |
| CL11523.Contig2_All | 31.89 | 0.01 | -11.64 | Actin-1 |
| Unigene178_All | 33.75 | 0.01 | -11.72 | Germin-like protein subfamily T member 2 |
| Unigene61537_All | 34.79 | 0.01 | -11.76 | Cytochrome c oxidase subunit 2 |
| CL8464.Contig2_All | 38.65 | 0.01 | -11.92 | Non-specific lipid-transfer protein 1 |
| CL15259.Contig1_All | 40.85 | 0.01 | -12.00 | Calmodulin |
| CL15259.Contig2_All | 63.25 | 0.01 | -12.63 | Calmodulin |
| Unigene61069_All | 18.39 | 0.01 | -10.84 | Cytochrome b-c1 complex subunit 7-2 |
| Unigene60236_All | 18.91 | 0.01 | -10.88 | Citrate synthase 4, mitochondrial |
| Unigene61680_All | 19.34 | 0.01 | -10.92 | Glycine-rich cell wall structural protein |
| Unigene61686_All | 20.02 | 0.01 | -10.97 | Proline-rich protein 4 |
| Unigene61360_All | 20.29 | 0.01 | -10.99 | Cytochrome c1 2, heme protein, mitochondrial |
| Unigene61533_All | 20.43 | 0.01 | -11.00 | NADH-ubiquinone oxidoreductase chain 1 |
| Unigene59583_All | 20.75 | 0.01 | -11.02 | Period clock protein |
| Unigene61112_All | 20.8 | 0.01 | -11.02 | ATP synthase subunit delta', mitochondrial |
| CL960.Contig1_All | 21.3 | 0.01 | -11.06 | Low-temperature-induced 65 kDa protein |
| CL8300.Contig1_All | 21.86 | 0.01 | -11.09 | WEB family protein At3g02930, chloroplastic |
| Unigene59949_All | 22.98 | 0.01 | -11.17 | ATP synthase subunit O, mitochondrial |
| Unigene61631_All | 24.05 | 0.01 | -11.23 | Calmodulin |
| Unigene58821_All | 24.29 | 0.01 | -11.25 | Period clock protein |

**Supplementary Table 12.** Top 50 most differentially expressed transcripts in MP stem tissue exposed to MeJA.

| Gene ID | MPSC | MRST | log2 Fold | Description |
|---|---|---|---|---|
| CL1013.Contig4_All | 0.01 | 46.29 | 12.18 | V-type proton ATPase subunit G1 |
| CL8982.Contig4_All | 0.01 | 35.36 | 11.79 | Proteasome subunit alpha type-3 |
| CL31.Contig3_All | 0.01 | 21.62 | 11.08 | serine/arginine-rich splicing factor 33 |
| CL5024.Contig5_All | 0.01 | 21.23 | 11.05 | Caffeoylshikimate esterase |
| CL5116.Contig5_All | 0.01 | 17.75 | 10.79 | Uridine kinase-like protein 5 |
| CL1755.Contig1_All | 0.01 | 17.65 | 10.79 | Limonoid UDP-glucosyltransferase |
| Unigene55563_All | 0.01 | 15.66 | 10.61 | Glucan endo-1,3-beta-glucosidase 14 |
| CL1738.Contig10_All | 0.01 | 14.68 | 10.52 | Diacylglycerol kinase 5 |
| Unigene36095_All | 0.01 | 13.95 | 10.45 | Glycine--tRNA ligase 1, mitochondrial |
| CL2805.Contig1_All | 0.01 | 13.45 | 10.39 | Glucan endo-1,3-beta-glucosidase |
| CL3689.Contig7_All | 0.01 | 11.5 | 10.17 | Myosin-15 |
| CL8331.Contig6_All | 0.01 | 10.93 | 10.09 | Glycine-rich protein A3 |
| CL6746.Contig1_All | 0.01 | 9.79 | 9.94 | EG45-like domain containing protein |
| Unigene27906_All | 0.01 | 9.79 | 9.94 | Cytochrome P450 94A2 |
| CL414.Contig10_All | 0.01 | 9.61 | 9.91 | Rae1-like protein At1g80670 |
| Unigene22980_All | 0.01 | 9.22 | 9.85 | F-box protein At4g19940 |
| CL639.Contig10_All | 0.01 | 9.13 | 9.83 | DEAD-box ATP-dependent RNA helicase 56 |
| Unigene21219_All | 0.01 | 8.78 | 9.78 | Cytochrome b6-f complex subunit 4 |
| CL1533.Contig4_All | 0.01 | 7.77 | 9.60 | Ferritin-3, chloroplastic |
| CL1828.Contig3_All | 0.01 | 7.75 | 9.60 | Pheophytinase, chloroplastic |
| Unigene3004_All | 0.01 | 7.47 | 9.54 | Hsp70-Hsp90 organizing protein 2 |
| CL16008.Contig2_All | 0.01 | 7.04 | 9.46 | Receptor-like protein kinase HSL1 |
| CL6746.Contig6_All | 0.01 | 6.87 | 9.42 | EG45-like domain containing protein |
| CL7783.Contig4_All | 0.01 | 6.79 | 9.41 | Aspartate aminotransferase 3, chloroplastic |
| Unigene21839_All | 0.01 | 6.77 | 9.40 | Probable galactinol--sucrose galactosyltransferase 2 |
| Unigene20567_All | 8.68 | 0.01 | -9.76 | Probable mediator of RNA polymerase II transcription subunit 37c |

| Gene ID | MPSC | MRST | log2 Fold | Description |
|---|---|---|---|---|
| CL16302.Contig4_All | 8.7 | 0.01 | -9.76 | Eukaryotic translation initiation factor 3 subunit A |
| CL2938.Contig3_All | 8.81 | 0.01 | -9.78 | 26S proteasome non-ATPase regulatory subunit 12 homolog A |
| CL639.Contig16_All | 8.87 | 0.01 | -9.79 | DEAD-box ATP-dependent RNA helicase 56 |
| CL2755.Contig3_All | 10.1 | 0.01 | -9.98 | Cytochrome P450 76C4 |
| CL236.Contig2_All | 10.29 | 0.01 | -10.01 | U-box domain-containing protein 72 |
| CL10949.Contig1_All | 10.47 | 0.01 | -10.03 | Xyloglucan endotransglucosylase/hydrolase protein 22 |
| CL2469.Contig2_All | 11.16 | 0.01 | -10.12 | Arginine/serine-rich-splicing factor RSP31 |
| CL1419.Contig1_All | 11.22 | 0.01 | -10.13 | UDP-glucuronic acid decarboxylase 6 |
| CL13479.Contig31_All | 11.33 | 0.01 | -10.15 | Ankyrin repeat-containing protein At3g12360 |
| CL10568.Contig1_All | 12.2 | 0.01 | -10.25 | Protein BUD31 homolog 2 |
| CL5053.Contig9_All | 12.41 | 0.01 | -10.28 | Protein RESTRICTED TEV MOVEMENT 2 |
| CL14228.Contig1_All | 14.31 | 0.01 | -10.48 | Boron transporter 4 |
| Unigene24917_All | 14.42 | 0.01 | -10.49 | Probable mediator of RNA polymerase II transcription subunit 37c |
| Unigene37189_All | 15.4 | 0.01 | -10.59 | Glucan endo-1,3-beta-glucosidase 14 |
| CL1654.Contig1_All | 17.04 | 0.01 | -10.73 | NA |
| CL14228.Contig5_All | 17.05 | 0.01 | -10.74 | Boron transporter 4 |
| CL1244.Contig8_All | 17.21 | 0.01 | -10.75 | MLP-like protein 43 |
| CL4768.Contig3_All | 18.67 | 0.01 | -10.87 | 60S ribosomal protein L36-2 |
| CL3716.Contig1_All | 19.71 | 0.01 | -10.94 | Thaumatin-like protein |
| CL7783.Contig1_All | 20.3 | 0.01 | -10.99 | Aspartate aminotransferase, cytoplasmic |
| CL3232.Contig1_All | 21.06 | 0.01 | -11.04 | Putative AC9 transposase |
| CL5816.Contig5_All | 22.42 | 0.01 | -11.13 | Molybdopterin synthase sulfur carrier subunit |
| CL2458.Contig1_All | 30.96 | 0.01 | -11.60 | Zinc finger A20 and AN1 domain-containing stress-associated protein 4 |
| CL4042.Contig1_All | 52.46 | 0.01 | -12.36 | UPF0057 membrane protein At4g30660 |

**Supplementary Table 13.** Top 50 most differentially expressed transcripts in MR capsule and stem tissues exposed to MeJA.

| Gene ID | MRCT | MRST | log2 Fold | Description |
|---|---|---|---|---|
| Unigene32119_All | 0.01 | 133.14 | 13.70 | Metallothionein-like protein type 2 |
| Unigene37766_All | 0.01 | 78.92 | 12.95 | Monothiol glutaredoxin-S7 |
| CL1738.Contig7_All | 0.01 | 62.56 | 12.61 | Protein THYLAKOID FORMATION1, chloroplastic |
| Unigene36493_All | 0.01 | 42.05 | 12.04 | Protein HEADING DATE 3A |
| Unigene44464_All | 0.01 | 40.03 | 11.97 | Formin-like protein 20 |
| CL7567.Contig4_All | 0.01 | 34.85 | 11.77 | 1-aminocyclopropane-1-carboxylate oxidase |
| Unigene39509_All | 0.01 | 10.67 | 10.06 | Anther-specific proline-rich protein APG (Fragment) |
| CL12466.Contig2_All | 0.01 | 10.32 | 10.01 | Aspartic proteinase CDR1 |
| CL5473.Contig2_All | 0.01 | 10.28 | 10.01 | MADS-box protein JOINTLESS |
| CL10988.Contig1_All | 0.01 | 10.27 | 10.00 | Auxin response factor 4 |
| Unigene44820_All | 0.01 | 10.2 | 9.99 | Ferritin-1, chloroplastic |
| CL666.Contig2_All | 0.01 | 10.18 | 9.99 | UDP-glycosyltransferase 85A3 |
| Unigene2274_All | 0.01 | 26.72 | 11.38 | Mannose-specific lectin |
| CL3234.Contig5_All | 0.01 | 23.78 | 11.22 | Cytochrome P450 87A3 |
| Unigene17574_All | 0.01 | 23.14 | 11.18 | Glutaredoxin-C13 |
| CL2134.Contig2_All | 0.01 | 21.45 | 11.07 | Dehydration-responsive protein RD22 |
| Unigene46011_All | 0.01 | 20.82 | 11.02 | NADH-ubiquinone oxidoreductase chain 3 |
| CL8075.Contig1_All | 0.01 | 19.72 | 10.95 | Glyceraldehyde-3-phosphate dehydrogenase GAPCP1, chloroplastic |
| CL4266.Contig1_All | 0.01 | 17.6 | 10.78 | MLP-like protein 34 |
| CL988.Contig4_All | 0.01 | 16.84 | 10.72 | Peroxidase 51 |
| CL2010.Contig2_All | 0.01 | 14.61 | 10.51 | Nudix hydrolase 8 |
| CL1579.Contig2_All | 0.01 | 14.08 | 10.46 | Glucan endo-1,3-beta-glucosidase |
| CL988.Contig2_All | 0.01 | 13.25 | 10.37 | Peroxidase 51 |
| Unigene1907_All | 0.01 | 12.15 | 10.25 | Pathogenesis-related protein PR-4B |
| Unigene1663_All | 0.01 | 11.39 | 10.15 | Ethylene-responsive transcription factor 1B |
| Unigene1201_All | 6.62 | 0.01 | -9.37 | Transcription factor MYB12 |

| Gene ID | MRCT | MRST | log2 Fold | Description |
|---|---|---|---|---|
| Unigene21839_All | 7.66 | 0.01 | -9.58 | Probable galactinol--sucrose galactosyltransferase 2 |
| Unigene29087_All | 7.79 | 0.01 | -9.61 | Cysteine-rich repeat secretory protein 3 |
| Unigene4577_All | 8.24 | 0.01 | -9.69 | Luminal-binding protein |
| Unigene18566_All | 8.26 | 0.01 | -9.69 | Serine/threonine-protein kinase AFC1 |
| CL1453.Contig2_All | 8.28 | 0.01 | -9.69 | Zinc finger protein CONSTANS-LIKE 4 |
| Unigene26612_All | 8.81 | 0.01 | -9.78 | Peroxidase 72 |
| CL1865.Contig3_All | 8.9 | 0.01 | -9.80 | Cation/H(+) antiporter 18 |
| CL2962.Contig4_All | 10.97 | 0.01 | -10.10 | Metal-nicotianamine transporter YSL3 |
| Unigene31999_All | 11.12 | 0.01 | -10.12 | Luminal-binding protein |
| Unigene49386_All | 11.42 | 0.01 | -10.16 | Proline-rich receptor-like protein kinase PERK8 |
| CL2468.Contig2_All | 11.9 | 0.01 | -10.22 | Monothiol glutaredoxin-S17 |
| Unigene246_All | 11.99 | 0.01 | -10.23 | Quinone-oxidoreductase homolog, chloroplastic |
| Unigene18938_All | 13.01 | 0.01 | -10.35 | Thaumatin-like protein 1 |
| Unigene24990_All | 13.04 | 0.01 | -10.35 | Aquaporin NIP2-2 |
| CL5990.Contig6_All | 13.51 | 0.01 | -10.40 | Cytochrome P450 71A1 |
| CL4586.Contig9_All | 13.79 | 0.01 | -10.43 | Serine carboxypeptidase-like 1 |
| CL13537.Contig3_All | 14.81 | 0.01 | -10.53 | Ornithine aminotransferase, mitochondrial |
| CL5154.Contig1_All | 16.51 | 0.01 | -10.69 | Extensin |
| CL4158.Contig5_All | 16.86 | 0.01 | -10.72 | 12S seed storage globulin 2 |
| Unigene2672_All | 21.57 | 0.01 | -11.07 | Non-legume hemoglobin |
| CL7680.Contig7_All | 29.29 | 0.01 | -11.52 | Protein Ycf2 |
| Unigene26993_All | 29.61 | 0.01 | -11.53 | MLP-like protein 43 |
| CL8331.Contig10_All | 42.59 | 0.01 | -12.06 | Glycine-rich protein A3 |
| CL8464.Contig2_All | 46.51 | 0.01 | -12.18 | Non-specific lipid-transfer protein 1 |

**Supplementary Table 14.** Top 50 most differentially expressed transcripts in MP capsule and stem tissues exposed to MeJA.

| Gene ID | MPCT | MPST | log2 Fold | Description |
|---|---|---|---|---|
| Unigene29504_All | 0.01 | 108.14 | 13.40 | MLP-like protein 43 |
| CL942.Contig1_All | 0.01 | 57.23 | 12.48 | Induced stolen tip protein TUB8 (Fragment) |
| CL4308.Contig10_All | 0.01 | 56.9 | 12.47 | Major latex protein 149 |
| CL5245.Contig4_All | 0.01 | 43.63 | 12.09 | Bidirectional sugar transporter SWEET3b |
| Unigene29132_All | 0.01 | 43.29 | 12.08 | Major latex protein 146 |
| CL8357.Contig3_All | 0.01 | 42.83 | 12.06 | Major latex protein 146 |
| CL3978.Contig2_All | 0.01 | 41.64 | 12.02 | Zinc finger A20 and AN1 domain-containing stress-associated protein 5 |
| CL2458.Contig2_All | 0.01 | 40.73 | 11.99 | Zinc finger A20 and AN1 domain-containing stress-associated protein 4 |
| Unigene30047_All | 0.01 | 38.74 | 11.92 | Major latex protein 146 |
| Unigene29131_All | 0.01 | 35.48 | 11.79 | Major latex protein 146 |
| CL8553.Contig2_All | 0.01 | 31.08 | 11.60 | Cysteine-rich repeat secretory protein 55 |
| CL7614.Contig2_All | 0.01 | 26.25 | 11.36 | Cytochrome P450 93A1 |
| CL7042.Contig1_All | 0.01 | 26.2 | 11.36 | Dirigent protein 23 |
| CL2010.Contig1_All | 0.01 | 23.55 | 11.20 | Nudix hydrolase 8 |
| Unigene31853_All | 0.01 | 23.47 | 11.20 | Zinc finger protein 4 |
| CL94.Contig6_All | 0.01 | 20.6 | 11.01 | Putative 3,4-dihydroxy-2-butanone kinase |
| CL3746.Contig4_All | 0.01 | 20.15 | 10.98 | Cytochrome P450 82C4 |
| CL12189.Contig1_All | 0.01 | 19.6 | 10.94 | Histidine-containing phosphotransfer protein 4 |
| CL1755.Contig1_All | 0.01 | 17.65 | 10.79 | Limonoid UDP-glucosyltransferase |
| Unigene27608_All | 0.01 | 14.09 | 10.46 | Extensin |
| CL14522.Contig1_All | 0.01 | 14.06 | 10.46 | Cytochrome P450 84A1 |
| CL3964.Contig1_All | 0.01 | 13.71 | 10.42 | Major allergen Pru ar 1 |
| CL7567.Contig4_All | 0.01 | 13.56 | 10.41 | 1-aminocyclopropane-1-carboxylate oxidase |
| CL8553.Contig4_All | 0.01 | 13.52 | 10.40 | Cysteine-rich repeat secretory protein 55 |
| CL12126.Contig3_All | 0.01 | 12.97 | 10.34 | Peroxidase 64 |

| Gene ID | MPCT | MPST | log2 Fold | Description |
|---|---|---|---|---|
| Unigene646_All | 29.28 | 0.01 | -11.52 | (S)-tetrahydroprotoberberine N-methyltransferase 1 |
| CL9555.Contig3_All | 32.01 | 0.01 | -11.64 | Cytochrome P450 71A1 |
| Unigene26258_All | 35.46 | 0.01 | -11.79 | Ethylene-responsive transcription factor RAP2-3 |
| CL610.Contig2_All | 37.09 | 0.01 | -11.86 | Aquaporin TIP3-1 |
| CL9555.Contig2_All | 39.52 | 0.01 | -11.95 | Cytochrome P450 71A1 |
| Unigene19857_All | 39.74 | 0.01 | -11.96 | Late embryogenesis abundant protein Dc3 |
| CL1478.Contig2_All | 41.47 | 0.01 | -12.02 | Extensin |
| CL3230.Contig4_All | 43.56 | 0.01 | -12.09 | Extensin |
| Unigene58032_All | 46.95 | 0.01 | -12.20 | EMB-1 protein |
| CL2092.Contig4_All | 52.33 | 0.01 | -12.35 | Late embryogenesis abundant protein B19.3 |
| CL7279.Contig3_All | 54.85 | 0.01 | -12.42 | Beta-glucosidase 25 |
| CL9221.Contig2_All | 63.28 | 0.01 | -12.63 | Glycine-rich protein HC1 |
| Unigene24806_All | 65.21 | 0.01 | -12.67 | Glycine-rich protein HC1 |
| CL2092.Contig2_All | 65.52 | 0.01 | -12.68 | Late embryogenesis abundant protein B19.4 |
| Unigene24811_All | 72.06 | 0.01 | -12.81 | Glycine-rich protein 3 short isoform |
| Unigene1830_All | 73.89 | 0.01 | -12.85 | Late embryogenesis abundant protein Lea14-A |
| Unigene37967_All | 74.7 | 0.01 | -12.87 | Non-specific lipid-transfer protein 3 |
| CL2738.Contig3_All | 75.85 | 0.01 | -12.89 | Non-specific lipid-transfer protein |
| Unigene36946_All | 100.76 | 0.01 | -13.30 | Late embryogenesis abundant protein B19.3 |
| CL2676.Contig2_All | 119.05 | 0.01 | -13.54 | Late embryogenesis abundant protein D-29 |
| Unigene17720_All | 127.69 | 0.01 | -13.64 | Antimicrobial peptide 1 |
| CL2738.Contig9_All | 136.15 | 0.01 | -13.73 | Non-specific lipid-transfer protein |
| Unigene44410_All | 140.6 | 0.01 | -13.78 | EMB-1 protein |
| Unigene45693_All | 142.45 | 0.01 | -13.80 | Non-specific lipid-transfer protein |
| Unigene39594_All | 239.81 | 0.01 | -14.55 | Non-specific lipid-transfer protein 1 |

**Supplementary Table 15.** Top 50 most differentially expressed transcripts in MR and MP capsule tissues exposed to MeJA.

| Gene ID | MRCT | MPCT | log2 Fold | Description |
|---|---|---|---|---|
| Unigene38886_All | 0.01 | 55.61 | 12.44 | NA |
| CL5401.Contig2_All | 0.01 | 44.67 | 12.13 | Probable S-adenosylmethionine-dependent methyltransferase At5g38780 |
| Unigene25473_All | 0.01 | 31.45 | 11.62 | Purothionin A-1 |
| Unigene10188_All | 0.01 | 29.72 | 11.54 | Chitin-binding lectin 1 |
| CL666.Contig2_All | 0.01 | 28.21 | 11.46 | UDP-glycosyltransferase 85A3 |
| CL15614.Contig5_All | 0.01 | 27.67 | 11.43 | Pyruvate decarboxylase 1 |
| CL599.Contig15_All | 0.01 | 26.9 | 11.39 | BTB/POZ and TAZ domain-containing protein 1 |
| Unigene23519_All | 0.01 | 24.39 | 11.25 | Extensin-2 |
| CL7783.Contig4_All | 0.01 | 21.23 | 11.05 | Aspartate aminotransferase 3, chloroplastic |
| CL1860.Contig5_All | 0.01 | 20.67 | 11.01 | Protein TRANSPARENT TESTA 12 |
| CL359.Contig8_All | 0.01 | 19.64 | 10.94 | Disease resistance protein RGA2 |
| CL4272.Contig1_All | 0.01 | 18.5 | 10.85 | Tubulin beta-4 chain |
| Unigene39056_All | 0.01 | 17.39 | 10.76 | Extensin-2 |
| CL3978.Contig3_All | 0.01 | 16.8 | 10.71 | Zinc finger A20 and AN1 domain-containing stress-associated protein 5 |
| CL7948.Contig3_All | 0.01 | 15.85 | 10.63 | S-norcoclaurine synthase |
| CL535.Contig3_All | 0.01 | 15.13 | 10.56 | Caffeoylshikimate esterase |
| CL7279.Contig2_All | 0.01 | 14.39 | 10.49 | Beta-glucosidase 25 |
| Unigene57779_All | 0.01 | 13.8 | 10.43 | Beta-glucosidase 29 |
| CL599.Contig4_All | 0.01 | 13.38 | 10.39 | BTB/POZ and TAZ domain-containing protein 1 |
| CL6638.Contig7_All | 0.01 | 12.63 | 10.30 | 6-phosphofructokinase 3 |
| CL13183.Contig2_All | 0.01 | 12.19 | 10.25 | Indole-3-acetic acid-amido synthetase GH3.6 |
| Unigene39507_All | 0.01 | 12.18 | 10.25 | MLP-like protein 34 |
| CL2750.Contig3_All | 0.01 | 12.16 | 10.25 | Formin-like protein 20 |
| CL3887.Contig2_All | 0.01 | 11.69 | 10.19 | Transcription factor MYB39 |
| CL2921.Contig4_All | 0.01 | 10.8 | 10.08 | Cytochrome P450 71A9 |
| CL3903.Contig6_All | 10.01 | 0.01 | -9.97 | RING-H2 finger protein ATL54 |

| Gene ID | MRCT | MPCT | log2 Fold | Description |
|---|---|---|---|---|
| Unigene31876_All | 10.14 | 0.01 | -9.99 | Transcription factor DIVARICATA |
| Unigene31323_All | 11.03 | 0.01 | -10.11 | LOB domain-containing protein 4 |
| CL6156.Contig1_All | 12.55 | 0.01 | -10.29 | Protein disulfide-isomerase 5-2 |
| Unigene44510_All | 12.91 | 0.01 | -10.33 | BTB/POZ and MATH domain-containing protein 3 |
| CL11375.Contig1_All | 12.97 | 0.01 | -10.34 | Monothiol glutaredoxin-S9 |
| CL9246.Contig2_All | 13.48 | 0.01 | -10.40 | Ras-related protein Rab5 |
| CL12992.Contig2_All | 13.75 | 0.01 | -10.43 | Aspartate aminotransferase, cytoplasmic |
| CL599.Contig5_All | 14.65 | 0.01 | -10.52 | BTB/POZ and TAZ domain-containing protein 1 |
| CL3978.Contig2_All | 14.96 | 0.01 | -10.55 | Zinc finger A20 and AN1 domain-containing stress-associated protein 5 |
| CL1084.Contig7_All | 15.86 | 0.01 | -10.63 | Proteasome subunit alpha type-5 |
| CL4158.Contig5_All | 16.86 | 0.01 | -10.72 | 12S seed storage globulin 2 |
| CL31.Contig25_All | 17.63 | 0.01 | -10.78 | Serine/arginine-rich splicing factor 33 |
| Unigene13736_All | 18.33 | 0.01 | -10.84 | Cyanidin 3-O-glucoside 7-O-glucosyltransferase (acyl-glucose) |
| Unigene38507_All | 20.55 | 0.01 | -11.00 | Probable S-adenosylmethionine-dependent methyltransferase At5g38780 |
| CL2458.Contig2_All | 24.84 | 0.01 | -11.28 | Zinc finger A20 and AN1 domain-containing stress-associated protein 4 |
| CL14345.Contig2_All | 28.58 | 0.01 | -11.48 | Vesicle-associated protein 1-3 |
| CL7680.Contig7_All | 29.29 | 0.01 | -11.52 | Protein Ycf2 |
| Unigene24180_All | 31.49 | 0.01 | -11.62 | Nuclear transcription factor Y subunit B-6 |
| CL8982.Contig4_All | 32.51 | 0.01 | -11.67 | Proteasome subunit alpha type-3 |
| CL9540.Contig1_All | 34.38 | 0.01 | -11.75 | 11-beta-hydroxysteroid dehydrogenase 1B |
| CL942.Contig1_All | 36.82 | 0.01 | -11.85 | Induced stolen tip protein TUB8 (Fragment) |
| CL11145.Contig3_All | 44.89 | 0.01 | -12.13 | 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 1 |
| CL8464.Contig2_All | 46.51 | 0.01 | -12.18 | Non-specific lipid-transfer protein 1 |
| CL13591.Contig5_All | 81.22 | 0.01 | -12.99 | Antimicrobial peptide 1 |

**Supplementary Table 16.** Top 50 most differentially expressed transcripts in MR and MP stem tissues exposed to MeJA.

| Gene ID | MRST | MPST | log2 Fold | Description |
|---|---|---|---|---|
| CL4308.Contig10_All | 0.01 | 56.9 | 12.47 | Major latex protein 149 |
| CL3746.Contig1_All | 0.01 | 56.47 | 12.46 | Cytochrome P450 82A1 (Fragment) |
| CL3978.Contig3_All | 0.01 | 48.23 | 12.24 | Zinc finger A20 and AN1 domain-containing stress-associated protein 5 |
| CL599.Contig3_All | 0.01 | 48.18 | 12.23 | BTB/POZ and TAZ domain-containing protein 1 |
| CL1013.Contig4_All | 0.01 | 46.29 | 12.18 | V-type proton ATPase subunit G1 |
| CL3746.Contig2_All | 0.01 | 37.96 | 11.89 | Cytochrome P450 82C2 |
| CL2750.Contig3_All | 0.01 | 23.33 | 11.19 | Formin-like protein 20 |
| CL7948.Contig3_All | 0.01 | 22.7 | 11.15 | S-norcoclaurine synthase |
| CL570.Contig17_All | 0.01 | 22.44 | 11.13 | Salutaridine reductase |
| CL3706.Contig4_All | 0.01 | 22.01 | 11.10 | Thiamine pyrophosphokinase 1 |
| CL3746.Contig4_All | 0.01 | 20.15 | 10.98 | Cytochrome P450 82C4 |
| CL599.Contig4_All | 0.01 | 18.94 | 10.89 | BTB/POZ and TAZ domain-containing protein 1 |
| CL832.Contig1_All | 0.01 | 18.6 | 10.86 | Protease Do-like 1, chloroplastic |
| CL5816.Contig3_All | 0.01 | 17.73 | 10.79 | Molybdopterin synthase sulfur carrier subunit |
| CL109.Contig5_All | 0.01 | 17.51 | 10.77 | Glycine--tRNA ligase 1, mitochondrial |
| Unigene36095_All | 0.01 | 13.95 | 10.45 | Glycine--tRNA ligase 1, mitochondrial |
| CL3903.Contig7_All | 0.01 | 12.67 | 10.31 | RING-H2 finger protein ATL54 |
| CL599.Contig15_All | 0.01 | 12.53 | 10.29 | BTB/POZ and TAZ domain-containing protein 1 |
| CL5442.Contig4_All | 0.01 | 12.25 | 10.26 | Heat shock cognate protein 80 |
| Unigene7968_All | 0.01 | 11.94 | 10.22 | F-box/kelch-repeat protein At3g06240 |
| CL3689.Contig7_All | 0.01 | 11.5 | 10.17 | Myosin-15 |
| CL12099.Contig3_All | 0.01 | 11.36 | 10.15 | Actin-related protein 8 |
| Unigene32387_All | 0.01 | 10.35 | 10.02 | Stromal 70 kDa heat shock-related protein, chloroplastic |
| Unigene17049_All | 0.03 | 28.96 | 9.91 | (S)-tetrahydroprotoberberine N-methyltransferase |
| Unigene6106_All | 0.01 | 9.34 | 9.87 | Disease resistance protein RGA2 |

| Gene ID | MRST | MPST | log2 Fold | Description |
|---|---|---|---|---|
| CL10453.Contig4_All | 14.96 | 0.01 | -10.55 | Q3Protein NRT1/ PTR FAMILY 7.1 |
| CL1607.Contig3_All | 14.96 | 0.01 | -10.55 | MATE efflux family protein 1 |
| CL4158.Contig16_All | 15.8 | 0.01 | -10.63 | 11S globulin seed storage protein G3 |
| CL2548.Contig5_All | 16.81 | 0.01 | -10.72 | E3 ubiquitin-protein ligase RGLG2 |
| Unigene36603_All | 16.82 | 0.01 | -10.72 | Leucine-rich repeat extensin-like protein 2 |
| CL1693.Contig2_All | 18.19 | 0.01 | -10.83 | Glucan endo-1,3-beta-glucosidase 8 |
| CL10967.Contig1_All | 18.59 | 0.01 | -10.86 | S-norcoclaurine synthase |
| CL1538.Contig1_All | 18.66 | 0.01 | -10.87 | Leaf-specific thionin |
| CL10568.Contig1_All | 19.18 | 0.01 | -10.91 | Protein BUD31 homolog 2 |
| CL599.Contig5_All | 19.21 | 0.01 | -10.91 | BTB/POZ and TAZ domain-containing protein 1 |
| CL8075.Contig1_All | 19.72 | 0.01 | -10.95 | Glyceraldehyde-3-phosphate dehydrogenase GAPCP1, chloroplastic |
| CL12547.Contig2_All | 20.34 | 0.01 | -10.99 | Non-specific lipid-transfer protein 4 |
| Unigene46011_All | 20.82 | 0.01 | -11.02 | NADH-ubiquinone oxidoreductase chain 3 |
| CL2134.Contig2_All | 21.45 | 0.01 | -11.07 | Dehydration-responsive protein RD22 |
| CL1755.Contig2_All | 21.73 | 0.01 | -11.09 | Cinnamate beta-D-glucosyltransferase |
| CL5053.Contig6_All | 22.51 | 0.01 | -11.14 | Protein RESTRICTED TEV MOVEMENT 2 |
| CL2738.Contig3_All | 22.83 | 0.01 | -11.16 | Non-specific lipid-transfer protein |
| CL1582.Contig23_All | 25.6 | 0.01 | -11.32 | Proline transporter 1 |
| CL5816.Contig5_All | 30.66 | 0.01 | -11.58 | Molybdopterin synthase sulfur carrier subunit |
| CL3149.Contig7_All | 31.03 | 0.01 | -11.60 | Aspartic proteinase-like protein 1 |
| CL1582.Contig3_All | 31.09 | 0.01 | -11.60 | Proline transporter 1 |
| CL1244.Contig8_All | 31.57 | 0.01 | -11.62 | MLP-like protein 43 |
| Unigene44464_All | 40.03 | 0.01 | -11.97 | Formin-like protein 20 |
| CL6646.Contig5_All | 41.17 | 0.01 | -12.01 | Tropinone reductase homolog At1g07440 |
| Unigene44712_All | 60.85 | 0.01 | -12.57 | S-norcoclaurine synthase |

# RESUME

**Name & Surname** Kuaybe Yücebilgili Kurtoğlu

**Date of Birth** 24.10.1986

**Education**

Marmara University, Biology Department, PhD (2011-2016) GPA: 4.00 /4.00

Fatih University, Biology Department, MS (2009-2011) GPA: 4.00 /4.00

Fatih University, Biology Department, BS (2004-2008) GPA: 3.9 /4.00

California State University, Northridge, USA, Exchange Program (2008 Spring)

**Experience**

Research Assistant, Department of Biology - Marmara University (2011-2016)

Research Assistant, Department of Biology - Fatih University (2009-2011)

**Honors & Awards**

TUBITAK BIDEB PhD Scholarship (2011-2016)
Government Education Scholarship (2004-2008)
Fatih University High Honour Certificates (2004-2008)

**Papers**

Ozyigit Ibrahim Ilker, Ertugrul Filiz, Vatansever Recep, Yucebilgili Kurtoglu Kuaybe, Koc Ibrahim, Ozturk Munir, Anjum Naser: *Identification and Comparative Analysis of H2O2-Scavenging Enzymes (Ascorbate Peroxidase and Glutathione Peroxidase) in Selected Plants Employing Bioinformatics Approaches*. Frontiers in Plant Science, 2016; 7(301), DOI:10.3389/fpls.2016.00301.

Turktas Mine, Yucebilgili Kurtoglu Kuaybe, Dorado Gabriel, Zhang Baohong, Hernandez Pilar, Unver Turgay: *Sequencing of plant genomes -A review*. Turkish Journal of Agriculture and Forestry 2014; 39(3). DOI:10.3906/tar-1409-93

Yucebilgili Kurtoglu Kuaybe, Kantar Melda, Budak Hikmet: *New wheat microRNA using whole-genome sequence*. Functional & Integrative Genomics 2014; 14(2). DOI:10.1007/s10142-013-0357-9

Yucebilgili Kurtoglu Kuaybe, Kantar Melda, Lucas Stuart J, Budak Hikmet: *Unique and Conserved MicroRNAs in Wheat Chromosome 5D Revealed by Next-Generation Sequencing*. PLoS ONE 2013; 8(7):e69801. DOI:10.1371/journal.pone.0069801

Budak Hikmet, Kantar Melda, Yucebilgili Kurtoglu Kuaybe: *Drought Tolerance in Modern and Wild Wheat.* The Scientific World Journal 2013; (10):548246. DOI:10.1155/2013/548246

Yucebilgili Kuaybe, Mehmetoglu Tugba, Gucin Zuhal, Salih Barik A: *Helicobacter pylori DNA in gallbladder tissue of patients with cholelithiasis and cholecystitis*. The Journal of Infection in Developing Countries 2009; 3(11):856-9. DOI:10.3855/jidc.334

**Presentations**

Yucebilgili Kurtoglu Kuaybe, Filiz Ertugrul, Ozyigit Ibrahim Ilker, Dogan Ilhan, Vatansever Recep, Tombuloglu Huseyin (2016). *Comparative in silico Analysis of Phosphoglycerate Kinase (PGK) Genes in Selected Plants*. Plant and Animal Genome Conference XXIV, San Diago, CA, USA; 2016 [Poster]

Yucebilgili Kurtoglu Kuaybe, Filiz Ertuğrul, Ozyigit Ibrahim Ilker, Vatansever Recep, Uras Mehmet Emin (2016). *Comparative Molecular Docking Analysis in Two Legume Species: Glycine max and Medicago truncatula.* Plant and Animal Genome Conference XXIV, San Diago, CA, USA; 2016 [Poster]

Yucebilgili Kurtoglu Kuaybe, Ozyigit Ibrahim Ilker, Filiz Ertugrul, Vatansever Recep: *Genome-Wide Analysis of Thionin Genes in Higher Plants*. Plant and Animal Genome Conference XXIII, San Diago, CA, USA; 2015 [Poster]

Yucebilgili Kurtoglu Kuaybe, Kantar Melda, Lucas Stuart James, Budak Hikmet: *Dissecting microRNAs in Wheat Chromosome 5D Using Next Generation Sequencing Data*. International Plant and Animal Genome Conference XXII 2014; San Diago, CA, USA; [Poster]

Lucas Stuart James, Yucebilgili Kurtoglu Kuaybe, Kantar Melda, Budak Hikmet: *Approaches to annotating miRNA genes in emerging genomic datasets of the Poaceae*. 2nd Plant Genomics Congress, 2014; London, UK [Poster]

Yucebilgili Kuaybe, Kaleintzi Emine, Muslumanoglu Omer: *Allele & Haplotype Analysis Of 17 Y-STR Loci & mtDNA in Turkish Population* 2nd International Biomedical Sciences and Archeological Methods Conference, 2011; Izmir, TURKEY [Oral]

Yucebilgili Kuaybe, Bolek Bora Kazım: *"Genotype analysis of Helicobacter Pylori strains present in stomack biopsies via multiplex PCR method"* National Biology Student Congress, 2006, Ege University, Izmir, TURKEY [Poster]