



University of Kentucky
UKnowledge

Theses and Dissertations--Plant and Soil
Sciences

Plant and Soil Sciences


2017

TRANSCRIPTIONAL AND POST-TRANSLATIONAL REGULATION OF TERPENOID INDOLE ALKALOID BIOSYNTHESIS IN *CATHARANTHUS ROSEUS*

Priyanka Paul

University of Kentucky, priyanka.kasabiyanka@gmail.com

Author ORCID Identifier:

 <https://orcid.org/0000-0001-8530-7877>

Digital Object Identifier: <https://doi.org/10.13023/ETD.2017.403>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Paul, Priyanka, "TRANSCRIPTIONAL AND POST-TRANSLATIONAL REGULATION OF TERPENOID INDOLE ALKALOID BIOSYNTHESIS IN *CATHARANTHUS ROSEUS*" (2017). *Theses and Dissertations--Plant and Soil Sciences*. 94.

https://uknowledge.uky.edu/pss_etds/94

This Doctoral Dissertation is brought to you for free and open access by the Plant and Soil Sciences at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Plant and Soil Sciences by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Priyanka Paul, Student

Dr. Ling Yuan, Major Professor

Dr. Mark Coyne, Director of Graduate Studies

TRANSCRIPTIONAL AND POST-TRANSLATIONAL REGULATION OF
TERPENOID INDOLE ALKALOID BIOSYNTHESIS IN *CATHARANTHUS ROSEUS*

DISSERTATION

A Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture, Food and Environment
at the University of Kentucky

By

Priyanka Paul

Director: Dr. Ling Yuan, Professor of Department of Plant and Soil Sciences

Lexington, Kentucky

2017

Copyright © Priyanka Paul 2017

ABSTRACT OF DISSERTATION

TRANSCRIPTIONAL AND POST-TRANSLATIONAL REGULATION OF TERPENOID INDOLE ALKALOID BIOSYNTHESIS IN *CATHARANTHUS ROSEUS*

Catharanthus roseus (Madagascar periwinkle) is the exclusive source of an array of terpenoid indole alkaloids (TIAs) that are used in the treatments of hypertension and certain types of cancer. TIA biosynthesis is under stringent spatiotemporal control and is induced by jasmonate (JA) and fungal elicitors. Tryptamine, derived from the indole branch, and secologanin from the iridoid branch are condensed to form the first TIA, strictosidine. Biosynthesis of TIA is regulated at the transcriptional level and several transcription factors (TFs) regulating the expression of genes encoding key enzymes in the pathway have been isolated and characterized. The JA-responsive APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF), ORCA3, and the basic helix-loop-helix (bHLH) factor, CrMYC2, are the key activators of the TIA biosynthesis. Recently, two other TFs, the bHLH IRIDOID SYNTHESIS 1 (BIS1) and BIS2 were also identified as regulators of TIA pathway. Analysis of *C. roseus* genome sequence has revealed that ORCA3 forms a physical cluster with two uncharacterized AP2/ERFs, ORCA4 and ORCA5. In plants, physically linked clusters of TFs are less characterized. Moreover, the regulation of TF clusters is relatively unexplored. My research uncovered that the *ORCA* gene cluster is differentially regulated. ORCA4 and ORCA5, while functionally overlapping with ORCA3, regulate an additional set of TIA pathway genes. *ORCA4* or *ORCA5* overexpression has resulted in significant increase of TIA accumulation in *C. roseus* hairy roots. In addition, ORCA5 directly regulates the expression of *ORCA4* and indirectly regulates *ORCA3*, likely via unknown factor(s). Interestingly, ORCA5 also activates the expression of *ZCT3*, a negative regulator of the TIA pathway. In addition CrMYC2 is capable of activating *ORCA3* and co-regulating pathway genes concomitantly with ORCA3.

Several lines of evidence suggest that, in addition to the transcriptional control, biosynthesis of TIAs is also controlled at the posttranslational level, such as protein phosphorylation. Available literature indicates that a mitogen-activated protein kinase (MAPK) cascade is involved in this process. Analysis of *C. roseus* MAP kinome, identified two independent MAPK cascades regulating the indole and iridoid branches of the TIA pathway. We showed that the ORCA cluster and CrMYC2 act downstream of a MAP kinase cascade consisting of *CrMAPKK1*, *CrMAPK3* and *CrMAPK6*.

Overexpression of *CrMAPKK1* in *C. roseus* hairy roots upregulates TIA pathway genes expressions and boosts TIA accumulation. The other cascade, consisting of *CrMAPKK6* and *CrMAPK13*, mostly regulates the iridoid branch of the TIA pathway. Overexpression of *CrMAPK13* in *C. roseus* hairy roots significantly upregulates iridoid pathway genes and boosts tabersonine accumulation. Moreover, we recently identified the third MAPK cascade, consisting of *CrMAPKK1* and *CrMAPK20*, that negatively regulates the indole branch of the TIA pathway. Overexpression of *CrMAPK20* in *C. roseus* hairy roots represses the genes regulated by *CrMYC2*-ORCAs and reduces catharanthine accumulation. These findings significantly advance our understanding of transcriptional and post-translational regulatory mechanisms that govern TIA biosynthesis in *C. roseus*.

KEY WORDS: AP2/ERF gene cluster, *Catharanthus roseus*, MAP kinome, terpenoid indole alkaloids, transcriptional and post-translational regulation.

Priyanka Paul

08-28-17

TRANSCRIPTIONAL AND POST-TRANSLATIONAL REGULATION OF
TERPENOID INDOLE ALKALOID BIOSYNTHESIS IN *CATHARANTHUS ROSEUS*

By

Priyanka Paul

Dr. Ling Yuan

Director of Dissertation

Dr. Mark Coyne

Director of Graduate Studies

08-28-17

*To my beloved husband
Sanjay
A man with love and support*

ACKNOWLEDGMENTS

I am grateful to so many people for their love and support over the past years. I would first like to express my sincere gratitude to my advisor Prof. Ling Yuan for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my Ph.D. committee: Prof. Arthur Hunt, Dr. Sharyn Perry, and Dr. Xuguo Zhou not only for their insightful comments and encouragement, but also for the hard questions which gave me incentive me to widen my research and expand my perspectives. I would also like to thank Prof. Subba Reddy Palli who served as my outside examiner.

The members of Prof. Yuan lab are a great group of people and I have enjoyed working with them for the past several years. It is just like my second home. I would like to thank Dr. Sitakanta Pattanaik who, with great patience in every situation, trained me in my research and with his great kindness he is always with me in every situation. I thank him for correcting and proofreading this dissertation. Without his precious support it would not have been possible to conduct this research. I also thank Kathy Shen who proofread and corrected the language of my research papers.

I sincerely thank Dr. Baruvana Patra and Dr. Sanjay Singh for their great help in conducting my research. The many discussions with Dr. Baruvana Patra, Dr. Sanjay Kumar Singh, Dr. Jayadri Shekhar Ghosh and Dr. Yongliang Liu were not only helpful but provided much encouragement. I sincerely thank Dr. Craig Schluttenhofer for proofreading this dissertation.

I thank all my lab mates and friends, Dr. Craig Schluttenhofer and Dr. Xueyi Sui for the stimulating discussions and for all the fun we have had in the last four years. I am also

grateful to the visiting scholar Guiping Ren, Xiaoyu Liu and former lab members Dr. Yongmei Wu, Dr. Chunmei Zhong, Dr. Tiandai Huang, Dr. Shiyuan Deng and Dr. Yuxiang Wu for a congenial environment.

Additionally, many other people from the KTRDC and UK Plant and Soil Sciences Department also deserve thanks, ranging from other graduate students to postdocs to technicians and professors.

My dream of doing a Ph.D. would not have been possible without the constant love, support, and encouragement from my life partner and lab mate Dr. Sanjay Singh. I especially thank my parents Barun Paul and Minati Paul, my sister Riyanka Paul as one of my best friends in life, and my in-laws Surendar Singh and Munera Devi who have continually endured in their support of my endeavors. I want to thank my little princess Ellora who is only 23 months old and giving all her loves to me. And I give thanks to the almighty God for the blessings and graces!

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	iii
TABLE OF CONTENTS.....	v
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
Chapter 1: Literature Review	
1.1 Introduction.....	1
1.2 TIA biosynthesis pathway.....	2
1.2.1 The MEP pathway.....	3
1.2.2 The iridoid pathway.....	3
1.2.3 The shikimate and indole pathways.....	6
1.2.4 The late TIA pathway.....	7
1.3 TIA biosynthesis in roots.....	10
1.4 Regulation of TIA biosynthesis in <i>C. roseus</i>	10
1.4.1 Transcriptional Regulation of TIA biosynthesis in <i>C. roseus</i>	11
1.4.2 Post-translational Regulation of TIA biosynthesis in <i>C. roseus</i>	16
1.5 Metabolic engineering by assembling different branches of TIA pathway.....	17
1.6 Conclusion.....	18
1.7 Outline of the dissertation.....	19
Chapter 2: A differentially regulated AP2/ERF transcription factor gene cluster modulate terpenoid indole alkaloid biosynthesis in <i>Catharanthus roseus</i>	
2.1 Introduction.....	21
2.2 Materials and Methods.....	24
2.2.1 Plant materials.....	24
2.2.2 RNA isolation and cDNA synthesis.....	24
2.2.3 Multiple sequence alignment, phylogenetic and co-expression analysis.....	24
2.2.4 Quantitative RT-PCR.....	25
2.2.5 Sub-cellular localization.....	26
2.2.6 Protoplast isolation and electroporation.....	26
2.2.7 Recombinant protein production and EMSA.....	27
2.2.8 Construction of plant expression vectors and generation of hairy roots.....	27
2.2.9 Alkaloid extraction and analysis.....	28
2.3 Results and Discussion.....	28
2.3.1 Two AP2/ERF genes form a cluster with <i>ORCA3</i>	28
2.3.2 ORCA3, ORCA4, and ORCA5 are preferentially induced in roots by MeJA.....	31
2.3.3 ORCA4 and 5 localized to nucleus.....	32
2.3.4 ORCAs are transcriptional activators.....	32

2.3.5 ORCA4 and 5 activate multiple TIA pathway genes.....	33
2.3.6 ORCA4 and 5 bind to the JRE of the <i>STR</i> promoter.....	36
2.3.7 Ectopic expression of <i>ORCA4</i> or <i>ORCA5</i> activates key TIA pathway genes and boosts metabolite accumulation in <i>C. roseus</i> hairy roots.....	37
2.3.8 ORCA3/4/5 are differentially regulated.....	43
2.3.9 ORCA TF cluster differentially activate TIA pathway genes.....	45
2.3.10 CrMYC2 co-regulates TIA pathway genes with ORCA3 by direct binding to a T/G-box.....	46
2.3.11 ORCA5 activates the <i>ORCA4</i> but not <i>ORCA3</i>	49
2.3.12 ORCA3 and 5 bind to the <i>ORCA4</i> promoter.....	50
2.3.13 ORCA5 activates the expression of <i>ZCT3</i> in <i>C. roseus</i> hairy roots and tobacco cells.....	52
2.3.14 <i>C. roseus</i> ORCAs and tobacco NIC2-locus ERFs are functionally interchangeable.....	52
2.3.15 AP2/ERF gene cluster regulation in other plants.....	54
2.4 Conclusion.....	55
Chapter 3: A MAP kinase cascade acts upstream of an AP2/ERF gene cluster to regulate the terpenoid indole alkaloid (TIA) biosynthesis in <i>Catharanthus roseus</i>	
3.1 Introduction.....	58
3.2 Materials and Methods.....	60
3.2.1 Plant materials.....	60
3.2.2 RNA isolation and cDNA synthesis.....	60
3.2.3 Multiple sequence alignment phylogenetic analysis.....	60
3.2.4 Identification, classification and co-expression of <i>C. roseus</i> protein kinases.....	60
3.2.5 Quantitative RT-PCR.....	61
3.2.6 Sub-cellular localization.....	62
3.2.7 Protoplast isolation and electroporation.....	62
3.2.8 Construction of plant expression vectors and generation of hairy roots.....	63
3.2.9 Recombinant protein production and in vitro phosphorylation assay.....	63
3.2.10 Alkaloid extraction and analysis.....	64
3.3 Results and Discussion.....	64
3.3.1 Analysis of <i>C. roseus</i> kinome identifies MAP kinases potentially involved in TIA pathway regulation.....	64
3.3.2 CrMAPKK1 is MeJA responsive, nucleus-localized, and autophosphorylated.....	68
3.3.3 CrMAPKK1 interacts with CrMAPK3 and CrMAPK6 in plant cells.....	69
3.3.4 The N-terminal kinase interaction motif is crucial for CrMAPKK1 function.....	72
3.3.5 CrMAPK3 and CrMAPKK1 significantly enhance the	

transactivation of TIA pathway gene promoters by ORCAs and CrMYC2.....	73
3.3.6 Ectopic expression of CrMAPKK1 activates key TIA pathway genes and boosts alkaloid accumulation in <i>C. roseus</i> hairy roots.....	77
3.4 Conclusion.....	80
Chapter 4: Genome wide identification of <i>Catharanthus roseus</i> MAP kinome and functional characterization of two MAP kinases involved in TIA regulation	
4.1 Introduction.....	82
4.2 Materials and Methods.....	87
4.2.1 Plant materials.....	87
4.2.2 RNA isolation and cDNA synthesis.....	87
4.2.3 Multiple sequence alignment and phylogenetic analysis.....	87
4.2.4 Identification of MAPK cascade genes in <i>C. roseus</i>	88
4.2.5 Quantitative RT-PCR (qRT-PCR).....	88
4.2.6 Protoplast isolation and electroporation.....	90
4.2.7 Yeast Two-hybrid Assay.....	90
4.2.8 Construction of plant expression vectors and generation of hairy roots.....	91
4.2.9 Alkaloid extraction and analysis.....	91
4.3 Results and Discussion.....	91
4.3.1 Identification of the MAP1K, MAP2K, MAP3K and MAP4K families in <i>C. roseus</i>	91
4.3.2 Phylogenetic relationship and conserved domain analysis.....	96
4.3.3 Expression profiles of CroMAP kinase cascade genes in different tissues.....	100
4.3.4 Multiple sequence alignment and motif analysis of CroMAP kinase cascade components.....	101
4.3.5 Identification of CroMAPK genes co-expressed with iridoid pathway genes.....	106
4.3.6 CrMAPK13 significantly enhances the transactivation of iridoid pathway gene promoters by BIS1.....	107
4.3.7 CrMAPK13 likely phosphorylates the amino acids S134 and T113 in BIS1.....	108
4.3.8 CrMAPK13 interacts with BIS1 and CrMAPKK6 in tobacco cells.....	110
4.3.9 CrMAPK13 interacts with CrMAPKK6 in yeast cells.....	112
4.3.10 Ectopic expression of CrMAPK13 activates key iridoid pathway genes and boosts alkaloid accumulation in <i>C. roseus</i> hairy roots.....	113
4.3.11 Spatial expression of CroMAPK genes potentially involved in regulation of indole branch of TIA pathway.....	116
4.3.12 CrMAPK20 significantly represses the transactivation potential of ORCAs on STR promoter in protoplasts.....	117
4.3.13 CrMAPK20 interacts with ORCAs in plant cells.....	118
4.3.14 MAPK20 interacts with ORCAs and CrMAPKK1 in yeast	

cells.....	121
4.3.15 Ectopic expression of <i>CrMAPK20</i> represses key indole pathway genes and reduces TIA accumulation in <i>C. roseus</i> hairy roots.....	122
4.3.16 CrMAPK20 interacts with CrMAPK3 in yeast cells.....	124
4.4 Conclusion.....	125
Chapter 5: Summary and Future directions.....	129
APPENDIX-A: LIST OF ABBREVIATIONS.....	134
References.....	136
Vita.....	156

LIST OF TABLES

Table 2.1 Oligonucleotides used in this study.....	26
Table 2.2 Position and sequences of GC-rich motifs in promoters of AP2/ERFs in <i>C. roseus</i> , tobacco, tomato and potato.....	55
Table 3.1 Oligonucleotides used in this study.....	62
Table 4.1 Oligonucleotides used in this study.....	89
Table 4.2 List of MAP1Ks identified in <i>C. roseus</i>	92
Table 4.3 List of MAP2Ks identified in <i>C. roseus</i>	93
Table 4.4 List of MAP3Ks identified in <i>C. roseus</i>	93
Table 4.5 List of MAP4Ks identified in <i>C. roseus</i>	95

LIST OF FIGURES

Figure 1.1 Schematic diagrams of the <i>C. roseus</i> TIA biosynthetic pathway.....	2
Figure 1.2 The iridoid pathway.....	5
Figure 1.3 The shikimate and indole pathways.....	7
Figure 1.4 The late TIA pathway.....	9
Figure 2.1 Phylogenetic analysis of AP2/ERF families and Amino acid sequence alignment of ORCA3, 4, and 5.....	30
Figure 2.2 MeJA induction of <i>ORCA3</i> , <i>ORCA4</i> , and <i>ORCA5</i>	31
Figure 2.3 Nuclear localization and transactivation assays of ORCA4 and ORCA5 in tobacco protoplasts.....	32
Figure 2.4 Co-expression of ORCA4 and ORCA5 with TIA regulatory and structural genes and transactivation of key TIA pathway gene promoters by ORCA3/4/5 in tobacco protoplasts.....	35
Figure 2.5 Binding of ORCA3/4/5 to the JRE of the <i>STR</i> promoter.....	37
Figure 2.6 PCR confirmation of the transgenic status of <i>C. roseus</i> hairy root lines overexpressing <i>ORCA4</i> or <i>CrMYC2</i>	39
Figure 2.7 Increase of key TIA pathway gene expression and alkaloid accumulation in <i>ORCA4</i> -overexpression <i>C. roseus</i> hairy roots.....	40
Figure 2.8 PCR confirmation of the transgenic status of <i>C. roseus</i> hairy root lines overexpressing <i>ORCA5</i>	41
Figure 2.9 Increase of key TIA pathway gene expression and alkaloid accumulation of <i>C. roseus</i> hairy root lines overexpressing ORCA5.....	42
Figure 2.10 Differential regulation of the <i>ORCA</i> gene cluster.....	44
Figure 2.11 Transactivation of <i>LAMT</i> and <i>SLS</i> gene promoters by ORCA3/4/5 in tobacco protoplasts.....	46
Figure 2.12 Schematic diagrams of three TIA pathway promoters regulated by CrMYC2 and ORCAs.....	48
Figure 2.13 CrMYC2 co-regulation of key TIA pathway genes with ORCA3 via binding to T/G-box.....	48
Figure 2.14 ORCA5 activates <i>ORCA4</i> by directly binding to the GC-rich sequence of the promoter and auto-regulates itself.....	51
Figure 2.15 ORCA5 activates the <i>ZCT3</i> promoter and ORCA and NIC2 locus are functionally interchangeable.....	53
Figure 2.16 A simplified model depicting the transcriptional regulation of the ORCA cluster and TIA pathway genes.....	57
Figure 3.1 Gene number distributions of MAP kinases (MAPK) in various plants and spatial expression analysis of the <i>C. roseus</i> MAPK cascade components.....	65
Figure 3.2 Multiple sequence alignment, phylogenetic analysis, nuclear localization, and bacterial expression-purification of CrMAPKK1.....	67
Figure 3.3 Co-expression of CrMAPKK1 with TIA regulatory and structural genes.....	68
Figure 3.4 MeJA induction and autophosphorylation of CrMAPKK1 and its interaction with CrMPK3 and 6.....	71
Figure 3.5 Sequence alignments of CrMAPK6 and AtMAPK6.....	72

Figure 3.6 Importance of the KIM motif for CrMAPKK1 function and phosphorylation of ORCA3 by CrMAPK3.....	75
Figure 3.7 Enhancement of the transactivation activity of ORCAs and CrMYC2 on TIA pathway gene promoters by CrMAPKK1.....	76
Figure 3.8 PCR confirmation of the transgenic status of CrMAPKK1-overexpressing <i>Catharanthus</i> hairy root lines.....	78
Figure 3.9 Increase of TIA pathway gene expression and alkaloid accumulation in <i>Catharanthus</i> hairy roots overexpressing <i>CrMAPKK1</i>	79
Figure 3.10 A simplified model depicting the transcriptional and posttranscriptional regulation of the ORCA cluster and TIA pathway genes....	80
Figure 4.1 Distribution of MAPK cascade kinase genes across the plant kingdom.....	96
Figure 4.2 Phylogenetic relationships, expression profiles, and protein structures of MAP1Ks in <i>C. roseus</i>	98
Figure 4.3 Phylogenetic relationships, expression profiles, and protein structures of MAP2Ks in <i>C. roseus</i>	98
Figure 4.4 Phylogenetic relationships, expression profiles, and protein structures of MAP3Ks in <i>C. roseus</i>	99
Figure 4.5 Phylogenetic relationships, expression profiles, and protein structures of MAP4Ks in <i>C. roseus</i>	100
Figure 4.6 Multiple sequence alignments of the CroMAPK cascade proteins in <i>C. roseus</i>	103
Figure 4.7 Co-expression analyses of CroMAPK cascade genes with iridoid pathway regulatory and structural genes.....	106
Figure 4.8 Sequence alignments of CrMAPK13 and AtMAPK13.....	107
Figure 4.9 Enhancement of the transactivation activity of BIS1 on TIA pathway gene promoters by CrMAPK13.....	108
Figure 4.10 CrMAPK13 likely phosphorylates BIS1.....	110
Figure 4.11 Interaction of CrMAPK13 with CrMPKK6 and BIS1.....	111
Figure 4.12 Physical interaction between CrMAPK13 and CrMAPKK6 in yeast cells.....	112
Figure 4.13 PCR confirmation of the transgenic status of CrMAPK13-overexpressing <i>C. roseus</i> hairy root lines.....	114
Figure 4.14 Increase of iridoid pathway gene expression and TIA accumulation in <i>C. roseus</i> hairy roots overexpressing CrMAPK13.....	115
Figure 4.15 Spatial expression analysis of the <i>C. roseus</i> MAPK cascade components and repression of CrMAPK20 by MeJA treatment.....	117
Figure 4.16 Co-expression of CrMAPK20 with TIA pathway genes and repression of the transactivation activity of ORCAs by CrMAPK20, phosphorylation of ORCA3 by CrMAPK20 and interaction of CrMAPK20 with ORCAs.....	119
Figure 4.17 Physical interaction of CrMAPK20 with the CrMAPKK1 and ORCAs detected in yeast two-hybrid assay.....	121
Figure 4.18 Repression of TIA pathway gene expression and alkaloid accumulation in <i>C. roseus</i> hairy roots overexpressing CrMAPK20.....	123

Figure 4.19 CrMAPK20 interacts with CrMAPK3 in yeast cell.....	125
Figure 4.20 A simplified model depicting the posttranslational regulation of the ORCA gene cluster and TIA pathway genes.....	127
Figure 5.1. Multiple sequence alignment of CrPP2C1 with AtPP2C1.....	132
Figure 5.2 CrPP2C1 significantly reduces the transactivation potential of ORCA3 on STR promoter.....	132

Chapter 1

Literature Review

1.1 Introduction

Plants produce thousands of specialized metabolites (a.k.a. natural products or secondary metabolites), such as phenols, terpenes and alkaloids. Alkaloids are large and structurally diverse group of plant specialized metabolites which are present in approximately 20% of plant species (Facchini and De Luca, 2008). The unique feature of alkaloids is the presence of a nitrogen atom within a heterocyclic ring (Ziegler and Facchini, 2008). Alkaloids protect plants against microbial, herbivore attack and/or UV irradiation. Many of these alkaloids are also beneficial to human health. Due to their low abundance in plants, many research projects have been developed to understand the molecular mechanism regulating these complex biosynthetic pathways. Further complicating the production, these molecules are usually synthesized in a plant-, an organ- or even cell-specific manner (St-Pierre et al., 2013).

Terpenoid indole alkaloids

Terpenoid indole alkaloids (TIAs) are one of the largest and most distinct groups of plant natural products with over 3000 special known structures found mostly in the Apocynaceae, Rubiaceae, Nyssaceae and Loganiaceae families (Facchini and De Luca, 2008). *Catharanthus roseus* (L.) G. Don., also known as Madagascar periwinkle of the family Apocynaceae, order Gentianiales (van der Heijden et al., 2004), produces more than 130 TIAs and several of them are used as pharmaceuticals, such as the anticancer drugs (vincristine and vinblastine), and antihypertensive agents (ajmalicine and serpentine). The biosynthesis of TIAs is under stringent spatiotemporal control and is induced by the phytohormone, jasmonate (JA), and fungal elicitors (Menke et al., 1999a). Despite the importance of TIAs to human health, the molecular mechanism of regulation of TIA biosynthesis is not well understood. The lack of appropriate genetic tools and highly complex nature of the biochemical pathway are the major bottlenecks that limit progress.

1.2 TIA biosynthesis pathway

The TIA biosynthesis pathway in *C. roseus* can be divided into five stages: (i) the MEP (2-C-methyl-D-erythritol-4-phosphate) pathway that provides the isoprenoid subunit isopentenyl diphosphate (IPP), (ii) the iridoid pathway that produces secologanin from IPP and provides the terpene moiety of TIAs, (iii) the shikimate and indole pathways provide the indole moiety, tryptamine, (iv) which undergoes a condensation reaction with the iridoid, secologanin to form strictosidine, the principal precursor of all TIAs; the strictosidine thereafter undergoes deglycosylation giving rise to many unstable intermediates for biosynthesis of various classes of TIAs and (v) the late pathway leads to the formation of crucial TIA precursor, vindoline. The biochemical coupling of vindoline and catharanthine results in the dimeric TIAs, such as vinblastine and vincristine (Figure 1.1) (Paul et al., 2017).

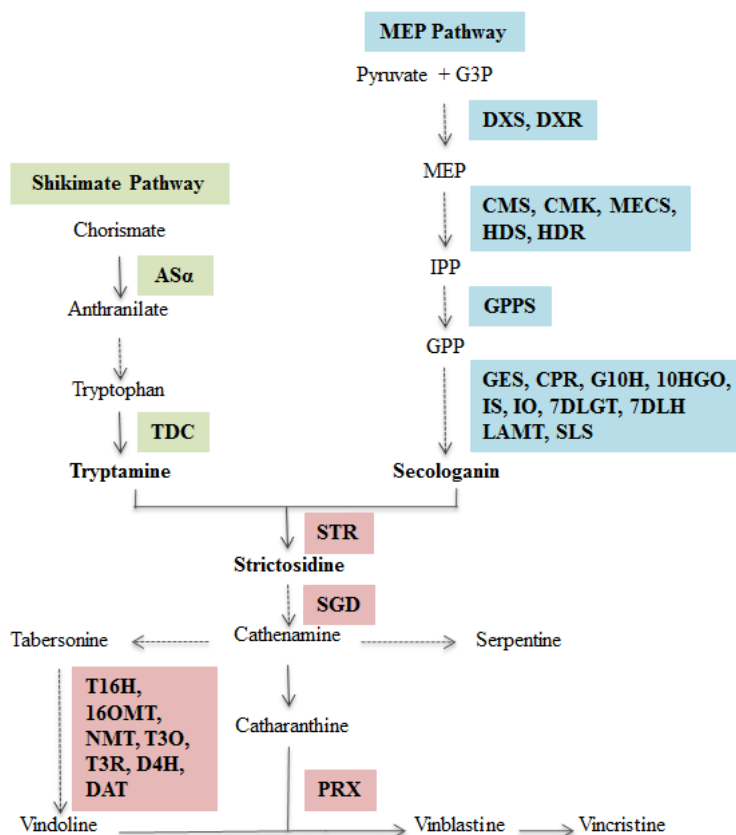


Figure 1.1 Schematic diagrams of the *C. roseus* TIA biosynthetic pathway.

As α , anthranilate synthase; TDC, tryptophan decarboxylase; DXS, 1-deoxy-5-xylulose-5-phosphate synthase; DXR, 1-deoxy-5-xylulose-5-phosphate reductase; CMS, 4-cytidine

5'-diphospho-2-C-methyl-D-erythritol synthase; CMK, 4-cytidyl-diphospho-2-C-methyl-D-erythritol kinase; MECS, 4-cytidyl-diphospho-2-C-methyl-D-erythritol synthase; HDS, (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate synthase; HDR, 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase; GPPS, geranyl diphosphate synthase; GES, geraniol synthase; CPR, cytochrome P450 reductase; G10H, geraniol 10-hydroxylase; 10HGO, 10-hydroxygeraniol oxidoreductase; 7DLGT, 7-deoxyloganic acid glucosyl transferase; 7DLH, 7-deoxyloganic acid hydroxylase; IS, iridoid synthase; LAMT, loganic acid O-methyltransferase; SLS, secologanin synthase; STR, strictosidine synthase; SGD, strictosidine b-glucosidase; T16H, tabersonine-16-hydroxylase; 16OMT, 16-hydroxytabersonine-O-methyltransferase; NMT, N-methyltransferase; NMT; T3O, tabersonine 3-oxygenase; T3R, tabersonine 3-reductase D4H, desacetoxyvindoline-4-hydroxylase; DAT, deacetylvindoline-4-O-acetyltransferase; PRX, peroxidases; G3P, glyceraldehyde 3-phosphate; MEP, 2-methyl-d-erythritol 4-phosphate; IPP, isopentenyl diphosphate; GPP, geranyl diphosphate.

1.2.1 The MEP pathway

In plants there are two separate routes for IPP biosynthesis, the plastidic MEP pathway and the cytosolic mevalonate (MVA) pathway (Lichtenthaler, 1999). Early feeding studies supported the role of the MEP pathway in providing IPP to TIA biosynthesis (Contin et al., 1998; Lichtenthaler, 1999). A number of MEP pathway enzymes have been cloned and characterized from *C. roseus*, such as deoxyxylulose 5-phosphate synthase (DXS) (Chahed et al., 2000), deoxyxylulose 5-reductase (DXR) (Veau et al., 2000), methylerythritol 2,4-diphosphate synthase (MECS) (Veau et al., 2000) and hydroxymethylbutenyldiphosphate synthase (HDS) (Oudin et al., 2007). The final product of the MEP pathway is IPP, which is converted to geraniol in the internal phloem-associated parenchyma (IPAP) cells of *C. roseus* leaves (Simkin et al., 2013).

1.2.2 The iridoid pathway

In *C. roseus*, the first step in iridoid biosynthesis is the oxidation of geraniol to 10-hydroxygeraniol, catalyzed by a cytochrome P450 monooxygenase (CYP76B6), geraniol 10-hydroxylase (G10H, a.k.a. G8O) (Collu et al., 2001). G10H interacts with a

membrane-bound cytochrome P450 reductase (CPR) (Meijer et al., 1993), which requires the cofactors FMN, FAD and NADPH for electrons transfer from NADPH to the cytochrome P450 monooxygenase. Further oxidation of 10-hydroxygeraniol to 10-oxogeraniol is catalyzed by 10-hydroxygeraniol oxidoreductase (10HGO, a.k.a. 8HGO) in both *Rauvolfia serpentina* (Ikeda et al., 1991) and in *Nepeta racemosa* (Hallahan et al., 1995). Two separate dehydrogenases, isolated from *C. roseus*, are probably involved in this conversion process. One *C. roseus* geraniol dehydrogenase like gene product catalyzed the conversion of 10-hydroxygeraniol to 10-oxogeraniol and the minor products 10-oxogeraniol plus 10-hydroxygeraniol in the presence of NADP (Keat et al., 2000; Krithika et al., 2015). Another dehydrogenase with a distinctive amino acid sequence partly converts 10-hydroxy-geraniol to 10-hydroxy-geraniol, and 10-oxogeraniol and 10-oxogeraniol in the presence of NAD in *C. roseus* (Miettinen et al., 2014). Interestingly, when incubated in the presence of iridoid synthase (IS) both proteins were capable to produce iridoidials and related nepetalactol (Miettinen et al., 2014; Krithika et al., 2015). IS which converts 10-oxogeraniol into iridodial, was recently cloned and functionally characterized (Geu-Flores et al., 2012; Miettinen et al., 2014; Krithika et al., 2015). IS enzyme is the first member of the progesterone 5 β -reductase (P5 β R) family shown to enantio-selectively convert progesterone to a cardenolide biosynthesis intermediate, called 5 β -pregnane-3, 20-dione in *C. roseus* (Munkert et al., 2015). Members of this gene family are found in a number of plant species where they are involved in ambiguous biological roles, including wound responses (Yang et al., 1997), leaf vascular strand formation and patterning (Jun et al., 2002), and also participate in undefined biochemical pathways. In *C. roseus* six members of this family including IS were cloned and functionally characterized (Munkert et al., 2015). Further oxidation of iridodial/nepetalactol by the 7-deoxyloganetic acid synthase (7-DLS) gene product yields deoxyloganetic acid (Salim et al., 2013; Miettinen et al., 2014). This cytochrome P450 catalyzes a 3-step oxidation to produce 7-deoxyloganetic acid, which is glucosylated to form deoxyloganic acid. Three different iridoid glucosyltransferases, UGT6, UGT7 and UGT8 are isolated from *C. roseus*. Because CrUGT8 (UDP-glucose iridoid glucosyltransferase/7-deoxyloganetic acid glucosyltransferase; 7-DLGT) has highest catalytic efficiency and high substrate specificity, it is considered to be the most likely

enzyme to perform this reaction *in vivo* (Asada et al., 2013). Deoxyloganic acid undergoes subsequent hydroxylation to loganic acid by the enzyme 7-deoxyloganic acid hydroxylase (7-DLH) (Salim et al., 2013). Loganic acid is then methylated by loganic acid methyl transferase (LAMT) to loganin (Murata et al., 2008). Cleavage of the cyclopentane ring by secologanin synthase (SLS) yields the iridoid secologanin (Irmeler et al., 2000; Yamamoto et al., 2004) (Figure 1.2). All steps including the glycosylation and succeeding hydroxylation are localized in the IPAP cells while LAMT and SLS are localized to epidermal cells (Irmeler et al., 2000; Burlat et al., 2004; Murata et al., 2008; Geu-Flores et al., 2012; Asada et al., 2013; Salim et al., 2014). These findings suggest the involvement of transporter(s) that may shuttle loganic acid from IPAP cells to the epidermal cells to produce secologanin. This assumption is validated by recently identified nitrate-peptide family (NPF) transporters, including CrNPF2.4, CrNPF2.5 and CrNPF2.6 in *C. roseus*. All three transporters were able to transport the iridoid glucosides 7-deoxyloganic acid, loganic acid, loganin and secologanin into *Xenopus* oocytes. Moreover, it is hypothesized that these three transporters are responsible for transporting multiple intermediates of the biosynthetic pathways that are present in both IPAP and epidermis cells (Larsen et al., 2017).

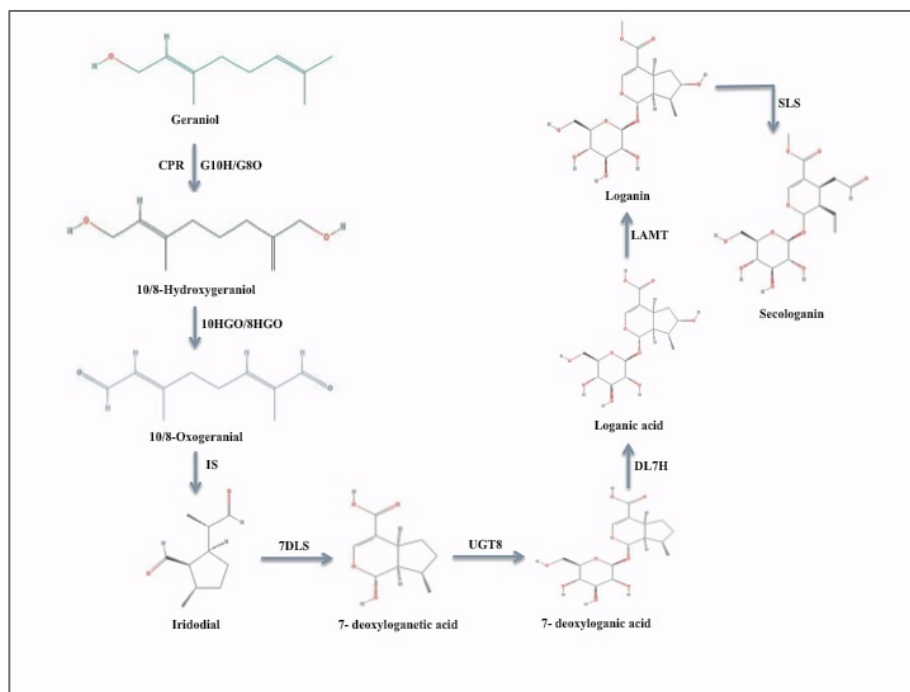


Figure 1.2 The iridoid pathway. CPR, cytochrome P450 reductase; G10H, geraniol-10-

hydroxylase; 10HGO, 10-hydroxygeraniol oxidoreductase; IS, iridoid synthase; 7DLS, 7-deoxyloganetic acid synthase; UGT8, iridoid glucosyltransferase; DL7H, deoxyloganic acid 7-hydroxylase; LAMT, loganic acid methyl transferase; SLS, secologanin synthase.

1.2.3 The shikimate and indole pathways

The indole moiety of TIAs is derived from tryptamine, which is produced from the decarboxylation of tryptophan by a pyridoxal-phosphate dependent enzyme tryptophan decarboxylase (TDC) (De Luca et al., 1989). The coupling between tryptamine and secologanin involves a stereo-selective Pictet–Spengler reaction catalyzed by strictosidine synthase (STR) to form the central precursor of TIAs, strictosidine (McKnight et al., 1990; McKnight et al., 1991). The glucose moiety of strictosidine is then removed by strictosidine β -D-glucosidase (SGD) resulting in a series of putative unstable hemiacetal intermediates by uncharacterized enzymes to form the distinctive classes of TIAs (Figure 1.3). The ring arrangements after the formation of these intermediates are usually species-specific (Zhu et al., 1990; Szabó, 2008). These distinct metabolites are produced in specific plant families (e.g. Apocynaceae, Rubiaceae and Loganiaceae), and each member produces a distinct set of compounds that provides variable biological function (Szabó, 2008). In *C. roseus*, varied arrangements of strictosidine aglycone produce three major classes of TIAs, such as corynanthe, iboga and aspidosperma (Qureshi and Scott, 1968). The biochemical reactions and genes involved in many of these main rearrangements remain to be characterized. The downstream steps such as synthesis of ajmalicine from strictosidine aglycone, have been poorly characterized. Carbinolamine, derived from strictosidine aglycone, functions as an intermediate to yield cathenamine, which is then reduced by cathenamine reductase (CR) to form ajmalicine. Two different CRs have been identified in *C. roseus* cell cultures (El-Sayed and Verpoorte, 2007). Cathenamine is reduced to ajmalicine and 19-epiajmalicine by one CR, whereas the other converts the iminium form of cathenamine into tetrahydroalstonine by tetrahydroalstonine synthase (THAS) (Hemscheidt and Zenk, 1985). Ajmalicine is converted into serpentine by a *C. roseus* vacuolar peroxidase (POD) (Blom et al., 1991).

In situ hybridization and immunological studies revealed that *TDC* and *STR* transcripts are localized to the epidermis of leaves, stems, flower-buds and cortical cells as well as protoderm of the apical meristems in root tips (St-Pierre et al., 1999). RT-PCR of laser-capture micro-dissected cells further supported that *TDC*, *STR* and *SGD* were particularly expressed in the epidermis of *C. roseus* (Murata et al., 2008). Furthermore, most of the TIA biosynthesis occurs in leaf epidermis (Thamm et al., 2016). Therefore, leaf epidermal cells are considered to be the primary active sites for TIA biosynthesis. Recently a tonoplast localized NPF transporter, CrNPF2.9 has been characterized from *C. roseus* that exports strictosidine from the vacuole into the cytosol (Payne et al., 2017). This finding underscores the importance of intracellular transport of TIA intermediates in *C. roseus*.

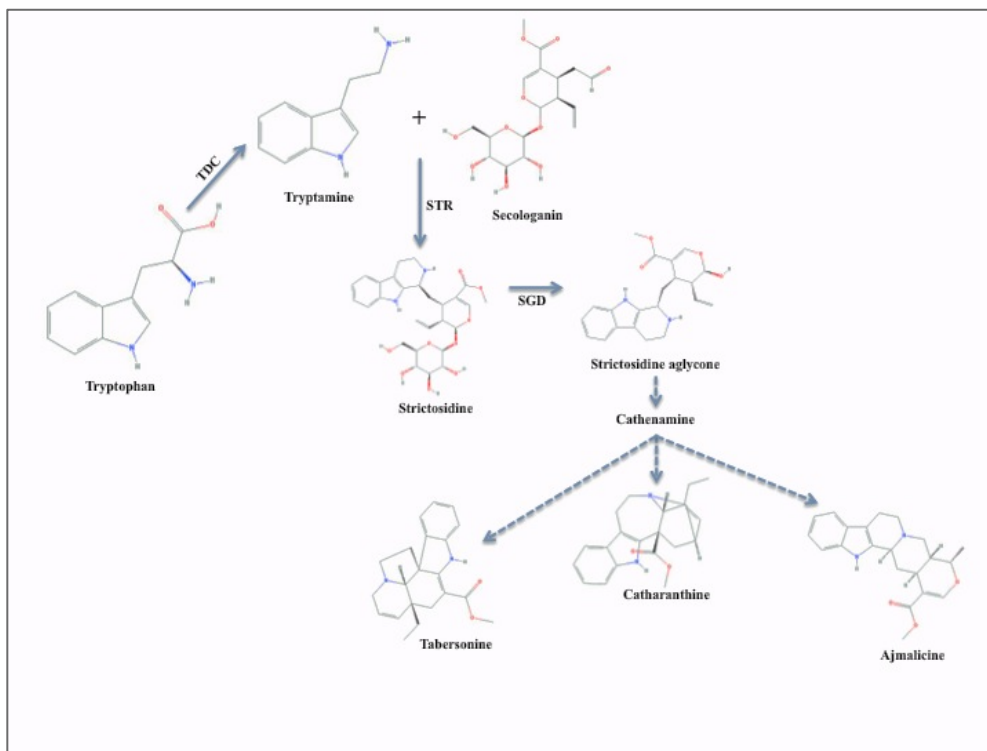


Figure 1.3 The shikimate and indole pathways. TDC, tryptophan decarboxylase; STR, strictosidine synthase; SGD, strictosidine b-glucosidase.

1.2.4 The late TIA pathway

The late pathway leads to the formation of crucial TIA precursors for bis-indole alkaloids. Vindoline is one of the monomers of the bisindole alkaloid vinblastine and is derived

from tabersonine by seven enzymatic steps. The first step comprises hydroxylation of tabersonine by tabersonine 16-hydroxylase 1 (T16H1; CYP71D12) (St-Pierre and De Luca, 1995; Schröder et al., 1999). T16H1 was first cloned from cell cultures. Later, it was reported that the enzyme occurs only in undifferentiated cells and in flowers. However, recently a gene encoding T16H2 (CYP71D351) has been identified and is shown to be involved in vindoline biosynthesis (Besseau et al., 2013). Subsequently, the hydroxyl group of 16-hydroxytabersonine is O-methylated by 16-hydroxytabersonine-16-O-methyltransferase (16-OMT) resulting 16-methoxytabersonine (Fahn et al., 1985; Levac et al., 2008). The 16-methoxytabersonine is then oxidized by a cytochrome P450, T3O (CYP71D1V2) at C3 position (Kellner et al., 2015; Qu et al., 2015) and then reduced by the cinnamyl alcohol dehydrogenase-like enzyme T3R to produce 3-hydroxy-16-methoxy-2,3-dihydrotabersonine (Qu et al., 2015). It was found that in the absence of T3R, the T3O produces a number of 2,3-epoxide and derivatives, which are no longer the substrates for T3R. Therefore, these two steps are likely to be catalyzed in a concerted manner. Next, the enzyme N-methyltransferase (NMT) (Liscombe et al., 2010) catalyzes n-methylation of 16-methoxy-2,3-dihydro-3-hydroxytabersonine to produce desacetoxyvindoline, which is then hydroxylated by the desacetoxyvindoline 4-hydroxylase (D4H) (De Carolis et al., 1990; De Carolis and De Luca, 1993; Vazquez-Flota et al., 1997) to form deacetylvindoline. The final step is the acetylation of deacetylvindoline by deacetylvindoline O-acetyltransferase (DAT) to yield vindoline (De Luca and Cutler, 1987; St - Pierre et al., 1998) (Figure 1.4). The last two steps in vindoline biosynthesis are light regulated and occur only in differentiated plant materials. Light is necessary for activation of D4H and DAT transcription. Localization of gene expression to idioblast and laticifer cells in the leaves implies that desacetoxyvindoline transport from leaf epidermis to the laticifer and idioblast cells is essential for completion of vindoline pathway (Vazquez-Flota et al., 1997; St-Pierre et al., 1999). The involvement of three different cell types (IPAP cells, leaf epidermis and idioblast/laticifers), the aerial tissue-specific expression of the vindoline biosynthetic pathway genes, and light regulation of the last two steps provide possible reasons for the failure of cell cultures to accumulate vindoline.

Finally, the dimerization of vindoline and catharanthine results in vinblastine and/or vincristine, the end products of TIAs biosynthesis in *C. roseus* (Roytrakul and Verpoorte, 2007; Costa et al., 2008) (Figure 1.4). The dimerization reaction is catalyzed by α -3', 4-anhydrovinblastine synthase (AVLBS), commonly known as PRX1, which is a member of class III basic peroxidase enzymes (Sottomayor et al., 1998; Costa et al., 2008). At least five isozymes of peroxidase have been implicated in this reaction (Sottomayor et al., 1998; Sottomayor and Barceló, 2003; Sottomayor et al., 2004).

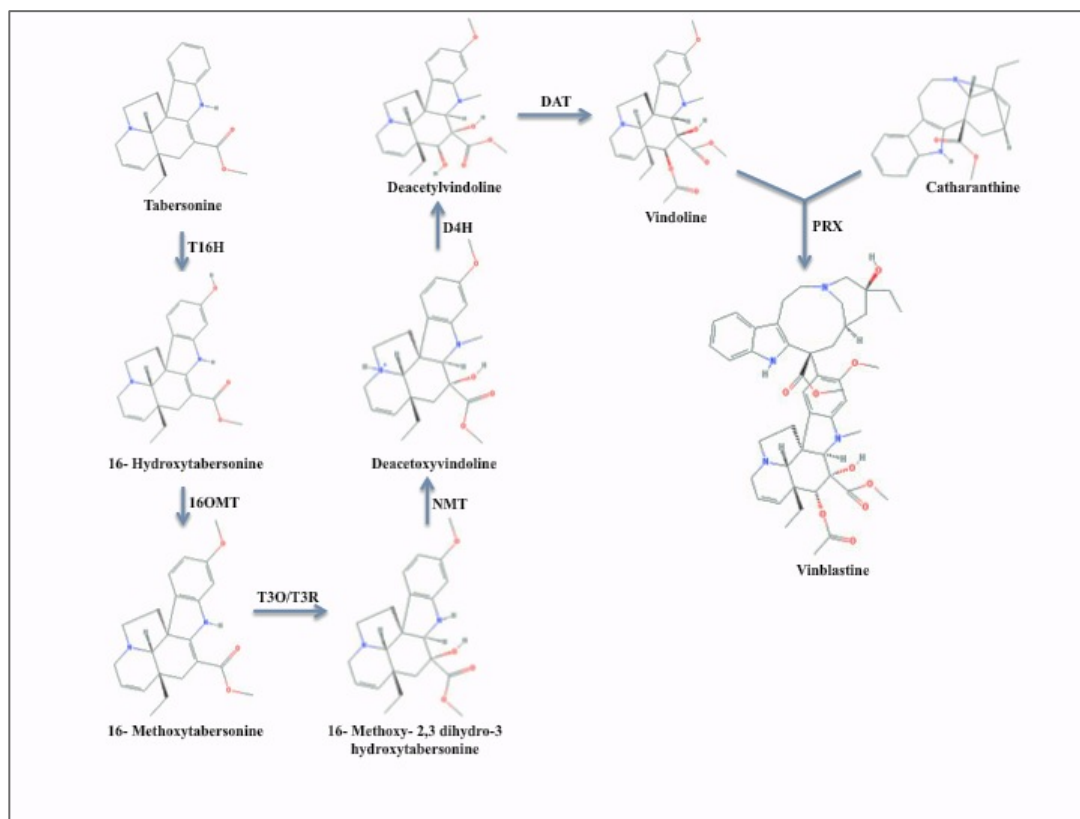


Figure 1.4 The late TIA pathway. T16H, tabersonine 16-hydroxylase; 16OMT, 16-hydroxytabersonine O-methyltransferase; T3O, tabersonine 3-oxidase; T3R, tabersonine 3-reductase; NMT, 2,3-dihydro 3-hydroxytabersonine N-methyltransferase; D4H, desacetylvindoline-4-hydroxylase; DAT, deacetylvindoline 4-Oacetyltransferase; PRX, peroxidase.

Although later steps of vindoline biosynthesis and accumulation occur in leaf idioblast or laticifers, catharanthine is stored on the leaf surface, specifically within the wax exudates

(Roepke et al., 2010). The spatial separation of two monomeric precursors, i.e. vindoline and catharanthine provides an explanation for low levels of dimeric TIAs in *C. roseus*. A TIA transporter, CrTPT2 belonging to the ATP-binding cassette (ABC) family has been cloned and characterized in *C. roseus* (Yu and De Luca, 2013). *CrTPT2* is preferentially expressed in leaf epidermal cells and yeast cells expressing *CrTPT2* export catharanthine into the media. *In planta* suppression of *CrTPT2* by VIGS results in a decrease catharanthine transport to the leaf surface as well as an increase of dimeric TIAs in the plant (Yu and De Luca, 2013). Though the reasons for the spatial separation of catharanthine and vindoline in *C. roseus* leaves is unclear, it has been hypothesized that leaf damage triggered by herbivory allows the two monomers to come together leading to the formation of the dimeric TIAs. These dimeric TIAs act as inhibitors against the herbivores to protect the plants (Roepke et al., 2010).

1.3 TIA biosynthesis in roots

In *C. roseus*, the TIAs profile in roots is different from that found in aerial parts. Though both ajmalicine and catharanthine are found in roots, vindoline is only accumulated in the aerial parts (van der Heijden et al., 2004). In roots, tabersonine can be converted into lochnericine or 19-hydroxytabersonine catalyzed by P450-dependent tabersonine-6,7-epoxidase and tabersonine 19-hydroxylase (T19H), respectively (Laflamme et al., 2001; Rodriguez et al., 2003). Both lochnericine and 19-hydroxytabersonine are proposed as intermediates in synthesis of 19-*O*-acetylhörhammericine and the reaction is catalyzed by minovincinine 19-*O*-acetyl transferase (MAT) (Giddings et al., 2011). However, the conversions of tabersonine are less studied in roots compared to leaves.

1.4 Regulation of TIA biosynthesis in *C. roseus*

The biosynthesis of TIAs is highly complex and under stringent spatio-temporal control (Thamm et al., 2016). Here, I discussed our current knowledge on the regulation of TIA biosynthesis, including transcriptional and posttranslational regulation, in *C. roseus*.

1.4.1 Transcriptional regulation of TIA biosynthesis in *C. roseus*

During the past decades significant progress has been made in isolating and

characterizing genes encoding key biosynthetic enzymes in the TIA pathway (Asada et al., 2013; Besseau et al., 2013; Qu et al., 2015). However, transcription factors (TFs) regulating the TIA pathway are relatively less characterized. Currently, TFs regulating the TIA pathway belong to two major families such as APETALA2/ETHYLENE RESPONSE FACTORS (AP2/ERFs; ORCA2/3) (Menke et al., 1999b; van der Fits and Memelink, 2000), and basic helix-loop-helix factors (bHLH; CrMYC2, BIS1/2) (Zhang et al., 2011; Van Moerkercke et al., 2015; Van Moerkercke et al., 2016). Additionally, WRKY family TF (CrWRKY1) (Suttipanta et al., 2011), C2H2 zinc fingers TFs (ZCT1/2/3) (Pauw et al., 2004; Rizvi et al., 2016), the basic leucine zipper (bZIP) factors (GBF1/2) (Sib eril et al., 2001), and a MYB-like factor (BPF1) (van der Fits et al., 2000), also regulate TIA biosynthesis.

Previous studies revealed that phytohormone, jasmonic acid (JA) and its methyl derivatives, called methyl jasmonates (MeJA), are major inducers of biosynthesis of various natural products, such as artemisinin (Shen et al., 2016), taxol (Mirjalili and Linden, 1996), and nicotine (Shoji et al., 2000). MeJA and fungal elicitors are also known to induce expression of the majority of structural and regulatory genes in the TIA pathway (Menke et al., 1999a). JA-responsive elements (JRE) are often found in the promoters of TIA pathway genes and regulators including *STR* (Menke et al., 1999b), *TDC* (Ouwkerk and Memelink, 1999) *CPR* (Cardoso et al., 1997) and *ORCA3* (Zhang et al., 2011). The JRE of *STR* promoter contains a core GCC-box elements known to bound by ORCA2/3 (Menke et al., 1999b). The *ORCA3* JRE contains a T/G-box (AACGTG) and an AT-rich sequence (Vom Endt et al., 2007). Generally, T/G-boxes are located in numerous promoters of stress- and light- induced genes in plants and are known binding sites for bHLH and bZIP family TFs.

AP2/ERF transcription factors: The AP2/ERF is a large TF family and is unique to plants. AP2/ERFs are characterized by the presence of ~60 conserved amino acids long DNA binding AP2 domain (Ohme-Takagi and Shinshi, 1995). Based on the number of AP2 domains and the presence of additional domains, including B3 DNA binding domain, AP2/ERF family can be classified into five subfamilies, such as AP2, ERF, DREB, RAV,

and the fifth (members not classified to the other four groups). Among them the ERF subfamily constitutes the largest group (Sakuma et al., 2002) and have been involved in responses to ethylene and different abiotic stresses (van der Fits and Memelink, 2000; Mizoi et al., 2012). In addition, previous studies have shown that AP2/ERFs are key regulators of biosynthesis of a number of plant specialized metabolites including nicotine, artemisinin and steroidal glycoalkaloid (SGA) biosynthesis in tobacco (*Nicotiana tabacum*), sweet wormwood/annual wormwood (*Artemisia annua* L.), tomato (*Solanum lycopersicum*) and potato (*S. tuberosum*), respectively (Shoji et al., 2010; Yu et al., 2012; Cárdenas et al., 2016; Thagun et al., 2016). The AP2/ERF-domain factors, ORCA1 and ORCA2 were the first TFs isolated from *C. roseus*. The JRE element in the *STR* promoter was used in a yeast one-hybrid (Y1H) assay to identify these two AP2/ERF TFs (Menke et al., 1999b). Unlike ORCA1, ORCA2 is a JA-responsive transcriptional activator, which binds to the JRE of the *STR* promoter (Menke et al., 1999b). Later, another AP2/ERF, ORCA3 was identified in *C. roseus* using transfer DNA (T-DNA) activation tagging (van der Fits and Memelink, 2000). ORCA3 has been shown to induce the expression of a number of TIA biosynthetic genes, including *ANTHRANILATE SYNTHASE α* (*AS α*), *TDC*, *DXS*, *CPR*, *STR* and *D4H*, but not *IS* and *G10H* (van der Fits and Memelink, 2000; Van Moerkercke et al., 2015). *ORCA3*-overexpressing cell suspension cultures accumulate strictosidine, ajmalicine and an unidentified lochnericine-type TIAs, only when the precursors loganin and tryptamine are added to the medium. This is possibly due to the fact that most of the iridoid pathway genes are not regulated by ORCA3 (van der Fits and Memelink, 2000). Recently, the *C. roseus* genome-sequence analysis has revealed two more *AP2/ERF* genes, which are physically clustered with *ORCA3* at the same genomic scaffold (Kellner et al., 2015). But experimental evidence is lacking on their biological functions.

Basic leucine zipper (bZIP) factors: The bZIPs are a large family of TFs reported in all eukaryotes. The bZIPs bind DNA as homo- or hetero-dimers and have a characteristic highly conserved bZIP domain that is ~40- to 80-amino acids long which consists of two motifs; a basic region that binds DNA, and a leucine-zipper domain which is involved in protein–protein interaction (Ellenberger et al.; Vinson et al., 1989; Vinson et al., 2006).

In plants, bZIP factors are involved in numerous biological processes, including growth and development, light signaling, pathogen defense and secondary metabolism (Jakoby et al., 2002; Alves et al., 2013; An et al., 2017). The bZIP TFs are known to bind T/G-box motifs in the target promoters to regulate their expression. The *C. roseus STR* and *TDC* promoters contain G-box (CACGTG) and T/G-box (AACGTG) elements, respectively (Menke et al., 1999b). Two bZIP factors, CrGBF1 and CrGBF2 have been identified by Y1H screening using the G-box of *STR* promoter as bait (Sib eril et al., 2001). Both GBFs were shown to bind the G-box in *STR* promoter. However, they have weak affinity to the T/G-box in the *TDC* promoter. Moreover, CrGBFs repressed *STR* expression in plant cells suggesting that they act as repressors of TIA biosynthesis genes (Sib eril et al., 2001).

Zinc finger *C. roseus* transcription factors (ZCTs): Three zinc finger *C. roseus* transcription factors (ZCTs) have been identified by a Y1H screening using the elicitor-responsive DB region of the *TDC* promoter (Pauw et al., 2004). Sequence analyses revealed that the ZCTs contain a LxLxL type repressor motif. The ZCTs bind to the RV fragment of the *STR* promoter and the DB region of the *TDC* promoter *in vitro* in a zinc-dependent manner. Moreover, ZCTs repressed the expression of *TDC* and *STR* in *C. roseus* cells (Pauw et al., 2004). The repressor motif in ZCTs is most likely responsible for the repressor activity of these proteins. The similarities in expression profile and binding affinity suggests that ZCTs are functionally redundant. Like *ORCA3*, ZCTs were also induced by MeJA and yeast elicitor. The binding site for ZCT proteins in the *STR* promoter overlaps yet distinct from that of *ORCA3*. Both *ORCA3* and ZCTs bind to the DB region in the *TDC* promoter (Pauw et al., 2004). To determine the exact nature of interaction and/or competition between ZCTs and *ORCA3*, more detailed studies need to be performed.

AT-hook proteins: The bipartite JRE in the *ORCA3* promoter comprises an AT-rich quantitative sequence responsible for high-level expression, and a qualitative sequence, T/G-box (AACGTG), that acts as an on/off switch (Vom Endt et al., 2007). Five AT-hook proteins have been isolated in *C. roseus* using the JRE of the *ORCA3* promoter in a Y1H screening. The AT-hook TFs bound to the AT-rich quantitative sequence in the

ORCA3 promoter to activate its expression. These observations suggested that AT-hook proteins co-regulate the expression of *ORCA3* along with other regulators and act as transcriptional enhancers (Vom Endt et al., 2007). The JRE of *STR* promoter contains a GCC-box like element that is bound by ORCA3; however, a GCC-box is not present in the *ORCA3* JRE, suggesting that ORCA3 is not regulated by an AP2/ERF TFs (Vom Endt et al., 2007).

bHLH transcription factors: The bHLH factors, one of the largest families of TFs in plants, are involved in numerous developmental and metabolic processes, including pathogen defense, plant growth and development, phytohormone signaling and biosynthesis of specialized metabolites in plants (MacAlister and Bergmann, 2011; Patra et al., 2013b; Song et al., 2013; Li and Sack, 2014; Zhou and Memelink, 2016). The bHLH domain is comprising of ~60 amino acids, with the N-terminal ~15–20 basic amino acids involved in DNA binding and the HLH region which forms the homo- or heterodimeric complexes (Atchley and Fitch, 1997). The bHLH TFs characteristically bind to the E-box sequences (CANNTG), including the G-box (CACGTG) in the promoter of their target genes with some variation in binding specificity (Carretero-Paulet et al., 2010; Fernández-Calvo et al., 2011). The bHLH TFs are divided into 12 sub-groups (Heim et al., 2003) and the bHLH TF CrMYC2, belongs sub-group IIIe, was isolated by Y1H screen using the G-box of the *STR* promoter in *C. roseus* (Zhang et al., 2011). CrMYC2 acts upstream of *ORCA3* and regulates its expression by directly binding to the T/G-box motif in the JRE of the *ORCA3* promoter (Zhang et al., 2011). Similar T/G-box motif was also found in the *TDC* promoter; however, whether CrMYC2 co-regulates *TDC* along with ORCA3 remains to be tested. RNAi-mediated suppression of CrMYC2 in *C. roseus* cell cultures repressed JA-dependent induction of *ORCA2/3*, suggesting CrMYC2 activates *ORCA3* expression directly in response to JAs. In addition, RNAi-mediated suppression of *CrMYC2* reduced alkaloid accumulation in *C. roseus* cells (Zhang et al., 2011).

While considered to be a master regulator, ORCA3 alone is not sufficient to activate the entire TIA pathway. ORCA2 and ORCA3 probably regulate TIA biosynthesis in

combination with additional TFs (van der Fits and Memelink, 2000). This hypothesis is validated by the recently identified and characterized sub-group IVa bHLH TFs, bHLH IRIDOID SYNTHESIS 1 (BIS1) and BIS2 (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016). These two TFs act in a complementary manner with ORCA3 to regulate the iridoid pathway genes. The expressions of both *BIS1/2* are induced by MeJA, and are specifically expressed in IPAP cells where iridoid biosynthesis takes place (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016). Moreover, the over-expression of *BIS1* induced MIA production in *C. roseus* cell cultures without precursor feeding or JA induction (Van Moerkercke et al., 2015)

Box P-binding factors (BPF-1): In the *STR* promoter a region present upstream of the G-box and the JRE, known as the BA region, is responsive to yeast elicitor but not to MeJA (van der Fits et al., 2000). Y1H screening using this region identified the parsley box P-binding factor (BPF-1) that contains a MYB-like domain. It has been demonstrated that BPF-1 binds specifically to the BA fragment of *STR* promoter at two different sites (van der Fits et al., 2000). The expression of BPF-1 is induced by yeast elicitor but not by MeJA (van der Fits et al., 2000). These findings suggest that TIA biosynthesis involves a JA-dependent pathway regulated by ORCAs/bHLHs and a JA-independent pathway via BPF-1 (Thamm et al., 2016).

WRKY transcription factors: WRKY TFs are primarily plant-specific and known to bind the W-boxes (TTGACC/T) elements in target promoters (Yang et al., 2013). WRKY TFs have emerged as key regulators of specialized metabolite biosynthesis in a number of plant species (Schluttenhofer and Yuan, 2015), including benzyloquinoline alkaloids in *Coptis japonica* (Kato et al., 2007), artemisinin in *Artemisia annua* (Ma et al., 2009), triterpenoid withanolide in *Withania somnifera* (Singh et al., 2017) and camalexin in *Arabidopsis thaliana* (Ma et al., 2009). W-box elements are common in TIA biosynthetic gene promoters, suggesting the potential role of WRKY proteins in the regulation of TIA biosynthesis (Suttipanta et al., 2011). A group III WRKY TF, CrWRKY1 was identified and functionally characterized in *C. roseus* (Suttipanta et al., 2011). Ectopic expression of *CrWRKY1* induced the accumulation of serpentine in hairy

roots. *CrWRKY1* expression was mostly observed in roots, less in fruits and mature leaves and was induced by MeJA. In addition, CrWRKY1 bound to W-box in the *TDC* promoter and regulate its expression. As CrWRKY1 is not an activator it needs other co-activators to induce expression of *TDC in vivo* (Suttipanta et al., 2011). Analysis of the *CrWRKY1* promoter revealed that it contains several E-boxes, MYB recognition motifs and JA-responsive as-1 motifs (TGACG); however, W-box motifs are not found in the *CrWRKY1* promoter. These findings suggest that TFs belonging to MYB, bHLH and TGA families are potentially involved in the regulation of *CrWRKY1* (Yang et al., 2013). Additional MeJA-responsive WRKYs have also been implicated in regulation of TIA biosynthesis (Schlottenhofer et al., 2014).

1.4.2 Post-translational regulation of TIA biosynthesis in *C. roseus*

Accumulating evidence suggest that besides transcriptional regulation, metabolic pathways are also regulated by posttranscriptional and post-translational mechanisms (Maier et al., 2013; Patra et al., 2013b; Shen et al., 2017). Post-translational regulatory mechanisms, including phosphorylation of TFs, have been studied in nicotine and camalexin biosynthesis in tobacco and Arabidopsis, respectively (Xu et al., 2008; De Boer et al., 2011; Mao et al., 2011) and protein kinases (PKs) involved in these processes have been identified. In plants, the PK gene family is considered to be one of the largest and most highly conserved gene families. PKs phosphorylate proteins leading to the functional changes and are involved in approximately all biological processes, including plant growth, development, and biotic and abiotic stress responses. The most common protein kinase-mediated phosphorylation involves the mitogen-activated protein kinase (MAPK) cascade, normally consisting of three sequentially activated kinases, MAPK kinase kinase (MAPKKK)-MAPK kinase (MAPKK)-MAPK, encoded by multiple genes (Smékalová et al., 2014). In MAPK cascade, MAPKs phosphorylate specific substrates, including TFs and enzymes leading to the specific cellular responses.

During the past few years genome sequencing of different plant species has made it possible to identify and characterize members of the MAPK cascades in a number of plants. To date, MAPK cascade components have been reported from a wide range of

plant species, including *Arabidopsis* (Ichimura et al., 2002), rice (Hamel et al., 2006; Rao et al., 2009), maize (Kong et al., 2013b; Liu et al., 2013), tomato (Kong et al., 2012; Wu et al., 2014), *Brachypodium distachyon* (Chen et al., 2012), *Capsicum annuum* (Liu et al., 2014), and cucumber (Wang et al., 2015). In *C. roseus*, JA-induced expression of TIA pathway gene transcripts are reduced significantly in the presence of protein kinase inhibitors but are not inhibited by the protein synthesis inhibitor, cycloheximide. This observation indicates the possible role of protein phosphorylation in the JA signal transduction (Menke et al., 1999a). Moreover, an undefined posttranslational modification of ORCA3 protein is required for its transcriptional activity, and phosphorylation by a protein kinase has been hypothesized in previous reports (Menke et al., 1999a; van der Fits and Memelink, 2000). In *C. roseus* a JA-inducible MAPK CrMAPK3, has been reported previously and proposed to play a role in abiotic stress signal transduction and regulation of TIA biosynthesis. Transient over-expression of CrMAPK3 showed increased expression of TIA biosynthesis pathway genes (*STR*, *TDC*, *D4H* and *DAT*) and accumulation of TIAs in *C. roseus* leaf tissue. (Raina et al., 2012).

1.5 Metabolic engineering by assembling different branches of TIA pathway

Recently, significant progress has been made in the molecular and biochemical characterization of genes encoding key enzymes in the iridoid branch of TIA pathway, including 10HGO, IS, 7DLS, DLGT, and DL7H (Geu-Flores et al., 2012; Asada et al., 2013; Besseau et al., 2013; Salim et al., 2013; Miettinen et al., 2014; Salim et al., 2014) which are required for sequential conversion of geraniol to secologanin. Importantly, strictosidine was successfully produced *de novo* in a yeast strain by reconstruction of 14 known TIA pathway genes. The engineered yeast strain accumulates greater than 0.5 mg/L strictosidine (Brown et al., 2015). This approach provides a basis for better engineering the production of strictosidine and serves as a platform for producing any number of complex TIA derivatives. Likewise, the complete pathway from IPP to strictosidine has also been successfully reconstructed in *Nicotiana benthamiana* by agro-infiltration (Miettinen et al., 2014). The molecular and biochemical characterization of T3O and T3R concluded the 7-step tabersonine to vindoline pathway (Qu et al., 2015). The reconstruction of this pathway in yeast led to the accumulation of vindorosine and

vindoline from tabersonine as well as of vindoline from 16-methoxytabersonine has also been demonstrated (Qu et al., 2015). Therefore, both yeast strains and the transient expression in *N. benthamiana*, can be further developed and optimized for targeted metabolic engineering. These two biotechnological advances lay the foundation for producing complex TIAs in heterologous systems in future. These approaches could be used for the synthesis of anticancer drugs vinblastine and additional medicinal TIAs.

1.6 Conclusion

In recent years, an advance in numerous next-generation sequencing (NGS) techniques, such as RNA-seq has enabled the scientific community to generate large-scale transcriptome data from a wide range of plant tissues. These transcriptomic resources have been used to identify genes for the missing steps in complex metabolic pathways, including iridoid biosynthesis and the conversion of tabersonine to vindoline in *C. roseus*. Successful reconstructions of the strictosidine biosynthesis pathway in both yeast and *N. benthamiana*, and the 7-step conversion of tabersonine to vindoline in yeast strain, are the major breakthroughs in TIA pathway. These findings may help in large-scale production of valuable TIAs. While remarkable progress had been made in the last few years in isolating and characterizing genes encoding key enzymes in TIA pathway, the regulation of TIA biosynthesis is relatively less understood. This is probably due to the lack of appropriate genetic tools and highly complex nature of the biochemical pathway. Understanding the molecular mechanism of TIA regulation at both transcriptional and posttranslational level will enhance our ability to engineer the biochemical pathway to over-produce these valuable compounds in plants.

1.7 Outline of the dissertation

Catharanthus roseus (Madagascar periwinkle) produces numerous bioactive TIAs, including the chemotherapeutics, vinblastine and vincristine. However, the transcriptional and posttranslational regulation of TIA biosynthesis is not fully understood.

Therefore, the major objectives of my research are

- to isolate and characterize TFs regulating the TIA pathway,
- to elucidate the molecular mechanism underlying TIA regulation,
- to identify all the *C. roseus* MAP kinases (MAP Kinome), and
- to isolate and characterize candidate MAPKs involved in TIA regulation.

Chapter two focuses on functional characterization of an AP2/ERF TF gene cluster that modulates TIA biosynthesis in *C. roseus*. ORCA3, a JA responsive AP2/ERF TF, and its regulator, CrMYC2, a bHLH TF play key roles in TIA biosynthesis. *ORCA3* forms a physical cluster with two functionally uncharacterized AP2/ERFs, *ORCA4* and *ORCA5*. My research revealed that (i) the *ORCA* gene cluster is differentially regulated and (ii) *ORCA4* and *ORCA5* while functionally overlapping with *ORCA3*, regulates an additional set of TIA genes. Unlike *ORCA3*, *ORCA4* or *ORCA5* overexpression resulted in significant increase of TIA accumulation in *C. roseus* hairy roots. (iii) *ORCA5* directly regulates the expression of *ORCA4* and indirectly regulates *ORCA3*, likely via an unknown factor(s). Moreover, *ORCA5* shows auto-regulation. (iv) *ORCA5* also activates the expression of *ZCT3*, a negative regulator of the TIA pathway. In addition, (v) CrMYC2 is capable of activating *ORCA3* as well as co-regulating TIA pathway genes concurrently with *ORCA3*.

Part of the content of this chapter is already published

Paul, P., Singh, S. K., Patra, B., Sui, X., Pattanaik, S., & Yuan, L. (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *New Phytologist* 213:1107-1123.

Chapter three My research revealed that the *ORCA* gene cluster and CrMYC2 act downstream of a MAP kinase cascade that comprises a MAPK kinase, *CrMAPKK1*, *CrMAPK3* and *CrMAPK6*. Here I describe the identification and characterization of *CrMAPKK1* that is involved in TIA pathway regulation. Overexpression of *CrMAPKK1* in *C. roseus* hairy roots up regulated TIA pathway genes and boosts TIA accumulation.

The content of this chapter is already published

Paul, P., Singh, S. K., Patra, B., Sui, X., Pattanaik, S., & Yuan, L. (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in Catharanthus roseus. New Phytologist 213:1107-1123.

Chapter four describes the analysis of *C. roseus* MAP kinome as well as isolation and characterization of two previously uncharacterized MAPKs, *MAPK20* and *MAPK13* that regulate the TIA biosynthesis in *C. roseus*. Overexpression of *CrMAPK13* in *C. roseus* hairy roots upregulated the BIS-regulated iridoid pathway genes and induced tabersonine accumulation. Overexpression of *CrMAPK20* in *C. roseus* hairy roots repressed the MYC2-ORCA regulated TIA pathway genes and reduced catharanthine accumulation.

Chapter five summarizes the overall findings of this research project and possible future directions of TIA pathway regulation.

Copyright © Priyanka Paul 2017

Chapter 2

A differentially regulated AP2/ERF transcription factor gene cluster modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*

2.1 Introduction

Terpenoid indole alkaloids (TIAs) are a group of diverse plant natural products. *Catharanthus roseus* (Madagascar periwinkle) produces over a hundred TIAs, including the anticancer drugs, vinblastine and vincristine (Kellner et al., 2015). TIAs are condensed products of tryptamine, derived from the shikimate pathway, and secologanin, from the terpenoid pathway (Figure 1.1). In recent years, significant progress has been made in isolating and characterizing genes encoding key biosynthetic enzymes in the TIA pathway (Asada et al., 2013; Besseau et al., 2013; Qu et al., 2015). By comparison, transcription factors (TFs) regulating the TIA pathway are less characterized. TFs belonging to the families of AP2/ERFs (Menke et al., 1999b; van der Fits and Memelink, 2000), bHLH (Zhang et al., 2011; Van Moerkercke et al., 2015; Van Moerkercke et al., 2016), WRKY (Suttipanta et al., 2011), and C2H2 zinc fingers (Pauw et al., 2004; Rizvi et al., 2016), have been shown to be regulators of the TIA pathway. In addition, the bZIP factors, GBF1 and GBF2 (Sibénil et al., 2001), and a MYB-like factor, BPF1 (van der Fits et al., 2000), also influence TIA biosynthesis.

Jasmonic acid (JA) and its methyl derivatives, methyl jasmonates (MeJA), are major inducers of biosynthesis of many natural products, including taxol (Mirjalili and Linden, 1996), artemisinin (Shen et al., 2016), and nicotine (Shoji et al., 2000). JA and fungal elicitors also induce expression of the majority of regulatory and structural genes in the TIA pathway (Menke et al., 1999b). JA responsive elements (JRE) are frequently found in the promoters of TIA pathway genes.

In *C. roseus*, among the first isolated TFs are AP2/ERF-domain type factors, ORCA1 and ORCA2 (Menke et al., 1999b). ORCA2 is a JA-responsive transcriptional activator that

binds to JRE of the *STR* promoter (Menke et al., 1999b). Later, another AP2/ERF, ORCA3 was identified using T-DNA activation tagging and shown to induce the expression of numerous TIA biosynthetic genes in the indole and vindoline branches, including *AS α* , *TDC*, *DXS*, *CPR*, *STR*, and *D4H*, but not those in the iridoid branch, such as *IS* and *G10H* (van der Fits and Memelink, 2000; Van Moerkercke et al., 2015). Although considered to be a master regulator, ORCA3 alone is insufficient to activate the complete TIA pathway. ORCA2 and ORCA3 likely regulate TIA biosynthesis combinatorically with other TFs (van der Fits and Memelink, 2000). This assumption is supported by the recent identification and characterization of two bHLH TFs, BIS1 and BIS2, which acts in a complementary manner with ORCA3 to regulate structural genes in the iridoid branch (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016). The recently sequenced *C. roseus* genome has revealed that two AP2/ERF genes, *CroERF90* and *CroERF92*, are physically clustered with *ORCA3* at the same genomic scaffold, and thus proposed to be involved in regulation of the TIA pathway in concert with ORCA3 (Kellner et al., 2015). However, experimental evidence is lacking on their biological functions.

Gene clusters have been identified in a number of plant species including maize (Frey et al., 1997), rice (Shimura et al., 2007), oat (Qi et al., 2004), Arabidopsis (Field et al., 2011), opium poppy (Winzer et al., 2012), and tomato (Itkin et al., 2013). Most of the plant gene clusters identified thus far comprise non-homologous structural genes encoding enzymes involved in specialized metabolite biosynthesis (Nutzmann et al., 2016). Unlike the biosynthetic gene cluster, the clustered TFs consist of homologous genes that have most likely arisen due to tandem duplication events. To date, physically clustered genes encoding TFs have been identified in a limited number of plant species, including Arabidopsis, tobacco tomato and potato (Duarte et al., 2006; Shoji et al., 2010; Cárdenas et al., 2016). The *NIC2* locus of *Nicotiana tabacum*, which comprises at least seven AP2/ERFs are close homologs of ORCAs in *C. roseus*. The *NIC2* locus ERFs regulates the expression of structural genes in nicotine biosynthetic pathway and nicotine accumulation in tobacco. The *nic2* mutant has a low nicotine phenotype (Shoji et al., 2010). Overexpression of the *NIC2* locus ERFs, e.g. ERF189 and ERF221/ORC1

significantly induced nicotine biosynthesis in tobacco (Shoji et al., 2010; De Boer et al., 2011). Recently, an AP2/ERF, GLYCOALKALOID METABOLISM 9 (GAME9) has been characterized from tomato and potato for its role in steroidal glycoalkaloids (SGAs) and upstream isoprenoid biosynthesis. Genome sequence analysis revealed that GAME9 is a member of an AP2/ERF TF gene cluster in both tomato and potato. GAME9 knockdown and overexpression in tomato and potato affects the expression of SGAs and upstream mevalonate pathway genes, such as the cholesterol biosynthesis gene STEROL SIDE CHAIN REDUCTASE 2 (SSR2) (Cárdenas et al., 2016). However, the regulatory processes that govern the functions of TF clusters are unexplored.

The bHLH factor, CrMYC2, activates *ORCA3*, which in turn induces the expression of several TIA biosynthetic genes (Zhang et al., 2011). In tobacco, MYC2 regulates the *AP2/ERFs* within the *NIC2* locus, and also co-regulates several nicotine pathway genes with the AP2/ERFs by direct binding to the pathway gene promoters (De Boer et al., 2011; Shoji and Hashimoto, 2011). Whether CrMYC2 also co-regulates the same TIA pathway genes with *ORCA3* remains to be investigated. Such possible dual regulation of pathway genes underlines a layered, complex regulatory network for fine-tuning gene expression.

To date, our knowledge on the transcriptional regulation of clustered TFs is limited. In this study, we cloned and characterized *ORCA4* (*CroERF90*) and *ORCA5* (*CroERF92*), the two members of the *ORCA* gene cluster. We provide experimental evidence that the *ORCA* gene cluster is differentially regulated. *ORCA4*- but not *ORCA3*- is directly regulated by *ORCA5*. *ORCA5* exhibits auto-regulation. Additionally, *ORCA5* directly activates the expression of a negative regulator, *ZCT3*. While *ORCA4* and *ORCA5* functionally overlap with *ORCA3*, they modulate an additional set of TIA genes. Interestingly, unlike *ORCA3* and 4, *ORCA5* significantly activates other TIA pathway genes, such as *LAMT* and *SLS*. *ORCA4* or *ORCA5* overexpression resulted in dramatic increase of TIA accumulation in *C. roseus* hairy roots. In addition, CrMYC2 activates *ORCA3* and co-regulates the *TDC* and *CPR* promoters concomitantly with *ORCA3*.

2.2 Materials and Methods

2.2.1 Plant materials

Catharanthus roseus (L.) G. Don var. 'Little Bright Eye' was used for gene expression and cloning, and generation of transgenic hairy roots. *Nicotiana tabacum* var. Xanthi cell line was used for protoplast-based transient expression assays.

2.2.2 RNA isolation and cDNA synthesis

Ten-day-old *C. roseus* seedlings treated with 100 μ M methyl jasmonate (MeJA) for 2 h were used for total RNA isolation and cDNA synthesis as described previously (Suttipanta et al., 2007). *ORCA4* and *ORCA5* were PCR amplified from cDNA using gene-specific primers (Table 2.1), cloned into the pGEM-T Easy vector (Promega), and sequenced.

2.2.3 Multiple sequence alignment, phylogenetic and co-expression analysis

Alignment of deduced amino acid sequences was performed using the ClustalW program (Larkin et al., 2007) with the default parameters. Phylogenetic trees were constructed and visualized using the neighbor-joining method through MEGA version 5.2 software (Tamura et al., 2011). The statistical reliability of individual nodes of the newly constructed tree was assessed by bootstrap analyses with 1,000 replications.

In order to analyze the expression patterns of *C. roseus* TIA pathway genes and TFs, transcriptome data for five different tissues (flower, mature leaf, immature leaf, stem and root) were obtained from the Sequence Read Archive (SRA) database at the National Center for Biotechnology Information (accession number SRA030483). Raw reads were processed and reads per kilobase of transcript per million mapped reads (RPKM) value was calculated as described before (Singh et al., 2015). Pairwise Pearson correlation coefficient for each transcript was calculated using RPKM. Matrix distances for expression heatmap were computed over Pearson correlations of gene expression values (RPKM) by HEATMAP.2 function of GGLOT (v.3.0.1) Bioconductor package in R (v3.2.2) (Wickham, 2009).

2.2.4 Quantitative RT-PCR

Quantitative real-time PCR (qRT-PCR) was performed as described previously (Suttipanta et al., 2011). The primers used in qRT-PCR are listed in Table 2.1. In addition to the *C. roseus* 40S Ribosomal Protein S9 (*RPS9*) gene, *Elongation Factor 1 α* (*EF1 α*) was used as a second internal control (Liscombe et al., 2010). All PCRs were performed in triplicate and repeated at least twice.

Table 2.1 Oligonucleotides used in this study

<i>RPS9</i>	5'-GAGGGCCAAAACAACTTA-3'	5'-CCCTTATGTGCCTTTGCCTA-3'
<i>EF1α</i>	5'-TACTGTCCCTGTTGGTCGTG-3'	5'-AAGAGCTTCGTGGTGCATCT-3'
<i>ASa</i>	GCGAACATTTGCAGATCCAT	GGCCGATTTGTTATTGTTCC
<i>TDC</i>	5'-ATCCGATCAAACCCATACCA-3'	5'-CGTCATCCTCGACCATTTTT-3'
<i>G10H</i>	5'-TTATTTCGATTCTGCCAAGG-3'	5'-TCCCCAAAGTGAATCGTCAT-3'
<i>CPR</i>	5'-TGGCAGAAAAGGCTTCTGAT-3'	5'-CTCAGCCTGTGTGCTATCCA-3'
<i>SLS</i>	5'-GTTTCCTTCTCACCGGAGTTG-3'	5'-CCCATTTGGTCAACATGTCA-3'
<i>STR</i>	5'-ACCATGTGTGGGAGGACAT-3'	5'-ATTTGAATGGCACTCCTTGC-3'
<i>LAMT</i>	5'-AGCCAAGGCAGTGATTGTTG-3'	5'-CTGCAATGCGGAAAGGTTTG-3'
<i>IS</i>	5'-CCACATGATTGCCTTTTACCG-3'	5'-AAACCCGAAAACCAGAGCTG-3'
<i>ORCA3</i>	5'-CGGGATCCGAAATACAGAAA-3'	5'-GCCCTTATACCGGTTCCAAT-3'
<i>ORCA4</i>	5'-ATAGTAGTACTGCCGCCGAAAG-3'	5'-ATCTCCGCCGCAAATTTTCC-3'
<i>ORCA5</i>	5'-TCTTCCAACGGAGGTTAACGG-3'	5'-AATGTTGTCTCCAGGGCTTG-3'
<i>ZCT1</i>	5'-AGCCGAAAACATGCTTGT-3'	5'-CGCCTTTGCAACAGGTTTAT-3'
<i>ZCT2</i>	5'-CGTCAATTTCCATCGTTTCA-3'	5'-CCGATAGCGAATTCAGTCC-3'
<i>ZCT3</i>	5'-GACAAGCTTTGGGAGGACAC-3'	5'-GGCAAGGCAGGTAAGTTCAA-3'
<i>ORCA4</i>	5'-gcgagatct ATGCTTGATGAAATTAATTT-3'	5'-atggagctc CATCTTATTCACCTTTTGATTA-3'
<i>ORCA5</i>	5'-gcggatcc ATGGAGTTTTTCAGCTACTAATT-3'	5'-atgctgcag TTACAATATTGTCTCCTGTTTCA-3'
<i>ZCT3</i>	5'-gcggatcc ATGGCACTTGAAGCTTTGAATTC-3'	5'-arggatcc CTAATTAATTTGATGGTTTTCAAC-3'
<i>ORCA3- Promoter</i>	5'-gccaattc CGGAGAGTA AAAACATACTAAATG-3'	5'-atgaagctt AGAAACTGATT TTTTATGAAGTTTT-3'
<i>ORCA4- Promoter</i>	5'-gccaattc CTCGTACACTATATTTATGTTATG-3'	5'-agaagctt GCTTTTCCTTAGAGAGATTGATTG-3'
<i>ORCA5- Promoter</i>	5'-gccaattc TCGGATTGTCTATAACCCTTTGTG-3'	5'-agaagctt ATCAAGA CGGGAAAGATCAGATTC-3'
<i>ZCT3- Promoter</i>	5'-gccaattc ATGAGTGCGAACTTGATAGTCG-3'	5'-atgaagctt AGCAGAAGAATTTGAAGATTGAAT-3'
<i>STR- Probe</i>	5'-ACATCACTCTTAGACCGCCTTCT T TGAAAGTGATTTCCCTTGGACCTT-3'	5'-AAGGTCCAAGGGAAATCACTTCA AAGAAGGCGGTCTAAGAGTGATGT- 3'

Table 2.1 Oligonucleotides used in this study (continued.)

<i>ORCA4</i> - Probe	5'-CCTTCATAGCCCGCCCAATTGG TAA ACGTGCCACCAACCTCC-3'	5'- GGAGGTTGGTGGCACGT TTA CCAATTGGGCGGGCTATGAAGG-3'
<i>rolB</i>	5'-CTTATGACAAACTCATAGATAAAA GGTT-3'	5'- TCGTAACTATCCAACACACA TCAC-3'
<i>rolC</i>	5'-CAACCTGTTTCCTACTTTGTTAAC- 3'	5'- AAACAAGTGACACACTCAGCTTC- 3'
<i>virC</i>	5'-TTTTGCTCCTTCAAGGGAGGTGCC- 3'	5'-GGCTTCGCCAACCAATTTG GAGAT-3'
<i>GUS</i>	5'-ATGGTAGATCTGAGGGTAAA TTTC-3'	5'- GCTAGCTTGTTCCTCCCTG-3'
<i>ntpII</i>	5'-ATGGGGATTGAACAAGATGGA-3'	5'-TCAGAAGAAGCTCGTCAAGAAG -3'

2.2.5 Sub-cellular localization

For sub-cellular localization, the full-length cDNA of *ORCA4* or *5* was fused to the N-terminus of the enhanced GFP (*eGFP*) in a pBS plasmid under *CaMV35S* promoter and *rbcS* terminator to generate pORCA4/5-eGFP. A pBlueScript (pBS) plasmid containing the only eGFP was used as a control. The plasmids containing either *eGFP* or *ORCA4/5-eGFP* were individually electroporated into tobacco protoplasts as described previously (Pattanaik et al., 2010a) and visualized after 20h incubation in the dark under a fluorescent microscope (Eclipse TE200, Nikon, Japan) (Pattanaik et al., 2010b).

2.2.6 Protoplast isolation and electroporation

For transient protoplast assays, the reporter plasmids were generated by cloning *STR*, *TDC*, *CPR*, *LAMT*, *SLS*, *ZCT3*, *ORCA3*, *ORCA4* or *ORCA5* promoters in a modified pUC vector containing a firefly *luciferase* (*LUC*) and *rbcS* terminator. The effector plasmids were constructed by cloning *ORCA3/4/5*, and *CrMYC2* into a modified pBS vector under the control of the *CaMV35S* promoter and *rbcS* terminator. For transactivation assay, *ORCA3*, *ORCA4* and *ORCA5* were fused to the GAL4 DNA binding domain (GAL4-DBD) in a pBS plasmid containing *mirabilis mosaic virus* (*MMV*) promoter and *rbcS* terminator. The reporter plasmid used in the assay contains firefly *LUC* driven by a minimal *CaMV35S* promoter with five tandem repeats of *GAL4 Response Elements* (*5X GALRE*), and *rbcS* terminator fused. Protoplast isolation from *N. tabacum* cell suspension cultures and electroporation with plasmid DNA were performed as described previously (Pattanaik et al., 2010b). The reporter, effector, and internal control plasmids were

activities in transfected protoplasts were measured as described previously (Suttipanta et al., 2007).

2.2.7 Recombinant protein production and EMSA

The *ORCA3*, *4* and *5* genes were cloned into the pGEX 4T-1 vector (GE Healthcare Biosciences, Pittsburgh, PA, USA) to generate GST-fusion proteins. The constructs were verified by DNA sequencing and transformed into *Escherichia coli* BL21 cells containing pRIL (Agilent, Santa Clara, CA, USA). Protein expression was induced by adding isopropyl β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.1mM to the cell cultures at A600 ~ 0.8 and induced for 2 h at 37°C. The cells were harvested by centrifugation and lysed using CellLytic B (Sigma, USA) according to the manufacturer's instructions. The GST fusion proteins were bound to Glutathione Sepharose 4B columns (Amersham) and eluted by using 10mM reduced glutathione in 50mM Tris-HCl (pH 8.0) buffer. For EMSA, biotin-labeled DNA probes were synthesized by Integrated DNA Technologies (IDT) and annealed to produce double-stranded probes. Complementary DNA probes were designed to include the jasmonate-responsive elements (JRE) of *STR* promoter (GenBank accession no. Y10182) (Menke et al., 1999a) and putative GCC-element of the *ORCA4* promoter. EMSA experiment was carried out using light shift chemiluminescent EMSA kit (Thermo Fisher Scientific). For the binding reactions 40 fmole of DS DNA was incubated with purified protein for 60 mins at room temperature. The protein-DNA binding specificity was further confirmed by performing competition experiment, where 1000 fold excess amount of cold probe (without Biotin label) was added to the binding reactions. The DNA-protein complexes were resolved by electrophoresis on 6% non-denaturing polyacrylamide gels and then transferred to Biotinylated membrane (0.45 mm; Pierce). The band shifts were detected by a chemiluminescent nucleic acid detection module (Pierce) and exposed to x-ray films.

2.2.8 Construction of plant expression vectors and generation of hairy roots

For plant transformation *ORCA4*, *ORCA5* and *CrMYC2* cDNAs were PCR-amplified and cloned into a modified pCAMBIA2301 vector containing the *CaMV35S* promoter and the *rbcs* terminator (Pattanaik et al., 2010b). The pCAMBIA2301 vector was used as an

empty vector (EV) control. The plasmids were mobilized into *Agrobacterium rhizogenes* R1000. Transformation of *C. roseus* seedlings and generation of hairy roots were performed using the protocol described previously (Suttipanta et al., 2011). Transgenic status of the hairy root lines was verified by PCR amplification of *rolB*, *rolC*, *virC*, *nptII* and *GUS* genes. Independent hairy root lines were selected for further analysis.

2.2.9 Alkaloid extraction and analysis

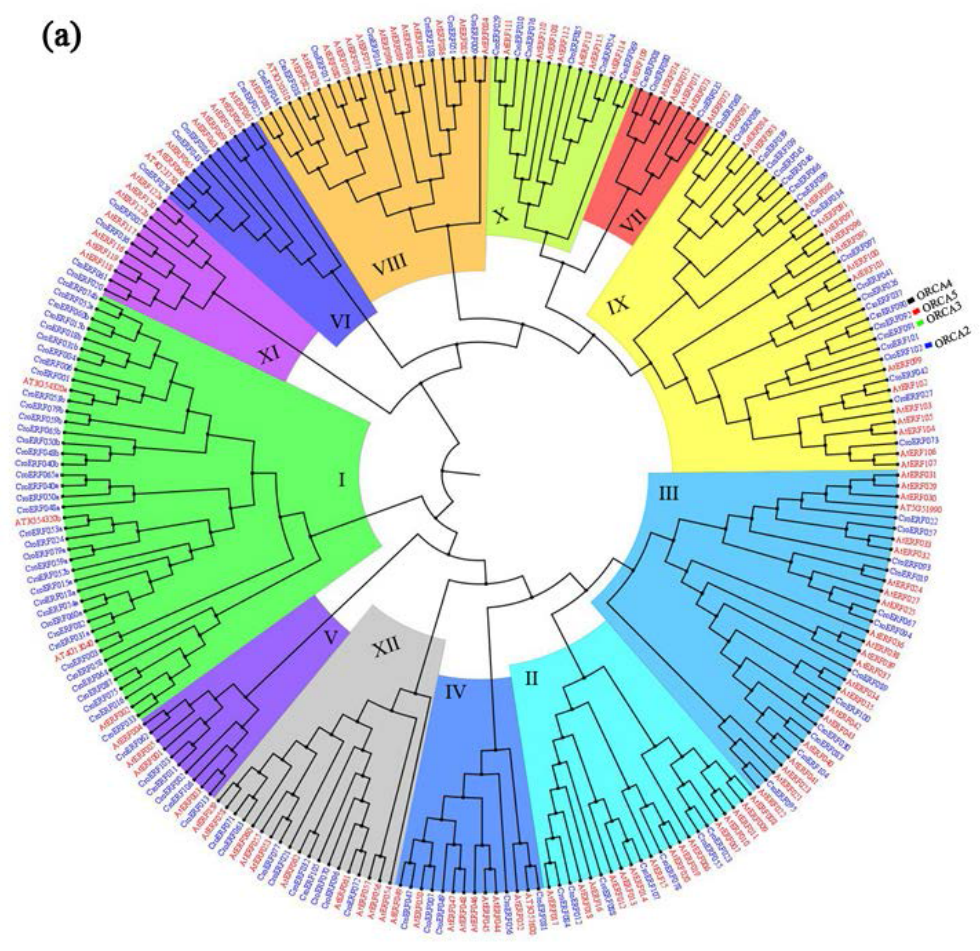
To extract alkaloids, transgenic hairy roots were frozen in liquid nitrogen and ground to powder. Samples were extracted in methanol (1:100 w/v) twice for 24 h on a shaker. Pooled extracts were then dried *via* a rotary evaporator and diluted in methanol 10 μ L/mg of the initial sample. The samples were then analyzed using high-performance liquid chromatography (HPLC), followed by electrospray-injection in a tandem mass spectrometry, as previously described (Suttipanta et al., 2011). The known alkaloid standards were run to identify elution times and mass fragments.

2.3 Results and Discussion

2.3.1 Two AP2/ERF genes form a cluster with *ORCA3*

AP2/ERFs have emerged as one of the major regulators of specialized metabolite biosynthesis in plants (Shoji et al., 2010; Yu et al., 2012). We identified 109 AP2/ERFs in *C. roseus* that were divided into 12 sub-groups by phylogenetic analysis. To conduct a cross-species comparison, a phylogenetic tree was constructed using the 109 AP2/ERF TFs from *C. roseus* and 122 from Arabidopsis (Figure 2.1a). Twelve main clades were formed, corresponding to subgroups I to XII of Arabidopsis. Two AP2/ERFs, CroERF90 and CroERF92, were of particular interest as they were homologous to ORCA3 in sub-group IX. The sub-group IX genes are involved in the regulation of defense responses and natural product biosynthesis facilitated by MeJA in numerous plant species, including Arabidopsis (McGrath et al., 2005), tobacco (Shoji et al., 2010), and *C. roseus* (van der Fits and Memelink, 2000). CroERF90 and CroERF92 have been recently identified as AP2 TF1 (030272) and AP2 TF2 (030274), respectively (Kellner et al., 2015). However, the two TFs have not been characterized and their biological functions remained unverified. Genome sequence analysis revealed that *CroERF90* (030272) and

CroERF92 (030274) are present adjacent to *ORCA3* (030273) in a 40 kb genomic scaffold, and the three physically clustered TFs do not contain any introns, suggesting that they may have arisen from tandem genome duplication events. Approximately 15% of the identified genes in all sequenced plant genomes contain significant numbers of tandemly arrayed gene families (Jander and Barth, 2007; Huang et al., 2009). Generally, tandem arrays contain two duplicated genes, and arrays with more than three genes are uncommon (Rizzon et al., 2006). TF families have higher expansion rates in plants than in animals (Shiu et al., 2005); however, the *Arabidopsis* genome contains only one region with more than five duplicated TF genes in tandem, although several such tandemly duplicated regions are present in soybean (Schmutz et al., 2010). The biological significance of such TF gene clusters has not been fully elucidated. The TF cluster, identified as the *NIC2* locus in tobacco, consists of seven homologous AP2/ERFs involved in the regulation of nicotine biosynthesis (Shoji et al., 2010). We cloned *CroERF90* (designated as *ORCA4*; GenBank accession no. KR703577) and *CroERF92* (*ORCA5*; GenBank accession no. KR703578). *ORCA4* is a 223 amino acid protein, having a calculated molecular mass of 24.4 kDa. The 311 amino acid *ORCA5* has a calculated molecular mass of 34.2 kDa. *ORCA4* and *ORCA5* share 53% amino acid sequence identity, and 38% and 47% sequence identity, respectively, with *ORCA3* (Figure 2.1b). The highest sequence identity was found within the conserved DNA binding AP2 domain.



(b)

```

ORCA4 1  --MDEITISASDLALLSEIEEHLSDS-----SEILENSDDDD-----SSNHLIIS
ORCA5 1  MEFSAINFISSDLSRLDSITQHLLSDSNFSDIFFEFFSESAPESSVISGDSSTKLFTEV
ORCA3 1  --MSEIISVSDRELLSEIEEHLSDS-----NSDDSS-----SELTSTE

ORCA4 46  SWEE-----TILKSNLPEPPAGSLSSSECSNSSTAT-----
ORCA5 61  NWSQKFSAASTEVEGNISPOLISVNIILELSESESSRISSPNDNIELSEVNGDYLEFL
ORCA3 39  NWE-----IFADFIN-----SGSEIC

ORCA4 77  ----AAG-----SLSSSECSNS-----TATAAGS
ORCA5 121 ATLTEAGNISPOLIAYVLSSEESSTISSPGDNIFIYTEANGGDLGFNYVDSNSPTSTES
ORCA3 57  ----KRG-----SPSSESC-----SNEMAES

ORCA4 97  PSSSCCN--SSTAASSSEENSSEWRYRGVRRRLWGKFAAEIRDPNKKGSRINLGTYE
ORCA5 181 QEDMAGKKKAGAAAEKGNAPIDCWRYRGVRRRFPWGFFAAEIRDPNKKGSRINLGTYE
ORCA3 75  CEDSVVG--PPEAAAGG-CSEKDWRYRGVRRRFPWGFFAAEIRDPKKKGSRINLGTYE

ORCA4 155  TPEDAALAYDGAAPKMRGSKAKLNFPHLIGSNISGPVRVNPRKRY---RKLTITSSSYSS
ORCA5 241  TPEDAALAYDRAAFQIRGAKALNFPHLIN-CVACQPRVGCRRRA-----MSSSSSESE
ORCA3 132  TPEDAALAYDAAAFNMRGAKARLNFPHLIGSNISGPVRVNPRKREPAEPSTISSSSSSS

AP2 Domain
ORCA4 212  EDNEGRKKRRN----
ORCA5 295  ESHGGKMKQETIL
ORCA3 192  SENSGRKKRRY----

```

Figure 2.1 Phylogenetic analysis of AP2/ERF families and Amino acid sequence alignment of ORCA3, 4, and 5. **a.** *Arabidopsis thaliana* and *Catharantus roseus* AP2/ERFs were used to generate the phylogenetic tree. **b.** Multiple sequence alignment of ORCA3, ORCA4 and ORCA5 of *C. roseus*. Identical amino acid residues in the alignment are shaded in black and similar amino acid residues in gray. The AP2 DNA binding domain is underlined.

2.3.2 ORCA3, ORCA4, and ORCA5 are preferentially induced in roots by MeJA

The expression of *ORCA3* is induced by MeJA (van der Fits and Memelink, 2000). To examine the JA-inducibility of *ORCA4* and *ORCA5* in various tissues, qRT-PCR was used to measure the expression of *ORCA3*, *ORCA4*, and *ORCA5* in MeJA-treated *C. roseus* seedlings, leaves, and roots (Figure 2.2). The three genes shared similar MeJA-induced expression patterns; however, the magnitudes of induction were different. *ORCA3* expression was induced by approximately 2.5-fold in leaves but 13-fold in roots. *ORCA4* and *ORCA5* were induced in leaves by 1.5- and 2.5-fold, respectively, and up to 4-5 fold in roots.

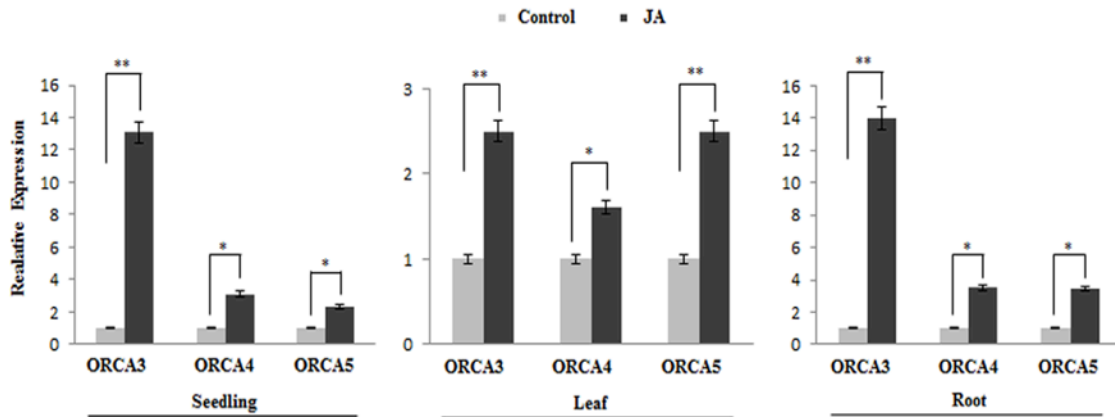


Figure 2.2 MeJA induction of ORCA3, ORCA4, and ORCA5. Ten day-old *C. roseus* seedlings were treated with 100 μ M MeJA for 2h, and gene expression in whole seedling, leaves, and roots were measured by qRT-PCR. Mock-treated seedlings were used as

controls. Data represent means \pm SDs of two biological samples. Statistical significance was calculated using the Student's t-test; * p value <0.05 , ** p value <0.01 .

2.3.3 ORCA4 and 5 localized to nucleus

To determine the sub-cellular localization, the ORFs of *ORCA4* and *ORCA5* were fused in-frame to the enhanced green fluorescent protein (eGFP) and expressed in tobacco protoplasts. While eGFP was detected throughout the cell, the ORCA4/5-eGFP fusion protein was localized to the nucleus (Figure 2.3a). These observations indicate that ORCA4 and 5 are nuclear proteins, which is consistent with their putative role as transcription factors.

2.3.4 ORCAs are transcriptional activators

To determine the transactivation activity, ORCA3, ORCA4 or ORCA5 fused to the GAL4-DBD were co-electroporated into tobacco protoplasts with a *luciferase* reporter driven by a minimal *CaMV35S* promoter with GAL4-responsive elements. Luciferase activities were significantly higher (approximately 6.5 to 12-fold) in tobacco protoplasts electroporated with either of these three ORCAs compared to the reporter only control (Figure 2.3b). Transactivation activities of ORCA3 (6.5-fold), ORCA4 (approximately 6.6-fold) and ORCA5 (12-fold) were significantly higher compared to the control. The significant inductions of reporter activity in plant cells suggest that three ORCAs are transcriptional activators.

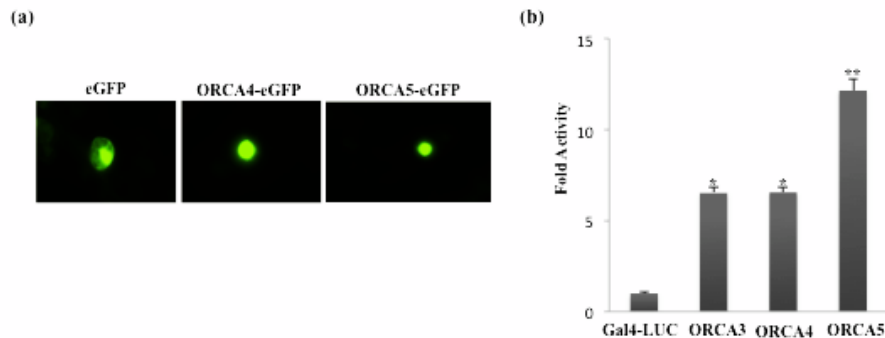


Figure 2.3 Nuclear localization and transactivation assays of ORCA4 and ORCA5 in tobacco protoplasts. a. Sub-cellular localization of ORCA4 and ORCA5. ORCA4/5-

eGFP is localized to the nucleus, whereas eGFP is accumulated throughout the cell in tobacco protoplasts. **b.** ORCA3, 4 and 5, fused to the GAL4 DBD (amino acid 1–147), were cloned into a modified pBS vector containing the *MMV* promoter and *rbcs* terminator to generate BD-ORCA3/4/5 plasmids, which were co-electroporated into tobacco protoplasts with a reporter vector containing the firefly *luciferase* gene under the control of a minimal *CaMV35S* promoter and 5X GAL response elements. The control represents the reporter alone. A plasmid containing the *GUS* reporter, driven by the *CaMV35S* promoter and *rbcs* terminator, was used as a control for normalization. Luciferase and GUS activities were measured 20 h after electroporation. Luciferase activity was normalized against GUS activity. Data presented here are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05, ** p value <0.01.

2.3.5 ORCA4 and 5 activate promoters of multiple TIA pathway genes

We next performed co-expression analysis of the ORCA gene cluster based on *C. roseus* transcriptomes available in the Medicinal Plant Genomic Resource (MPGR) (Góngora-Castillo et al., 2012). TIA pathway genes are regulated by at least two independent transcriptional cascades. The CrMYC2-ORCA3 cascade regulates key genes in the indole pathway including *TDC*, as well as some genes involved in the conversion of the loganic acid to downstream TIAs, including *LAMT*, and *STR*. The majority of the iridoid pathway genes, including *GES* and *G10H*, are regulated by another cascade featuring BIS1/BIS2 (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016). Co-expression analysis of key structural and TF genes in the TIA pathway revealed two major clusters: one contains the *ORCA* gene cluster and its target genes; the other includes *BIS1*, *IS*, and *G10H*. The *ORCA*-containing cluster is further divided into two sub-clusters. In one, *ORCA4* was grouped with *CrMYC2*, *ORCA3*, *STR*, *TDC*, *LAMT* and *CPR*; in the other, *ORCA5* was grouped with *ORCA2* and several negative regulators of *ORCA3*, such as *CrWRKY1* and *ZCTs* (Figure 2.4a).

To experimentally verify the involvement of ORCA4 and ORCA5 in TIA biosynthesis, we used a protoplast-based transactivation system (Pattanaik et al., 2010b) to determine the ability of the TFs to activate selected TIA pathway genes. The *TDC*, *CPR*, and *STR*

promoters, fused to a firefly *luciferase* reporter gene, were co-electroporated into tobacco protoplasts with/without the constructs expressing *ORCA4* or *ORCA5*. As a positive control, *ORCA3*, known to activate the three promoters, was included in the assays (Figure 2.4b). *ORCA3*, 4, and 5 strongly activated the *STR* promoter by 10-, 8-, and 4-fold, respectively, compared to the control. The activations of the *CPR* and *TDC* promoters by the three factors were relatively moderate, ranging from 1.5- to 2-fold. Previous work also showed similar promoter preference for *ORCA3* (van der Fits and Memelink, 2000). Collectively, our results suggest that *ORCA4* and *ORCA5*, individually, are activators of the three TIA pathway genes with varying magnitudes. The sequence divergence outside of the conserved AP2 DNA-binding domain of ORCAs may be partly responsible for the differential activation of the target promoters.

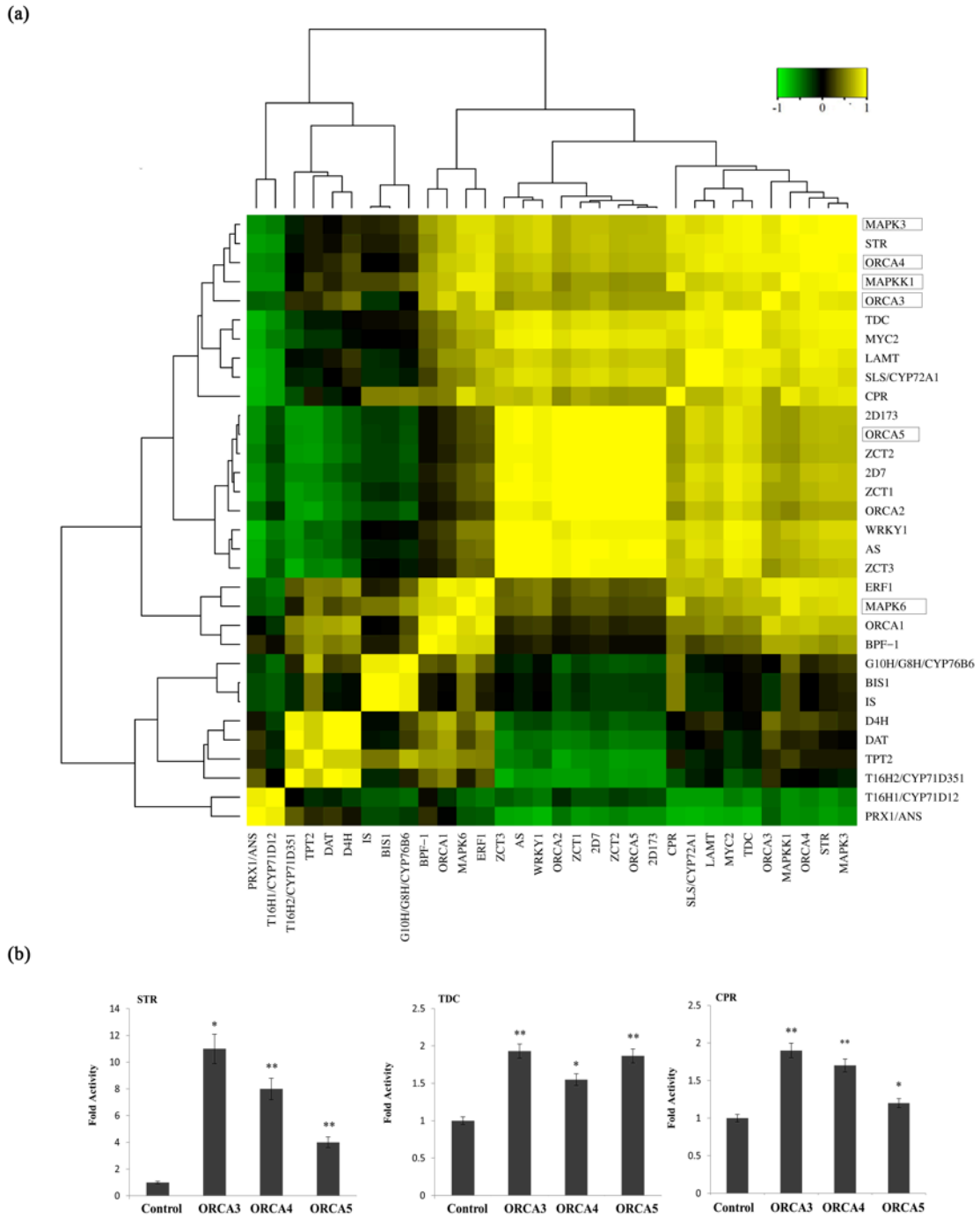


Figure 2.4 Co-expression of *ORCA4* and *ORCA5* with TIA regulatory and structural genes and transactivation of key TIA pathway gene promoters by *ORCA3/4/5* in tobacco protoplasts. a. Hierarchical clustering and heat-map show that *ORCA4* was co-expressed with *CrMYC2* and *ORCA3*, and their targets, *STR*, *TDC*, *LAMT*, *SLS*, *CPR*;

while *ORCA5* was clustered with *CrWRKY1*, *ORCA2*, and *ZCTs*. **b.** Transactivation of gene promoters of *TDC*, *CPR*, and *STR* by *ORCA3*, *ORCA4*, and *ORCA5*. *TDC*, *CPR* and *STR* promoters fused to *luciferase* reporter were electroporated into tobacco protoplasts either alone or with an effector plasmid (*ORCA3*, *ORCA4* or *ORCA5*). A plasmid containing the *GUS* reporter, controlled by the *CaMV35S* promoter and *rbcs* terminator, was used as a control for normalization. Luciferase and GUS activities were measured 20 h after electroporation. Luciferase activity was normalized against GUS activity. Control represents the reporter alone without effectors. Data presented are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05, ** p value <0.01.

2.3.6 ORCA4 and ORCA5 bind to the JRE of the *STR* promoter

We have shown that *ORCA4* and *5* activate the promoters of key TIA pathway genes, including *STR*, *TDC* and *CPR* in tobacco cells (Figure 2.4b). A previous report has shown that *ORCA3* binds to the JRE of the *STR* promoter in a sequence-specific manner (Van Der Fits and Memelink, 2001). To determine whether *ORCA4* and *5* bind the JRE in the *STR* promoter, electrophoretic mobility shift assays (EMSAs) was used. Recombinant, glutathione S-transferase (GST)-tagged *ORCA3/4/5* proteins (GST-*ORCA3/4/5*) were purified from *E. coli* using GST affinity chromatography (Figure 2.5a). Purified GST-tagged *ORCA3*, *ORCA4* or *ORCA5* proteins were incubated with the 5' biotin-labeled probes covering the JRE region of the *STR* promoter. Similar to *ORCA3*, *ORCA4* and *ORCA5* bind to the JRE resulting in a mobility shift (Figure 2.5b). The binding ability of GST-*ORCA3/4/5* proteins was further confirmed by a competition experiment using unlabeled (cold) probes. The binding of *ORCA3/4/5* proteins to the JRE was competed out the labeled probe by increasing the concentrations (1000X in excess) of unlabeled probe (Figure 2.5b), suggesting that the *ORCA* gene cluster binds to a JA-responsive GC-rich element in *STR* promoter.

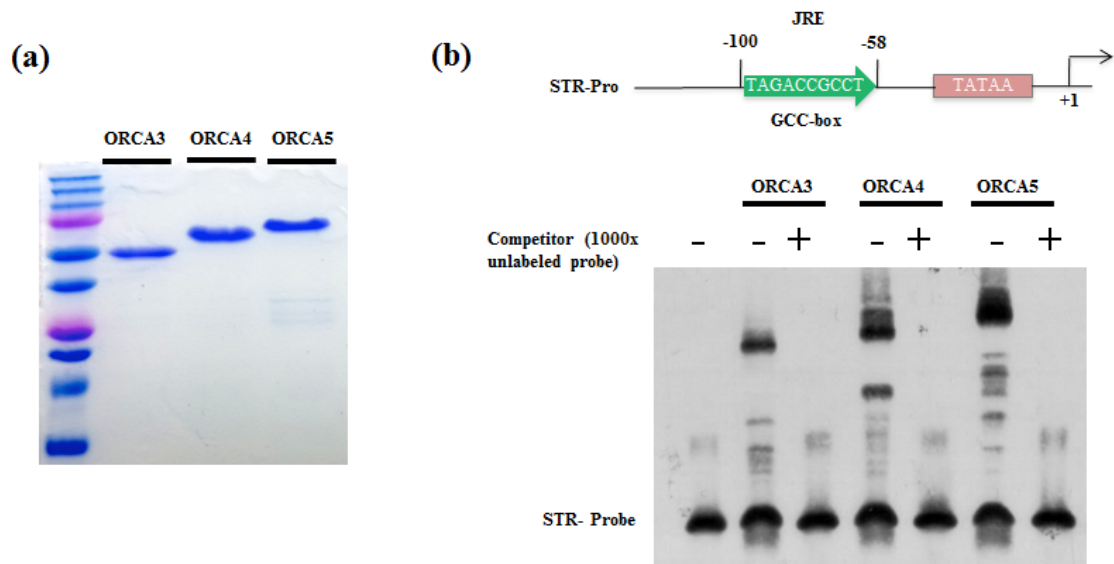


Figure 2.5 Binding of ORCA3/4/5 to the JRE of the *STR* promoter. **a.** Bacterial expression-purification of ORCA3, 4 and 5 proteins. **b.** Autoradiograph shows up-shifted bands in EMSA using purified GST-ORCA3/4/5 proteins and biotin-labeled *STR* probe (containing JRE). The DNA-protein binding was further confirmed by performing competition experiment using 1000 fold excess amount of unlabeled probe.

2.3.7 Ectopic expression of *ORCA4* or *ORCA5* activates key TIA pathway genes and boosts metabolites accumulation in *C. roseus* hairy roots

To further substantiate our findings and to demonstrate the role of ORCA4 and 5 in TIA production, we generated transgenic *C. roseus* hairy roots overexpressing *ORCA4* or *ORCA5*. The transgenic status of the hairy roots was confirmed by PCR (Figure 2.6a and 2.8a), and two transgenic lines (OE-1 and OE-2) for each construct were selected for further analysis. Compared to the empty vector (EV) control, expressions of *ORCA4* and *ORCA5* were 100-120 fold and 24-45 fold higher in the transgenic lines, respectively (Figure 2.7a and 2.8b). Expression of *ASα*, *TDC*, *CPR*, *G10H*, *IS*, *LAMT*, *SLS*, and *STR*, were 2 to 40 fold higher in the *ORCA4*-overexpression lines compared with the EV control (Figure 2.7a). Expression of *ASα*, *TDC*, *CPR*, *G10H*, *IS*, *SLS*, *STR*, and *SGD* were also significantly higher in the *ORCA5*-overexpression lines compared with the EV control (Figure 2.9a). Expression of the genes encoding C2H2 zinc finger TFs, *ZCT1*,

ZCT2 and *ZCT3* repressors of TIA biosynthesis, were increased in both *ORCA4* and *ORCA5* overexpression lines (Figure 2.7a and 2.9a). The higher expression of the *ZCT* repressors in *ORCA4/ORCA5*-overexpression lines indicates the existence of a negative regulatory loop that is probably involved in the fine-tuning of TIA biosynthesis. Interestingly, expression of the genes encoding the TIA pathway activators, such as CrMYC2, *ORCA3* and *ORCA4* were also significantly increased in *ORCA5* overexpression lines (Figure 2.9a).

Previous studies show that overexpression of *ORCA3* in *C. roseus* hairy roots only induces the expression of a subset of TIA pathway genes. Consequently, TIA production does not increase in the transgenic *ORCA3* hairy roots (Peebles et al., 2009). However, overexpression of *ORCA4* or *ORCA5* significantly increased the transcripts levels of genes in both indole (i.e. *ASa*, *TDC*) and iridoid pathways (i.e., *CPR*, *G10H*, *IS*, *LAMT* and *SLS*). In addition, expression of *STR* and *SGD* were also significantly increased. *STR* catalyzes the condensation of indole (tryptamine) and iridoid (secologanin) precursors to form the first TIA (strictosidine). *SGD* removes glucose moiety from strictosidine to form stictosidine aglycone that subsequently yields cathenamine, which acts as precursor for catharanthine and tabersonine. To determine the metabolic outcomes of *ORCA4/ORCA5*-overexpression, we measured the alkaloids in two independent hairy root lines using HPLC-ESI-MS/MS as described previously (Suttipanta et al., 2011). Tabersonine, ajmalicine, and catharanthine accumulation increased significantly both in *ORCA4* and *ORCA5* hairy root lines compared with EV lines (Figure 2.7b and 2.9b). Notably, the production of tabersonine was increased by greater than 40-fold in *ORCA4* and approximately 7-15-fold in *ORCA5* overexpression lines. In *C. roseus* roots, tabersonine is converted to the hörhammericine and echitovenine catalyzed by minovincinine 19-O-acetyltransferase (*MAT*) (Laflamme et al., 2001) and tabersonine 19-hydroxylase (*T19H*) (Giddings et al., 2011). We found that expressions of both *MAT* and *T19H* are significantly induced in *ORCA5*-overexpressing transgenic hairy roots lines, indicating the increased accumulation of hörhammericine in hairy roots (Figure 2.9a). Collectively, these findings suggest that *ORCA4* and *ORCA5* have a wide range of target

genes and only a subset of which are regulated by ORCA3. Therefore, the individual components of the ORCA gene cluster are functionally overlapping as well as divergent.

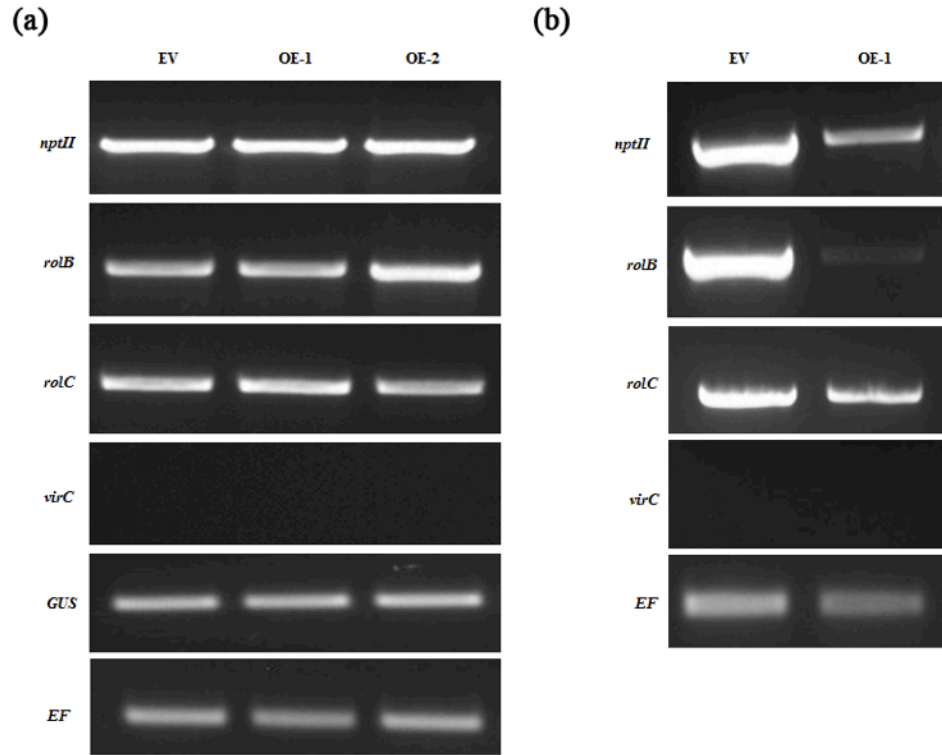


Figure 2.6 PCR confirmation of the transgenic status of *C. roseus* hairy root lines overexpressing *ORCA4* or *CrMYC2*. **a.** Gene-specific primers amplified various fragments in the T-DNA from the Empty Vector (EV) control and *ORCA4*-overexpression hairy root lines (OE-1 and OE-2). **b.** Transgenic status of the hairy roots of EV control and *CrMYC2*-overexpression were confirmed by PCR amplification. *rolB* and *rolC*, protein-tyrosine phosphatase RolB and C, respectively; *virC*, virulence protein C; EV, empty vector; *EF1 α* , elongation factor 1 α ; *GUS*, β -glucuronidase; *NPTII*, neomycin phosphotransferase II; OE, overexpression.

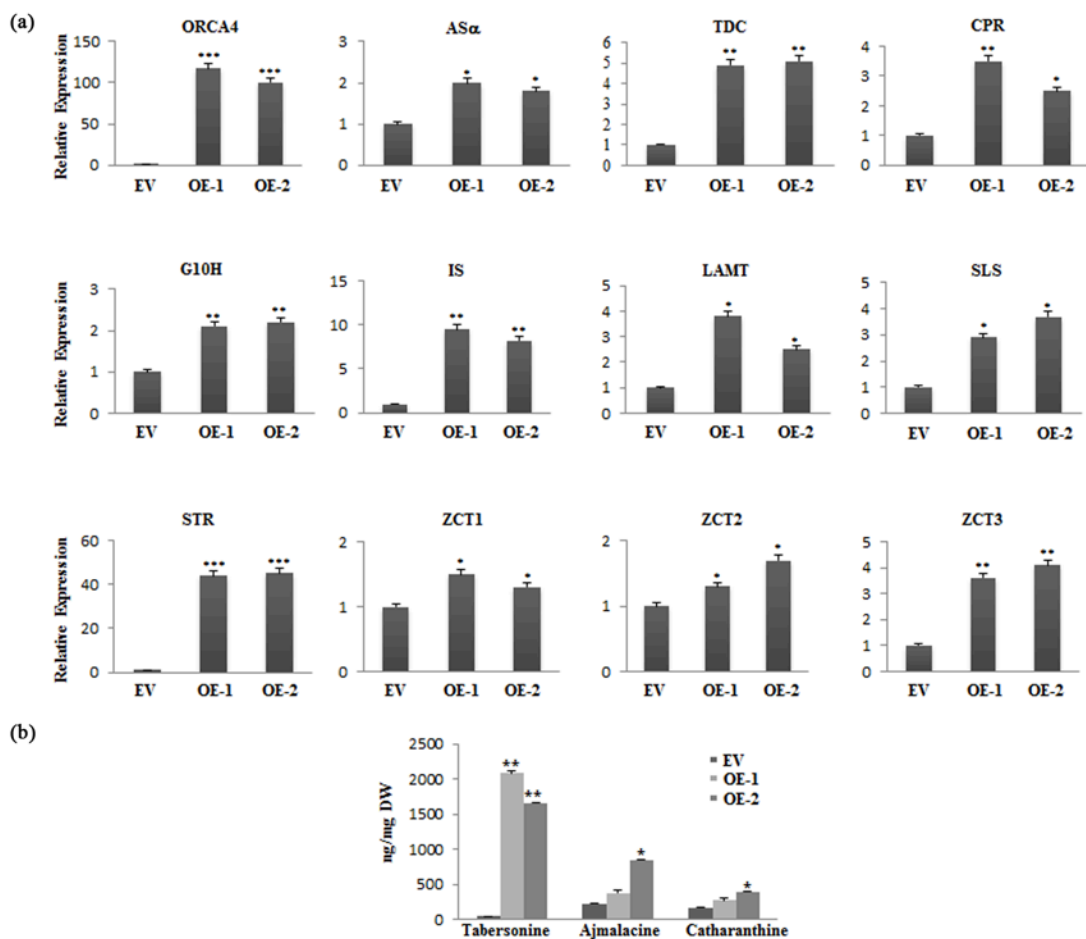


Figure 2.7 Increase of key TIA pathway gene expression and alkaloid accumulation in *ORCA4*-overexpression *C. roseus* hairy roots. **a.** Relative expression of TIA pathway genes in empty vector (EV) control and two *ORCA4*-overexpression (OE-1 and OE-2) hairy root lines measured by qRT-PCR. **b.** Measurement of tabersonine, ajmalicine, and catharanthine in EV control and OE-1 and OE-2. Alkaloids were extracted and analyzed by HPLC-MS/MS, and the levels of the alkaloids were estimated based on peak areas compared to standards. Alkaloid levels are indicated in ng/mg dry weight (DW). Data presented here are the means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test, * p value <0.05, ** p value <0.01, *** p <0.001.

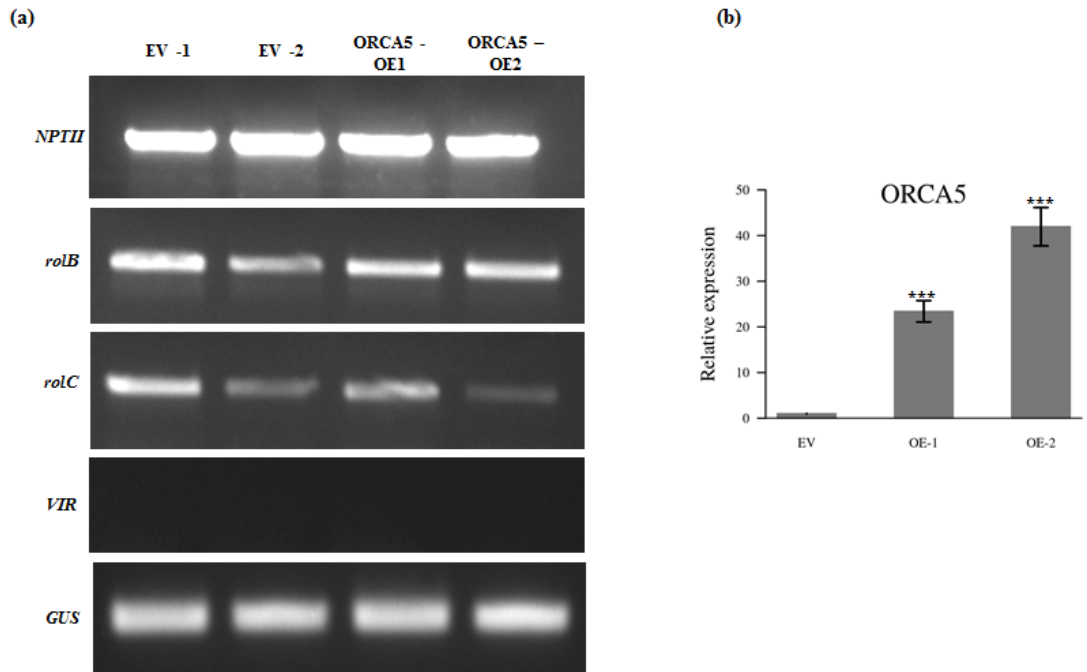


Figure 2.8 PCR confirmation of the transgenic status of *C. roseus* hairy root lines overexpressing *ORCA5* **a.** Gene-specific primers amplified various fragments in the T-DNA from the Empty Vector (EV) control and *ORCA5*-overexpression hairy root lines (OE-1 and OE-2). **b.** Relative expression of *ORCA5* in empty vector (EV) control and two overexpression (OE-1 and OE-2) hairy root lines measured by qRT-PCR. *rolB* and *rolC*, protein-tyrosine phosphatase RolB and C, respectively; *virC*, virulence protein C; EV, empty vector; *EF1 α* , elongation factor 1 α ; *GUS*, β -glucuronidase; *NPTII*, neomycin phosphotransferase II; OE, overexpression.

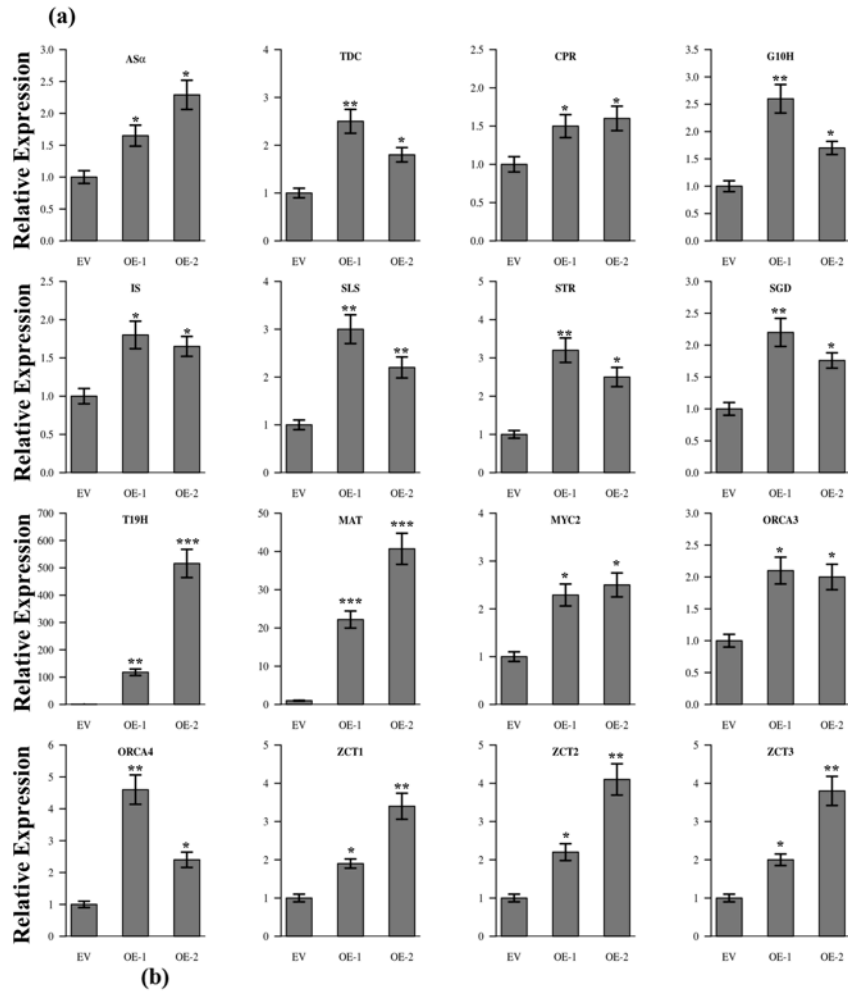


Figure 2.9 Increase of key TIA pathway gene expression and alkaloid accumulation of *C. roseus* hairy root lines overexpressing *ORCA5*. a. Relative expression of TIA pathway genes in empty vector (EV) control and two *ORCA5*-overexpression (OE-1 and OE-2) hairy root lines measured by qRT-PCR. b. Measurement of tabersonine,

ajmalicine, and catharanthine in EV control and OE-1 and OE-2. Alkaloids were extracted and analyzed by HPLC-MS/MS, and the levels of the alkaloids were estimated based on peak areas compared to standards. Alkaloid levels are indicated in ng/mg dry weight (DW). Data presented here are the means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test, * p value <0.05, ** p value <0.01, *** p <0.001.

2.3.8 *ORCA3/4/5* are differentially regulated

The bipartite JRE in the *ORCA3* promoter comprises an AT-rich quantitative sequence responsible for high-level expression, and a qualitative sequence, T/G-box (AACGTG), that acts as an on/off switch (Vom Endt et al., 2007). CrMYC2 binds to the qualitative sequence to activate *ORCA3*, which in turn activates several TIA pathway genes, including *TDC*, *CPR*, and *STR*. Our analysis of the *ORCA4* and *5* promoters revealed the presences of T/G boxes in a similar sequence organization found in the *ORCA3* promoter (Figure 2.10a). The question thus arose as to whether or not CrMYC2 also regulates *ORCA4* and *ORCA5*. We used two independent approaches to address this question. We cloned the promoter sequences of *ORCA4* and *ORCA5* and used the protoplast-based transactivation assay to determine the ability of CrMYC2 to activate luciferase-reporter gene expression driven by the promoters of *ORCA3*, *4*, or *5*. The reporter plasmids were individually co-electroporated into tobacco protoplasts with or without the *CrMYC2*-expression vector. CrMYC2 strongly activated the *ORCA3* promoter, but not that of *ORCA4* or *ORCA5* (Figure 2.10b). These findings suggest that TF(s) other than CrMYC2 may be involved in the regulation of *ORCA4* and *ORCA5*. We also generated hairy root lines overexpressing *CrMYC2* (Figure 2.6b) and examined the expression of *ORCA4* and *ORCA5*. Transcriptomic analysis of the *CrMYC2*-overexpression lines showed increased expression of *ORCA4* and *ORCA5*, which were further verified by qPCR using independently isolated mRNA from the hairy roots (relative expression of *ORCA4* and *ORCA5* increased 2.5- and 2.0-fold, respectively, compared to the EV control) (Figure 2.10c). Our results suggest that CrMYC2 directly regulates *ORCA3* by binding to its promoter, and indirectly controls *ORCA4* and *ORCA5*, likely via the activation of a yet to be identified transcriptional activator that binds to the *ORCA4/5* promoters.

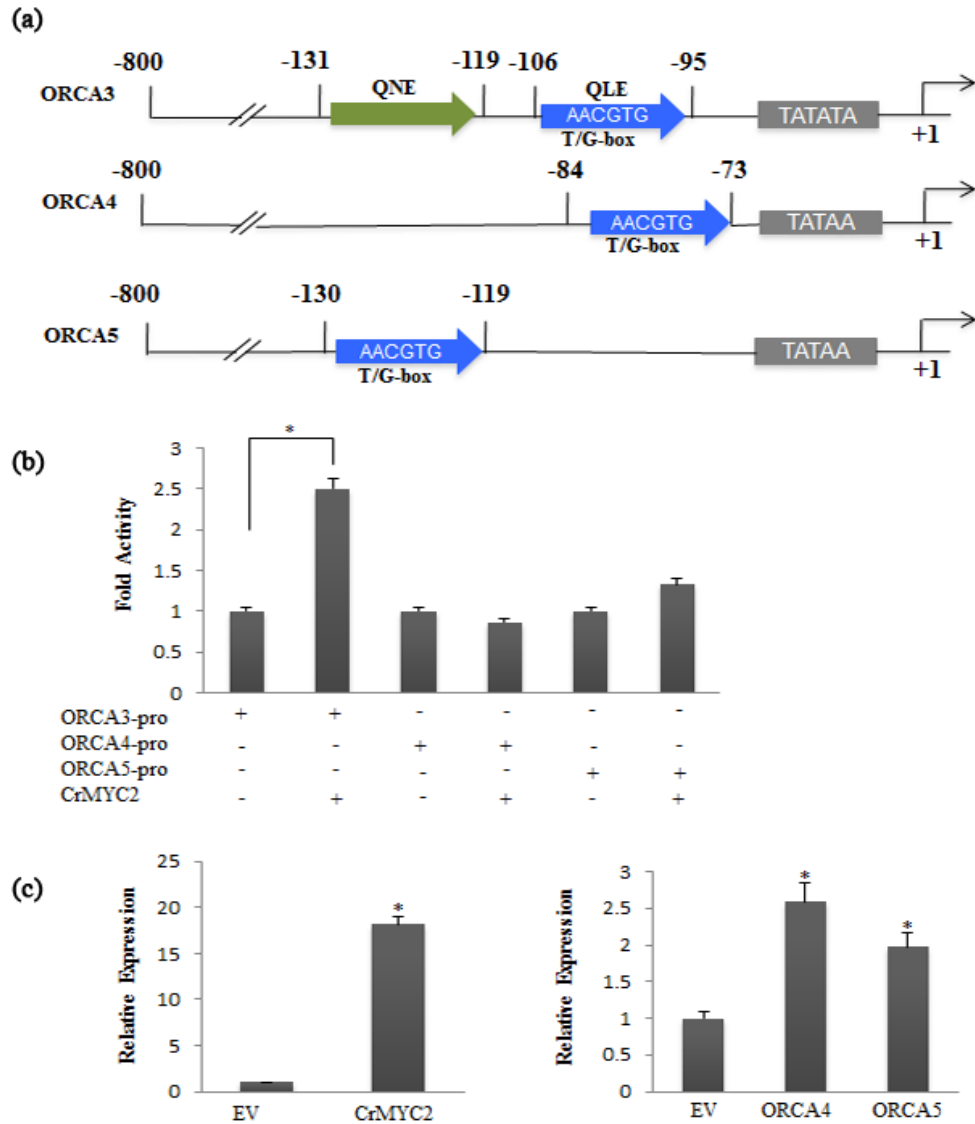


Figure 2.10 Differential regulation of the *ORCA* gene cluster. a. Schematic diagrams show the key elements in the promoters of *ORCA3*, 4, and 5. The positions of the T/G-box (qualitative elements; QLE), and quantitative elements (QNE) are indicated. b. CrMYC2 transactivation of the *ORCA3*, 4, and 5 gene promoters (-pro), fused to the *luciferase* reporter, were carried out by co-electroporation of the CrMYC2-expression vector with an *ORCA*-pro-reporter construct into tobacco protoplasts. A plasmid containing *GUS* reporter, driven by the *CaMV35S* promoter and *rbcS* terminator, was used as a control for normalization. Luciferase and GUS activities were measured 20 h

after electroporation. Luciferase activity was normalized against GUS activity. Control represents the reporter alone without effectors. Data presented are means \pm SDs of three biological replicates. **c.** Relative expression of *CrMYC2*, *ORCA4*, and *ORCA5* in EV control and transgenic *CrMYC2*-OE hairy roots were measured using qRT-PCR. Data represent means \pm SDs of two biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05.

2.3.9 ORCA TF cluster differentially activate TIA pathway genes

The ORCAs are known to directly regulate a number of genes in indole and down stream branches in TIA pathway (van der Fits and Memelink, 2000; Paul et al., 2017). In *C. roseus*, synthesis of secologanin requires nine enzymes and seven of them are regulated by BIS1 and BIS2 (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016). To determine whether ORCA TF cluster directly regulates the expression of *LAMT* and *SLS*, we used protoplast-based transactivation system (Pattanaik et al., 2010b). The *LAMT* and *SLS* gene promoters fused to a firefly *luciferase* reporter gene were co-electroporated into tobacco protoplasts with/without the constructs expressing *ORCA3*, *ORCA4* or *ORCA5* (Figure 2.11). *ORCA3*, *ORCA4* and *ORCA5* activated the *LAMT* promoter significantly compared to the control. Interestingly, *ORCA5* significantly activated the *SLS* promoter (approximately 2.5-fold) compared to the control; however, no significant activation was observed for *SLS* promoter by *ORCA3* and *ORCA4* (Figure 2.11). Collectively, our results suggest that *ORCA3*, *ORCA4* and *ORCA5* differentially activate the TIA biosynthetic pathway genes in *C. roseus*.

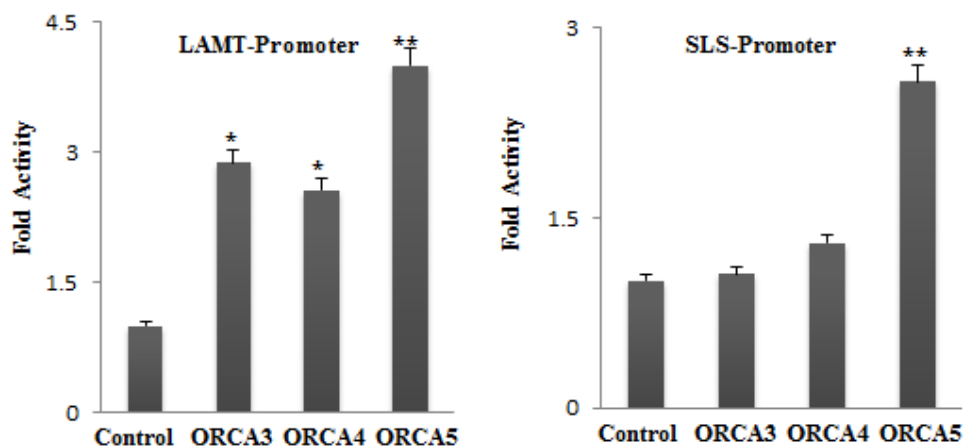


Figure 2.11 Transactivation of TIA pathway gene promoters by ORCA3/4/5 in tobacco protoplasts. Transactivation of *LAMT* and *SLS* promoters by ORCA3, ORCA4, and ORCA5. *LAMT* and *SLS* promoters fused to *luciferase* reporter were electroporated into tobacco protoplasts either alone or with an effector plasmid (*ORCA3*, *ORCA4* or *ORCA5*). A plasmid containing the *GUS* reporter was used as a control for normalization. Luciferase and GUS activities were measured 20 h after electroporation. Luciferase activity was normalized against GUS activity. Control represents the reporter alone without effectors. Data presented are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05, ** p value <0.01.

2.3.10 CrMYC2 co-regulates TIA pathway genes with ORCA3 by direct binding to a T/G-box

It is unclear whether CrMYC2 directly targets the TIA pathway genes, thus dually regulating the genes with ORCA3/4/5. We analyzed the *cis*-elements in promoters of ORCA-regulated TIA pathway genes, *TDC*, *CPR*, and *STR*. The *TDC* promoter contains two elicitor-responsive elements between -99 and -87 and between -87 and -37, relative to transcription start site (TSS) (Ouwerkerk and Memelink, 1999). A T/G box (AACGTG) is present in the elicitor-responsive region between -99 and -87 (Figure 2.14). Moreover, the T/G box in the *TDC* promoter is identical to the one present in the JRE of the *ORCA3* promoter to which CrMYC2 binds. A G-box (CACGTG) and a T/G box-like

element (TACGTG) are also found adjacent to the elicitor responsive region in the *STR* and *CPR* promoters, respectively (Figure 2.12). Guided by this observation, we hypothesized that *TDC* and other TIA pathway genes are dually regulated by both ORCA3 and CrMYC2. We then tested the effects of *ORCA3-CrMYC2* co-expression on the *TDC*, *CPR*, and *STR* promoters, using the protoplast-based transactivation assay. While ORCA3 and CrMYC2 individually activated both the *TDC* and *CPR* promoters, *ORCA3-CrMYC2* co-expression further enhanced the activation of the two promoters, suggesting that CrMYC2 not only directly activates *TDC* and *CPR*, but also acts synergistically with ORCA3 (Figure 2.13a). Although a G-box exists in the *STR* promoter, CrMYC2 by itself did not activate the promoter, and, consequently, no additive enhancement of *STR* activation was observed in *ORCA3-CrMYC2* co-expression (Figure 2.13a). Our result is in agreement with a previous report that CrMYC2 does not activate the *STR* promoter (Zhang et al., 2011). Moreover, a GCC-motif in the *STR* promoter is sufficient to confer ORCA-mediated JA response that is not affected by deletion of the G-box from the *STR* promoter (Menke et al., 1999b). We next addressed the question of whether CrMYC2 activates the *TDC* promoter through binding to the T/G box. We mutated the T/G box sequence in the *TDC* promoter by site-directed mutagenesis, and tested for activation of the mutated promoter (mTDC) by ORCA3 and CrMYC2, individually or in combination. ORCA3, but not CrMYC2, activated mTDC (Figure 2.13b), indicating that the T/G box is necessary for CrMYC2 binding. We also converted CrMYC2 to a dominant repressor by fusing CrMYC2 with the EAR-motif repression domain (SRDX) (Hiratsu et al., 2003), and expected CrMYC2-SRDX, that binds the T/G box, to repress *TDC* promoter activity. In the protoplast assay, CrMYC2-SRDX repressed the CrMYC2-induced *TDC* promoter activation by 3-fold (Figure 2.13b). *TDC* promoter activity in the CrMYC2-SRDX assay was 50% lower than that of the promoter-only control, indicating that the dominant repressor is capable of suppressing background activities, presumably induced by the heterologous factors in the protoplasts. In contrast to the synergistic effect of *ORCA3-CrMYC2* co-expression on the *TDC* promoter, co-expression of *ORCA3/CrMYC2-SRDX* did not show any additive enhancement. Our findings further clarified the compositional makeup of the CrMYC2-mediated regulatory network controlling TIA biosynthesis.

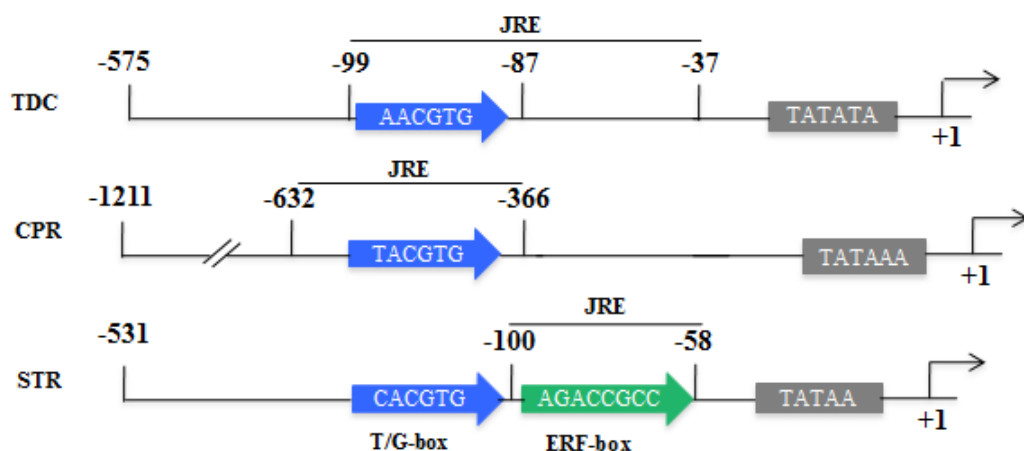


Figure 2.12 Schematic diagrams of three TIA pathway promoters regulated by **CrMYC2** and **ORCA3**. The *TDC*, *CPR* and *STR* promoters contain one or more jasmonate-responsive elements (JRE). The *TDC* promoter contains two JRE between -99 to -87 and -87 to -37. One of them contains a T/G box (AACGTG). The JRE in the *CPR* promoter resides between -632 to -366 and contains a T/G box-like (TACGTG) element present close to it. The JRE in the *STR* promoter resides between -100 to -58 with a G-box adjacent to it.

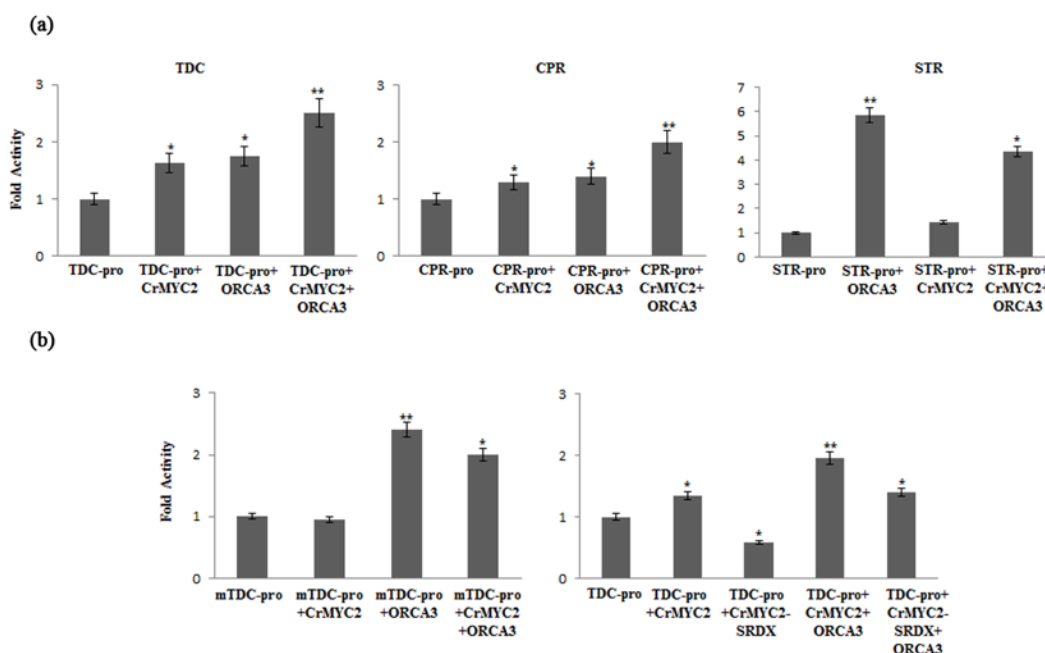


Figure 2.13 CrMYC2 co-regulation of key TIA pathway genes with ORCA3 via binding to T/G-box. a. Transactivation of gene promoters of the *TDC*, *CPR*, and *STR* by CrMYC2, ORCA3, and the combination of the two factors. **b.** Transactivation of the mutant (*mTDC*) and wild-type (*TDC*) promoters by CrMYC2, ORCA3, and the combination of ORCA3 with either CrMYC2 or CrMYC2-SRDX. The *TDC*, *mTDC*, *CPR*, and *STR* promoters, fused to the luciferase reporter, were electroporated alone or with an effector plasmid (ORCA3, CrMYC2, or CrMYC2-SRDX) into tobacco protoplasts. A plasmid containing *GUS* reporter, controlled by the *CaMV35S* promoter and *rbcs* terminator, was used as a control for normalization. Luciferase and *GUS* activities were measured 20 h after electroporation. Luciferase activity was normalized against *GUS* activity. Control represents the promoters alone without effectors. Data presented here are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test, * p value <0.05, ** p value <0.01.

2.3.11 ORCA5 activates the *ORCA4* but not *ORCA3*

Expressions of *ORCA3* and *ORCA4* were increased significantly in *ORCA5*-overexpressing transgenic hairy root lines (Figure 2.9a). We thus hypothesized that *ORCA5* regulates the expression of *ORCA3* and *ORCA4*. To test our hypothesis, *ORCA3* and *ORCA4* promoters fused to a firefly *luciferase* reporter were co-electroporated into tobacco protoplasts with/without the constructs expressing *ORCA3*, *ORCA4* or *ORCA5*. While there was no significant effect of *ORCA4* on its own promoter activation, *ORCA5* significantly activated *ORCA4*. In addition *ORCA4* was weakly activated by *ORCA3* (Figure 2.14b). On the other hand, transcriptional activity of *ORCA3* promoter was not induced by any of these three ORCAs (Figure 2.14a). Next, we asked whether *ORCA5* regulates itself (auto-regulation) or regulated by *ORCA3* or 4. To test this possibility, the *ORCA5* promoter fused to a firefly *luciferase* reporter gene, were co-electroporated with/without *ORCA3*, *ORCA4* or *ORCA5* into tobacco protoplasts. Our results showed that *ORCA5* moderately activated itself compared to the control (Figure 2.14c). *ORCA3* or *ORCA4* had no effects on the transcriptional activity of the *ORCA5* promoter. Taken together, these findings suggest that *ORCA4* is directly activated by *ORCA5* and

indirectly by ORCA3 via yet to be identified TF(s). Moreover, self-regulation of ORCA5 suggests the presence of a positive auto-regulatory loop in TIA gene regulatory network.

2.3.12 ORCA3 and 5 bind to the *ORCA4* promoter

We next analyzed the *ORCA4* promoter and identified a GC-rich motif (AGCCCGCCC), a putative binding site for AP2/ERFs and mutated it to AGCAAAACC by site-directed mutagenesis. The mutant promoter *mORCA4-pro* fused to *luciferase* reporter was used as a reporter in protoplast assay for activation by ORCA5. Mutation in the putative AP2/ERF binding site significantly reduced the activation of the *ORCA4* promoter by ORCA5 (Figure 2.14b), suggesting ORCA3 and ORCA5 can activate the *ORCA4* possibly by binding to the GC-rich motif in its promoter in the plant cell.

To determine whether ORCA3 and 5 binds the GC-rich element, EMSA was used. The purified recombinant, GST-tagged ORCA3 and 5 proteins (GST- ORCA3/5) were incubated with 5' biotin-labeled probes covering the GC-rich element of the *ORCA4* promoter. Figure 2.14d shows that ORCA3 and ORCA5 proteins interacted with the GC-rich element resulting in a mobility shift. The binding ability of GST-ORCA3/5 proteins was further confirmed by a competition experiment using unlabeled (cold) probes. The binding of the biotin-labeled probes by the ORCA3, or 5 could be eliminated by excess concentrations (1000x) of the unlabeled (cold) probe (Figure 2.14d), suggesting that ORCA3 and 5 regulate *ORCA4* by directly binding to its promoter.

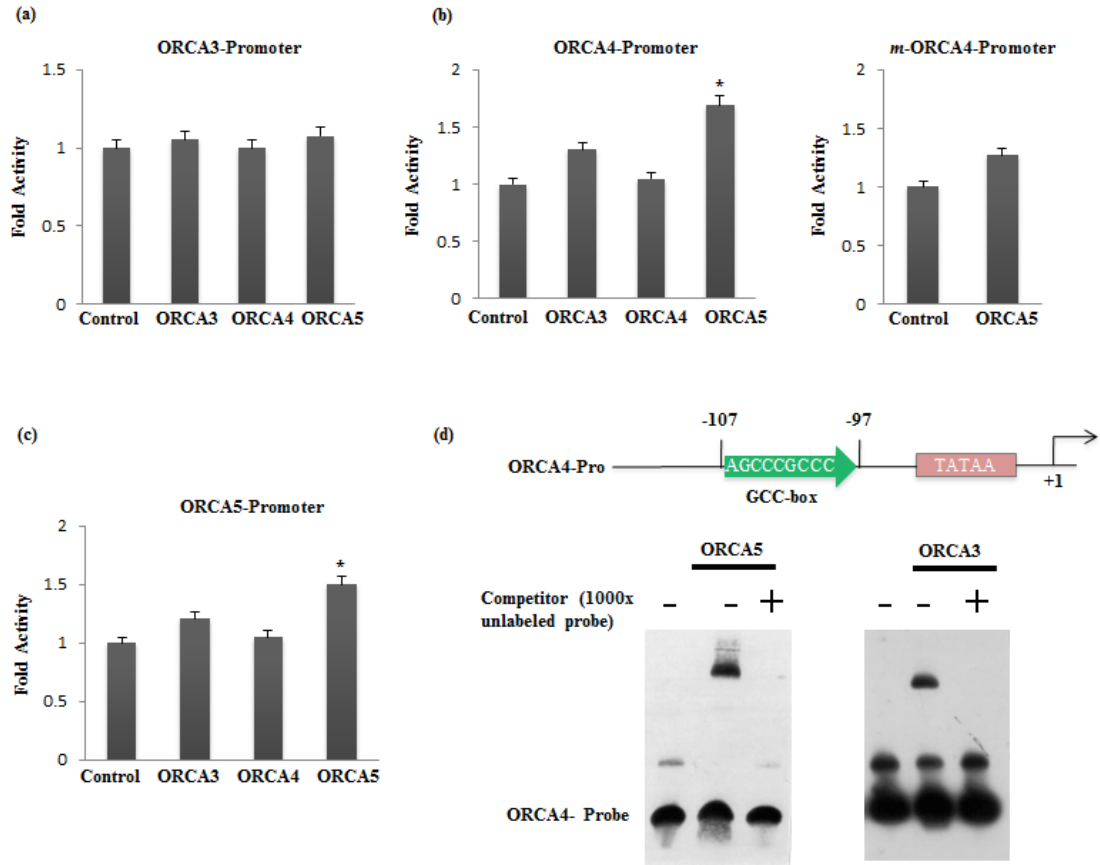


Figure 2.14 ORCA5 activates ORCA4 by binding to the putative GC-rich motif in the *ORCA4* promoter, and auto-regulates in protoplast transient assay. Activation of *ORCA3* (a), *ORCA4*, *ORCA4* mutant (b) and *ORCA5* (c) by ORCAs, promoters fused to *luciferase* reporter were electroporated into tobacco protoplasts either alone or with an effector plasmid (*ORCA3*, *ORCA4* or *ORCA5*). A plasmid containing the *GUS* reporter, controlled by the *CaMV35S* promoter and *rbcS* terminator, was used as a control for normalization. Luciferase and GUS activities were measured 20 h after electroporation. Luciferase activity was normalized against GUS activity. Control represents the reporter alone without effectors. Data presented are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05. **d.** Autoradiograph shows up-shifted bands in EMSA using purified GST-ORCA3 and 5 proteins and biotin-labeled *ORCA4* probe (containing GC- rich motif). The DNA-protein binding was further confirmed by performing competition experiment using 1000 fold excess amount of unlabeled *ORCA4* probe.

2.3.13 ORCA5 activates the expression of ZCT3 in *C. roseus* hairy roots and tobacco cells

The ZCT proteins are known as negative regulators of TIA pathway (Pauw et al., 2004). ZCTs bind to *TDC* and *STR* promoter and repress their expression in *C. roseus* cell line. In both *ORCA4* and *ORCA5* overexpression hairy root lines, expressions of ZCTs were increased significantly (Figure 2.7a and 2.9a). To determine whether the expression of these repressors are directly regulated by ORCAs, we first analyzed the *cis*-elements in promoters of *ZCT1*, *ZCT2* and *ZCT3* *in silico* and a putative GC-rich motif was found in the *ZCT3* promoter. To test whether *ZCT3* is regulated by ORCAs, *ZCT3* promoter fused to a firefly *luciferase* reporter gene was co-electroporated into tobacco protoplasts with/without the constructs expressing *ORCA3*, *ORCA4* or *ORCA5* (Figure 2.15a). Only *ORCA5* significantly activated the *ZCT3* promoter, compared to the control. To test whether *ORCA5* is regulated by *ZCT3*, *ORCA5* promoter fused to a firefly *luciferase* reporter gene, were co-electroporated into tobacco protoplasts with construct expressing *ZCT3*. However, no significant difference was observed (Figure 2.15b). Activation of *ZCT3* by *ORCA5*, suggests that ZCT is possibly involved in controlling magnitude of activation of ORCA targets and modulates accumulation of TIAs in plants.

2.3.14 *C. roseus* ORCAs and tobacco *NIC2*-locus ERFs are functionally interchangeable

C. roseus ORCA3/4/5 are closely related to the tobacco *NIC2* locus AP2/ERFs, ERF189 and ERF221 (a.k.a. ORC1) (Shoji et al., 2010). In addition, both ORCAs and *NIC2* locus ERFs are induced by MeJA and recognize GC-rich *cis*-elements in target promoters in two distinct metabolic pathways. For instance, a P-box element (CCGCCCTCCA) in *PMT* gene promoter is recognized by ERF189, ERF179, ERF163, and ERF115 in tobacco (Shoji et al., 2010). Here, we have shown that all three members of ORCA cluster bind to the GCC-like element (TAGACCGCCT) in *STR* promoter (Van Der Fits and Memelink, 2001). It is, therefore, intriguing to know whether the *C. roseus* ORCAs and tobacco *NIC2*-locus ERFs are functionally interchangeable. To address this question, we used a protoplast-based transactivation system (Pattanaik et al., 2010b). The promoter of *putrescine N-methyltransferase (PMT)*, a rate-limiting enzyme in nicotine biosynthetic

pathway in tobacco, fused to a firefly *luciferase* reporter gene was co-electroporated into tobacco protoplasts with/without the plasmids expressing *ERF221*, *ORCA3*, *ORCA4* or *ORCA5* (Figure 2.15c). *ERF221*, *ORCA3* and *ORCA5* significantly activated the *PMT* promoter, compared to the control. Similarly, the *STR* promoter fused to a firefly *luciferase* reporter gene, was co-electroporated into tobacco protoplasts with/without the construct expressing *ORCA3*, *ERF189* and *ERF221*. Both *ERF189* and *ERF221* significantly activated the *STR* promoter (Figure 2.15c), suggesting that the tobacco *ERF189* and *ERF221* are functional equivalents of ORCAs in *C. roseus* (Figure 2.15d).

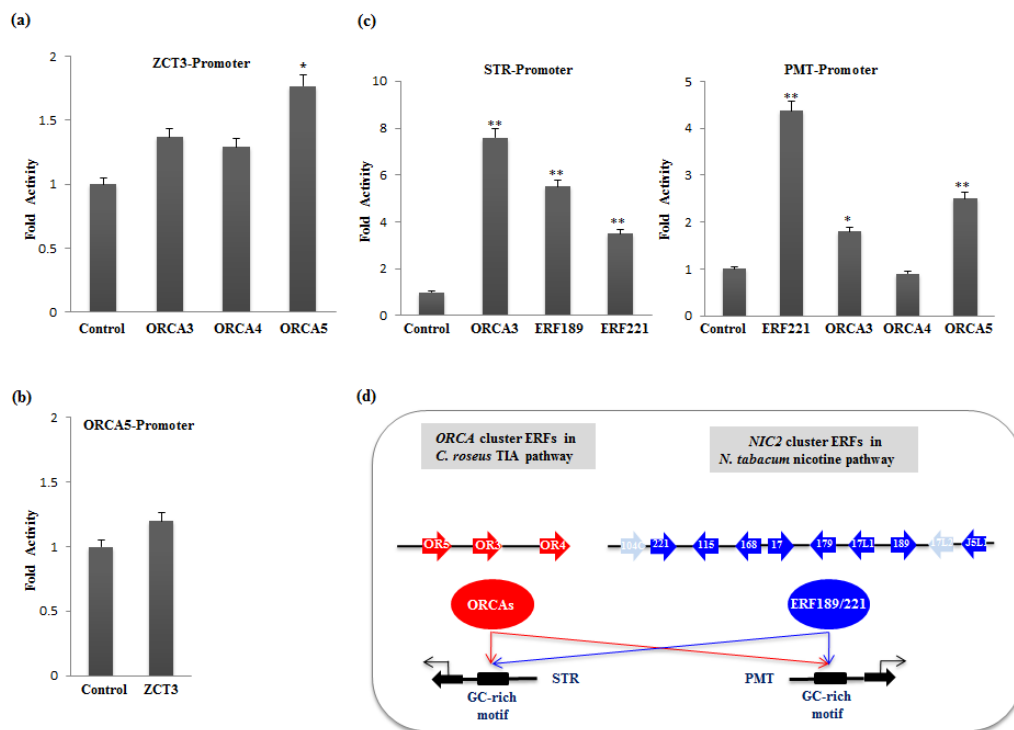


Figure 2.15 ORCA5 activates the *ZCT3* promoter yet *ZCT3* could not activate *ORCA5*. *C. roseus* ORCAs and tobacco *NIC2*-ERFs are functionally interchangeable

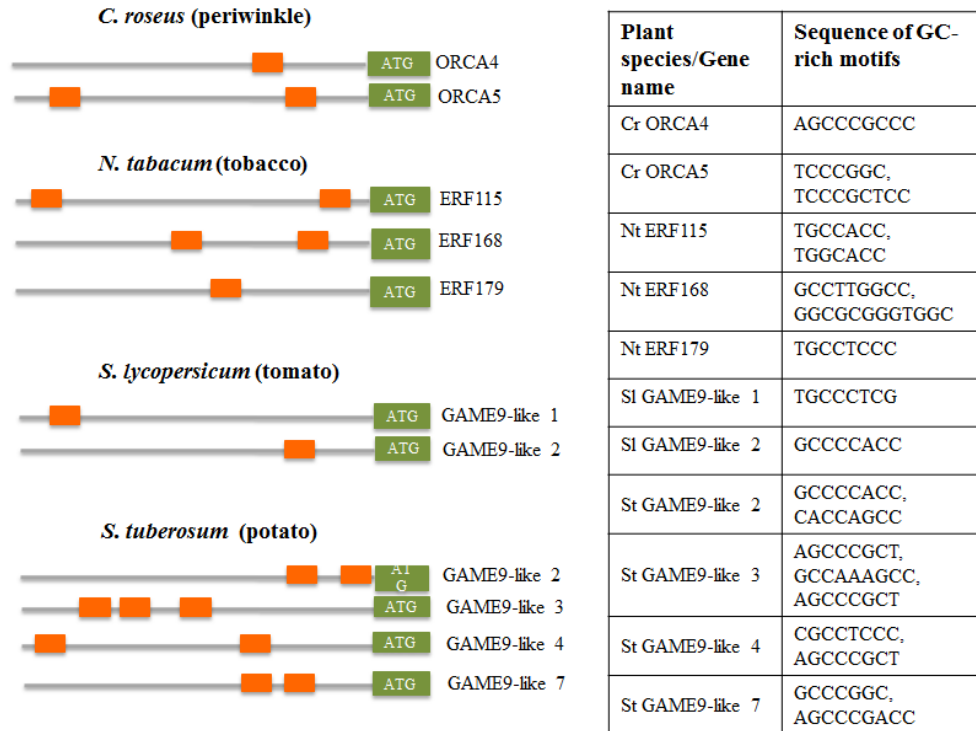
a. Transactivation of the *ZCT3* promoter by three ORCAs. *ZCT3* promoter fused to *luciferase* reporter was electroporated into tobacco protoplasts either alone or with an effector plasmid (*ORCA3*, *ORCA4* or *ORCA5*). **b.** Transactivation of the *ORCA5* promoter by *ZCT3*. *ORCA5* promoter fused to luciferase reporter was electroporated into tobacco protoplasts with an effector plasmid *ZCT3*. **c.** Transactivation of *STR/PMT*

promoter by ORCA3/4/5/ERF189/221. *STR/PMT* promoter fused to *luciferase* reporter was electroporated into tobacco protoplasts with an effector plasmid ORCA3/4/5/ERF189/221. A plasmid containing the *GUS* reporter, controlled by the *CaMV35S* promoter and *rbcS* terminator, was used as a control for normalization. Luciferase and GUS activities were measured 20 h after electroporation. Luciferase activity was normalized against GUS activity. Control represents the reporter alone without effectors. Data presented are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05, ** p value < 0.001. **d.** Schematic diagram showing cross-regulation between *C. roseus* ORCAs and tobacco NIC2-ERFs.

2.3.15 AP2/ERF gene cluster regulation in other plants

In addition to *C. roseus* and tobacco, the AP2/ERF gene clusters have also been identified in tomato and potato and shown to be involved in the biosynthesis of SGAs (Shoji et al., 2010; Cárdenas et al., 2016). Phylogenetic analysis revealed that ORCA3/4/5 (Paul et al., 2017), NIC2 locus ERFs in tobacco (Shoji et al., 2010) and GAME9 and GAME9-like proteins in tomato and potato (Cárdenas et al., 2016) are a member of the ERF IXa sub-family (Nakano et al., 2006). The IXa sub-family AP2/ERFs are known to bind the GC-rich elements in the promoters of target genes (Fujimoto et al., 2000; Shoji et al., 2013). Moreover, the ORCA and NIC2 locus ERFs are functionally interchangeable (Figure 2.15). Furthermore, in *C. roseus*, ORCA5 regulates its own expression. The question thus arose whether the members of AP2/ERF gene clusters from other plant species are involved in intra- and/or inter-cluster regulation. To address this question, we searched for GC-rich motifs in AP2/ERFs promoters in tobacco, tomato and potato. Out of the eight NIC2 locus ERFs, *ERF115* and *ERF168* contain two and *ERF179* contains one GC-rich sequence in the promoter. *GAME9-like 1* and *2* in tomato contain single GC-rich sequences. Potato *GAME9-like 2, 3, 4* and *7* containing several GC-rich motifs in their promoters (Table 2.2). These observations suggest that similar to the ORCA cluster in *C. roseus*, the AP2/ERF gene clusters across the plant species are likely involved in self- or cross-regulation.

Table 2.2: Position and sequences of GC-rich motifs in promoters of AP2/ERFs in *C. roseus*, tobacco, tomato and potato.



2.4 Conclusion

In this report, we illustrate the functions of ORCA4 and ORCA5, two previously uncharacterized AP2/ERFs, and demonstrate their overlapping and divergent functions with ORCA3 (Figure 2.16). ORCA4 and 5 have broad transactivation specificity. We demonstrated that like ORCA3, ORCA4 and ORCA5 show sequence-specific binding to JRE and trans-activation of *STR* promoter. Ectopic expression of *ORCA4* or 5 in *C. roseus* hairy roots led to dramatic increase in TIA accumulation (e.g. more than 40-fold and a 15-fold increase of tabersonine in *ORCA4* and *ORCA5* overexpression lines, respectively), a result that has not been achieved previously by expressing a single gene encoding either a TF or pathway enzyme. The *ORCA* gene cluster is differentially regulated and has overlapping yet distinct functions. It was showed that in addition to the genes from indole branch, ORCA5 also activated the genes of iridoid branch, including *LAMT* and *SLS*. It was demonstrated that ORCA5 directly regulates *ZCT3* to control the

TIA gene regulation. CrMYC2, a key regulator in JA-signaling, directly regulates *ORCA3* but indirectly affects *ORCA4* and 5, likely through an uncharacterized TF. These observations suggest a sub-functionalization of *ORCA4* and 5. The sequence divergence outside of the conserved AP2 DNA-binding domain of ORCAs may be partly responsible for the differential activation of the target promoters. We also experimentally verified that CrMYC2 dually controls TIA gene regulation as an activator of *ORCA3* and a co-activator of the *TDC* and *CPR* genes regulated by the ORCA gene cluster. However, CrMYC2 does not bind the G-box elements on the promoters of *G10H*, and *STR* genes, that are regulated by *ORCA4* and 5, but not *ORCA3*.

The *C. roseus* ORCAs and tobacco NIC2 locus ERFs are JA inducible and are positive regulators of species-specific specialized metabolites biosynthesis pathways, such as TIAs and nicotine, respectively. In this study, we showed that these homologous TFs are functionally interchangeable between the species; ORCAs significantly activated the *PMT* promoter and transcriptional activity of *STR* promoter was significantly induced by both ERF189 and ERF221.

We also showed that *ORCA5* directly regulates *ORCA4* by binding a GC-rich sequence of its promoter but indirectly affects *ORCA3*, likely through an uncharacterized TF. In addition, *ORCA5* regulates its own expression. To determine, whether such self-/cross-regulatory mechanism is conserved in other plant species, we analyzed the promoters of AP2/ERFs clusters in tobacco, tomato and potato for GC-rich sequences. Interestingly, several ERFs contain a number of the GC-rich motifs, suggesting possible self- or cross-regulation of AP2/ERFs across the plant kingdom. However, future work is needed to experimentally verify the presence of such regulatory network in plants. This study represents a detailed characterization of transcriptional regulatory mechanisms that govern a plant TF cluster. The knowledge of the ORCA gene cluster can be exploited for biotechnological and synthetic biology applications.

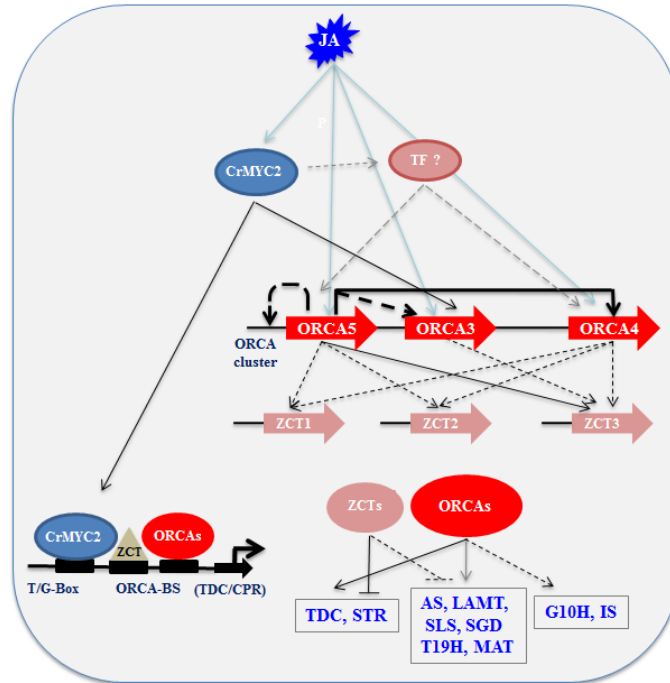


Figure 2.16 A simplified model depicting the transcriptional regulation of the ORCA cluster and TIA pathway genes. Upon perception of JA, CrMYC2 activates *ORCA3* and possibly another unidentified TF, which in turn activates *ORCA4* and *ORCA5*. Solid black arrows indicate direct activation; broken arrows represent indirect or undetermined activation. The ORCA3/4/5 activate three subsets of TIA pathway genes. ORCA3 and ORCA4/5 regulate several genes in the indole and downstream TIA pathways, whereas ORCA4 and 5 activates an additional set of genes in the seco-iridoid pathway (*IS* and *G10H*). ORCA5 directly regulates ORCA4 and indirectly ORCA3. Moreover, ORCA5 shows auto-regulation. ZCT3, the key negative regulator of TIA pathway, was directly regulated by ORCA5 and indirectly by ORCA3 and 4. CrMYC2, in addition to activating *ORCA3*, co-regulates *TDC* and *CPR* with ORCA3/4 by binding to the T/G-Box of the gene promoters.

Chapter 3

A MAP kinase cascade acts upstream of *ORCA* gene cluster to regulate TIA biosynthesis in *Catharanthus roseus*

3.1 Introduction

The protein kinase (PK) gene family is considered to be one of the largest and most highly conserved gene families in plants. PKs phosphorylate proteins leading to the functional changes and are likely involved in approximately all biological processes, including plant growth, development, and biotic and abiotic stress responses. Genome-wide survey of a number of plant species reveals that more than 3% of the genes are annotated as PKs, indicating their functional importance (Lehti-Shiu et al., 2009; Lehti-Shiu and Shiu, 2012). Interestingly, the plant kinome (complete set of PKs) is significantly larger than that of other eukaryotes. The intense expansion of plant kinomes over other eukaryotes could be the result of current duplication events and a high retention rate of duplicates (Hanada et al., 2008; Lehti-Shiu and Shiu, 2012). The plant PK superfamily consists of typical and atypical PKs. Typical PKs have been further classified into nine groups, AGC (Protein Kinase A; PKA–Protein Kinase G; PKG–Protein Kinase C; PKC), CAMK (calcium- and calmodulin-regulated kinases), CK1 (casein kinase 1), CMGC (include cyclin-dependent kinases), MAPK (Mitogen-activated protein kinases), STE (include many kinases functioning in MAP kinase cascades), GSK (glycogen synthase kinase), TK (tyrosine kinases), and TKL (tyrosine kinase-like kinases) based on their amino acid sequence similarities, domain organization, and modes of regulation (Lehti-Shiu and Shiu, 2012). PKs which do not come under those nine groups are classified as the ‘other’ group of PKs. Atypical PKs lack significant sequence similarity with typical PKs (Hanks and Hunter, 1995). The most common and extensively studied PKs include the MAPK cascade. A MAPK cascade consists of three sequentially activated kinases, including MAPK kinase kinase (MAPKKK), MAPK kinase (MAPKK), and MAPK, encoded by multiple genes (Smékalová et al., 2014). The MAPKs phosphorylate specific substrates, including transcription factors and enzymes leading to

specific cellular responses. MAPKs are activated due to the phosphorylation of both threonine (T) and tyrosine (Y) residues in their TXY motif, located in the activation loop between the catalytic subdomains VII and VIII by active MAPKKs, which in turn are activated by MAPKKKs by phosphorylation of the conserved S/TX₃₋₅S/T motif at their catalytic domain (Ichimura et al., 2002).

In plants, MAPK cascades are involved in a number of biological processes, such as cell division (Jiménez et al., 2007), growth and development (Xu and Zhang, 2015), hormonal responses (Tena et al., 2001), and environmental stress responses (Samuel et al., 2000; Pitzschke et al., 2009; Chávez Suárez and Ramírez Fernández, 2010; Heinrich et al., 2011; Parages et al., 2013).

Catharanthus roseus of the family Apocynaceae produces a group of natural products, known as terpenoid indole alkaloids (TIAs), including the anticancer drugs, vinblastine and vincristine (Kellner et al., 2015). Compared to the transcriptional regulation of TIA pathway, the posttranslational regulation is relatively less explored. A previous study has shown that the JA-induced accumulation of TIA pathway gene transcripts was significantly attenuated by the addition of protein kinase inhibitors, indicating the probable involvement of protein phosphorylation in JA signal transduction (Menke et al., 1999a). Moreover, it has been hypothesized that posttranslational modification of ORCA3 protein is essential to activate transcriptional activity of its targets, and phosphorylation by a protein kinase has been proposed to be involved in this process; however, it has never been experimentally verified (Menke et al., 1999a; van der Fits and Memelink, 2000). A JA-inducible MAPK, CrMAPK3, has been previously isolated from *C. roseus* (Raina et al., 2012). Transient overexpression of *CrMAPK3* in *C. roseus* leaves resulted in the upregulation of regulatory and structural genes, including *ORCA3*, *TDC*, *STR* and *D4H* in TIA pathway (Raina et al., 2012). However, nothing is known about the MAPKK and MAPKKK of the *C. roseus* JA-inducible protein kinase cascade.

In this study, we provide experimental evidence that ORCA gene cluster and CrMYC2 act downstream of a MAP kinase cascade that comprises of a previously uncharacterized

MAP kinase kinase, CrMAPKK1, and CrMAPK3/6. Overexpression of *CrMAPKK1* in *C. roseus* hairy roots up-regulates TIA pathways genes and boosts TIA accumulation.

3.2 Materials and Methods

3.2.1 Plant materials

Catharanthus roseus (L.) G. Don var. ‘Little Bright Eye’ was used for gene expression and cloning, and generation of transgenic hairy roots. *Nicotiana tabacum* var. Xanthi cell line was used for protoplast-based transient expression assays.

3.2.2 RNA isolation and cDNA synthesis

Ten-day-old *C. roseus* seedlings treated with 100 μ M methyl jasmonate (MeJA) for 2 h were used for total RNA isolation and cDNA synthesis as described previously (Suttipanta et al., 2007). *CrMAPKK1*, *CrMAPK3*, and *CrMAPK6* were PCR amplified from cDNA using gene-specific primers (Table 3.1), cloned into the pGEM-T Easy vector (Promega), and sequenced.

3.2.3 Multiple sequence alignment and phylogenetic analysis

Alignment of deduced amino acid sequences was performed using the ClustalW program (Larkin et al., 2007) with the default parameters. Phylogenetic trees were constructed and visualized using the neighbor-joining method through MEGA version 5.2 software (Tamura et al., 2011). The statistical reliability of individual nodes of the newly constructed tree was assessed by bootstrap analyses with 1,000 replications.

3.2.4 Identification, classification and co-expression analysis of *C. roseus* protein kinases

To identify *C. roseus* protein kinases (PKs), all protein-coding genes were downloaded from the latest version of the *C. roseus* genome from Dryad Digital Repository (Kellner et al., 2015). Hidden Markov Models (HMMs) of the ‘typical’ Pkinase clan [Pkinase (PF00069) and Pkinase_Tyr (PF07714)] (Finn et al., 2010) were used to search for putative PKs using HMMER v3.1b2 (Finn et al., 2011) with an E-value cut-off of <1.0. After this initial screen, 775 sequences containing the PK domain were identified. All

sequences were further aligned with the PFAM kinase domain models to confirm the presence of kinase domains (Pkinase and Pkinase_Tyr domains) and eliminate pseudogenes as described previously (Lehti-Shiu and Shiu, 2012). In this analysis, the putative PKs were considered typical PKs only if the domain alignments covered at least 50% of the PFAM domain models. Finally, a total of 691 proteins containing at least one kinase domain were identified and assigned as *C. roseus* ‘typical’ PKs. We then identified MAP kinases in the *C. roseus* kinome using the previously described signature motifs (Jonak et al., 2002).

To analyze the expression patterns of *C. roseus* PKs and TIA pathway genes, transcriptome data for five different tissues (flower, mature leaf, immature leaf, stem, and root) were obtained from the Sequence Read Archive (SRA) database at the National Center for Biotechnology Information (accession number SRA030483). Raw reads were processed and the reads per kilobase of transcript per million mapped reads (RPKM) value was calculated as described before (Singh et al., 2015). Pairwise Pearson correlation coefficient for each transcript was calculated using RPKM. Matrix distances for expression heatmap were computed over Pearson correlations of gene expression values (RPKM) by the heatmap.2 function of gplots (version 3.0.1) Bioconductor package in R (version 3.2.2).

3.2.5 Quantitative RT-PCR

Quantitative real-time PCR (qRT-PCR) was performed as described previously (Suttipanta et al., 2011). The primers used in qRT-PCR are listed in Table 3.1. In addition to the *C. roseus* 40S Ribosomal Protein S9 (*RPS9*) gene, *Elongation Factor 1 α* (*EF1 α*) was used as a second internal control (Liscombe et al., 2010). All PCRs were performed in triplicate and repeated at least twice.

Table 3.1 Oligonucleotides used in this study

<i>RPS9</i>	5'-GAGGGCCAAAACAACTTA-3'	5'-CCCTTATGTGCCTTTGCCTA-3'
<i>EF1α</i>	5'-TACTGTCCCTGTTGGTTCGTG-3'	5'-AAGAGCTTCGTGGTGCATCT-3'
<i>TDC</i>	5'-ATCCGATCAAACCCATACCA-3'	5'-CGTCATCCTCGACCATTTTT-3'
<i>G10H</i>	5'-TTATTCGGATTCTGCCAAGG-3'	5'-TCCCCAAAGTGAATCGTCAT-3'
<i>CPR</i>	5'-TGGCAGAAAAGGCTTCTGAT-3'	5'-CTCAGCCTGTGTGCTATCCA-3'
<i>SLS</i>	5'-GTTCCCTTCTCACCGGAGTTG-3'	5'-CCCATTTGGTCAACATGTCA-3'
<i>STR</i>	5'-ACCATTGTGTGGGAGGACAT-3'	5'-ATTTGAATGGCACTCCTTGC-3'
<i>IS</i>	5'-CCACATGATTTCGCTTTTACCG-3'	5'-AAACCCGAAAACCAGAGCTG-3'
<i>CrMAPKK1</i>	5'-AAGGGCAAAGACCTGATTGG-3'	5'-TCAGGCAAACCTGGTGGTTC-3'
<i>ZCT1</i>	5'-AGCCGAAAACCTCATGCTTGT-3'	5'-CGCCTTTGCAACAGGTTTAT-3'
<i>ZCT2</i>	5'-CGTCAATTTCCATCGTTTCA-3'	5'-CCGATAGCGAATTCAAGTCC-3'
<i>ZCT3</i>	5'-GACAAGCTTTGGGAGGACAC-3'	5'-GGCAAGGCAGGTAAGTTCAA-3'
<i>CrMAPKK1</i>	5'-gcggatccATGGCATTAGTTAAAGAC CGTC-3'	5'-atgctgcagTTATGGCCTTTTAACCT TTTGC-3'
<i>CrMAPK3</i>	5'-gccaattgATGGTTGATGCAAATAT GGCC-3'	5'-atgctgcagTTATGCATATTCTGGATT TAGAG-3'
<i>CrMAPK6</i>	5'-gcgtcgactcATGGACGGTGGTGCAG CTCAG-3'	5'-atgctgcagTCATATTTGCTGATATT CGGGATT-3'
<i>rolB</i>	5'CTTATGACAAACTCATAGATAAAA GGTT-3'	5'-TCGTA ACTATCCA ACTCACATC AC-3'
<i>rolC</i>	5'-CAACCTGTTTCTACTTTGTAAAC- 3'	5'-AAACAAGTGACACACTCAGCT TC-3'
<i>virC</i>	5'-TTTTGCTCCTTCAAGGGAGGTG CC-3'	5'-GGCTTCGCCAACCAATTTGGAG AT-3'
<i>GUS</i>	5'-ATGGTAGATCTGAGGGTAAATTT C-3'	5'-GCTAGCTTGTTGCCTCCCTG-3'
<i>ntpII</i>	5'-ATGGGGATTGAACAAGATGGA-3'	5'-TCAGAAGA ACTCGTCAAGAAG -3'

3.2.6 Sub-cellular localization

For sub-cellular localization, two plasmids were constructed, one containing the enhanced GFP (eGFP) and the other containing CrMAPKK1-eGFP fusion. The expression of both genes was under the control of the *CaMV35S* promoter and the *rbcS* terminator.

3.2.7 Protoplast isolation and electroporation

For transient protoplast assays, the reporter plasmids were generated by cloning *STR* and *TDC* promoters in a modified pUC vector containing a firefly *luciferase* (*LUC*) and *rbcS* terminator. The (*5XRE*)-*35S_{pro}:LUC* was described previously (Patra et al., 2013a). The effector plasmids were constructed by cloning *ORCA3/4/5*, *CrMYC2*, *CrMAPK3*, and

CrMAPKK1 into a modified pBlueScript (pBS) vector under the control of the *CaMV* 35S promoter and *rbcS* terminator. For the protoplast-based two-hybrid assay, *CrMAPK3* and *CrMAPK6* were fused to the GAL4 DNA binding domain in a pBS plasmid containing *MMV* promoter and *rbcS* terminator. β -glucuronidase (*GUS*) reporter, driven by the *CaMV*35S promoter and *rbcS* terminator, was used as an internal control in the protoplast assay. Protoplast isolation from tobacco cell suspension cultures and electroporation with plasmid DNA were performed using previously described protocols (Pattanaik et al., 2010b). The reporter, effector, and internal control plasmids were electroporated into tobacco protoplasts in different combinations; luciferase and GUS activities in transfected protoplasts were measured as described previously (Suttipanta et al., 2007). For sub-cellular localization, the plasmids containing either *eGFP* or *CrMAPKK1-eGFP* were independently electroporated into tobacco protoplasts and visualized after 20 h of incubation at room temperature using a fluorescence microscope (Eclipse TE200, Nikon, Japan).

3.2.8 Construction of plant expression vectors and generation of hairy roots

For plant transformation *CrMAPKK1*, cDNA was PCR-amplified and cloned into a modified pCAMBIA2301 vector containing the *CaMV*35S promoter and *rbcS* terminator (Pattanaik et al., 2010b). The pCAMBIA2301 vector was used as an empty vector (EV) control. The plasmids were mobilized into *Agrobacterium rhizogenes* R1000. Transformation of *C. roseus* seedlings and generation of hairy roots were performed using the protocol described previously (Suttipanta et al., 2011). Transgenic status of the hairy root lines was verified by PCR amplification of *rolB*, *rolC*, *virC*, *nptII* and *GUS* genes. Three independent hairy root lines were selected for further analysis.

3.2.9 Recombinant protein production and *in vitro* phosphorylation assay

To produce recombinant CrMAPKK1 protein, the corresponding gene was PCR amplified and cloned into the pGEX4T1 vector (GE Healthcare Biosciences, USA). The resulting pGEX4T1-CrMAPKK1 plasmid was used as templates for site-directed mutagenesis to generate pGEX4T1- CrMAPKK1^{K85M} plasmid. Mutagenesis was carried out as previously described (Zheng et al., 2004), and mutations were verified by

sequencing. Wild-type and mutant plasmids were transformed into BL21 cells containing pRIL (Agilent, USA). Protein expression was induced by adding isopropyl β -D-thiogalactopyranoside to a final concentration of 1 mM to the cell cultures at $A_{600} \sim 0.6$ and incubated for 3 hours at 30°C. The cells were harvested by centrifugation and lysed using CelLytic B (Sigma). The GST fusion proteins were bound to Glutathione Sepharose 4B columns (Amersham) and then eluted by 10 mM reduced glutathione in 50 mM Tris-HCl (pH7.5) buffer. The phosphorylation assay was performed using the protocols of Kinase-Glo® luminescent kinase assay platform (Promega).

3.2.10 Alkaloid extraction and analysis

To extract alkaloid, transgenic hairy roots were frozen in liquid nitrogen and ground to powder. Samples were extracted in methanol (1:100 w/v) twice for 24 h on a shaker. Pooled extracts were then dried *via* a rotary evaporator and diluted in methanol 10 μ L/mg of the initial sample. The samples were then analyzed using high-performance liquid chromatography (HPLC), followed by electrospray-injection in a tandem mass spectrometry, as previously described (Suttipanta et al., 2011). The known alkaloid standards were run to identify elution times and mass fragments.

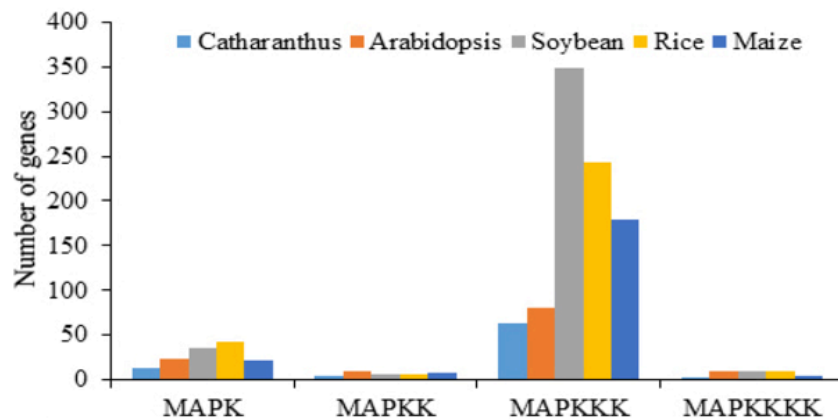
3.3 Results and Discussion

3.3.1 Analysis of *C. roseus* kinome identifies MAP kinases potentially involved in TIA pathway regulation

JA-responsive expression of the structural and TF genes is a hallmark of the TIA pathway. JA-induced expression of TIA structural genes, such as *TDC* and *STR*, is not inhibited by the protein synthesis inhibitor, cycloheximide, but by a PK inhibitor (Menke et al., 1999a; Van Der Fits and Memelink, 2001). These findings suggest that ORCA proteins are stable *in vivo* and phosphorylation is involved in post-translational modification of the ORCA proteins (Menke et al., 1999a). However, there is no comprehensive study on *C. roseus* PKs and their roles in the regulation of TIA pathway. Prior to this study, MAPK3 was the only described PK in *C. roseus* (Raina et al., 2012). We thus analyzed the *C. roseus* genome for genes encoding PKs and identified 83 putative MAPKs (Figure 3.1a), which represent 0.25% of all protein-coding genes in the genome and 12% of the *C.*

roseus kinome. We performed co-expression analysis for all putative MAPKs in different *C. roseus* tissues (Figure 3.1b), and identified a MAPKK, which together with MAPK3 and 6 presented in same clade, designated here as *CrMAPKK1* (GenBank accession no. KR703579). *CrMAPKK1* encodes a 327-amino acid protein, with a calculated molecular mass of 36.7 kDa. Sequence alignment revealed that CrMAPKK1 shares 75% and 61% amino acid sequence identity with tobacco NtMAPKK1/JAM1 and Arabidopsis AtMAPKK9, respectively (Figure 3.2a). Phylogenetic analysis of MAPKK proteins from different plant species placed CrMAPKK1 close to NtMAPKK1 (Figure 3.2b). NtMAPKK1 has been shown to modulate expression of nicotine biosynthesis genes in tobacco (De Boer et al., 2011), whereas the AtMAPKK9 cascade in Arabidopsis plays a key role camalexin biosynthesis (Xu et al., 2008). We next performed co-expression analysis of *CrMAPKK1* with TIA biosynthetic and regulatory genes. *CrMAPKK1* was clustered with *TDC*, *STR*, *LAMT*, *SLS*, *CrMYC2*, *ORCAs*, and *CrMAPK3* (Figure 3.3), suggesting that CrMAPKK1 may be involved in regulation of the TIA biosynthetic pathway. We also identified two MAPKKKs (designated as *CrMAPKKK1* and 2), which were present in the same clade with CrMAPKK1. Only *CrMAPKKK1* is JA-responsive, and we thus believe that the MAPK cascade comprises CrMAPKKK1, CrMAPKK1, and CrMAPK3/6.

(a)



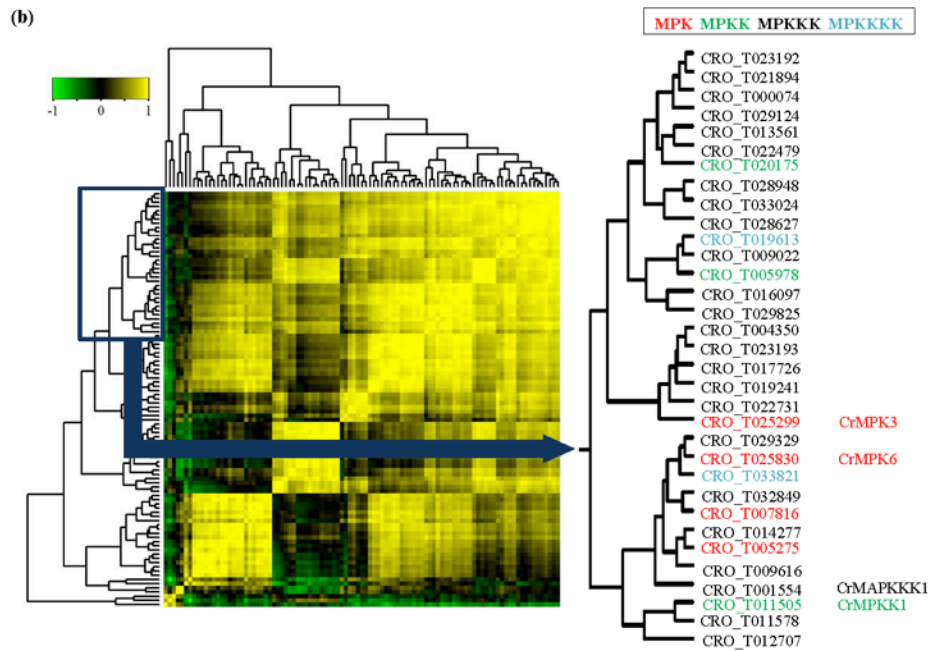


Figure 3.1 Gene number distribution of MAP kinases (MAPK) in various plant species (a) and spatial expression analysis of the *C. roseus* MAPK cascade components (MAPKs, MAPKKs, MAPKKKs and MAPKKKKs) using transcriptome of different *C. roseus* tissues, including roots, stems, flowers, immature and mature leaves (b).

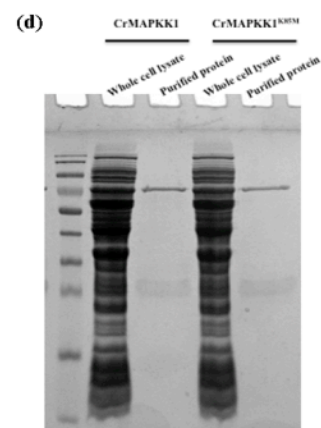
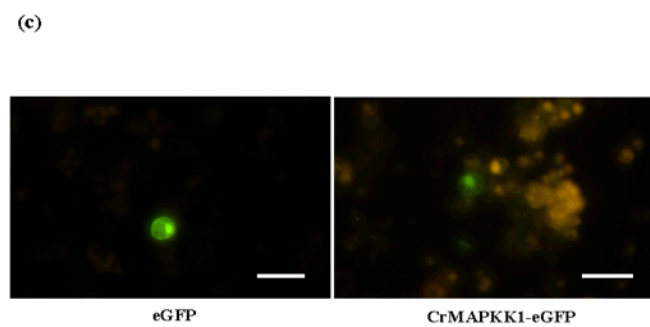
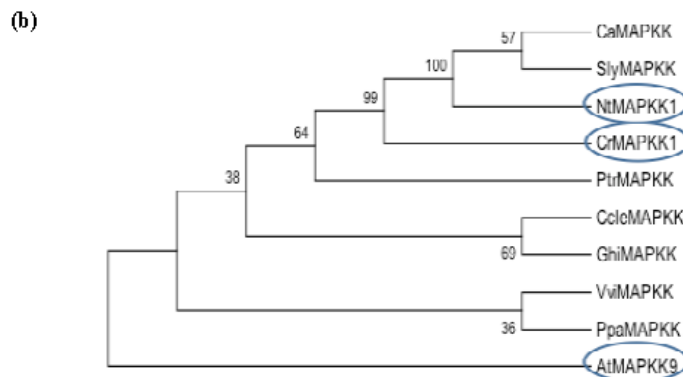
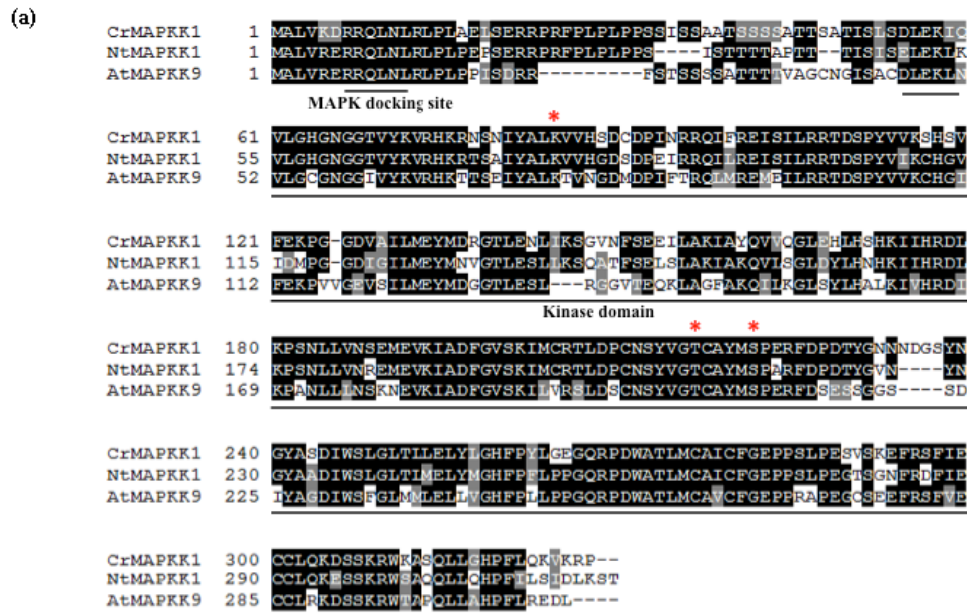


Figure 3.2 Multiple sequence alignment of CrMAPKK1 with NtMAPKK1 and AtMAPKK9. Identical amino acid residues in the alignment are shaded in black and similar amino acid residues in gray. The kinase domain and KIM motif (MAPK-docking site) are underlined. The asterisks represent conserved lysine (K) residue in ATP-binding

site and serine (S)/ threonine (T) residues in kinase domain (a). Phylogenetic analysis of different plant MAPKKs (b). Nuclear-localization of CrMAPKK1 in tobacco protoplast (c), and bacterial expression-purification of CrMAPKK1 (d).

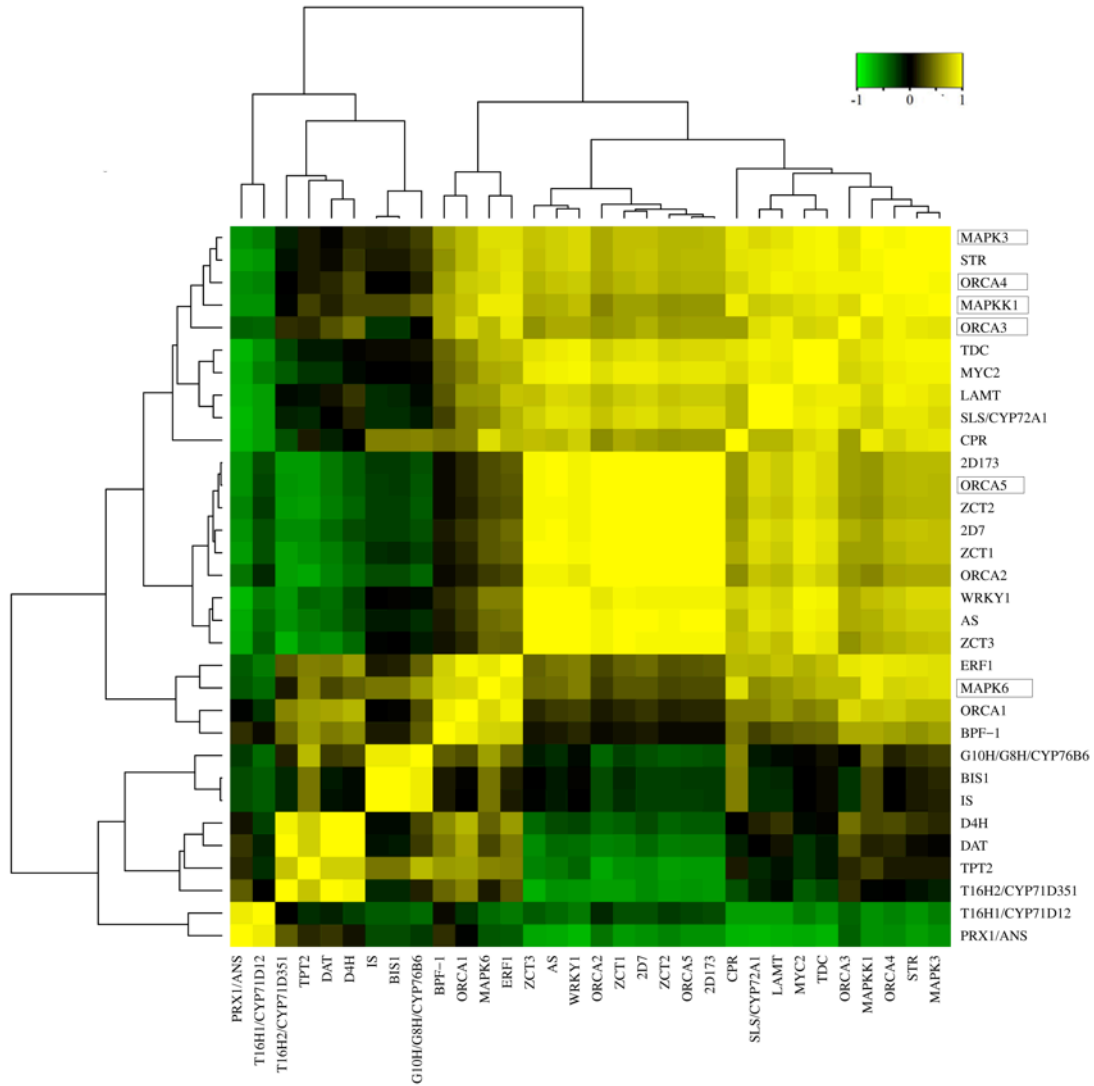


Figure 3.3 Co-expression of *CrMAPKK1* with TIA regulatory and structural genes.

Hierarchical clustering and heat-map show that *CrMAPKK1* was co-expressed with *CrMYC2* and *ORCA3/4*, and their targets, *STR*, *TDC*, *LAMT*, *SLS*, *CPR* and *CrMAPK3*.

3.3.2 *CrMAPKK1* is MeJA responsive, nucleus-localized, and autophosphorylated

To examine the effect of MeJA on *CrMAPKK1* expression, 10 day-old *C. roseus* seedlings were treated with 100 μ M MeJA for 2h, and gene expression in whole

seedlings, leaves, and roots were measured by qRT-PCR. Expression of *CrMAPKK1* increased by approximately 3-fold in MeJA-treated seedlings compared with the untreated control (Figure 3.4a). In addition, similar to what was observed for *ORCA3*, *ORCA4*, and 5, the MeJA-mediated induction of *CrMAPKK1* transcripts was higher in roots compared to leaves. To visualize sub-cellular localization, *CrMAPKK1* was fused in-frame to the enhanced *GFP* (*eGFP*) and expressed in tobacco protoplasts. While eGFP was detected throughout the cell, the CrMAPKK1-eGFP fusion protein was localized to the nucleus (Figure 3.2c). In a MAP kinase cascade, a kinase activates substrates via phosphorylation of Ser/Thr (S/T) residues in a conserved (S/T)XXX(S/T) motif located between VII and VIII kinase sub-domains (Widmann et al., 1999). Some kinases are capable of self-activating (autophosphorylation). We generated an inactive CrMAPKK1 by site-directed mutagenesis of the conserved Lys85 to a Met (M85) at the ATP-binding site (Iyer et al., 2005). *CrMAPKK1* and the mutant gene (*CrMAPKK1^{K85M}*) were expressed in *E. coli* and the recombinant proteins were purified to homogeneity (Figure 3.2d). Recombinant proteins (1µg) were used in a kinase autophosphorylation assay using the ADP-Glo™ Kinase Assay kit (Promega, USA). Autophosphorylation of CrMAPKK1 was strong; by comparison, that of CrMAPKK1^{K85M} was extremely weak (Figure 3.4b), suggesting that CrMAPKK1 is an enzymatically active MAPKK and belongs to the class of autoactive MAPKK (Kiegerl et al., 2000).

3.3.3 CrMAPKK1 interacts with CrMAPK3 and CrMAPK6 in plant cells

The activation of MAPKs by upstream MAPKKs, in response to specific stimuli, is both complex and redundant. Moreover, a single MAPKK can activate a number of MAPKs (Lee et al., 2008; Andreasson and Ellis, 2010). MAPKKs have been shown to interact with MAPKs in yeast two-hybrid (Y2H) assay (Lee et al., 2008; Wankhede et al., 2013). CrMAPKK1 is homologous to AtMAPKK9, which acts upstream of AtMAPK3/AtMAPK6 to regulate camalexin biosynthesis and salt stress sensitivity (Xu et al., 2008). We identified and cloned the potential AtMAPK6 ortholog, designated as CrMAPK6 (GenBank accession no. KR703580), that encodes a 360-amino acid protein with a calculated molecular mass of 45.6 kDa. The amino acid sequence of CrMAPK6 exhibits a molecular structure typical of plant MAPKs with a conserved kinase domain

and C-terminal MAPKK docking site (*LHDISDEPTC*). CrMAPK6 shares 86.7% amino acid sequence identity with AtMAPK6 (Figure 3.5). We expected CrMAPK3 and CrMAPK6 to interact with CrMAPKK1 as substrates. To establish a kinase interaction, we first used Y2H assay but found extremely high self-activation that hindered the experiment with the three kinases in yeast cells. We then applied a plant cell-based two-hybrid system (Patra et al., 2013a). Plasmids expressing *CrMAPK3* or *CrMAPK6*, fused to the GAL4 DNA binding domain, were co-electroporated into tobacco protoplasts with a reporter vector expressing the firefly luciferase gene under the control of a minimal *CaMV35S* promoter, and 5XGAL response elements. *CrMAPK3* or *CrMAPK6* expression in protoplasts also induced luciferase activity (approximately 11- and 20-fold, respectively, compared to the basal level activity of the reporter plasmid); however, the luciferase activity was significantly increased in protoplasts co-expressing *CrMAPKK1* and *CrMAPK3* or *CrMAPK6* (approximately 40- and 62-fold, respectively) (Figure 3.4c). Therefore, CrMAPKK1 interacts with and likely phosphorylates, CrMAPK3 and CrMAPK6 in plant cells.

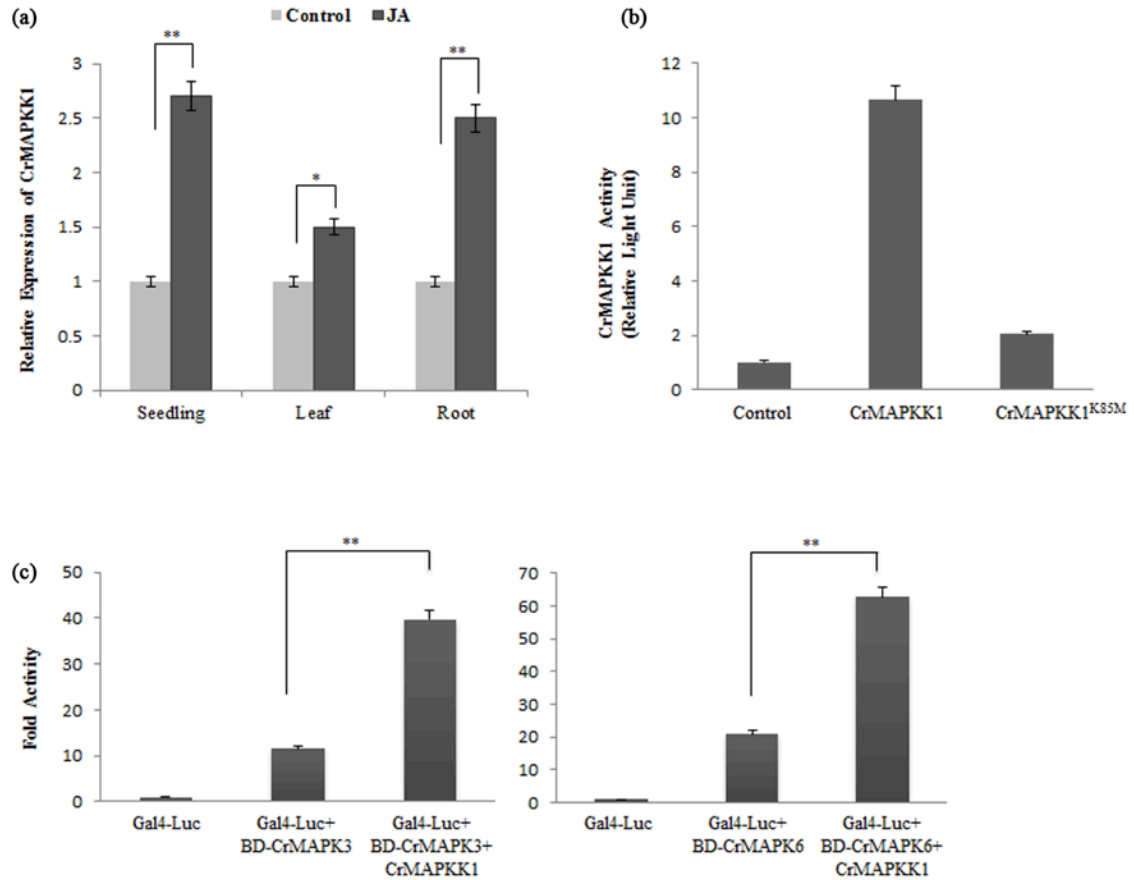


Figure 3.4 MeJA induction and autophosphorylation of CrMAPKK1 and its interaction with CrMAPK3 and 6. **a.** Ten day-old *C. roseus* seedlings were treated with 100 μ M MeJA for 2 h, and relative gene expression in different tissues were measured by qRT-PCR. Mock-treated seedlings were used as a control. Data represent means \pm SDs of two biological replicates with three technical replicates. **b.** Recombinant wild-type (CrMAPKK1) and mutant (CrMAPKK1^{K85M}) proteins were purified from *E. coli* cells. Autophosphorylation assays were performed using the non-radioactive ADP-Glo™ Kinase Assay System (Promega). Lysate of empty-vector transformed *E. coli* cell was used as a control. The phosphorylation activity was expressed in light units relative to the control. Data presented are means \pm SDs of three replicates. **c.** *CrMAPK3* and *CrMAPK6*, fused to the GAL4 DNA binding domain (amino acid 1-147), were cloned into a modified pBS vector containing the *MMV* promoter and *rbcS* terminator to generate BD-CrMAPK3 and BD-CrMAPK6, respectively. *CrMAPKK1* was cloned into a

modified pBS vector containing the *CaMV35S* promoter and *rbcS* terminator. Effector plasmids were co-electroporated into tobacco protoplasts with a reporter vector containing the firefly luciferase gene under the control of a minimal *CaMV35S* promoter and 5XGAL response elements. Control represents the reporter alone. A plasmid containing the *GUS* reporter, driven by *CaMV35S* promoter and *rbcS* terminator, was used as a control for normalization. Luciferase and GUS activities were measured 20 hours after electroporation. Luciferase activity was normalized against GUS activity. Data presented here are means \pm SDs of three biological replicates with four technical replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05, ** p value <0.01.

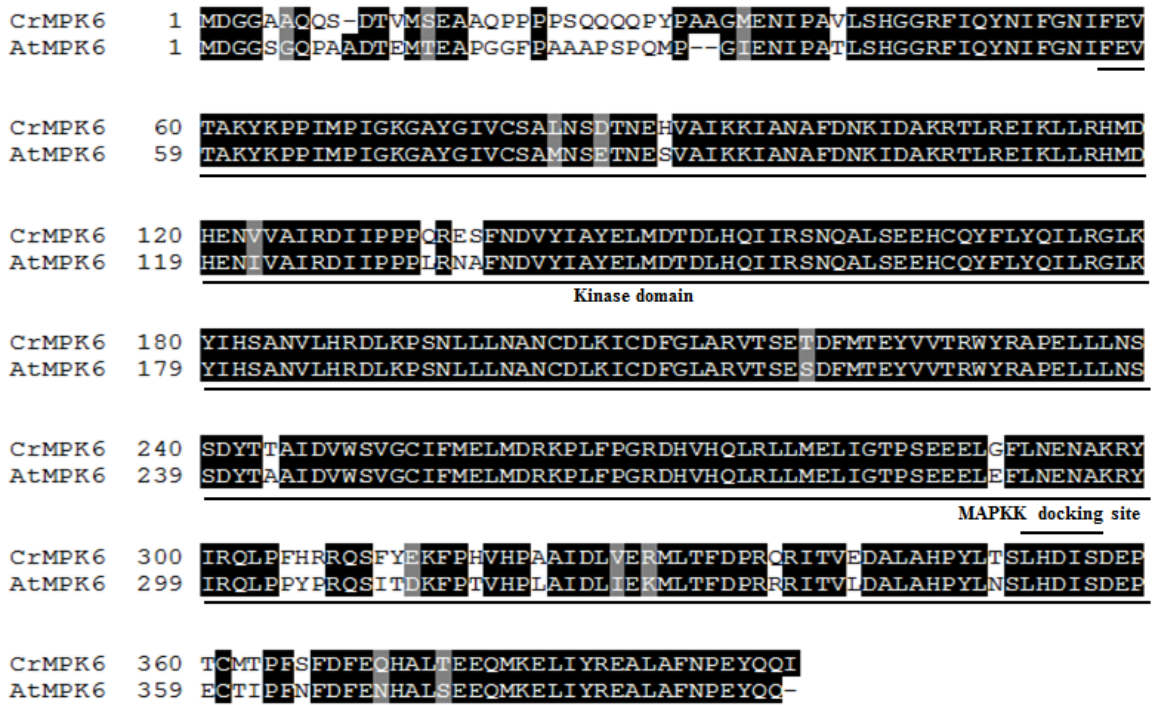


Figure 3.5 Sequence alignment of CrMAPK6 and AtMAPK6. Identical amino acid residues in the alignment are shaded in black and similar amino acid residues in gray. The kinase domain and MAPKK- docking site (CD) are underlined.

3.3.4 The N-terminal kinase interaction motif is crucial for CrMAPK1 function

The specificity and responsiveness of MAPK activation are associated with a complex that is formed by kinase-kinase interactions, or by docking to a scaffold protein (Kolch, 2000; Tanoue et al., 2000). The common docking (CD) domain, initially identified in MAPKs from mammals and yeast, is also conserved in plant MAPKs (Zhang and Liu, 2001). The CD domain is required for the interaction of MAPKs with upstream kinases, MAPK phosphatases, and substrates (Kolch, 2000; Tanoue et al., 2000). Moreover, the kinase interaction motif (KIM) of MAPKK is involved in interaction with the MAPK CD domain. Basic amino acids in the center of the KIM domain electrostatically interact with the acidic residues in the MAPK CD domain (Kolch, 2000; Tanoue et al., 2000). Another characteristic of the plant MAPKKs is the presence of an N-terminal Leu/Pro-rich region adjacent to the KIM domain (Jin et al., 2003). We identified a KIM domain (RRQLNL) at the CrMAPKK1 N-terminus (Figure 3.6a). To establish the importance of both the KIM domain and the Leu/Pro-rich region in the interaction of CrMAPKK1 with CrMAPK3/6, we generated two CrMAPKK1 mutants. In one, the di-arginine (RR) in the KIM domain was substituted with alanine (*CrMAPKK1^{AA}*); in the other, the N-terminal 49 amino acids, containing both the KIM motif and the Leu/Pro-rich region, were deleted (*CrMAPKK1^N*; Figure 3.6a). The ability of these two mutants to activate CrMAPK3/6 was tested in plant protoplasts. The interaction of CrMAPK3 or CrMAPK6 with CrMAPKK1^{AA} was reduced by approximately 50% compared to that with the wild-type CrMAPKK1. The residual activity is likely due to the presence of the Leu/Pro-rich region because CrMAPKK1^N was unable to interact with CrMAPK3 or CrMAPK6 (Figure 3.6b). These results suggest that the conserved KIM domain is required for CrMAPKK1 function.

3.3.5 CrMAPK3 and CrMAPKK1 significantly enhances the transactivation of TIA pathway gene promoters by ORCAs and CrMYC2

A previous study has shown that transient overexpression of *CrMAPK3* in *C. roseus* leaves elevates the expression of ORCA3-regulated genes such as *STR* (Raina et al., 2012). We used the tobacco protoplast assay to determine whether CrMAPK3 can enhance the transactivation activity of ORCA3 or ORCA4 on TIA pathway gene promoters. The results showed that, while ORCA3 or ORCA4 activated the *STR*

promoter, the activation was significantly higher when *CrMAPK3* was co-expressed with *ORCA3* or *ORCA4* (Figure 3.6c). Therefore, the components of the ORCA gene cluster are potential substrates for the CrMAPKK1-CrMAPK3 cascade. We next used *ORCA3*, the best-characterized member of the ORCA gene cluster, to determine whether phosphorylation events are responsible for the enhanced activities of the clustered ORCAs on *STR*. First, we scanned *ORCA3* for the minimal MAPK phosphorylation motifs (Ser or Thr followed by Pro; SP or TP) (Sharrocks et al., 2000). Using the NetPhos 2.0 Server, we found that *ORCA3* contains two potential MAPK phosphorylation sites, at Ser60 and Thr132. We substituted both Ser60 and Thr132 with Ala by site-directed mutagenesis and tested the *ORCA3* mutant (*ORCA3*^{S60A/T132A}) for activation of the *STR* promoter in the presence or absence of CrMAPK3. In the protoplast assay, activation of the *STR* promoter by CrMAPK3-*ORCA3* was significantly higher compared to CrMAPK3-*ORCA3*^{S60A/T132A} (Figure 3.6d), suggesting that CrMAPK3 is capable of phosphorylating *ORCA3* and modulates its transactivation activity. Because similar phosphorylation sites are also found in *ORCA4* and *5*, it is reasonable to suggest that they are also substrates of CrMAPK3.

We subsequently examined whether CrMAPKK1, which likely phosphorylates CrMAPK3/6, can enhance the activities of the ORCA gene cluster and CrMYC2. The *TDC* or *STR* promoter-reporter vector was electroporated into protoplasts with various combinations of vectors expressing *ORCA3*, *ORCA4*, *ORCA5*, *CrMYC2*, and *CrMAPKK1*. Individually, all four TFs activated the expression of the *TDC* or *STR* promoters; however, co-expression with *CrMAPKK1* significantly increased activation (Figure 3.7a). To determine whether the enhanced transactivation resulted from CrMAPKK1 activity, we substituted CrMAPKK1 with the inactivated mutant, CrMAPKK1^{K85M}, in the *STR* activation assay. CrMAPKK1^{K85M} was unable to enhance the promoter activation by *ORCA3* (Figure 3.7b). These results suggest that the CrMAPKK1-MAPK3/6 cascade can potentially enhance activation of the ORCA gene cluster and CrMYC2 in TIA biosynthesis.

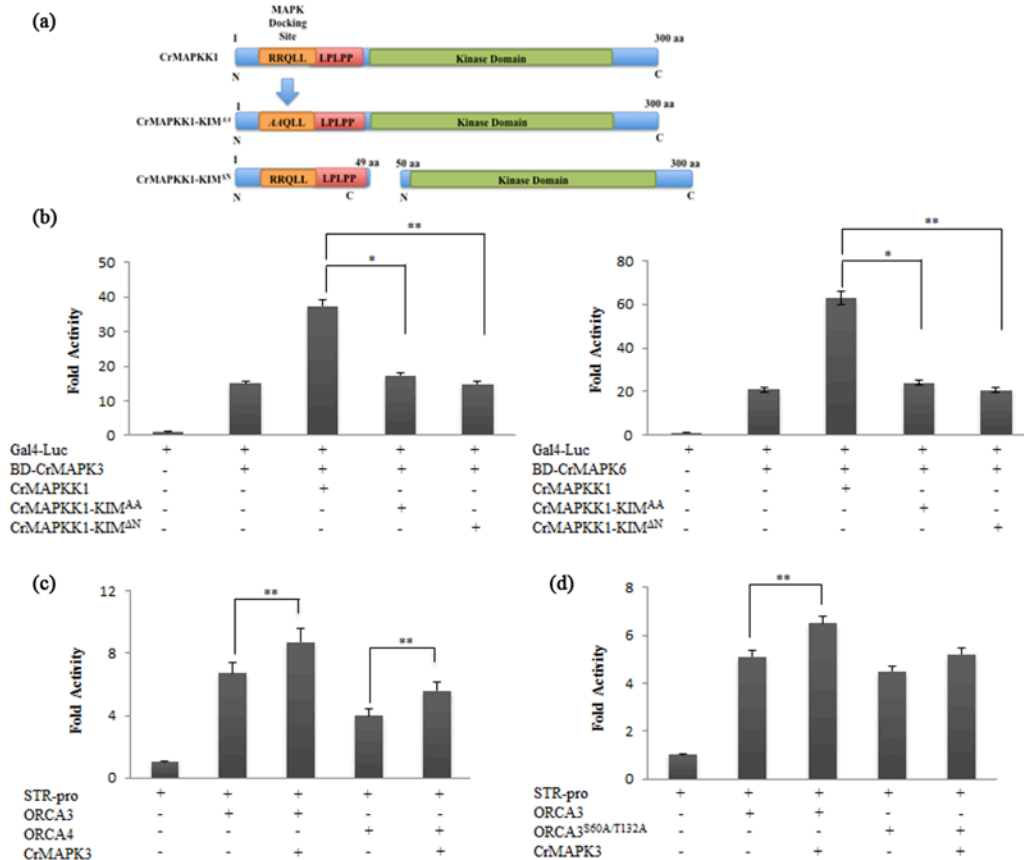


Figure 3.6 Importance of the KIM motif for CrMAPKK1 function and phosphorylation of ORCA3 by CrMAPK3. **a.** Schematic diagrams showing different *CrMAPKK1* plasmids. In *CrMAPKK1*^{AA}, the di-arginine (RR) in the kinase interaction motif (KIM; RRQLNLL) was substituted with two alanine (AA). In *CrMAPKK1*^{DN}, the N-terminal 49 amino acids containing KIM and the Leu/Pro-rich region were deleted. **b.** *CrMAPK3* and *CrMPK6*, fused to the GAL4 DNA binding domain, were cloned into a modified pBS plasmid containing the *MMV* promoter and *rbcS* terminator. *CrMAPKK1* expression was driven by the *CaMV35S* promoter and *rbcS* terminator. Effector plasmids were co-electroporated into tobacco protoplasts with a reporter vector containing the firefly luciferase gene under the control of the minimal *CaMV35S* promoter and 5X GAL response elements. The control represents the reporter alone. **c.** The *STR* promoter-reporter vector (*STR-pro*) was co-electroporated with ORCA3 or ORCA4 with or without CrMAPK3. **d.** The *STR* promoter-reporter vector was co-electroporated with ORCA3 or

the inactive mutant, ORCA3^{S60A/T132A}, with or without CrMAPK3. Luciferase and GUS activities were measured 20 hours after electroporation. Luciferase activity was normalized against GUS activity. Data presented here are means ± SDs of three biological replicates with three technical replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05, ** p value <0.01.

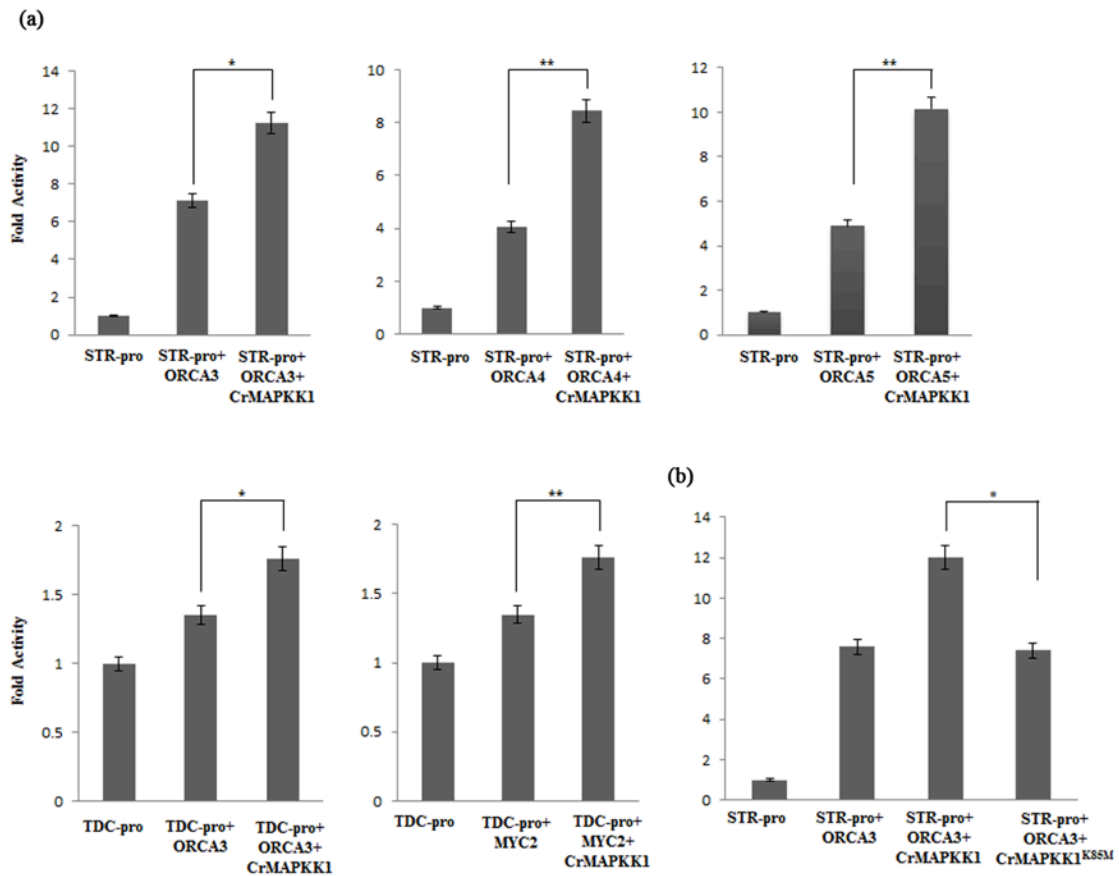


Figure 3.7 Enhancement of the transactivation activity of ORCAs and CrMYC2 on TIA pathway gene promoters by CrMAPKK1. a. Transactivation of the *STR* and *TDC* promoters (*STR*-pro and *TDC*-pro, respectively) by ORCA3, ORCA4, ORCA5, and CrMYC2 in the presence or absence of CrMAPKK1. **b.** Influence of inactive mutant, CrMAPKK1^{K85M}, on transactivation activity of ORCA3 on the *STR* promoter. The *STR* or *TDC* promoter, fused to the luciferase reporter, was electroporated alone or with different effectors into tobacco protoplasts. The control represents the promoters alone. A plasmid

containing *GUS* reporter, driven by the *CaMV35S* promoter and *rbcS* terminator, was used as a control for normalization. Luciferase and GUS activities were measured 20 hours after electroporation. Luciferase activity was normalized against GUS activity. Data presented here are means \pm SDs of three biological replicates with three technical replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05, ** p value <0.01.

3.3.6 Ectopic expression of *CrMAPKK1* activates key TIA pathway genes and boosts alkaloid accumulation in *C. roseus* hairy roots

The ability of the CrMAPKK1-MAPK3/6 cascade to enhance transactivation by the ORCA gene cluster and CrMYC2 prompted us to investigate its effects on gene expression and TIA biosynthesis in *C. roseus*. We generated transgenic *C. roseus* hairy root lines overexpressing *CrMAPKK1*. The transgenic status of the hairy roots was verified by PCR (Figure 3.8a), and three transgenic lines (OE-a, -b, and -c) were selected for further analysis. *CrMAPKK1* expressions in transgenic lines were 1.5 to 3-fold higher compared to the EV control (Fig. 3.8b). Next, we measured the expression of TIA regulatory and structural genes. *STR* expression was 16-18 fold higher in the CrMAPKK1-overexpression lines compared to the EV control; expression of other ORCA3/4-target genes, including *TDC*, *CPR*, *IS*, and *ZCTs*, were also increased significantly, although to lesser extents (Fig. 3.9a). To determine whether the changes in gene expression were also accompanied by increased TIA accumulation, alkaloids were extracted from independent transgenic hairy root lines and measured using LC-ESI-MS/MS as described previously (Suttipanta et al., 2011). The concentrations of tabersonine, ajmalicine, and catharanthine were significantly increased in transgenic lines compared to the EV control (Fig. 3.9b). Taken together, our findings support our hypothesis that the CrMAPKK1-MAPK3/6 cascade acts upstream of the ORCA gene cluster and CrMYC2 to modulate the expression of TIA pathway genes.

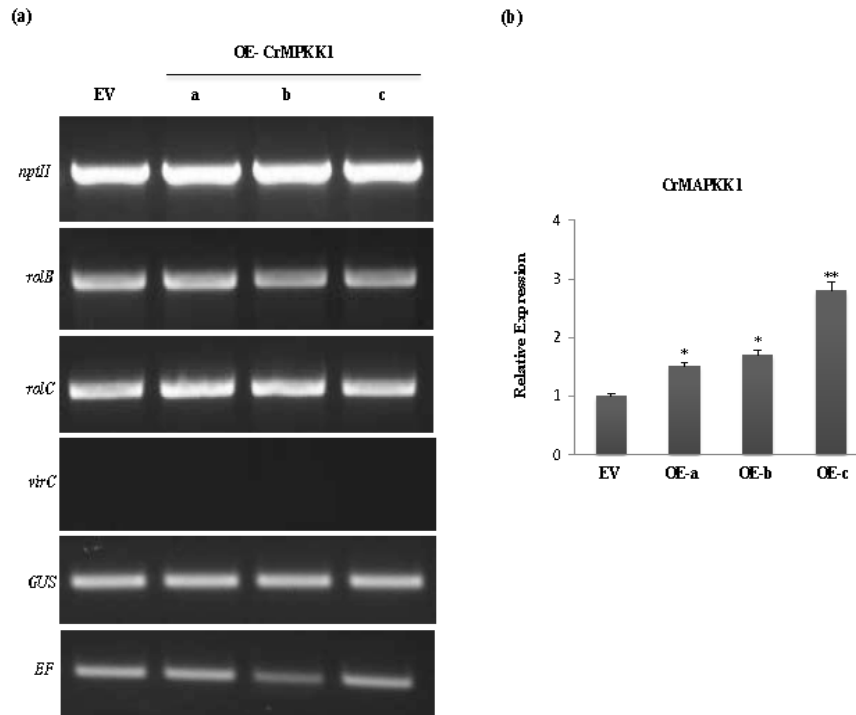


Figure 3.8 PCR confirmation of the transgenic status of *CrMAPKK1*-overexpressing *C. roseus* hairy root lines. **a.** Gene-specific primers amplified various fragments in the T-DNA from the Empty Vector (EV) control and *CrMAPKK1*-overexpression hairy root lines (OE-a, OE-b, and OE-c). **b.** Relative expression of *CrMAPKK1* in empty vector (EV) control and three overexpression (OE-a, OE-b, and OE-c) hairy root lines measured by qRT-PCR. *rolB* and *rolC*, protein-tyrosine phosphatase RolB and C, respectively; *virC*, virulence protein C; EV, empty vector; *EF1 α* , elongation factor 1 α ; *GUS*, β -glucuronidase; *NPTII*, neomycin phosphotransferase II; OE, overexpression

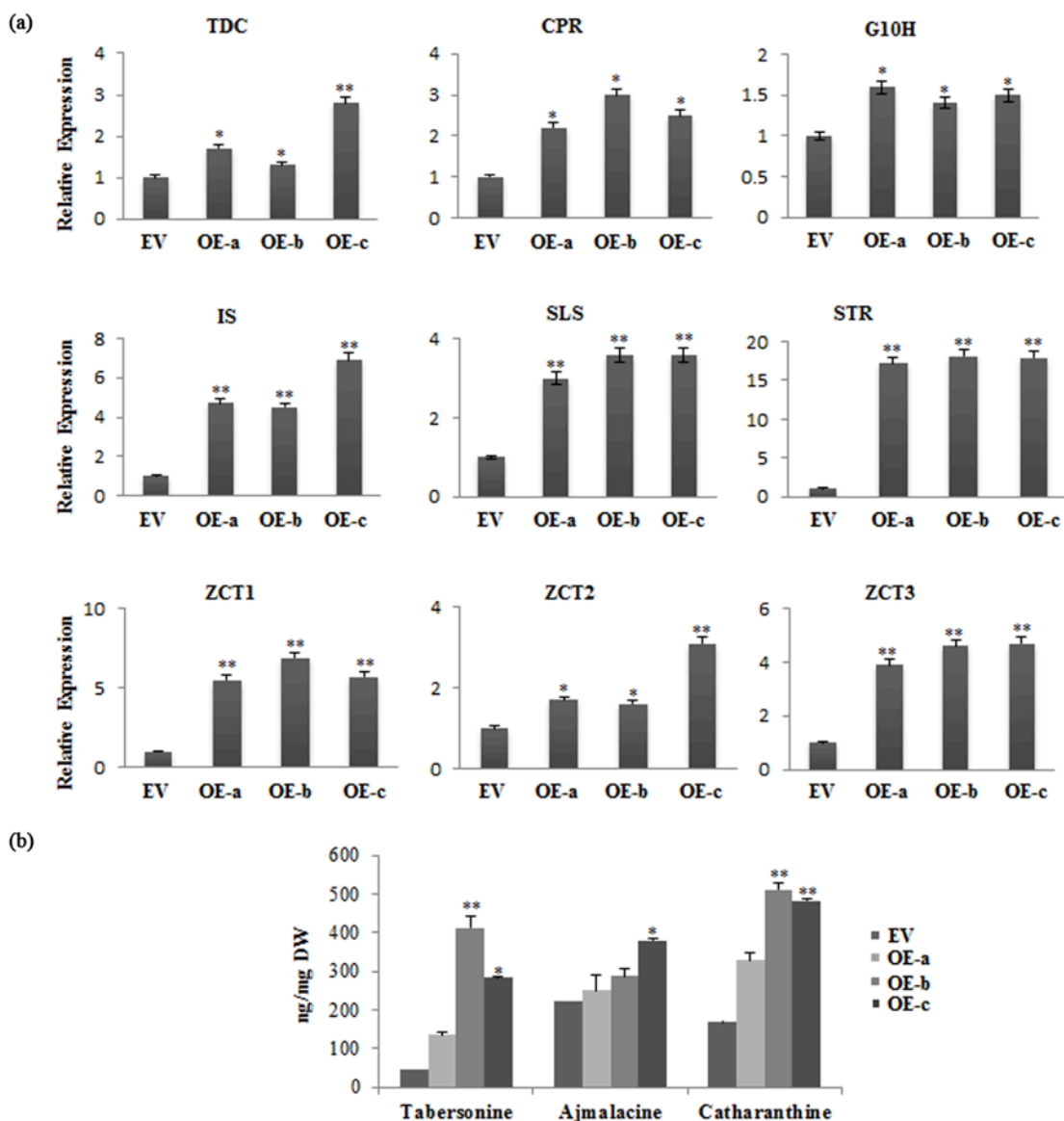


Figure 3.9 Increase in TIA pathway gene expression and alkaloid accumulation in *C. roseus* hairy roots overexpressing *CrMAPKK1*. **a.** Relative expression of TIA pathway genes, in EV control and three *CrMAPKK1*-overexpression (OE-a, OE-b, and OE-c) hairy root lines, were measured by qRT-PCR. **b.** Tabersonine, ajmalicine, and catharanthine were extracted from the EV control and the three *CrMAPKK1*-OE lines and analyzed by LC-MS/MS. The levels of tabersonine, ajmalicine, and catharanthine were estimated based on peak areas compared to the standards. Alkaloid levels are indicated in ng/mg dry weight (DW). Data represented here are the means \pm SDs from two biological

replicates. Statistical significance was calculated using the Student's t-test, * p value <0.05, ** p value <0.01.

3.4 Conclusion

In this report, we illustrate that both the *ORCA* gene cluster and CrMYC2 are subject to an additional layer of regulation by a CrMAPK cascade that consists of CrMAPK3/6, CrMAPKK1, and, possibly, CrMAPKKK1. The involvement of the CrMAPK cascade in TIA biosynthesis is substantiated by overexpression of *CrMAPKK1* in *C. roseus* hairy roots. This study represents a detailed characterization of post-translational regulatory mechanisms that govern a plant TF cluster and is summarized in this model (Figure 3.10).

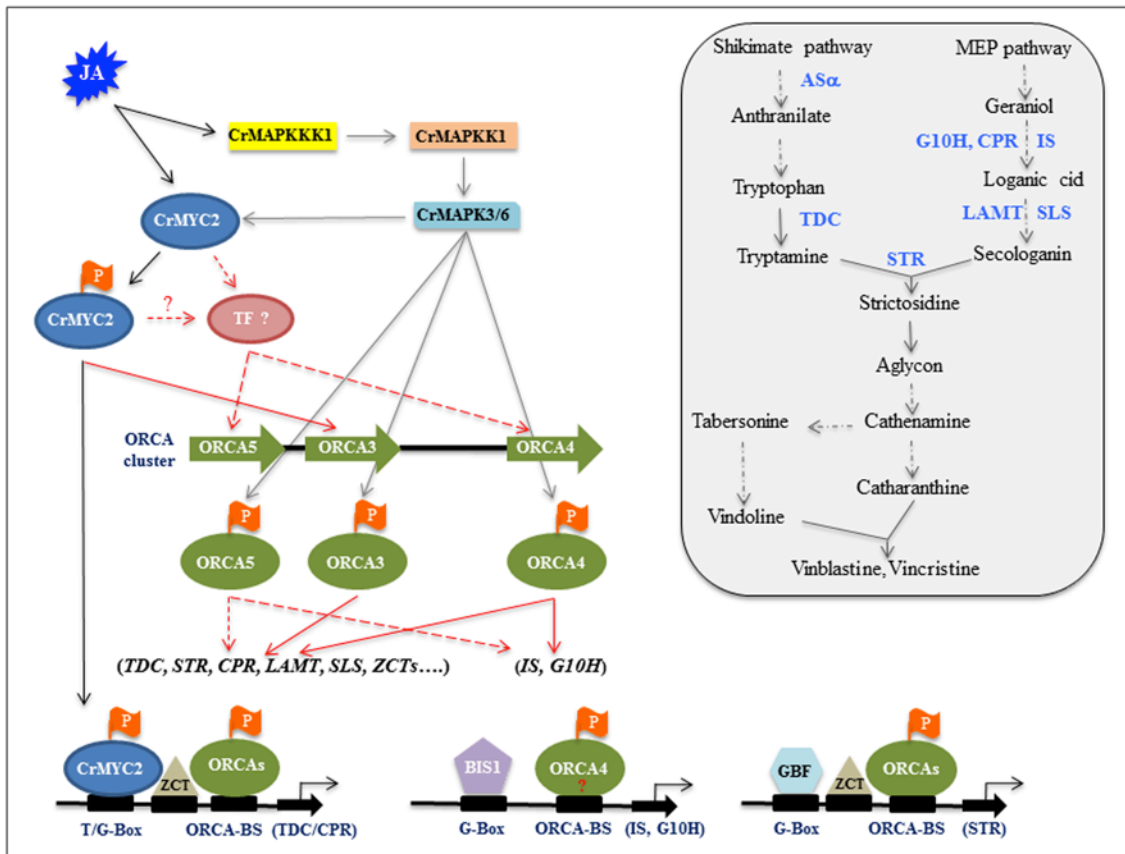


Figure 3.10 A simplified model depicting the posttranslational regulation of the CrMYC2, ORCA gene cluster, and TIA pathway genes. Upon perception of JA, the

MAP kinase cascade, consisting of CrMAPKKK1, CrMAPKK1, and CrMAPK3/6, phosphorylates CrMYC2 and the ORCA gene cluster (gray arrows). Phosphorylated CrMYC2 (labeled by orange flags) activates *ORCA3* and possibly another unidentified TF, which in turn activates *ORCA4* and *ORCA5*. Solid red arrows indicate direct activation; broken arrows represent indirect or undetermined activation. The phosphorylated ORCA 3/4/5 (labeled by orange flags) activate two subsets of TIA pathway genes (in parentheses). ORCA3, ORCA4 and ORCA5 regulate several genes in the indole and downstream TIA pathways, whereas ORCA4 and 5 activate an additional set of genes in the seco-iridoid pathway (*IS* and *G10H*). CrMYC2, in addition to activating *ORCA3*, co-regulates *TDC* and *CPR* with ORCAs by binding to the T/G-Box of the gene promoters. BIS1, binding to the G-Box, activates *IS* and *G10H*. ORCA4/5 also activates the same genes, likely by binding to the ORCA-binding sites (BS). A G-box binding factor (GBF) along with ORCA3/4/5 co-regulates *STR*. Inset: A simplified TIA pathway with key genes highlighted in blue.

Copyright © Priyanka Paul 2017

Chapter 4

Genome-wide identification of the *Catharanthus roseus* MAP kinome and functional characterization of two MAP kinases involved in TIA regulation

4.1 Introduction

Plants are constantly subjected to a number of environmental challenges, including drought, cold, salinity, pathogen and herbivore attacks. To cope with such challenges plants evolved sophisticated mechanisms to adapt their metabolism, growth, and development. Genes encoding proteins involved in signal perception and transduction have been identified and characterized in a number of plant species. Among them, protein kinases (PKs) and protein phosphatases (PPs) are significantly abundant in number. Protein phosphorylation mediated by PKs is one of the most important post-translational modifications, to transduce external signals into internal responses. One of the most common protein kinase cascades is the MAPK cascade. The MAPK cascade is typically comprising of three successively acting protein kinases, such as MAPK Kinase Kinases (MAPKKKs/MAP3Ks/MEKKs/MPKKK), MAPK Kinases (MAPKKs/MAP2Ks/ MEKs/MPKKs) and MAP Kinases (MAPKs/MAP1Ks/MPKs), connected to each other by phosphorylation (Smékalová et al., 2014). MAPKs phosphorylate specific substrates, including transcription factors (TFs), enzymes, or other kinases in response to external stimuli resulting in specific cellular responses (Cristina et al., 2010). MAPKs are activated by dual phosphorylation of conserved threonine (T) and tyrosine (Y) residues in the TXY motif located at their activation loop (or T-loop) by upstream dual specificity MAPKKs, which are in turn activated by MAPKKKs via phosphorylation of conserved serine (S) and / or threonine (T) residues in S/T-X₃₋₅-S/T motif (Cristina et al., 2010).

The MAP1Ks are divided into four groups based on protein structure and sequence of the TXY sequence motif. The MAP1Ks with TEY motif are grouped into A, B and C, whereas that having motif TDY are into group D (Ichimura et al., 2002). The possible number of MAP1Ks in group D varies between different species, such as 10 in rice (Rao

et al., 2010), 8 in Arabidopsis (Cardinale et al., 2002), 14 in soybean (Neupane et al., 2013), and 11 in maize (Kong et al., 2013b; Liu et al., 2013). Group A MAP1Ks are well characterized in plants and are mostly involved in hormonal and environmental stress responses. For instance, tobacco SIPK and WIPK, which are orthologous to AtMAPK6 and AtMAPK3 of Arabidopsis respectively, are activated during both biotic and abiotic stress responses (Seo et al., 1995; Zhang and Klessig, 2001; Meng et al., 2013; Vogel et al., 2014). On the contrary, group B MAP1Ks have been relatively less studied and are reported to be involved in both environmental stress responses and cell division (Kosetsu et al., 2010; Takahashi et al., 2010; Berriri et al., 2012). The function of Group C MAP1Ks is also less studied. For instance, in Arabidopsis, MAPK7 is reported to be involved in circadian rhythm whereas in rice the orthologue mediates resistance against *Xanthomonas oryzae* (Schaffer et al., 2001; Jalmi and Sinha, 2016). Very little is known about the biological roles of the Group D MAPKs. In rice, the OsMAPK20-4 has been shown to be involved in both developmental regulation and environmental stress responses (Agrawal et al., 2003).

Phylogenetic analysis has grouped plant MAP2K genes into four groups: A, B, C, and D (Ichimura et al., 2002) and are involved in a wide range of biological processes, including pathogen defense and secondary metabolism. The Group A MAP2Ks, including MAPKK1 and MAPKK2 act upstream of MAPK4 in Arabidopsis (Ichimura et al., 1998; Huang et al., 2000; Cardinale et al., 2002) to negatively regulate the defense responses against biotrophic pathogens (Petersen et al., 2000; Qiu et al., 2008). Arabidopsis MAPKK6 is reported to be involved in cell division (Takahashi et al., 2010). Group B MAP2Ks, such as Arabidopsis MAPKK3 and tobacco NPK2 are involved in nuclear transport in yeast cells (Quimby et al., 2000). Group C MAP2Ks, such as AtMEK4 and 5 are involved in various environmental stress responses (Ren et al., 2002; Liu et al., 2008). Group D MAP2Ks, such as Arabidopsis MAPKK9 activates the MAPK3/MAPK6 to induce camalexin and ethylene biosynthesis (Liu and Zhang, 2004; Xu et al., 2008; Yoo et al., 2008); MAPKK9-MAPK3/MAPK6 cascade is also known to regulate leaf senescence in Arabidopsis (Zhou et al., 2009). CrMAPKK1 and JAM1, which are

orthologous to AtMAPKK9, are involved in regulation of TIA and nicotine biosynthesis in *C. roseus* and tobacco, respectively (De Boer et al., 2011; Paul et al., 2017).

MAP3Ks constitute a large and diverse family of kinases that have been classified into three subfamilies, MEKK-like, Raf-like, and ZIK-like family, based on the sequence of the catalytic domain (Wrzaczek and Hirt, 2001). Similar to the MEKKs in yeast (*e.g.* STE11 and BCK1) and mammal (*e.g.*, MEKK1), the plant MEKK-like subfamily members also have a conserved catalytic domain, and the majority of them are involved in stress responses by activating downstream MAP2Ks (Champion et al., 2004). In Arabidopsis, AtMEKK1 activates MAPKK4/MAPKK5-MAPK3/MAPK6 cascade to regulate innate immunity in response to bacterial and fungal pathogens (Asai et al., 2002). Similar to mammalian RAF1, the plant Raf-like MAP3Ks also share the GTXX (W/Y) MAPE signature motif (Ichimura et al., 2002; Jonak et al., 2002). In Arabidopsis, the Raf-like MAP3Ks including Enhanced Disease Resistance 1 (EDR1) and Constitutive Triple Response 1 (CTR1) are reported to be involved in defense responses and ethylene signaling, respectively (Kieber et al., 1993; Frye and Innes, 1998; Frye et al., 2001; Huang et al., 2003). Though ZIK-like kinases are identified in plant MAP3Ks phylogenetic analyses, these enzymes have not been reported to phosphorylate the downstream MAP2Ks (Champion et al., 2004; Xu et al., 2005).

Plant MAPK cascades are involved in numerous biological processes, such as cell division (Jiménez et al., 2007), growth and development (Xu and Zhang, 2015), hormonal responses (Tena et al., 2001), and environmental stress responses, including pathogen infection (Pitzschke et al., 2009), wounding (Heinrich et al., 2011), drought, salinity (Chávez Suárez and Ramírez Fernández, 2010), UV irradiation (Parages et al., 2013), reactive oxygen species (ROS) (Pitzschke et al., 2009), and ozone (Samuel et al., 2000). To date, numerous MAPK cascades have been identified and characterized in plants. In Arabidopsis, the MEKK1-MAPKK4/5-MAPK3/6-WRKY22/WRKY29 cascade is involved in plant innate immunity (Asai et al., 2002; Galletti et al., 2011), the MEKK1-MAPKK1/2-MAPK4 positively regulates defense responses against necrotrophic fungi and negatively control defenses against biotrophic pathogens

(Petersen et al., 2000; Qiu et al., 2008), and YDA-MAPKK4/5-MAPK3/6 regulates stomatal development (Eckardt, 2007). In rice, the OsMAPKK4-OsMAPK1-OsWRKY53 cascade is involved in the wounding signaling pathway (Yoo et al., 2014), and OsMAPKK3-OsMAPK7-OsWRK30 module acts against *Xanthomonas oryzae*, which causes leaf blight disease in rice (Jalmi and Sinha, 2016). Recently, a MAPK cascade consisting of GhMAPKKK15–GhMAPKK4–GhMPK6 has been reported to directly phosphorylates and activates GhWRKY59 that regulates drought response in cotton (Li et al., 2017). The genome sequencing has made it possible to identify and characterize MAPK families in a number of other plant species, including maize (Kong et al., 2013b; Liu et al., 2013), wheat (Wang et al., 2016), tomato (Kong et al., 2012; Wu et al., 2014), *Brachypodium distachyon* (Chen et al., 2012), *Vitis vinifera* (Çakır and Kılıçkaya, 2015), cucumber (Wang et al., 2015), and *Capsicum annuum* (Liu et al., 2014).

Catharanthus roseus of the family Apocynaceae is the exclusive source of an array of bioactive TIAs, including the anticancer drugs, vincristine and vinblastine (Kellner et al., 2015). However, the natural production of TIAs in *C. roseus* is extremely limited. In recent years, significant progress has been made in studying the transcriptional regulation of the TIA biosynthesis. TFs belonging to AP2/ERF (Menke et al., 1999b; van der Fits and Memelink, 2000; Paul et al., 2017), bHLH (Zhang et al., 2011; Van Moerkercke et al., 2015; Van Moerkercke et al., 2016) and WRKY (Suttipanta et al., 2011) families have been shown as the key regulators of the TIA pathway. Previously, two members of AP2/ERF gene family such as ORCA2 and ORCA3 were reported to regulate the TIA biosynthesis in *C. roseus* (Menke et al., 1999b; van der Fits and Memelink, 2000). Recently, I cloned and characterized two additional AP2/ERF genes, *ORCA4* and *ORCA5*, which are physically clustered with *ORCA3* at the same genomic scaffold (Kellner et al., 2015). I have shown that though *ORCA4* and *ORCA5* functionally overlap with *ORCA3*, they regulate an additional set of TIA pathway genes. *ORCA4* or *ORCA5* overexpression resulted in significant increase of TIA accumulation in *C. roseus* hairy roots (Paul et al., 2017). The genes of indole branch are mostly regulated by the ORCA gene cluster (Menke et al., 1999b; van der Fits and Memelink, 2000; Paul et al., 2017) and CrMYC2 (Zhang et al., 2011); however the majority of iridoid pathway genes are controlled by two

newly identified bHLH TFs, BIS1 and BIS2 (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016).

Post-transcriptional and post-translational modulations of TFs add another layer of regulatory complexity to the biosynthesis of specialized metabolites in plants (Maier et al., 2013; Patra et al., 2013a; Patra et al., 2013b; Shen et al., 2017). Post-translational regulatory mechanisms, including phosphorylation of TFs have been implicated in the biosynthesis of plant specialized metabolites, including camalexin and nicotine biosynthesis in *Arabidopsis* and tobacco, respectively (Xu et al., 2008; De Boer et al., 2011; Mao et al., 2011). In *C. roseus*, JA-induced expressions of TIA pathway gene transcripts are reduced significantly in the presence of protein kinase inhibitor, K-252a, but are not inhibited by the protein synthesis inhibitor, cycloheximide (Menke et al., 1999a). This observation indicates the possible role of protein phosphorylation in the JA signal transduction cascade (Menke et al., 1999a). Moreover, a JA-inducible MAPK, CrMAPK3, has been reported in *C. roseus* (Raina et al., 2012). Recently, we demonstrated that in *C. roseus*, the ORCA gene cluster and CrMYC2 act downstream of a MAP kinase cascade that includes CrMAPKKK1, CrMAPKK1, CrMAPK3 and 6. Overexpression of *CrMAPKK1* in *C. roseus* hairy roots upregulated TIA pathway gene expression and boosted TIA accumulation (Paul et al., 2017). To date, no systematic and genome-wide investigations of the MAPK gene families have been reported in *C. roseus*.

In this study, we performed a genome-wide analysis to identify all MAPKs in *C. roseus*. A total of 83 putative *C. roseus* MAPKs, including 12 CrMAP1Ks, 6 CrMAP2Ks, 59 CrMAP3Ks, and 6 CrMAP4Ks were identified. Their domain organization and expression profiles in different tissues were analyzed. We next identified and functionally characterized two previously uncharacterized MAPK cascades that regulate two different branches of TIA pathway. The MAPKK6-MAPK13 positively regulates the iridoid branch of TIA pathway in *C. roseus*. Overexpression of *CrMAPK13* in *C. roseus* hairy roots upregulated the BIS-regulated iridoid pathway genes and increased tabersonine accumulation. The other MAPK cascade consists of CrMAPKK1 and CrMAPK20, which regulates the genes in the indole branch of TIA pathway. Overexpression of *CrMAPK20*

in *C. roseus* hairy roots down-regulated the CrMYC2-ORCA regulated indole pathway genes and reduced catharanthine accumulation. In conclusion, this work provided a detailed understanding of the post-translational regulatory network in TIA biosynthesis in *C. roseus*.

4.2 Materials and Methods

4.2.1 Plant materials

C. roseus (L.) G. Don var. ‘Little Bright Eye’ was used for gene expression and cloning, and generation of transgenic hairy roots. Tobacco var. Xanthi cell line was used for protoplast transient expression assays.

4.2.2 RNA isolation and cDNA synthesis

C. roseus var. ‘Little Bright Eye’ seeds were surface-sterilized with 30% commercial bleach, then washed five times with sterile water and afterward germinated on half-strength Murashige and Skoog (MS) medium (Caisson Labs, USA). Ten-day-old *C. roseus* seedlings were next treated with 100 μ M methyl jasmonate (MeJA) for 2 h and mock-treated samples were used as a control. Both samples were used for total RNA isolation and cDNA synthesis as described previously (Suttipanta et al., 2007). CrMAPKK6, CrMAPK13, and CrMAPK20 were PCR amplified from cDNA using gene-specific primers (Table 4.1), and then cloned into the pGEM-T Easy vector (Promega, USA) followed by sequencing confirmation.

4.2.3 Multiple sequence alignment and phylogenetic analysis

Alignment of deduced amino acid sequences was performed using the ClustalOmega program (Larkin et al., 2007) with the default parameters. Phylogenetic trees were constructed and visualized using the neighbor-joining (NJ) method through MEGA version 6 software (Tamura et al., 2011). The statistical reliability of individual nodes of the newly constructed tree was assessed by bootstrap analyses with 1,000 replications.

4.2.4 Identification and co-expression analysis of MAPK cascade genes in *C. roseus*

The method used to identify all the putative MAP1K, MAP2K, MAP3K and MAP4K family genes were similar to that used previously for other plants (Chen et al., 2012; Kong et al., 2013a). Briefly, to identify *C. roseus* MAPK cascade genes, all protein-coding genes were obtained from the *C. roseus* draft genome, the Dryad Digital Repository (Kellner et al., 2015). Next, the MAPK cascade genes from Arabidopsis and rice (*Oryza sativa*) were used as queries to search against the *C. roseus* protein databases with the BLASTP program (Wang et al., 2015) with an e-value of $1e^{-10}$ as the threshold. Putative MAP kinases in the *C. roseus* proteome were identified using the signature motifs previously described (Jonak et al., 2002) and a total of 91 putative MAPK cascade genes were identified. The Conserved Domain Database of The National Center for Biotechnology Information (NCBI) (Wang et al., 2015) and Pfam (Lehti-Shiu and Shiu, 2012) databases were then used to confirm further and to predict the conserved domains in MAPK cascade genes. Finally, a custom PERL script was used to calculate the isoelectric points (pIs) and molecular weights (MWs) of MAPK cascade proteins.

To study the expression patterns of *C. roseus* MAPK cascade genes and TIA pathway genes, transcriptome data for five different tissues, including flower, mature leaf, immature leaf, stem, and root were obtained from the Sequence Read Archive (SRA) database at the NCBI (accession number SRA030483). Raw reads were further processed and reads per kilobase of transcript per million mapped reads (RPKM) value was calculated as described previously (Singh et al., 2015). Pairwise Pearson correlation coefficient for each transcript was calculated using RPKM. Matrix distances for expression heatmap were computed over Pearson correlations of gene expression values (RPKM) by the heatmap.2 function of gplots (version 3.0.1) Bioconductor package in R (version 3.2.2) (Wickham, 2009).

4.2.5 Quantitative RT-PCR (qRT-PCR)

The qRT-PCR was performed as described previously (Suttipanta et al., 2011). The primers used in qRT-PCR are listed in Table 4.1. The *C. roseus* 40S Ribosomal Protein

S9 (*RPS9*) gene was used as an internal control (Liscombe et al., 2010). All PCRs were performed in triplicate and repeated twice.

Table 4.1 Oligonucleotides used in this study

<i>RPS9</i>	5'-GAGGGCCAAAACAAACTTA-3'	5'-CCCTTATGTGCCTTTGCCTA-3'
<i>EF1α</i>	5'-TACTGTCCCTGTTGGTCGTG-3'	5'-AAGAGCTTCGTGGTGCATCT-3'
<i>ASa</i>	GCGAACATTTGCAGATCCAT	GGCCGATTTGTTATTGTTCC
<i>TDC</i>	5'-ATCCGATCAAACCCATACCA-3'	5'-CGTCATCCTCGACCATTTTT-3'
<i>G10H</i>	5'-TTATTCCGATTCTGCCAAGG-3'	5'-TCCCAAAGTGAATCGTCAT-3'
<i>CPR</i>	5'-TGGCAGAAAAGGCTTCTGAT-3'	5'-CTCAGCCTGTGTGCTATCCA-3'
<i>SLS</i>	5'-GTTCTTCTCACCGGAGTTG-3'	5'-CCCATTTGGTCAACATGTCA-3'
<i>STR</i>	5'-ACCATTGTGTGGGAGGACAT-3'	5'-ATTTGAATGGCACTCCTTGC-3'
<i>LAMT</i>	5'-AGCCAAGGCAGTGATTGTTG-3'	5'-CTGCAATGCGGAAAGGTTTG-3'
<i>IS</i>	5'-CCACATGATTCGCCTTTTACCG-3'	5'-AAACCCGAAAACCAGAGCTG-3'
<i>ORCA3</i>	5'-CGGGATCCGAAATACAGAAA-3'	5'-GCCCTTATACCGGTTCCAAT-3'
<i>ORCA4</i>	5'-ATAGTAGTACTGCCGCCGAAAG-3'	5'-ATCTCCGCCGCAAATTTTCC-3'
<i>ORCA5</i>	5'-TCTTCCAACGGAGGTTAACGG-3'	5'-AATGTTGTCTCCAGGGCTTG-3'
<i>CrMAPKK1</i>	5'-AAGGGCAAAGACCTGATTGG-3'	5'-TCAGGCAAACCTGGTGGTTC-3'
<i>ZCT1</i>	5'-AGCCGAAAACATCATGCTTGT-3'	5'-CGCCTTTGCAACAGGTTTAT-3'
<i>ZCT2</i>	5'-CGTCAATTTCCATCGTTTCA-3'	5'-CCGATAGCGAATTCAGTCC-3'
<i>ZCT3</i>	5'-GACAAGCTTTGGGAGGACAC-3'	5'-GGCAAGGCAGGTAAGTTCAA-3'
<i>MAPK13</i>	5'-ATCAATCCTGTTGGCAGTGG-3'	5'-TCTCGAAGGGTTCTTTTGGC-3'
<i>MAPKK6</i>	5'-AGCCATTGACGGAACCTCAAG-3'	5'-TAAGGCGCAATCCTTTCTGG-3'
<i>MAPK19</i>	5'-TCGCCAAAACCACTGATTGC-3'	5'-TGCATATTCTGAGCCAAGGC-3'
<i>MAPK20</i>	5'-GCTTGGCCAGAGTTGCATTC-3'	5'-AGATGGCTCCCTCTCGACTT-3'
<i>CrMAPKK1</i>	5'-gcggatcc ATGGCATTAGTTA AAGACCGTC-3'	5'-atgctgcag TTATGGCCTTTTA ACCTTTTGC-3'
<i>CrMAPKK6</i>	5'-gcgaattcATGAAGCTTAAGAAGCCAT TG-3'	5'-atggtcgacTTACCTCGGAAAATTG TAGG-3'
<i>CrMAPK13</i>	5'-gcggatccATGGAAATGGAAAAGG AAAAG-3'	5'-atgtctagaATCATGCCATCTCATGA GCTTG-3'
<i>CrMAPK20</i>	5'-gcaaggatcc ATGCAGCCCCGATCACAGGA-3'	5'-accggtgcagTTACCTA GGAAGAGAGACATGTTTTTCTC-3'
<i>rolB</i>	5'-CTTATGACAAACTCATAGATAAAA GGTT-3'	5'-TCGTAACTATCCAACCTCACATCA C-3'
<i>rolC</i>	5'-CAACCTGTTTCTACTTTGTTAAC- 3'	5'-AAACAAGTGACACACTCAGCTT C-3'
<i>virC</i>	5'-TTTTGCTCCTTCAAGGGAGGTGCC- 3'	5'-GGCTTCGCCAACCAATTTGGAGA T-3'
<i>GUS</i>	5'-ATGGTAGATCTGAGGGTAAATTT C-3'	5'-GCTAGCTTGTTCCTCCCTG-3'
<i>ntpII</i>	5'-ATGGGGATTGAACAAGATGGA-3'	5'-TCAGAAGAACTCGTCAAGAAG- 3'

4.2.6 Protoplast isolation and electroporation

For transient protoplast assays, the reporter plasmids were generated by cloning the *STR*, *G10H*, or *IS* promoter into a modified pUC vector containing a firefly *luciferase* (*LUC*) reporter and *rbcS* terminator. The (*5XGALRE*)-*35S_{pro}:LUC* was described previously (Patra et al., 2013a). The effector plasmids were constructed by cloning *ORCA3/4/5*, *BIS1*, *CrMAPK13*, *CrMAPK20* and *CrMAPKK6* into a modified pBS vector under the control of the *CaMV35S* promoter and *rbcS* terminator. For the protoplast-based two-hybrid assay, *CrMAPK13* and *CrMAPK20* were fused to the GAL4 DNA binding domain in a pBS plasmid containing *MMV* promoter and *rbcS* terminator. A *GUS* reporter, driven by the *CaMV35S* promoter and *rbcS* terminator, was used as an internal control in all protoplast assays. Protoplast isolation from tobacco cell suspension cultures and electroporation with plasmid DNA were performed as previously described (Pattanaik et al., 2010b; Pattanaik et al., 2010a). The reporter, effector, and internal control plasmids were electroporated into tobacco protoplasts in different combinations; luciferase and GUS activities in transfected protoplasts were measured as described previously (Suttipanta et al., 2007; Paul et al., 2017). Each experiment repeated three times.

4.2.7 Yeast two-hybrid Assay

In order to detect the kinase-kinase or kinase-substrate interaction, yeast two-hybrid (Y2H) assay was performed. The plasmids, pAD-GAL4-2.1 and pBD-GAL4-Cam (Stratagene, USA) containing the GAL4 activation and GAL4 DNA-binding domains, respectively were used. The full-length coding region of *CrMAPK13* and *CrMAPKK6* were cloned into both pAD-GAL4 and pBD-GAL4 vectors. Similarly, the full-length coding region of *CrMAPK20* was cloned into pBD-GAL4 vector, whereas the *CrMAPKK1*, *ORCA3*, *ORCA4*, *ORCA5* and *CrMAPK3* were cloned into pAD-GAL4 vector. The resultant plasmids were co-transformed into yeast strain AH109 using the PEG/LiAC method (Pattanaik et al., 2010b). Co-transformation of the empty vector pAD-GAL4 and pBD-*CrMAPK13/CrMAPKK6/CrMAPK20* were used as negative control. Transformed colonies were selected on synthetic dropout (SD) medium lacking leucine and tryptophan (SD-Leu-Trp). Colonies from double selection plates were then screened for growth on triple selection SD medium lacking histidine, leucine and tryptophan (SD-

His-Leu-Trp) with a required concentration of 3-amino-1,2,4-triazole (3AT) (Pattanaik et al., 2010b). Each experiment was repeated at least two times.

4.2.8 Construction of plant expression vectors and generation of hairy roots

For plant transformation, *CrMAPK20* and *CrMAPK13* cDNAs were PCR-amplified from *C. roseus* seedling cDNA and then cloned into a modified pCAMBIA2301 vector containing the *CaMV35S* promoter and the *rbcS* terminator (Pattanaik et al., 2010b). The pCAMBIA2301 vector only was used as an empty vector (EV) control. The plasmids were mobilized into *Agrobacterium rhizogenes* R1000 using freeze-thaw method (Weigel and Glazebrook, 2006). Transformation of *C. roseus* seedlings and generation of hairy roots were performed using the protocol described previously (Suttipanta et al., 2011; Paul et al., 2017). Transgenic status of the hairy root lines was verified by PCR amplification of the *rolB*, *rolC*, *virC*, *nptII* and *GUS* genes. Primers used in this study are listed in Table 4.1. Three independent hairy root lines were selected for further analysis.

4.2.9 Alkaloid extraction and analysis

To extract alkaloid, transgenic hairy roots were frozen in liquid nitrogen and ground to a powder. Samples were extracted in methanol (1:100 w/v) twice for 24 h on a shaker. Pooled extracts were then dried *via* a rotary evaporator and diluted in methanol 10 µL/mg of the initial sample. The samples were then analyzed using high-performance liquid chromatography (HPLC), followed by electrospray-injection in a tandem mass spectrometry, as previously described (Suttipanta et al., 2011; Paul et al., 2017). The known alkaloid standards were run to identify elution times and mass fragments.

4.3 Results and Discussion

4.3.1 Identification of the MAP1K, MAP2K, MAP3K and MAP4K families in *C. roseus*

MAPK cascades are ubiquitous in all eukaryotes and are responsible for various cellular responses (Yoo et al., 2008). We took advantage of the *C. roseus* draft genome (Kellner et al., 2015) to identify MAP1K, MAP2K, MAP3K and MAP4K genes in its genome. We conducted BLASTP searches against the *C. roseus* protein database using MAP kinase

cascade proteins from Arabidopsis and rice, as a query. The candidate sequences were further evaluated for the putative functional domains and motifs. Sequences that did not contain the known conserved motifs of the MAP1K, MAP2K MAP3K and MAP4K family proteins were excluded from further analysis. We, finally, identified 12 CrMAP1Ks, 6 CrMAP2Ks, 67 CrMAP3Ks and 6 CrMAP4Ks genes in *C. roseus*. Out of 67 MAP3Ks, we could retrieve full-length sequences for 59 and these were used for further analysis.

The CrMAP1K genes were predicted to encode proteins from 219 to 753 amino acids in length with calculated molecular weights (MWs) ranging from 25.2 to 85.0 kDa and theoretical isoelectric points (pIs) ranging from 5.19 to 9.2. The sub-cellular localization predication indicated that most of the members of CrMAP1K are located in the cytoplasm, except CRO_T006826 and CRO_T008379, which were present in mitochondria (Table 4.2). The 6 CrMAP2Ks proteins are 341–606 amino acid in length, with calculated MWs ranging from 38.26 to 68.0 kDa and pIs from 5.57 to 8.81. They were predicted to be localized in the cytoplasm and nucleus (Table 4.3). The polypeptide lengths of the CrMAP3K genes ranged from 138 to 1426 amino acids with MWs ranging from 15.5 to 154.4 kDa and theoretical pIs ranging from 4.66 to 9.97. Similarly, CrMAP3K proteins were located in the various cellular compartments, including cytoplasm, nucleus, plasma membrane, mitochondria, or chloroplast (Table 4.4). The polypeptide lengths of the MAP4K genes ranged from 507 to 883 amino acids, and their predicted MWs ranged from 56.68 to 97.96 kDa. The theoretical pIs ranged from 5.77 to 8.88 (Tables 4.5).

Table 4.2. List of MAP1Ks identified in *C. roseus*.

Gene	Full length	MAP1K group	T-loop	Length (amino acids)	Molecular weight (kDa)	Isoelectric point	Best Arabidopsis Match	Subcellular localization:score
CRO_T025299	Yes	A	TEY	375	43.06	5.26	AT3G45640	cytoplasmic: 0.83
CRO_T025830	Yes	A	TEY	398	45.69	5.51	AT2G43790	cytoplasmic: 0.86
CRO_T003948	Yes	B	TEY	340	39.21	6.06	AT4G01370	cytoplasmic: 0.83
CRO_T006826	Yes	B	TEY	387	43.91	6.27	AT4G01370	mitochondrial: 0.63
CRO_T026682	Yes	B	TEY	380	43.27	5.19	AT1G07880	cytoplasmic: 0.85
CRO_T024779	Yes	C	TEY	630	71.95	6.37	AT3G18040	cytoplasmic: 0.84

Table 4.2. List of MAP1Ks identified in *C. roseus* (continued.)

CRO_T022398	Yes	C	TEY	371	42.06	7.16	AT2G18170	cytoplasmic: 0.84
CRO_T032753	Yes	D	TDY	568	64.37	8.98	AT5G19010	cytoplasmic: 0.74
CRO_T008379	Yes	D	TDY	373	42.92	6.53	AT1G10210	mitochondrial: 0.62
CRO_T007816	Yes	D	TDY	445	51.75	9.2	AT2G42880	cytoplasmic: 0.77
CRO_T005275	Yes	D	TDY	753	85.00	8.9	AT3G14720	cytoplasmic: 0.42
CRO_T033572	Yes	D	TDY	219	25.19	7.7	AT3G18040	cytoplasmic: 0.64

Table 4.3. List of MAP2Ks identified in *C. roseus*.

Gene	Full length	MAP2K group	Length (amino acids)	Molecular weight (kDa)	Isoelectric point	Best Arabidopsis Match	Subcellular localization:score
CRO_T004930	Yes	A	355	39.76	6.01	AT5G56580	cytoplasmic: 0.85
CRO_T005978	Yes	A	354	39.25	5.57	AT4G29810	cytoplasmic: 0.66
CRO_T019539	Yes	B	606	68.03	5.78	AT5G40440	cytoplasmic: 0.66
CRO_T030701	Yes	C	361	40.94	6.03	AT1G32320	nuclear: 0.67
CRO_T011505	Yes	D	341	38.26	8.04	AT1G73500	nuclear: 0.51
CRO_T020175	Yes	D	377	41.60	8.81	AT1G51660	cytoplasmic: 0.65

Table 4.4. List of MAP3Ks identified in *C. roseus*.

Gene	Full length	MAP3K group	Length (amino acids)	Molecular weight (kDa)	Isoelectric point	Best Arabidopsis Match	Subcellular localization:score
CRO_T022055	No	MEKK	361	38.35	10.41	AT1G63700	cytoplasmic: 0.72
CRO_T025162	No	MEKK	332	36.90	8.08	AT3G50310	nuclear: 0.82
CRO_T025486	No	MEKK	181	20.06	4.83	AT3G07980	cytoplasmic: 0.52
CRO_T003904	Yes	MEKK	294	32.39	9.08	AT1G09000	cytoplasmic: 0.67
CRO_T007266	Yes	MEKK	553	60.73	5.32	AT4G08500	nuclear: 0.75
CRO_T008488	Yes	MEKK	353	38.23	8.96	AT1G63700	cytoplasmic: 0.64
CRO_T009969	Yes	MEKK	437	48.90	8.87	AT5G66850	cytoplasmic: 0.57
CRO_T012002	Yes	MEKK	1031	113.73	5.72	AT5G55090	cytoplasmic: 0.69
CRO_T014277	Yes	MEKK	711	77.82	5.57	AT4G08500	nuclear: 0.77
CRO_T014486	Yes	MEKK	280	31.36	4.98	AT4G36950	nuclear: 0.39
CRO_T015682	Yes	MEKK	296	32.28	4.66	AT4G36950	mitochondrial: 0.46
CRO_T017735	Yes	MEKK	266	29.42	4.92	AT1G07150	nuclear: 0.66
CRO_T020201	Yes	MEKK	722	80.10	7.23	AT3G06030	cytoplasmic: 0.46
CRO_T020390	Yes	MEKK	206	23.07	4.88	AT5G50180	plasma membrane: 0.32
CRO_T020775	Yes	MEKK	707	76.47	9.16	AT1G53570	nuclear: 0.71
CRO_T022731	Yes	MEKK	436	48.27	5.5	AT2G32510	nuclear: 0.42
CRO_T024393	Yes	MEKK	404	45.32	4.85	AT4G36950	nuclear: 0.9
CRO_T027542	Yes	MEKK	138	15.50	6.27	AT1G63700	cytoplasmic: 0.52
CRO_T029124	Yes	MEKK	1297	143.95	7.99	AT5G66850	nuclear: 0.38
CRO_T031663	Yes	MEKK	286	31.27	6.01	AT3G13530	cytoplasmic: 0.66
CRO_T001547	No	Raf	150	16.92	8.05	AT5G01850	cytoplasmic: 0.54
CRO_T013561	No	Raf	981	108.96	6.64	AT5G57610	nuclear: 0.55

Table 4.4. List of MAP3Ks identified in *C. roseus* (continued.)

CRO_T025685	No	Raf	688	76.76	5.66	AT2G31010	cytoplasmic: 0.62
CRO_T032003	No	Raf	597	66.62	5.99	AT4G24480	cytoplasmic: 0.6
CRO_T000074	Yes	Raf	830	91.10	5.62	AT5G03730	nuclear: 0.47
CRO_T000665	Yes	Raf	218	24.82	9.17	AT1G08720	cytoplasmic: 0.79
CRO_T001188	Yes	Raf	732	82.29	7.3	AT4G23050	cytoplasmic: 0.55
CRO_T001963	Yes	Raf	1426	154.40	5.28	AT3G46920	cytoplasmic: 0.54
CRO_T002980	Yes	Raf	388	43.92	9.14	AT1G62400	mitochondrial: 0.91
CRO_T006759	Yes	Raf	402	44.71	7.92	AT3G22750	cytoplasmic: 0.81
CRO_T007171	Yes	Raf	315	36.53	8.51	AT4G14780	cytoplasmic: 0.89
CRO_T007789	Yes	Raf	1287	142.50	5.58	AT5G11850	nuclear: 0.53
CRO_T008177	Yes	Raf	469	52.81	4.71	AT4G35780	cytoplasmic: 0.81
CRO_T008654	Yes	Raf	425	47.42	7.53	AT5G50000	cytoplasmic: 0.78
CRO_T009390	Yes	Raf	511	57.91	8.68	AT2G43850	mitochondrial: 0.8
CRO_T009758	Yes	Raf	225	25.59	8.46	AT5G11850	cytoplasmic: 0.28
CRO_T009765	Yes	Raf	467	53.14	6.27	AT4G18950	cytoplasmic: 0.56
CRO_T011578	Yes	Raf	1299	143.72	5.12	AT1G16270	ER: 0.39
CRO_T012915	Yes	Raf	841	94.65	6.34	AT3G58640	nuclear: 0.49
CRO_T013026	Yes	Raf	352	40.45	8.42	AT1G62400	cytoplasmic: 0.8
CRO_T013539	Yes	Raf	1281	141.06	5.39	AT1G79570	nuclear: 0.57
CRO_T013653	Yes	Raf	1237	137.80	5.07	AT5G57610	cytoplasmic: 0.63
CRO_T014965	Yes	Raf	459	52.09	5.69	AT3G58760	cytoplasmic: 0.77
CRO_T015407	Yes	Raf	1364	150.86	5.7	AT3G24715	nuclear: 0.52
CRO_T016949	Yes	Raf	370	42.36	8.32	AT1G14000	cytoplasmic: 0.34
CRO_T017479	Yes	Raf	200	21.96	7.66	AT5G01850	cytoplasmic: 0.83
CRO_T017535	Yes	Raf	365	40.50	6.55	AT5G66710	cytoplasmic: 0.82
CRO_T018707	Yes	Raf	437	49.03	7.78	AT1G14000	cytoplasmic: 0.65
CRO_T018906	Yes	Raf	571	64.69	5.81	AT4G38470	cytoplasmic: 0.84
CRO_T019753	Yes	Raf	672	72.36	5.17	AT5G11850	cytoplasmic: 0.5
CRO_T021479	Yes	Raf	281	32.04	9.26	AT4G38470	mitochondrial: 0.66
CRO_T023512	Yes	Raf	282	31.53	9.97	AT4G24480	mitochondrial: 0.76
CRO_T025952	Yes	Raf	421	47.08	6.52	AT3G01490	cytoplasmic: 0.75
CRO_T028356	Yes	Raf	428	47.58	9.08	AT2G24360	cytoplasmic: 0.52
CRO_T029878	Yes	Raf	1184	133.01	6.1	AT1G04700	nuclear: 0.6
CRO_T031633	Yes	Raf	505	56.83	5.84	AT4G35780	cytoplasmic: 0.67
CRO_T032564	Yes	Raf	464	52.52	9.2	AT5G58950	cytoplasmic: 0.63
CRO_T032849	Yes	Raf	812	89.01	7.11	AT3G06620	cytoplasmic: 0.52
CRO_T015959	No	ZIK	772	88.80	5.61	AT3G04910	cytoplasmic: 0.74
CRO_T002574	Yes	ZIK	708	80.69	5.31	AT3G04910	cytoplasmic: 0.77
CRO_T003421	Yes	ZIK	304	34.70	5.95	AT5G55560	cytoplasmic: 0.77
CRO_T009022	Yes	ZIK	631	71.48	5.48	AT5G58350	nuclear: 0.34
CRO_T015596	Yes	ZIK	411	46.55	5.24	AT3G18750	cytoplasmic: 0.85
CRO_T016097	Yes	ZIK	780	87.40	5.4	AT3G18750	ER: 0.46
CRO_T018206	Yes	ZIK	588	67.58	6.09	AT3G51630	cytoplasmic: 0.76
CRO_T028948	Yes	ZIK	303	34.49	5.48	AT5G55560	cytoplasmic: 0.72
CRO_T033582	Yes	ZIK	589	66.75	5.31	AT5G41990	cytoplasmic: 0.57

Table 4.5. List of MAP4Ks identified in *C. roseus*.

Gene	Full length	Length (amino acids)	Molecular weight (kDa)	Isoelectric point	Best Arabidopsis Match	Subcellular localization:score
CRO_T001030	Yes	696	77.20	6.14	AT5G14720	cytoplasmic: 0.76
CRO_T019613	Yes	682	75.70	5.77	AT1G79640	cytoplasmic: 0.76
CRO_T015367	Yes	507	56.68	5.97	AT4G14480	nuclear: 0.8
CRO_T033821	Yes	883	97.96	6.89	AT4G10730	cytoplasmic: 0.66
CRO_T029825	Yes	853	93.48	5.22	AT1G69220	nuclear: 0.61
CRO_T005088	Yes	847	93.61	8.88	AT3G15220	mitochondrial: 0.84

To investigate the distribution of MAPK cascade genes in the plant kingdom, we chose plant species from different evolutionarily important groups, including monocot, dicot, and primitive plants, such as *Physcomitrella patens* (moss), *Selaginella moellendorffii* (spikemoss) and *Amborella trichopoda*. The results suggest that genes encoding MAPK cascade components are widely present in the plant kingdom and, therefore, may have potential roles in the regulation of various cellular functions (Figure 4.1). Typically, MAP2K gene members constitute the smallest fraction of MAPK cascade (Figure 4.1). For instance, MAP2Ks in Arabidopsis and *C. roseus* is roughly half the number of the MAP1Ks. This observation suggests that MAP2Ks probably activate multiple MAP1Ks downstream, and the cross talk between different signal transduction pathways might be converged at this level in plant MAPK cascades (Ichimura et al., 2002). MAP3K genes transmit every type of upstream signals to the MAPK cascade. MAP3K showed greater variations in structures and domain composition (Yin et al., 2013). This is probably the reason that the MAP3K family constitutes the largest members. In yeast, humans and other mammals, MAP3Ks are activated either through phosphorylation by MAP4Ks (Posas and Saito, 1997; Sells et al., 1997) or by G protein and G protein-coupled receptors (Fanger et al., 1997; Sugden and Clerk, 1997). Several MAP4Ks have been identified in plant genomes based on phylogenetic analyses of their kinase domain. For instance, a MAP4K, MIK, was characterized from *Zea mays* (Wang et al., 2014b), and ScMAP4K1, from *Solanum chacoense*, was reported to play important roles in the ovule, seed, and fruit development (Major et al., 2009).

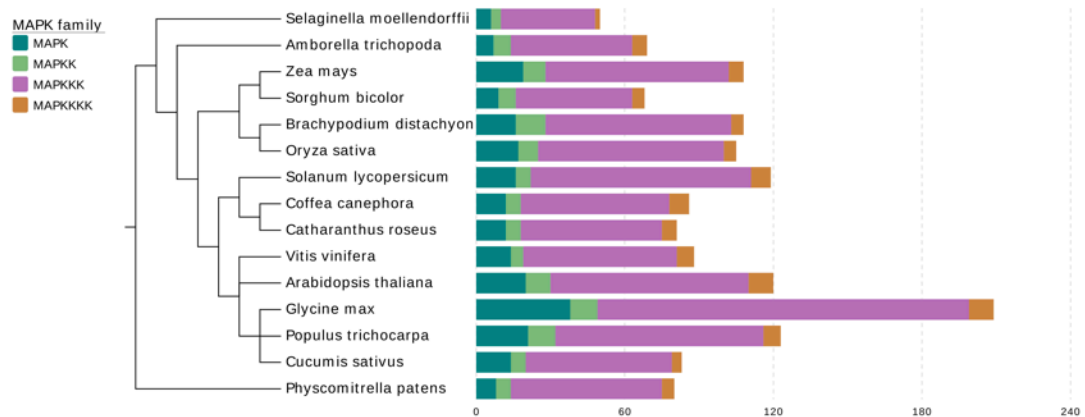


Figure 4.1 Distribution of MAPK cascade kinase genes across the plant kingdom. The total number of MAP1Ks, MAP2Ks, MAP3Ks and MAP4Ks found in *C. roseus*, as well as in other monocot, dicot, *Physcomitrella patens* (moss), *Selaginella moellendorffii* (spikemoss) and *Amborella trichopoda*.

4.3.2 Phylogenetic relationship and conserved domain analysis

To further characterize the MAPK cascade genes, un-rooted phylogenetic trees were produced by aligning the full-length protein sequences of all 12 CrMAP1Ks, 6 CrMAP2Ks, 59 CrMAP3Ks, and 6 CrMAP4Ks using the NJ method. Plant MAP1Ks are commonly grouped into four major subfamilies (A, B, C, and D) based on their phosphorylation motifs and the phylogenetic relationships between amino acid sequences (Ichimura et al., 2002). Similarly, all the 12 CrMAP1Ks were also classified into groups A (2), B (3), C (2) and D (5) (Figure 4.2a). Furthermore, the groups A, B, and C CrMAP1Ks harbor the TEY (Thr-Glu-Tyr) motif at the phosphorylation site, whereas the members of group D contain a TDY (Thr-Asp-Tyr) motif (Table 4.2). The group D MAP1Ks form a distant clade, which is consistent with that described previously in other plants including Arabidopsis and rice (Ichimura et al., 2002; Hamel et al., 2006). All the CrMAP1Ks contained a single Pkinase domain (Figure 4.2c).

Likewise, the six CrMAP2Ks were also divided into four groups, namely, the A (2), B (1), C (2), and D (1) subfamilies, which is consistent with the MAP2Ks in other plant species including *Arabidopsis* and rice (Hamel et al., 2006; Chen et al., 2012; Liang et al., 2013) (Figure 4.3a). Like CrMAP1Ks, all the CrMAP2Ks also contain a single Pkinase domain (Figure 4.3c). Consistent with previous reports, a proline-rich region and extended C-terminal region was also found in the members of group B of CrMAP2Ks subfamily (Wang et al., 2015).

The MAP3K gene family in plants is usually divided into three subfamilies, namely MEKK, Raf, and ZIK (Wrzaczek and Hirt, 2001). Similarly, the 59 CrMAP3K genes in *C. roseus* were clustered into three groups (17 MEKKs, 34 Rafs, and 8 ZIKs) (Figure 4.4a), which is consistent with those reported for other plant species (Rao et al., 2010; Kong et al., 2013a; Sun et al., 2014; Wang et al., 2014a; Wu et al., 2014). All the CrMAP3K proteins have at least a kinase domain, and most of them have a serine/threonine protein kinase active site. In *C. roseus*, most of the CrMAP3Ks in Raf subfamily have a long N-terminal regulatory domain and C-terminal Pkinase_Tyr domain (Figure 4.4c). It has been proposed that the long regulatory domain in the N-terminal region of the Raf subfamily may be involved in protein-protein interactions and regulate or specify the kinase activity (Jouannic et al., 1999). By contrast, the majority of the members of the ZIK and MEKK subfamilies have single Pkinase domain (Figure 4.4a). These results are consistent with the previous findings in *Arabidopsis*, rice, and tomato (Rao et al., 2010; Wu et al., 2014). The CrMAP4K gene family in *C. roseus* formed two major cluster and harbor single Pkinase domain (Figure 4.4a,b).

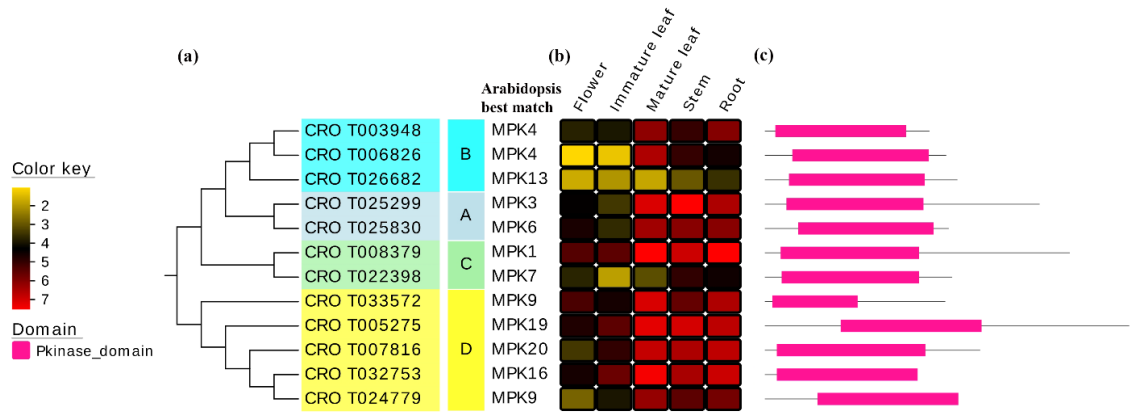


Figure 4.2 Phylogenetic relationships (a), expression profiles in five different tissues (flower, immature leaf, mature leaf, stem and root) (b), and protein structures (c) of CrMAP1Ks in *C. roseus*.

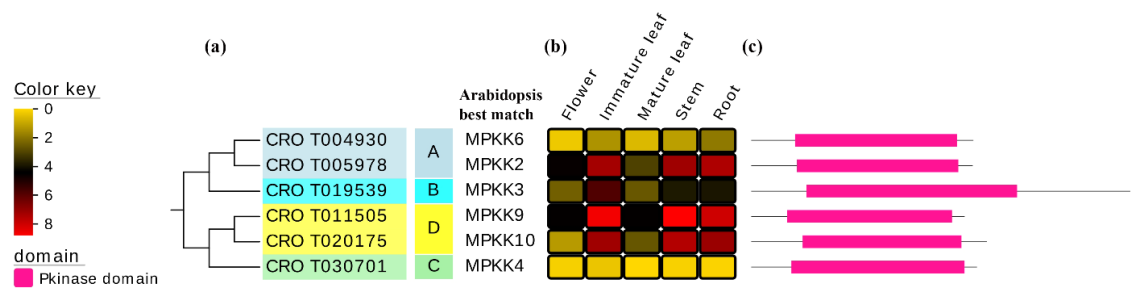


Figure 4.3 Phylogenetic relationships (a), expression profiles in five different tissues including (flower, immature leaf, mature leaf, stem and root) (b), and protein structures (c) of CrMAP2Ks in *C. roseus*.

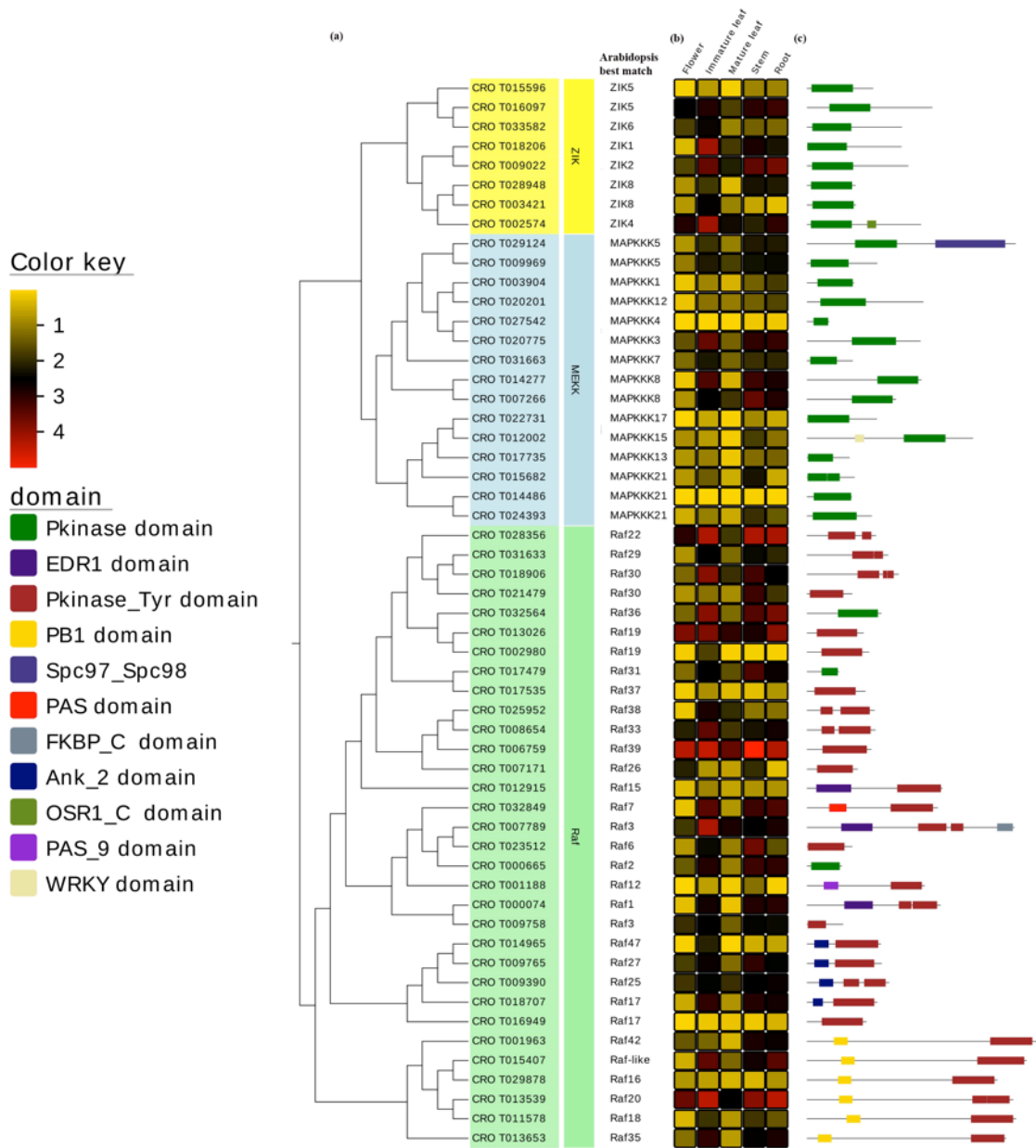


Figure 4.4 Phylogenetic relationships (a), expression profiles in five different tissues (flower, immature leaf, mature leaf, stem and root) (b), and protein structures (c) of CrMAP3Ks in *C. roseus*.

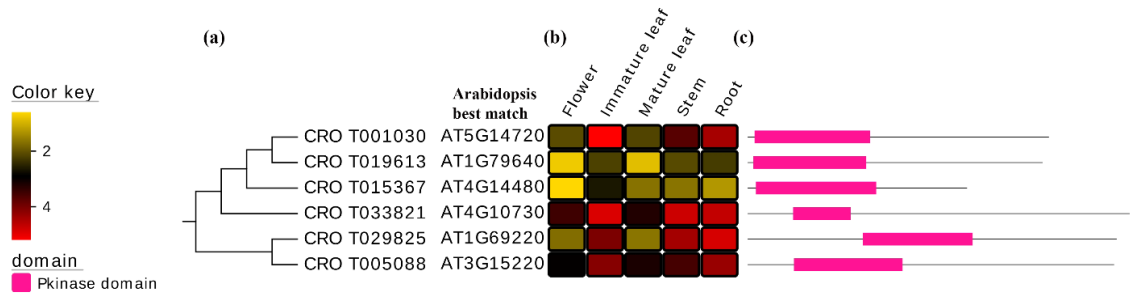


Figure 4.5 Phylogenetic relationships (a), expression profiles in five different tissues (flower, immature leaf, mature leaf, stem and root) (b), and protein structures (c) of CrMAP4Ks in *C. roseus*.

4.3.3 Expression profiles of CrMAP kinase cascade genes in different tissues

To gain insight into the spatial expression patterns and putative functions of CrMAPK cascade genes in *C. roseus*, the expression pattern of CrMAPK cascade genes was investigated using available RNA-seq data (Góngora-Castillo et al., 2012). FPKM values were obtained from five different tissues, including the root, stem, leaf and flower of plants. Based on the log₂-transformed (FPKM + 1) values, expression levels of *CrMAPK* cascade genes were classified as low (≤ 1 FPKM), moderate (≥ 1 FPKM and ≤ 6 FPKM), and high (≥ 6 FPKM) in all tissues as described before (Lu et al., 2015; Shen et al., 2017). The expression levels of these genes are presented in heat-maps (Figure 4.2b, 4.3b, 4.4b and 4.5b). In general, higher numbers of transcripts were detected in mature leaf, stem and root. All the CrMAPK cascade members were expressed in at least one tissue (Figure 4.2b, 4.3b, 4.4b and 4.5b). However, four genes, including CRO_T030701 (CrMP2K), CRO_T014486, CRO_T016949 and CRO_T027542 (CrMP3Ks), showed very low expression in all the tested tissues (Figure 4.2b, 4.3b, 4.4b and 4.5b). Eleven genes, including the homologs of Arabidopsis AtMAPK1, AtMAPK9, AtMAPK19, and AtMAPK16 were found to be highly expressed in all tissues suggesting their vital role in plant growth and development (Figure 4.2b, 4.3b, 4.4b and 4.5b). Some of the members of CrMAP1K (e.g. CRO_T006826), CrMAP2K (e.g. CRO_T019539, CRO_T030701), CrMAP3K (e.g. CRO_T014486, CRO_T015682, CRO_T012002), and CrMAP4K

(CRO_T018206, CRO_T033582, CRO_T002574) families were preferentially expressed in specific tissues (Figure 4.2b, 4.3b, 4.4b and 4.5b), suggesting their function in tissue-specific processes in *C. roseus*.

Previous studies showed that MAPK cascade genes are widely involved in the growth and development of higher plants (Xu and Zhang, 2015). In Arabidopsis, AtMAPK3 and AtMAPK6 play important roles in anther cell differentiation and normal anther lobe formation (Hord et al., 2008). The MAPK cascade AtMAPKK9-AtMAPK6 in Arabidopsis plays an important role in leaf senescence (Zhou et al., 2009). In soybean, MAPK4 and MAPK7 regulate plant growth and development (Shi et al., 2010; Liu et al., 2011). The NPK1 in tobacco has been suggested to regulate cell size, cytokinesis, and plant growth (Soyano et al., 2003). Duan *et al.* identified OsMAPKK4 as a factor for grain size in rice and suggested a possible link between the brassinosteroids (BRs) and MAPK pathways during grain growth (Duan et al., 2014). The spatial expression profile of *C. roseus* MAPKs suggests their involvement in a number of biological processes. Recently, CrMAPKK1, CrMAPK3 and CrMAPK6 are reported to be involved in regulation of TIA in *C. roseus* (Raina et al., 2012; Paul et al., 2017). However, the involvement of MAPK cascades in the regulation of plant growth and development has not been demonstrated in *C. roseus*.

4.3.4 Multiple sequence alignment and motif analysis of CrMAP kinase cascade components

MAPKs from several plant species, including Arabidopsis and *Brassica rapa* (Ichimura et al., 2002; Lu et al., 2015) have been identified based on the presence of specific features, such as the conserved MAPK motifs and the phosphorylation-activation TXY motifs (Cristina et al., 2010). The amino acid sequence alignment of the 12 CrMAP1Ks identified in this study is shown in Figure 4.6a. Most of these CrMAP1Ks contain the conserved motifs, such a catalytic loop (C-loop), phosphate-binding loop (P-loop), activation loop (T-loop), and a T[D/E]Y motif in the activation loop, which are known to be important for function of MAP1Ks. Of the 12 MAP1Ks, five contain the TDY motif and the other seven contain the TEY motif. Furthermore, the N-terminal regions of most of the CrMAP1Ks have a glycine-rich motif (G[R/K]G[A/S]YG) located in the P-loop

region, which acts as an ATP- and GTP-binding site (Saraste et al., 1990) (Figure 4.6a). The conserved C-loop, composed of the HRDLKP[G/S/K]N motif which contains an Asp (D) residue critical for phosphorylation reaction. This signature motif is present in most of the CrMAP1Ks (Figure 4.6a). The MAPK common docking (CD) domain consists of an acidic amino acid cluster (LH[D/E]XX[D/E]EPXC), which establishes electrostatic interactions with upstream factors and functions as a docking site for MAP2Ks, substrates, and negative regulators (Tanoue et al., 2000; Paul et al., 2017). This conserved CD domain, was detected in the extended C-terminal region of the CrMAP1K members belonging to the A and B groups but was not found in members of group C and D CrMAP1Ks (Figure 4.6a).

Multiple sequence alignment of the CrMAP2Ks showed that each CrMAP2K contains the active site motif (D [L/I/V] K), as well as a highly conserved phosphorylation target site (S/T-XXXXX-S/T) within the activation loop. Additionally, a conserved amino acid signature motif, VGTxxYMSPER, is present in the catalytic domain (Figure 4.6b) as previously described in Arabidopsis and other plant MAP2K proteins (Ichimura et al., 2002; Chen et al., 2012; Liang et al., 2013).

As reported in Arabidopsis and other plant MAP3Ks, the MEKK subfamily in *C. roseus* harbors the conserved (G[T/S]Px[W/Y/F]MAPEV) motif. The Raf subfamily in *C. roseus* has the (GTXX[W/Y]MAPE) motif, whereas the ZIK subfamily has the (GTPEFMAPE[L/V]Y) signature motif (Jonak et al., 2002; Rao et al., 2010) (Figure 4.6c). The ZIK proteins have been reported to be involved in circadian rhythm. The ZIK protein, with-no-lysine (K)1 (WNK1) in Arabidopsis is involved in the control of circadian rhythms (Murakami-Kojima et al., 2002). The WNK2/5/8 proteins are found to regulate the Arabidopsis flowering time *via* modulating the photoperiod pathway (Wang et al., 2008). A rice WNK1 was reported to be involved in circadian rhythm, and also differentially responded to diverse abiotic stress responses (Kumar et al., 2011). However, the role of CrZIK proteins in *C. roseus* has yet to be elucidated.

Little is known about the roles of MAP4Ks in plants. Six ORFs showing strong similarity with the Arabidopsis MAP4Ks were identified in *C. roseus* genome (Figure 4.6d). Most

of the CrMAP4K also have similar signature sequence (TFVGGTPXWMAPEV) described previously for Arabidopsis (Champion et al., 2004).

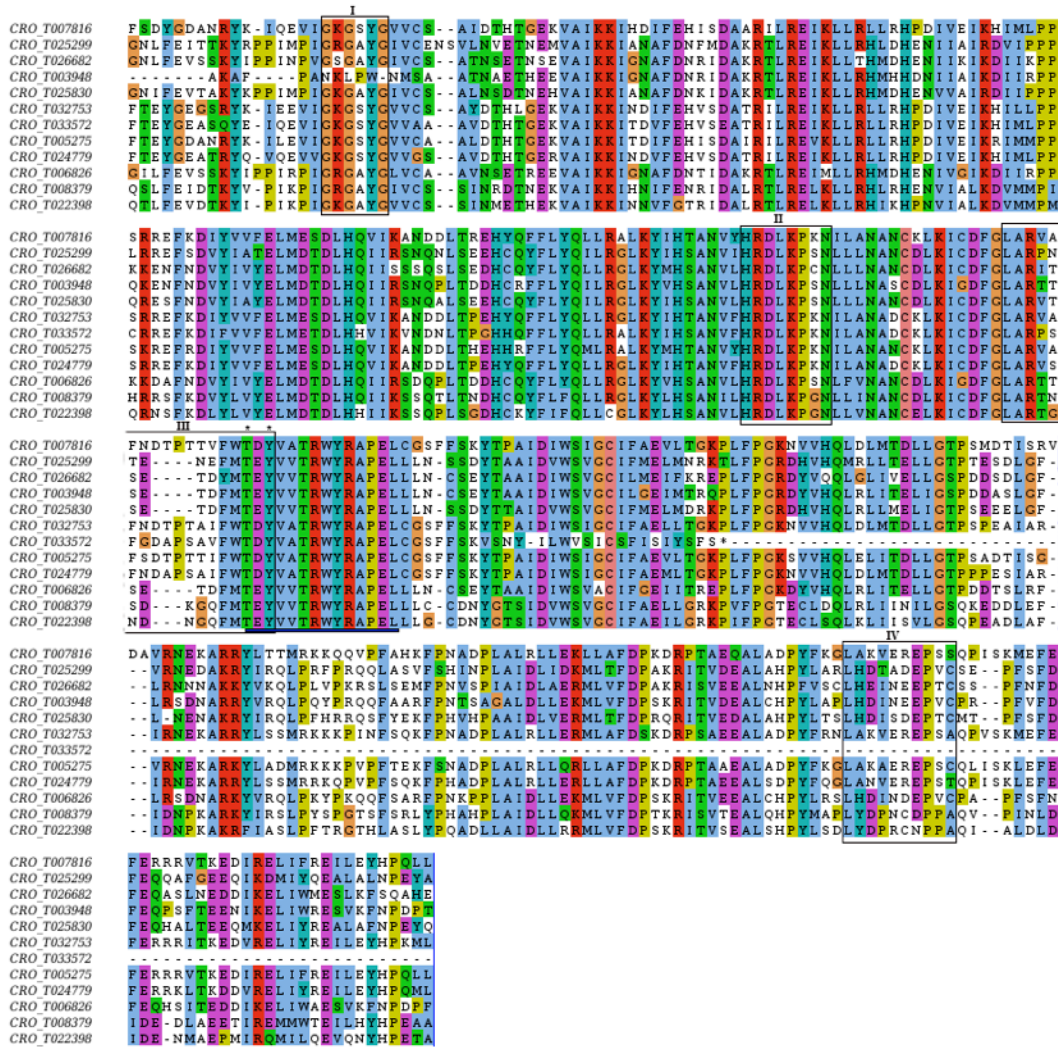


Figure 4.6a Multiple sequence alignment of the 12 CrMAP1K proteins in *C. roseus*.

Multiple sequence alignment was performed using Clustal Omega and displayed using Jalview. Sequence regions with highly conserved and encompassing important motifs are shown. The alignment is color-coded by ClustalX color scheme that integrates amino acid properties and evolutionary conservation. The P-loop, C-loop, activation-loop motifs and common docking (CD) domain are indicated by roman numbers I to IV, respectively.

The conserved motifs described previously (Bailey et al., 2006) are indicated by lines below the alignments. The phosphorylation-activation motif TXY is indicated by asterisks.

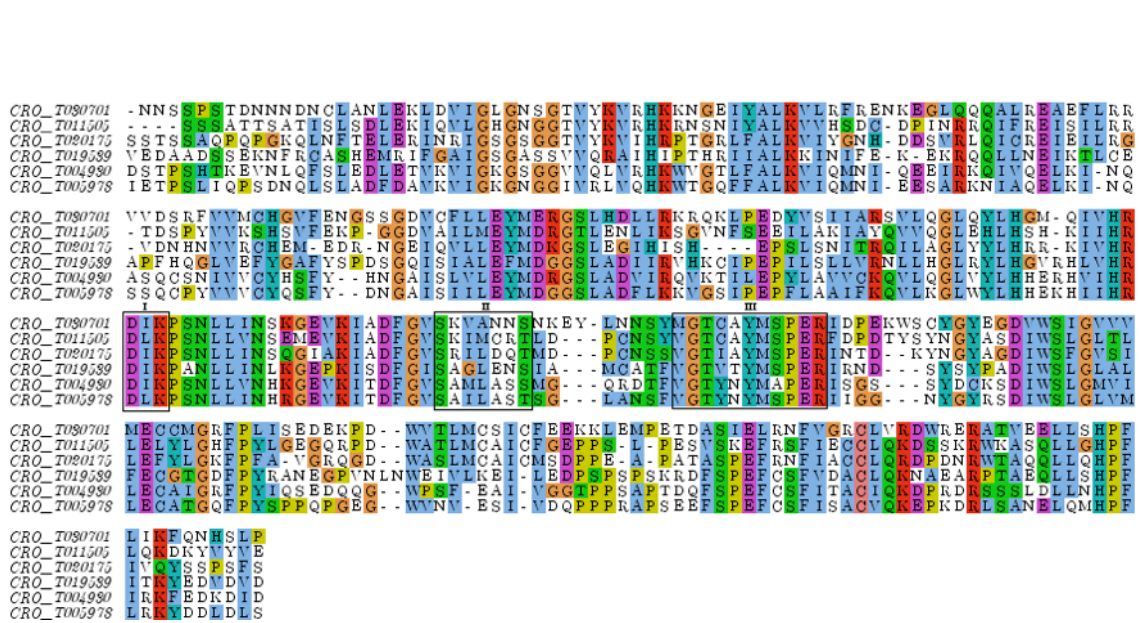


Figure 4.6b Multiple sequence alignment of the 6 CroMAP2K proteins.

Alignment was performed using Clustal Omega and displayed using Jalview. Sequence regions, which are highly conserved and encompassing important motifs are shown here. The alignment is color-coded by ClustalX color scheme that integrates amino acid properties and evolutionary conservation. The active site motif (D [L/I/V] K), conserved phosphorylation target site (S/T-XXXXX-S/T) within the activation loop and signature VGTxxYMSPER motif described previously (Bailey et al., 2006) are indicated by roman numbers I to III, respectively.

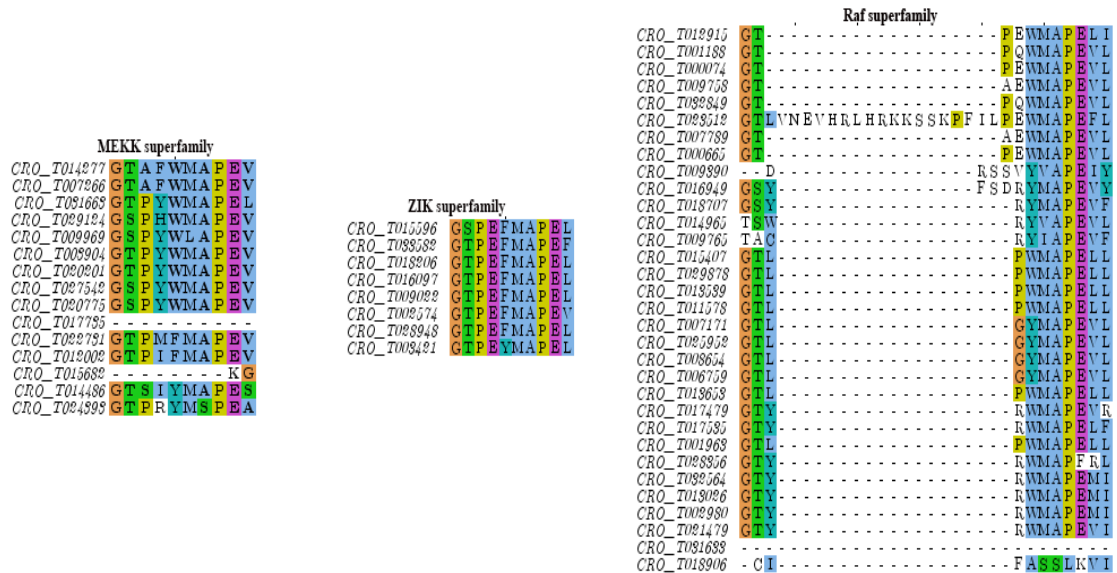


Figure 4.6c Multiple sequence alignment of the CroMAP3K proteins.

Alignment was performed using Clustal Omega and displayed using Jalview. Only sequence regions, which are highly conserved and encompassing important motifs are shown. The alignment is color-coded by a ClustalX color scheme that integrates amino acid properties and evolutionary conservation. The signature motifs described previously (Bailey et al., 2006) are shown.

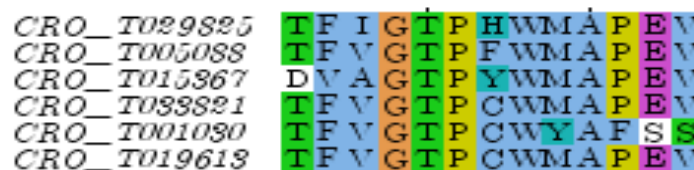


Figure 4.6d Multiple sequence alignment of the conserved motif in CroMAP4K proteins.

Alignment was performed using Clustal Omega and displayed using Jalview. Only sequence regions, which are highly conserved, and encompassing important motifs are shown. The alignment is color-coded by a ClustalX color scheme that integrates amino acid properties and evolutionary conservation. The signature motifs described by Bailey et al., 2006 for MAP4K proteins are shown.

4.3.5 Identification of a CrMAPK cascade, CrMAPKK6-CrMAPK13, involved in the regulation of iridoid branch of TIA pathway

I next performed co-expression analysis of the all CrMAPK members with key TIA pathway structural genes and regulators using *C. roseus* transcriptomes data (Góngora-Castillo et al., 2012). Co-expression analysis revealed that CRO_T026682, CRO_T004930, CRO_T003904 and CRO_T020201 were tightly co-expressed with iridoid pathway genes and TFs, including *G10H*, *IS*, *IO*, *7DLGT*, and *BIS1/BIS2* (Figure 4.7). Of these four candidates, CRO_T026682 and CRO_T004930 are homologous to the Arabidopsis AtMAPK13 and AtMAPKK6, respectively. CRO_T026682 is a member of group B CrMAP1Ks and CRO_T004930 belongs to the group A CrMAP2Ks (Figure 4.2a, 4.3a). On the other hand, CRO_T003904 and CRO_T020201 belong to the MEKK subfamily of *C. roseus* CrMAP3Ks and are homologous to the Arabidopsis AtMAPKKK1 and AtMAPKKK12, respectively (Figure 4.4a). To investigate their biological function in *C. roseus*, I cloned CRO_T026682 and CRO_T004930 and designated them as CrMAPK13 and CrMAPKK6, respectively. CrMAPK13 is 379 amino acid long with a calculated MW of 43.28 kDa. The CrMAPKK6 is 354 amino acid long and calculated MW is 39.77 kDa.

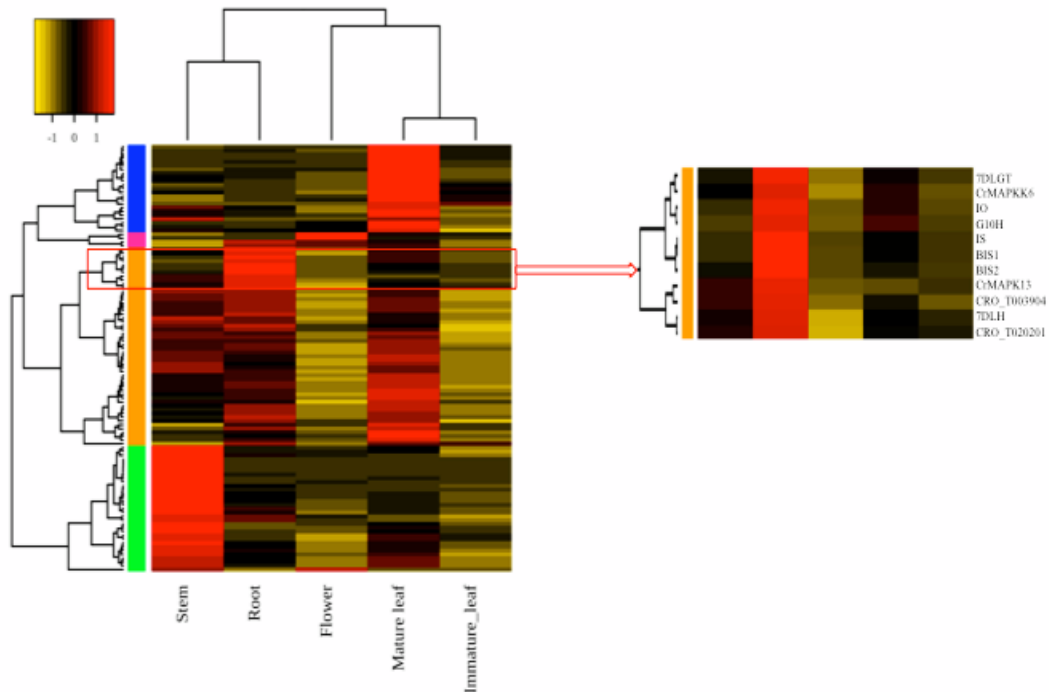


Figure 4.7 Co-expression analyses of *CrMAPK* cascade genes with iridoid pathway regulatory and structural genes. Hierarchical clustering and heat-map show that *CrMAPK13* and *CrMAPK6* are co-expressed with *BIS1* and *BIS2*, and their targets, *G10H*, *IS*, *IO*, *7DLH* and *7DLGT*.

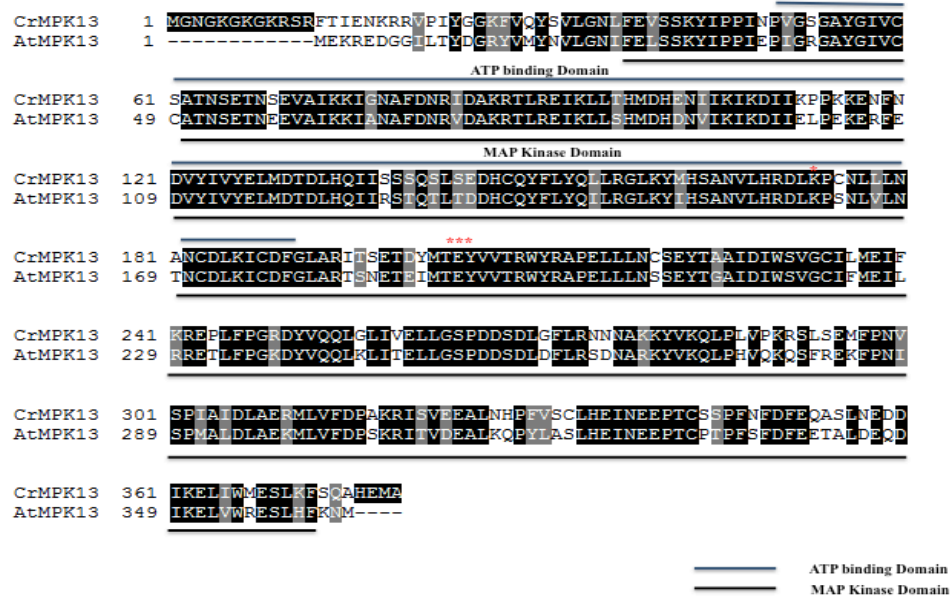


Figure 4.8 Sequence alignments of CrMAPK13 and AtMAPK13. Black colors showing the amino acids, which are identical and gray colors showing the amino acids that are similar among the sequences. ATP-binding and Kinase domains are underlined. The asterisks represent TEY motif in kinase domain.

4.3.6 CrMAPK13 significantly enhances the transactivation of iridoid pathway gene promoters through BIS1

BIS1 is a key regulator of the iridoid branch of the TIA pathway in *C. roseus* (Van Moerkercke et al., 2015). We thus examined whether CrMAPK13 can enhance the activities of BIS1 on iridoid pathway gene promoters. The *IS* or *G10H* promoter-reporter vector was electroporated into protoplasts with various combinations of vectors expressing *BIS1* and *CrMAPK13*. Individually, BIS1 activated the expression of the *IS* or *G10H* promoters; however, co-expression with *CrMAPK13* significantly increased their

activities (Figure 4.9a). To determine whether the enhanced transactivation resulted from CrMAPK13 activity, we substituted CrMAPK13 with the inactivated mutant, CrMAPK13^{K173M}, in the *IS* promoter activation assay. CrMAPK13^{K173M} was unable to enhance the promoter activation by BIS1 (Figure 4.9b). These results suggest that the CrMAPK13 can potentially enhance activation of the BIS1 in TIA biosynthesis.

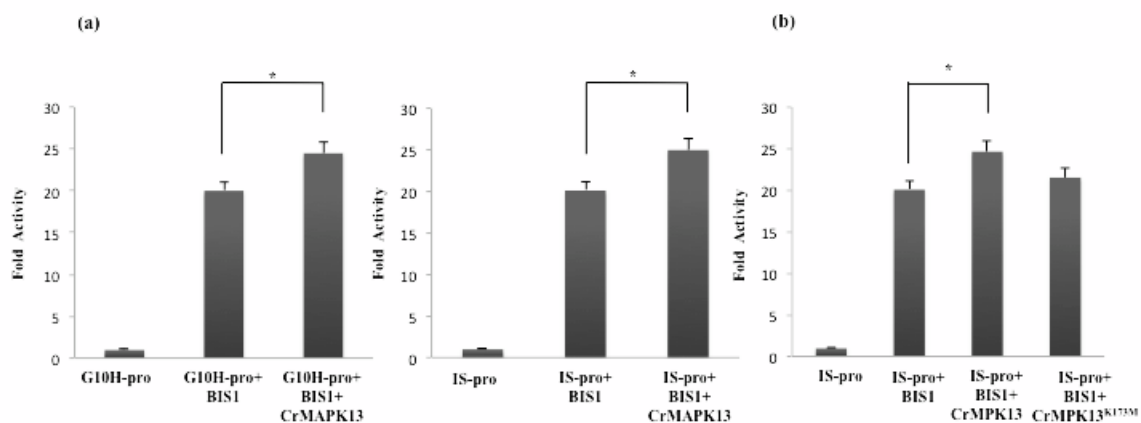


Figure 4.9 Enhancement of the transactivation activity of BIS1 on TIA pathway gene promoters by CrMAPK13. **a.** Transactivation of the *G10H* and *IS* promoters (*G10H*-pro and *IS*-pro, respectively) by BIS1 in the presence or absence of CrMAPK13. **b.** Influence of the inactive mutant, CrMAPK13^{K173M} on transactivation activity of BIS1 on the *IS* promoter. The *IS* or *G10H* promoter, fused to the luciferase reporter, was electroporated alone or with different effectors into tobacco protoplasts. Control represents the promoters alone. A plasmid containing *GUS* reporter, driven by the *CaMV35S* promoter and *rbcS* terminator, was used for normalization. Luciferase and *GUS* activities were measured 20 h after electroporation. Luciferase activity was normalized against *GUS* activity. Data presented here are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05 .

4.3.7 CrMAPK13 likely phosphorylates the amino acids S134 and T113 in BIS1

Next, we used BIS1 to determine whether phosphorylation events are responsible for the enhanced activities of the BIS1 on *IS*. First, we scanned BIS1 for the minimal MAPK phosphorylation motifs (Ser or Thr followed by Pro; SP or TP motif) (Sharrocks et al., 2000; Paul et al., 2017) using the NetPhos 2.0 Server (<http://www.cbs.dtu.dk/services/NetPhos-2.0/>). We found that BIS1 contains two potential MAPK phosphorylation sites, at Ser134 and Thr113. We substituted both Ser134 and Thr113 with Ala independently and in combination by site-directed mutagenesis, and tested the BIS1 mutants (BIS1^{S134A}, BIS1^{T113A}, and BIS1^{S134A/T113A}) for activation of the *IS* promoter in the presence or absence of CrMAPK13. In the protoplast assay, activation of the *IS* promoter by CrMAPK13-BIS1 (WT) was significantly higher than for CrMAPK13-BIS1^{S134A}/BIS1^{T113A}/BIS1^{S134A/T113A} (Figure 4.10), suggesting that CrMAPK13 likely phosphorylates BIS1 and modulates its transactivation activity.

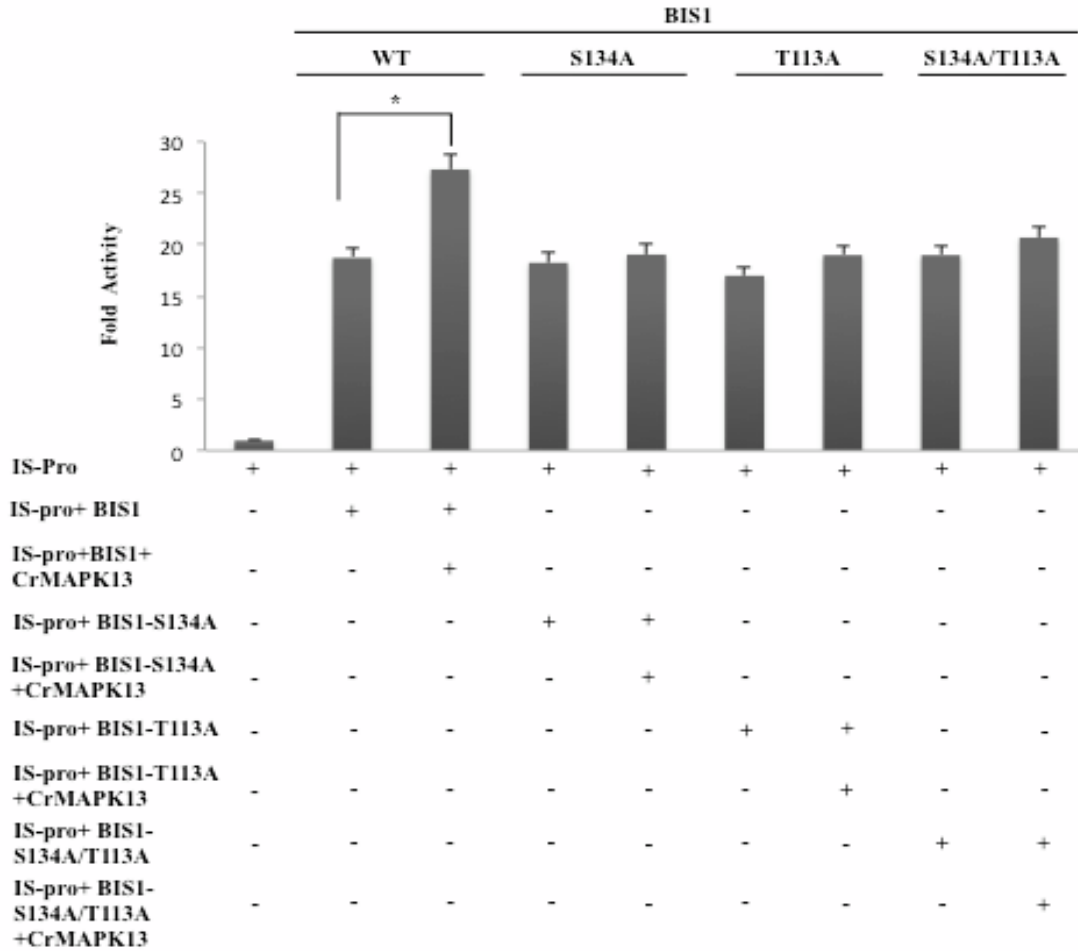


Figure 4.10 CrMAPK13 likely phosphorylates BIS1. The *IS* promoter-reporter vector was co-electroporated with wild-type (WT) BIS1, BIS1^{S134A}, BIS1^{T113A} and BIS1^{S134A/T113A}, with or without CrMAPK13. Luciferase and GUS activities were measured 20 h after electroporation. Luciferase activity was normalized against GUS activity. Data presented here are means ± SDs of three biological replicates. Statistical significance was calculated using the Student’s t-test; * p value <0.05.

4.3.8 CrMAPK13 interacts with BIS1 and CrMAPKK6 in tobacco cells

MAP1Ks have been shown to interact with their substrates as well as upstream MAP2Ks in yeast (Lee et al., 2008; Ding et al., 2009) and plant cells (Paul et al., 2017). To determine whether CrMAPK13 interacts with CrMAPKK6 and BIS1, we used a plant cell-based two-hybrid system (Patra et al., 2013a). Plasmids expressing *CrMAPK13* or

BIS1 fused to the GAL4 DNA binding domain were co-electroporated into tobacco protoplasts with a reporter vector expressing the firefly *luciferase* with 5XGAL response elements (*5XGALRE-35S_{pro}:LUC*). *DB-CrMAPK13* or *DB-BIS1* expression in protoplasts also induced luciferase activity (approximately 3- and 14- fold, respectively, compared to the basal level activity of the reporter plasmid). However, the luciferase activity was significantly increased in protoplasts co-expressing *CrMAPKK6* and *CrMAPK13* (approximately 5- and 16.5- fold, respectively) (Figure 4.11). These results suggest that *CrMAPKK6* interacts with *CrMAPK13*, and *CrMAPK13* interact with and likely phosphorylates *BIS1* in plant cells.

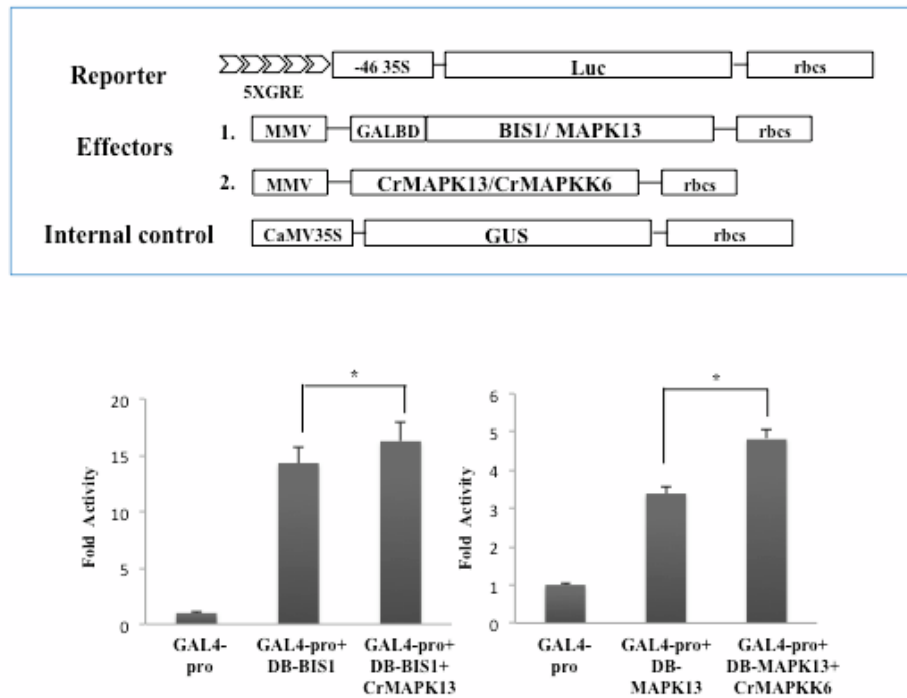


Figure 4.11 Interaction of CrMAPK13 with CrMPKK6 and BIS1. CrMAPK13 and BIS1 fused to the GAL4 DNA binding domain (DB) (amino acid 1–147), were cloned into a modified pBS vector containing the *MMV* promoter and *rbcS* terminator to generate DB-CrMAPK13 and DB-BIS1, respectively. CrMAPKK6 and CrMAPK13 were cloned into a modified pBS vector containing the *CaMV35S* promoter and *rbcS* terminator. Effector plasmids were co-electroporated into tobacco protoplasts with a

reporter vector containing the firefly luciferase gene under the control of a minimal *CaMV35S* promoter and 5XGAL response elements. The control represents the reporter alone. A plasmid containing the GUS reporter was used for normalization. Luciferase activity was normalized against GUS activity. Data presented here are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05.

4.3.9 CrMAPK13 interacts with CrMAPKK6 in yeast cells

We next used yeast two-hybrid (Y2H) assay, to establish interaction between CrMAPK13 and CrMAPKK6. AD-MAPK13 and AD-MAPKK6 constructs were co-transformed into yeast cells with the BD-MAPKK6 and BD-MAPK13 constructs, respectively. Yeast cells containing any one of these two combinations, grew well on both double (SD/-Leu/-Trp) and triple (SD/-His/-Leu/-Trp) selection medium. Yeast cells containing BD-CrMAPK13 or BD-CrMAPKK6 with pAD-GAL4 (AD) did not grow on triple selection medium (Figure 4.12). The Y2H result demonstrated that CrMAPK13 could interact with CrMAPKK6 in yeast cells.

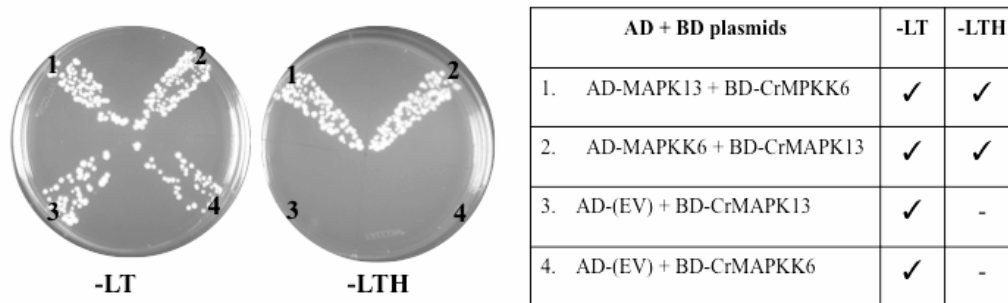


Figure 4.12 Physical interaction between CrMAPK13 and CrMAPKK6 in yeast cell.

AH109 yeast strains were transformed with the corresponding plasmids 1. pAD-CrMAPK13 + pBD-CrMAPKK6, 2. pAD-CrMAPKK6 + pBD-CrMAPK13, 3. pAD (EV) + pBD-CrMAPK13, and 4. pAD (EV) + pBD-CrMAPKK6. The yeast transformants were grown in SD/-Leu/-Trp (-LT) and SD/-His/-Leu/-Trp (-LTH) selection media. EV, empty vector

4.3.10 Ectopic expression of *CrMAPK13* activates key iridoid pathway genes and boosts alkaloid accumulation in *C. roseus* hairy roots

The ability of MAPK13 to enhance transactivation of BIS1 on iridoid pathway gene promoters in tobacco cells (Figure 4.9), prompted us to investigate its effects on gene expression and TIA biosynthesis in *C. roseus*. We generated transgenic *C. roseus* hairy root lines overexpressing *CrMAPK13*. The transgenic status of the hairy roots was verified by PCR (Figure 4.13a), and three transgenic lines (OE-1, -2, and -3) were selected for further analysis. *CrMAPK13* expressions in transgenic lines were approximately 50 to 150-fold higher compared to the EV control (Figure 4.13b). Next, we measured the expression of regulatory and structural genes from the iridoid pathway. The expression of seven BIS1-targeted genes, such as *GES*, *G10H*, *8HGO*, *IS*, *IO*, *7DLGT*, and *7DLH*, were significantly increased in the *CrMAPK13*-overexpression lines compared to the EV control. Interestingly, expressions of both *BIS1* and *BIS2* were also increased significantly in *CrMAPK13* overexpression lines (Figure 4.14a). To determine whether the changes in gene expression were also accompanied by increased TIA accumulation, alkaloids were extracted from independent transgenic hairy root lines and measured using LC-ESI-MS/MS as described previously (Suttipanta et al., 2011; Paul et al., 2017). While accumulation of ajmalicine and catharanthine were not significantly affected, the accumulation of tabersonine was significantly increased in all three transgenic lines compared to the EV control (Figure 4.14b). Taken together, our findings support our hypothesis that the CrMAPKK6-MAPK13 cascade acts upstream of BIS1 to modulate the expression of iridoid pathway genes.

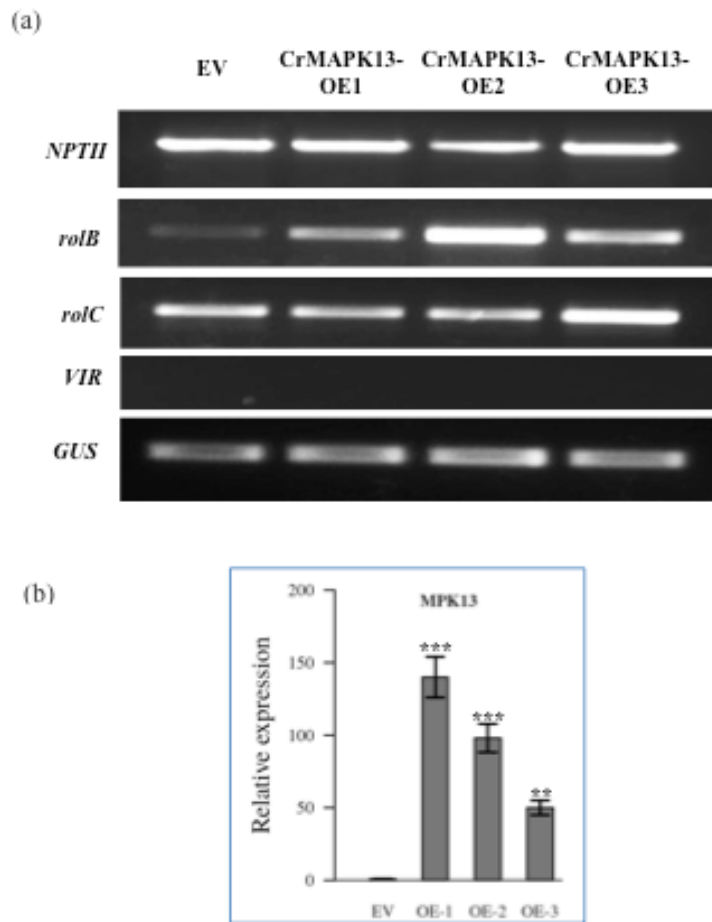
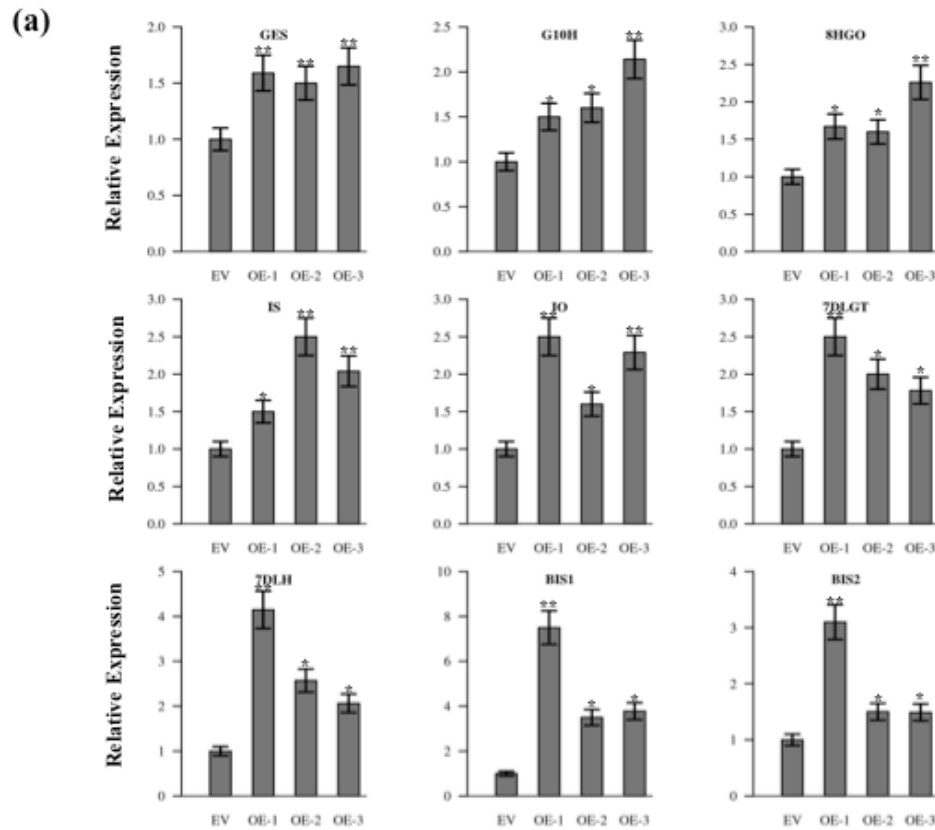


Figure 4.13 PCR confirmation of the transgenic status of *CrMAPK13*-overexpressing *C. roseus* hairy root lines. **a.** Gene-specific primers amplified various fragments in the T-DNA from the Empty Vector (EV) control and *CrMAPK13*-overexpression hairy root lines (OE-1, OE-2, and OE-3). **b.** Relative expression of *CrMAPK13* in empty vector (EV) control and three overexpression (OE-1, OE-2, and OE-3) hairy root lines measured by qRT-PCR. *rolB* and *rolC*, protein-tyrosine phosphatase RolB and C, respectively; *virC*, virulence protein C; EV, empty vector; *EF1 α* , elongation factor 1 α ; *GUS*, β -glucuronidase; *NPTII*, neomycin phosphotransferase II; OE, overexpression



(b)

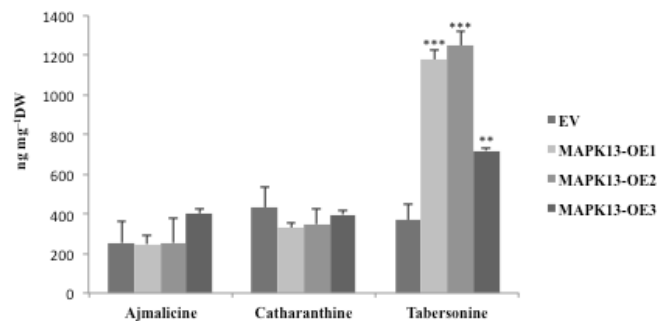


Figure 4.14 Increase of iridoid pathway gene expression and TIA accumulation in *C. roseus* hairy roots overexpressing *CrMAPK13*. a. Relative expression of iridoid pathway genes; *GES*, *G10H*, *8HGO*, *IS*, *IO*, *7DLGT* and *7DLH* in EV control and three *CrMAPK13*-overexpression (OE-1, OE-2, and OE-3) hairy root lines were measured by qRT-PCR. b. Ajmalicine, catharanthine, and tabersonine were extracted from the EV

control and the three *CrMAPK13*-OE lines and analyzed by LC-MS/MS. The levels of ajmalicine, catharanthine, and tabersonine were estimated based on peak areas compared to the standards. Alkaloid levels are indicated in ng/mg dry weight (DW). Data represented here are the means \pm SDs from three biological replicates. Statistical significance was calculated using the Student's t-test, * p value <0.05, ** p value <0.01, and *** p value <0.001.

4.3.11 Identification of a CrMAPK cascade, CrMAPKK1-CrMAPK20, involved in the regulation of indole branch of TIA pathway.

I have previously demonstrated that the CrMAPKK1-CrMAPK3 cascade positively regulates the indole branch of TIA pathway in *C. roseus* (Paul et al., 2017) (Chapter 3). Based on the co-expression analysis of MAPK cascade components in different tissues, including root, stem, flower, immature and mature leaves (Figure 4.15a) we selected two additional MAP1K genes, *CrMAPK19*, and *CrMAPK20*. These two kinases were co-expressed with *CrMAPKK1*. To examine the effects of MeJA on *CrMAPK19*, and *CrMAPK20* expression, a key inducer of TIA biosynthesis (Menke et al., 1999a), *C. roseus* seedlings were treated with 100 μ M MeJA for 2h, and then gene expression was measured in whole seedlings using qRT-PCR. Compared to the control, *CrMAPK19* transcripts did not change significantly in MeJA-treated seedlings. Interestingly, expression of *CrMAPK20* was reduced significantly (about 7-fold) in MeJA-treated seedlings compared with the untreated control (Figure 4.15b). We cloned the *CrMAPK20* (CRO T007816; Figure 4.2a) for further characterization. CrMAPK20, a member of group D CrMAP1K, is 444 amino acid long with calculated MW of 51.76 kDa.

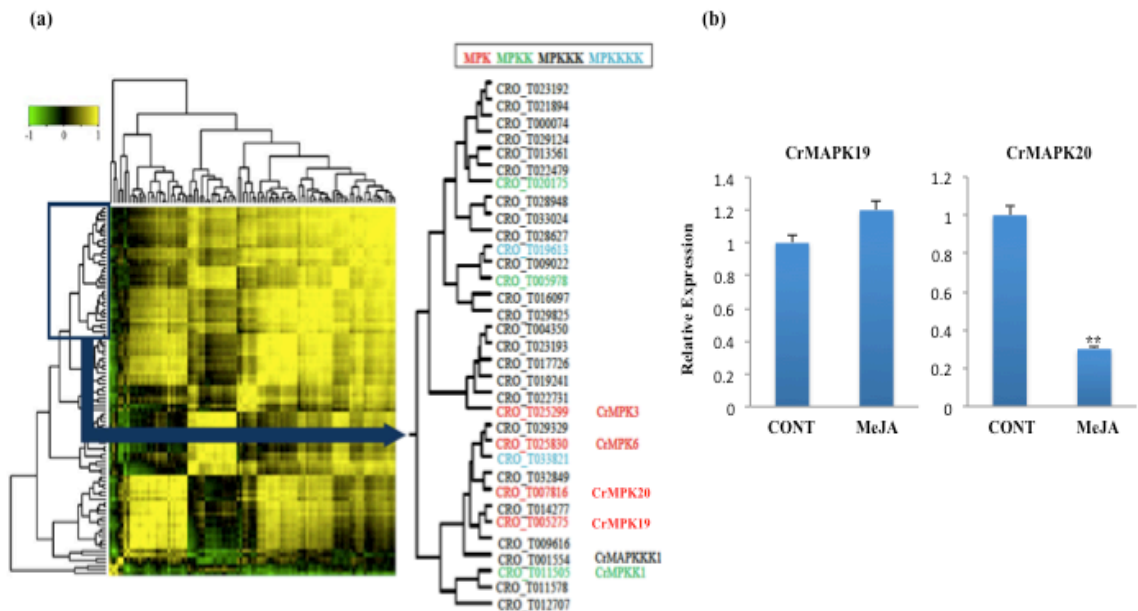


Figure 4.15 Spatial expression analysis of the *C. roseus* MAPK cascade components and repression of *CrMAPK20* by MeJA treatment. **a.** Co-expression heat-map of the *C. roseus* MAPK in different tissues, including roots, stems, immature leaves, mature leaves, and flowers. **b.** Ten-day-old *C. roseus* seedlings were treated with 100 μ M MeJA for 2 h, and relative gene expression was measured by qRT-PCR. Mock-treated seedlings were used as a control. Data represent means \pm SDs of two biological replicates.

4.3.12 *CrMAPK20* significantly represses the transactivation potential of ORCAs on *STR* promoter in protoplasts

We next performed a co-expression analysis of *CrMAPK20* with TIA biosynthetic and regulatory genes. At least two independent transcriptional cascades are known to regulate the TIA pathway genes (Paul et al., 2017). BIS1 and BIS2 regulate the seco-iridoid pathway genes, including *G10H* and *IS* (Van Moerkercke et al., 2015; Van Moerkercke et al., 2016). *CrMYC2* and ORCAs regulate genes in the indole pathway, such as *TDC*, as well as genes downstream in TIAs pathway, such as *LAMT*, *SLS*, and *STR*. The co-expression analysis showed that *CrMAPK20* was clustered with *ASa*, *ZCT3*, *BPF-1*, and

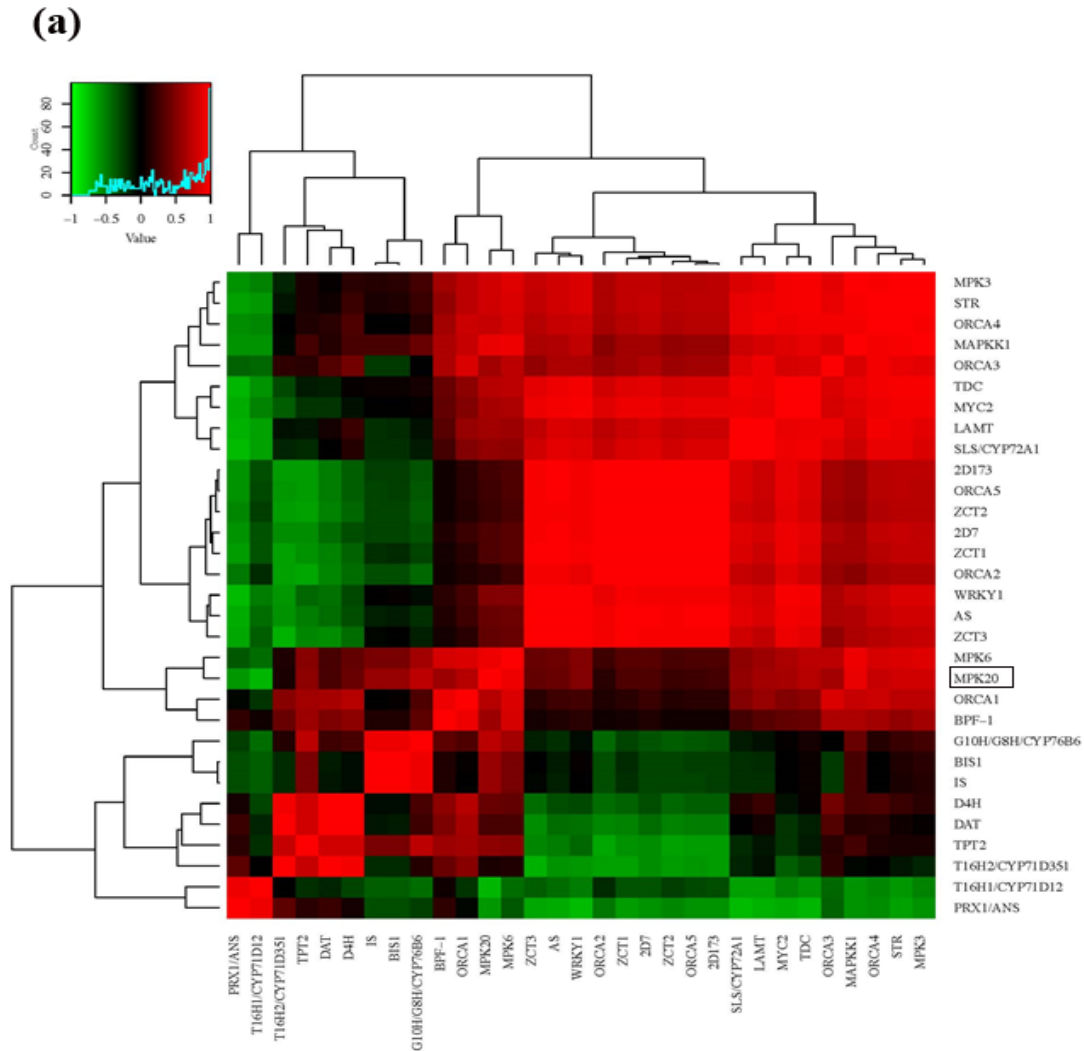
CrMAPK6 (Figure 4.16a), suggesting that *CrMAPK20* may be involved in regulation of the indole branch of the TIA pathway.

To test this hypothesis, the *STR* promoter-reporter vector was electroporated into protoplasts with various combinations of vectors expressing *ORCA3*, *ORCA4*, and *CrMAPK20*. As expected, individually, both TFs activated the expression of the *STR* promoter; however, co-expression with *CrMAPK20* significantly repressed activation of the *STR* promoter (Figure 4.16b). These results suggest that the *CrMAPK20* can repress activation of the ORCA cluster on TIA pathway gene promoter. We next used *ORCA3*, a member of the ORCA cluster, to determine whether phosphorylation events are responsible for the repressed activities ORCAs on *STR* promoter. Using the NetPhos 2.0 server (<http://www.cbs.dtu.dk/services/NetPhos-2.0/>), we previously showed that *ORCA3* contains three putative MAPK phosphorylation sites at Ser60, Ser83 and Thr132. Putative phosphorylation sites were substituted individually with Ala by site-directed mutagenesis. We also showed that the Ser60 and Thr132 sites are likely phosphorylated by *CrMAPK3* (Paul et al., 2017). Here, we evaluated the activity of *ORCA3*^{S83A} on *STR* promoter in the presence or absence of *CrMAPK3* or *CrMAPK20*. In a protoplast assay, *CrMAPK3* significantly induced transactivation activity *ORCA3*^{S83A} on *STR* promoter; however, the activation *STR* promoter by *ORCA3*^{S83A} was not significantly affected by *CrMAPK20* (Figure 4.16c), suggesting that *CrMAPK20* likely phosphorylates *ORCA3* at S83 and modulates its transactivation activity. Differential phosphorylation may be one of the methods for fine-tuning *ORCA3* activation of target promoters. However, more experiments are needed to validate this hypothesis. Because similar phosphorylation sites are also found in *ORCA4* and *ORCA5*, it is assumed that they are also substrates of *CrMAPK20*.

4.3.13 *CrMAPK20* interacts with ORCAs in plant cells

The activation of MAPKs cascade, in response to specific stimuli, is both complex and often redundant. Moreover, a single MAP1K can activate a number of substrates (Lee et al., 2008; Andreasson and Ellis, 2010). MAP1Ks have been shown to interact with their substrates in yeast and plant cells (Lee et al., 2008; Wankhede et al., 2013; Paul et al.,

2017). We speculated that ORCAs are substrates of CrMAPK20. To establish a kinase interaction, we first used a plant cell-based two-hybrid system (Paul et al., 2017). Plasmids expressing *CrMAPK20*, fused to the GAL4 DNA binding domain, were co-electroporated into tobacco protoplasts with a reporter vector expressing the firefly *luciferase* with 5XGAL response elements (*5XGALRE-35S_{pro}:LUC*). CrMAPK20 expression in protoplasts induced luciferase activity (approximately 2-fold, compared to the basal level activity of the reporter plasmid); however, the luciferase activity was significantly increased in protoplasts co-expressing ORCA3/4/5 and CrMAPK20 (approximately 3.5-, 3.25-, and 5.5-fold, respectively) (Figure 4.16d). These results suggest that CrMAPK20 interacts with, and likely phosphorylates, ORCAs in plant cells.



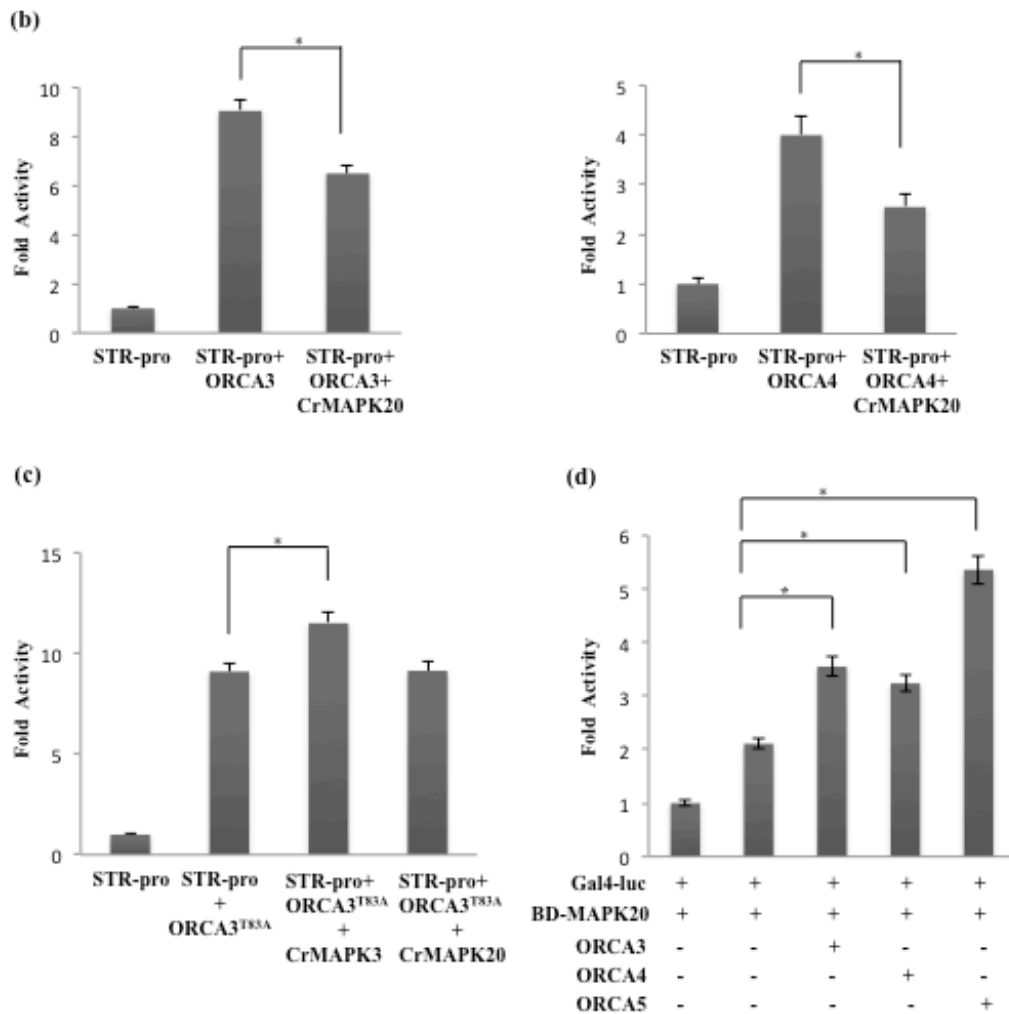


Figure 4.16 Co-expression of *CrMAPK20* with TIA regulatory and structural genes, *CrMAPK20* repression of the ORCAs transactivation activity on TIA pathway gene promoter, phosphorylation of ORCA3 by *CrMAPK20*, and interaction of *CrMAPK20* with ORCAs. **a.** A hierarchical clustering and heat-map show that *CrMAPK20* was co-expressed with *BPF-1*, *ZCT3 AS α* and *CrMAPK6*. **b.** Transactivation of the *STR* promoter (*STR-pro*) by ORCA3 and ORCA4 in the presence or absence of *CrMAPK20*. **c.** The *STR* promoter-reporter vector was co-electroporated with ORCA3 (wt) or the mutant, ORCA3^{T83A} with or without *CrMAPK3*/*CrMAPK20*. **d.** *CrMAPK20*, fused to the GAL4 DNA binding domain, was cloned into a modified pBS vector containing the *MMV* promoter and *rbcS* terminator to generate BD-*CrMAPK20*.

ORCA3/4/5 were cloned into a modified pBS vector containing the *CaMV35S* promoter and *rbcS* terminator. Effector plasmids were co-electroporated into tobacco protoplasts with a reporter vector containing the firefly *luciferase* gene under the control of a minimal *CaMV35S* promoter and 5XGAL response elements. Luciferase activity was normalized against GUS activity. Data presented here are means \pm SDs of three biological replicates. Statistical significance was calculated using the Student's t-test; * p value <0.05 .

4.3.14 MAPK20 interacts with ORCAs and CrMAPKK1 in yeast cell

To further confirm the interaction of CrMAPK20 and CrMAPKK1, ORCA3, ORCA4 and ORCA5, we used Y2H assay. The BD-MAPK20 construct was co-transformed into yeast cells with the AD-CrMAPKK1, AD-ORCA3, AD-ORCA4, and AD-ORCA5 constructs. Yeast cells containing any one of four combinations except ORCA4, grew well on both double (SD/-Leu/-Trp) and triple (SD/-His/-Leu/-Trp) selection medium with 0.5mM 3AT. The negative control, BD-CrMAPK20, and pAD, did not grow on triple selection medium, but grew well in double selection medium (Figure 4.17). The Y2H results demonstrated that CrMAPK20 could interact with CrMAPKK1, ORCA3 and 5 in yeast cells. Therefore, the components of the ORCA cluster are potential substrates of the CrMAPKK1-CrMAPK20 cascade.

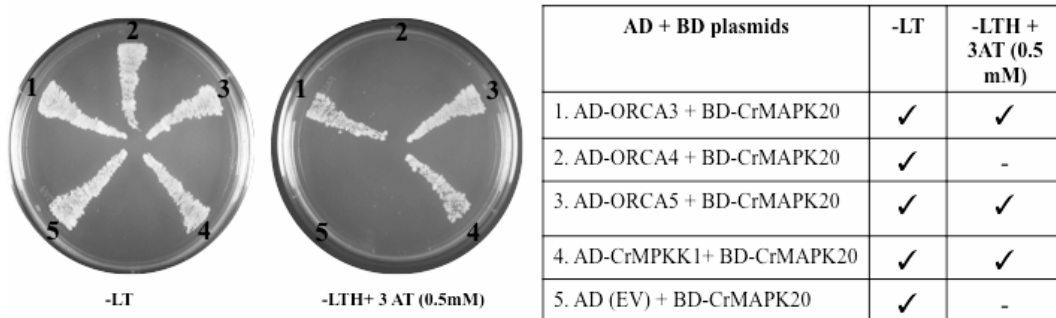


Figure 4.17 Physical interaction of CrMAPK20 with the CrMAPKK1 and ORCAs detected in yeast two-hybrid assay. AH109 yeast strains were transformed with plasmids pAD-CrMAPKK1, ORCA3, ORCA4, or ORCA5 + pBD-CrMAPK20, and

control i.e. pAD (EV) + pBD-CrMAPK20. The yeast transformants were grown in SD/-Leu/-Trp (-LT), and SD/-His/-Leu/-Trp (-LTH) selection media with 0.5mM 3AT (3-amino-1,2,4-triazole). EV, empty vector.

4.3.15 Ectopic expression of *CrMAPK20* represses key indole pathway genes and reduces TIA accumulation in *C. roseus* hairy roots

To further substantiate our findings and to demonstrate the role of CrMAPK20 in TIA production, we generated transgenic *C. roseus* hairy roots overexpressing *CrMAPK20*. The transgenic status of the hairy roots was confirmed by PCR (Figure 4.18a) and two transgenic lines (OE-1 and OE-2) were selected for further analysis. Expression of *CrMAPK20* was 150-200 fold higher in the transgenic lines compared to the empty vector (EV) control (Figure 4.18b). Expression of structural genes, including *TDC*, *LAMT*, *SLS*, *STR*, and *SGD* as well as regulators, such as *ORCAs* and *CrMYC2* were reduced significantly in the *CrMAPK20*-overexpression lines compared with the EV control.

STR catalyzes the condensation of indole and iridoid precursors, such as tryptamine and secologanin respectively to form the core TIA precursor, strictosidine. SGD removes the glucose moiety from strictosidine to form the stictosidine aglycone which subsequently yields cathenamine, the alleged precursor of key TIAs, including catharanthine, tabersonine and ajmalicine. Therefore, to determine the metabolic outcomes of *CrMAPK20*-overexpression, we measured the alkaloids in two independent hairy root lines using HPLC-ESI-MS/MS as described previously (Suttipanta et al., 2011; Paul et al., 2017). Interestingly, the catharanthine accumulation reduced significantly (about 5-25 fold respectively) in *CrMAPK20* hairy root lines compared with EV lines; however, no significant changes were observed for tabersonine and ajmalicine accumulation (Figure 4.18c). Thus, our findings supported our hypothesis that the CrMAPK1-MAPK20 cascade acts upstream of the ORCA cluster to negatively modulate the expression of indole pathway genes and alkaloid accumulation.

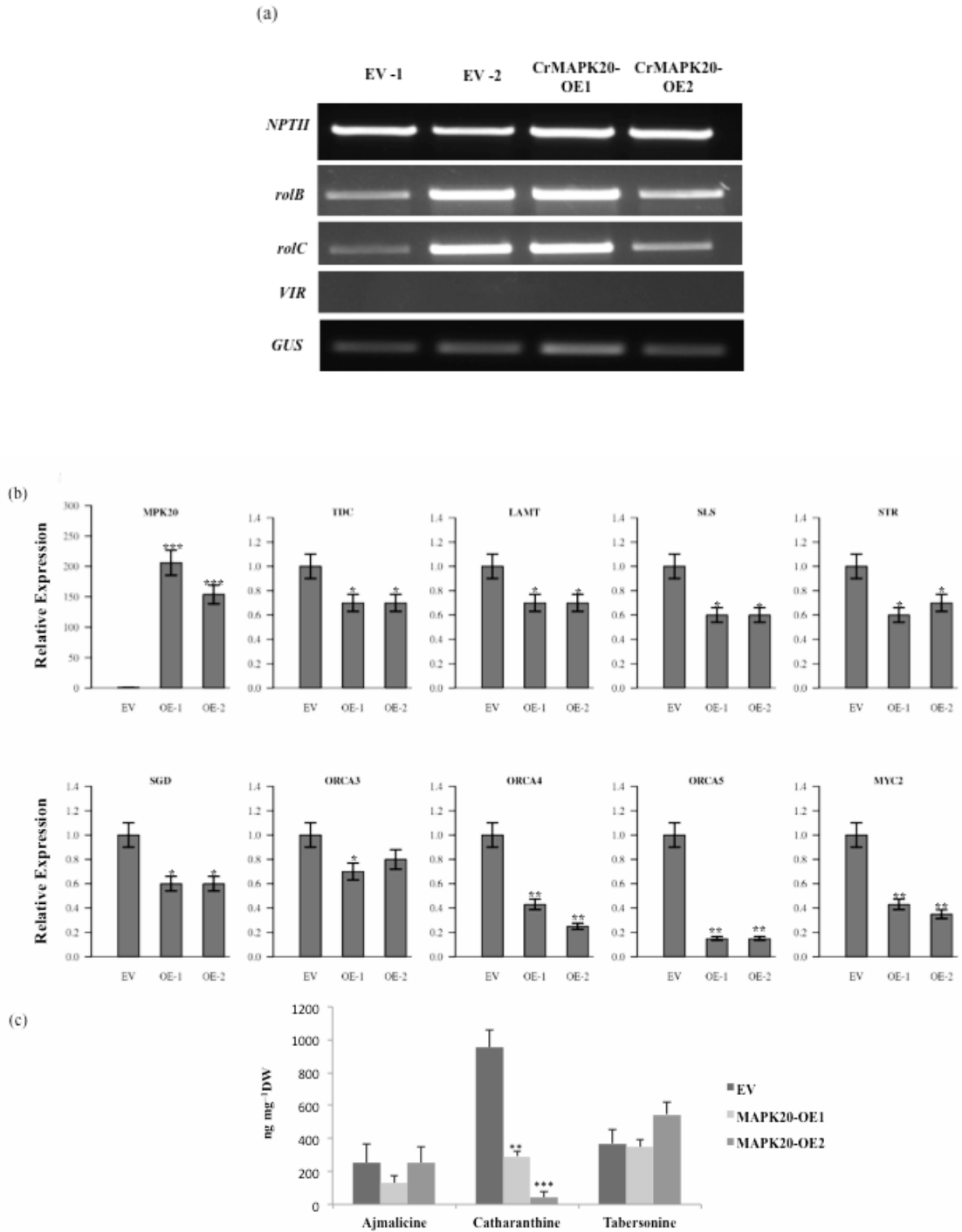


Figure 4.18 Repression of TIA pathway gene expression and alkaloid accumulation in *C. roseus* hairy roots overexpressing *CrMAPK20*. a. PCR confirmation of the transgenic status of *CrMAPK20*-overexpressing *C. roseus* hairy root lines. Gene-specific primers amplified various fragments in the T-DNA from the Empty Vector (EV) control

and *CrMAPK20*-overexpression hairy root lines (OE-1 and OE-2). rolB and rolC, protein-tyrosine phosphatase RolB and C, respectively; virC, virulence protein C; EV, empty vector; *EF1 α* , elongation factor 1 α ; *GUS*, β -glucuronidase; *NPTII*, neomycin phosphotransferase II; OE, overexpression. **b.** Relative expression of *CrMAPK20* and indole pathway genes: *TDC*, *LAMT*, *SLS*, *STR* and *SGD* in EV control and two *CrMAPK20*-overexpression (OE-1 and OE-2) hairy root lines, were measured using qRT-PCR. **c.** Ajmalicine, catharanthine, and tabersonine were extracted from the EV control and the two *CrMAPK20*-OE lines and analyzed by HPLC-MS/MS. The levels of ajmalicine, catharanthine, and tabersonine were estimated based on peak areas compared to the standards. Alkaloid levels are indicated in ng/mg dry weight (DW). Data represented here are the means \pm SDs from two biological replicates. Statistical significance was calculated using the Student's t-test, * p value <0.05, ** p value <0.01, and *** p value <0.001.

4.3.16 CrMAPK20 interacts with CrMAPK3 in yeast cell

We demonstrated that CrMAPK20 interacts with CrMAPKK1 and ORCA gene cluster and negatively affects the TIA biosynthesis. Our previous study showed that CrMAPKK1 interacts with CrMAPK3 and positively regulates the TIA biosynthesis by affecting transactivation activity of *ORCA* gene cluster (Paul et al., 2017). We speculated that CrMAPK20 and CrMAPK3 act as molecular switch to modulate TIA biosynthesis. However, the interconnection between these two kinases and underlying molecular mechanism are not known. A recent report showed the physical interaction between OsMPK20-4 and OsMPK3 in rice and indicated their possible role in stomatal defense (Sheikh et al., 2013). To test whether CrMAPK20 interacts with CrMAPK3, we performed Y2H assay. The BD-MAPK20 construct was co-transformed into yeast cells with the AD-MAPK3. Yeast cells containing BD-MAPK20 and AD-MAPK3, grew on both double (SD/-Leu/-Trp) and triple (SD/-His/-Leu/-Trp) selection medium. However, yeast cell harboring BD-MAPK20 with empty pAD vector did not grow on triple selection medium, but grew well on double selection medium (Figure 4.19). The Y2H result demonstrated that CrMAPK20 could interact with CrMAPK3 in yeast cells.

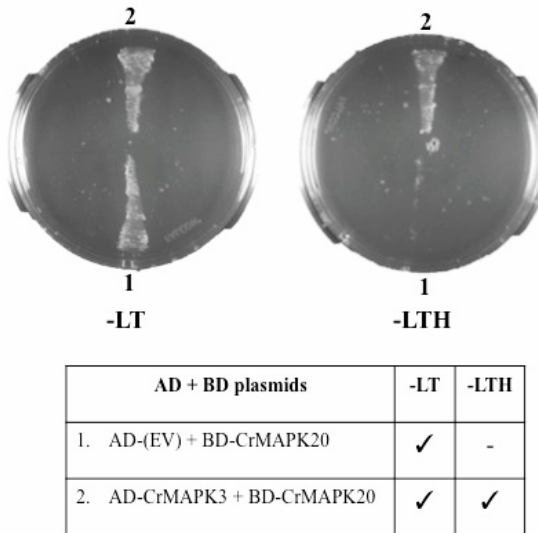


Figure 4.19 Physical interaction of CrMAPK20 and CrMAPK3 detected in yeast two-hybrid assay. AH109 yeast strains were transformed with plasmids pAD-CrMAPK3 + pBD-CrMAPK20 and control, i.e., pAD-GAL4 + pBD-CrMAPK20. The yeast transformants were grown in SD/-Leu/-Trp (-LT), and SD/-His/-Leu/-Trp (-LTH) selection media.

4.4 Conclusion

The posttranslational regulation of TIA pathway has been poorly explored in *C. roseus*. Previous studies speculated the involvement of protein kinases, especially MAPKs, in TIA biosynthesis (Menke et al., 1999a). Moreover, Raina *et al.* have isolated and characterized a MAPK, CrMAPK3 from *C. roseus* (Raina et al., 2012). *CrMAPK3* expression was induced by wounding and UV treatment. In addition, transient overexpression of *CrMAPK3* upregulated TIA biosynthetic pathway genes, such as *TDC*, *STR*, *D4H* and *DAT*, and regulators (*ORCA3*) and resulted in higher TIAs accumulation (Raina et al., 2012). In our previous report, we described a previously uncharacterized MAPK cascade consisting of CrMAPK3/6, CrMAPKK1, and, possibly CrMAPKKK1 that act upstream of a bHLH TF, CrMYC2 and an AP2/ERF gene cluster (*ORCA* gene cluster) to positively regulate the TIA biosynthesis (Paul et al., 2017). We took advantage of the available *C. roseus* draft genome sequences (Kellner et al., 2015) to identify all

possible members of MAPK cascade. Moreover, we performed phylogenetic analysis, domain organization and expression profile in different organs of all the identified MAPKs in *C. roseus*. In addition, we functionally characterized two unique MAPK cascades that regulate two different branches of TIA biosynthesis. The MAPKK6-MAPK13-BIS1 positively regulates the genes of iridoid branch of TIA pathway in *C. roseus*. The involvement of this cascade in TIA biosynthesis was validated by overexpression of CrMAPK13 in *C. roseus* hairy roots. We also identified a negative MAPK cascade comprising of MAPKK1-MAPK20-ORCAs that regulates the indole branch in TIA biosynthesis. As shown in the proposed model (Figure 4.20), (1) the MAPK cascade comprising of CrMAPKK6-CrMAPK13 activates BIS1, which regulates the key iridoid pathway genes, (2) in the presence of MeJA, the CrMAPKK1-CrMAPK3/6 activates the key TFs of indole branch, such as ORCA cluster and CrMYC2, which in turn activates the TIA pathway genes (Paul et al., 2017). However, (3) in the absence of MeJA, a third MAPK cascade consisting of CrMAPKK1-CrMAPK20 was activated and repressed the transactivation activity of *ORCA* gene cluster. CrMAPK20 and CrMAPK3 act as a molecular switch. The physical interaction between the CrMAPK20 and CrMAPK3 was the possible mechanism behind this molecular switch of these two MAPKs; however, a detailed work is needed in future to confirm this unique mechanism. In conclusion this work described a detailed understanding of posttranslational regulation in TIA biosynthesis in *C. roseus*.

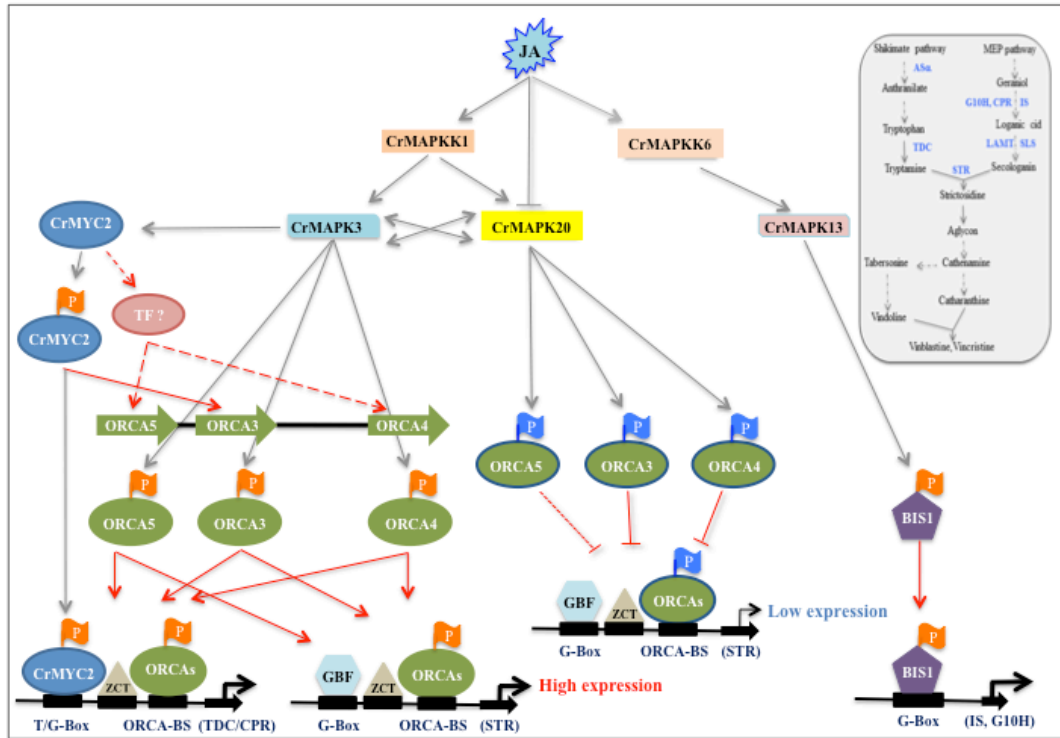


Figure 4.20 A simplified model depicting the posttranslational regulation of the *ORCA* gene cluster and TIA pathway genes. Upon perception of jasmonic acid (JA), (1) the MAP kinase cascade, consisting of CrMAPKKK1, CrMAPKK1 and CrMAPK3/6, phosphorylates CrMYC2 and the *ORCA* gene cluster (gray arrows). Phosphorylated CrMYC2 (labeled by orange flags) activates ORCA3 directly and ORCA4 and ORCA5 indirectly. The phosphorylated ORCA3/4/5 (labeled by orange flags) activate subsets of TIA pathway genes (in parentheses). (2) The MAP kinase cascade, consisting of CrMAPKK6 and CrMAPK13, phosphorylates BIS1 (gray arrows). Phosphorylated BIS1 (labeled by orange flags), by binding to the G-Box activates *IS* and *G10H* gene promoters. (3) In the absence of jasmonic acid (JA), the MAP kinase cascade consisting of CrMAPKK1 and CrMAPK20, phosphorylates the *ORCA* gene cluster (gray arrows). The phosphorylated ORCA3/4/5 (labeled by blue flags and blue parentheses) represses genes in indole branch of TIA pathway (in parentheses). CrMAPK3 interacts and possibly represses the CrMAPK20. Solid red arrows indicate direct activation; dashed arrows represent indirect or undetermined activation. Solid T-bars represents direct repression;

dashed T-bars indicate indirect or unknown repression. The X- indicates protein-protein interaction. Inset: a simplified TIA pathway with key genes highlighted in blue.

Copyright © Priyanka Paul 2017

Chapter 5

Summary and Future directions

Summary: *Catharanthus roseus* (a.k.a Madagascar periwinkle), of the family Apocynaceae, produces over a hundred terpenoid indole alkaloids (TIAs) including, ajmalicine, serpentine, vinblastine, vincristine, which are used in the treatment of hypertension and cancer (Kellner et al., 2015). TIAs are produced in a very limited amount in plants. Although a significant progress has been made in isolation and functional characterization of genes encoding major biosynthetic enzymes in the TIA pathway (Liu et al., 2007; Asada et al., 2013; Besseau et al., 2013; Salim et al., 2014; Qu et al., 2015), knowledge on the transcriptional and posttranslational regulation of TIA pathway is still fragmented. The current knowledge on TIA biosynthesis and its regulation, including transcriptional and posttranslational regulation, are discussed in details in **Chapter 1**.

Previous studies showed that the TIA pathway is regulated by three major class of transcription factors (TFs), such as AP2/ERFs (ORCA2 and 3) (Menke et al., 1999b; van der Fits and Memelink, 2000), bHLH (CrMYC2, BIS1 and BIS2) (Zhang et al., 2011; Van Moerkercke et al., 2015; Van Moerkercke et al., 2016), and WRKY (CrWRKY1) (Suttipanta et al., 2011) TFs. Recently, analysis of *C. roseus* draft genome sequence revealed that the JA-responsive AP2/ERF TF, ORCA3 forms a physical cluster with two uncharacterized AP2/ERFs, AP2 TF1 030272 and AP2 TF2 030274 (Kellner et al., 2015). In plants, physically linked clusters of TFs are less characterized. Moreover, the regulation of TF clusters is relatively unexplored. In **Chapter 2**, my research uncovered that the *ORCA* gene cluster is differentially regulated. ORCA4 and 5, though functionally overlapping with ORCA3, regulate an additional set of TIA pathway genes. *ORCA4* or *ORCA5* overexpression has resulted in significant increase of TIA accumulation in *C. roseus* hairy roots. In addition, ORCA5 directly regulates the expression of *ORCA4* and indirectly *ORCA3*, likely via an unknown factor(s). ORCA3 and ORCA5 bind to a GC-rich element in *ORCA4* promoter. Interestingly, ORCA5 also activates the expression of *ZCT3*, a negative regulator of the TIA pathway. In addition CrMYC2 shows dual

regulation: directly activating *ORCA3* and co-regulating pathway genes concomitantly with *ORCA3*.

Previous studies suggest that biosynthesis of TIAs is also regulated at the posttranslational level, including protein phosphorylation (Menke et al., 1999a; Van Der Fits and Memelink, 2001). A recent report also indicates that a mitogen-activated protein kinase (MAPK) is involved in this process (Raina et al., 2012). In **Chapter 3**, my research revealed that both the *ORCA* gene cluster and *CrMYC2* act downstream of a MAP kinase cascade consisting of *CrMAPKK1*, *CrMAPK3* and 6. Overexpression of *CrMAPKK1* in *C. roseus* hairy roots upregulates key TIA pathway genes expressions and boosts TIA accumulation.

Taking advantage of the *C. roseus* draft genome sequence, I identified all the putative MAPK members and determined their phylogenetic classification, domain organization and expression patterns in different organs in *C. roseus* in **Chapter 4**. Moreover, I functionally characterized two previously uncharacterized MAPK cascades that regulate two main branches of the TIA pathway: *CrMAPKK6-CrMAPK13* regulates the iridoid branch of the TIA pathway and overexpression of *CrMAPK13* in *C. roseus* hairy roots significantly up-regulates iridoid pathway genes and increases tabersonine accumulation. The other MAPK cascade, *CrMAPKK1-CrMAPK20* negatively regulates the indole branch of the TIA pathway. *CrMAPK20* overexpression resulted in repression of *CrMYC2-ORCA*s regulated genes and reduced catharanthine accumulation in *CrMAPK20* overexpressing hairy roots.

In conclusion, the overall findings of my research project significantly advance our understanding of the transcriptional and post-translational regulatory mechanisms that govern TIA biosynthesis in *C. roseus*. Moreover, this knowledge could be utilized to engineer *in vivo* or heterologous systems to produce these valuable TIAs for the benefits of mankind.

Future direction 1: Protein phosphorylation and de-phosphorylation facilitated by protein kinases (PKs) and protein phosphatases, respectively is a common signal

transduction mechanism in a number of biological processes. In contrast to PKs, protein phosphatases show more diversity and are divided into two major classes: protein tyrosine phosphatases (PTPs) and protein serine (S)/threonine (T) phosphatases. The PTP family includes tyrosine-specific and dual-specificity phosphatases (DSPTPs) (Denu et al., 1996; Schweighofer et al., 2004). The protein serine/threonine phosphatases, on the other hand, are classified into the protein phosphatases P (PPP) and protein phosphatases M (PPM) gene families (Schweighofer et al., 2004). The PPP family includes type 1 phosphatases (PP1), and type 2 phosphatases that can be further divided into PP2A and PP2B. The PPM family contains PP2C and pyruvate dehydrogenase phosphatase (Cohen, 1997). In *Arabidopsis*, 76 genes encoding PP2C-type phosphatases have been identified. These PP2Cs are again classified into 10 groups (A–J) (Kerk et al., 2002). A previous report suggested that in *Arabidopsis* a type 2C Ser/Thr phosphatase, AtPP2C1, changes the levels of JA and ethylene during wounding, negatively modulates MAPK4 and MAPK6 to regulate the innate immunity. Therefore, AtPP2C1 is also considered as MAPK phosphatases (MKP) (Schweighofer et al., 2007).

A potential homolog of AtPP2C1 was identified from *C. roseus* and designated here as CrPP2C1. I cloned the CrPP2C1 that shares 51.74% amino acid sequence identity with AtPP2C1 (Figure 5.1). To experimentally verify the involvement of CrPP2C1 in TIA biosynthesis, I used a protoplast-based transactivation system (Pattanaik et al., 2010b). The *STR* promoters, fused to a firefly *luciferase* reporter gene, was co-electroporated into tobacco protoplasts with/without the constructs expressing *ORCA3* and/or *CrPP2C1*. As expected *ORCA3* activated the *STR* promoter; however the activation was significantly reduced when *CrPP2C1* was co-expressed with *ORCA3* (Figure 5.2). Thus, CrPP2C1 significantly reduces the transactivation potential of *ORCA3* on *STR* promoter in a plant cell. More experimental evidence are needed in future to demonstrate whether CrPP2C1 can dephosphorylate the CrMAPKK1-CrMAPK3/6 cascade to regulate the TIA biosynthesis in *C. roseus*.

CrPP2C1	1	MSEITVAPISNSPVFSPSNRV...STSLFCKTSSSSASASTSTSSPETLKNPVSPS...SP
AtPP2C1	1	MSCSVAIVCNSPVFSPPSSSLFCNKSSIISSPQESLSLTLSHRKPQT.SSPSSPSTTVGSSP
CrPP2C1	54	LRLFRITOKLPHSPSGLIRSSGGDSLGCDSVLASTSTSPTVLKRKRPARLDIPIVSMGFGN
AtPP2C1	59	KSPFRLRFQKPPSGF..APGFLSFGSESVSAS.SPPGGVLKRRPTRLDDIPIGVAGVVA
CrPP2C1	114	IPATPFAGSAREREEMFEVE...GDGYAVYCKKGRREAMEDRYSAVVDLOGEISKQVCFGIF
AtPP2C1	115	PISSAAVAATPREECREVEREGDGYSVYCKRGRREAMEDRESAITNLHGDRKQALFGVY
CrPP2C1	172	DHGGAFAAFAAKNLNKNINELE.KMENDDIKEAVKNGYLDTDSEFLKQQL.GGSCS
AtPP2C1	175	DHGGAFAAFAAKNLNKNIVEVVGKRESEIAEAVKHGYLATDASFLKEEDVKGSSC
CrPP2C1	230	VTAVIRKGNLIISNAGDCRAVVSRRGVAEALTSDFRPSRVDEKDRIEALGGYVCCFNGVW
AtPP2C1	235	VTALVNEGNLVVSNAGDCRAVMSVGGVAKALSSDHRPSRDDERKRIETTGGYVDTFHGVW
CrPP2C1	290	RVHGSLAVSRGIGDQYKQWIAEPDTRILDLNPELEFLILASDGLWDKVSNOEADIVR
AtPP2C1	295	RIQGLAVSRGIGDAQLKQWIAEPETKISRIEHDHEFLILASDGLWDKVSNOEAVDIAR
CrPP2C1	350	PMLTHLGRF.....
AtPP2C1	355	PLCLGTEKPLLLAACKKLVLDLSASRGSSDDISVMLIPLRQFI

Figure 5.1. Multiple sequence alignment of CrPP2C1 with AtPP2C1 (AP2C1). The identical amino acid residues in the alignment are in black shade and similar residues are in the gray shade.

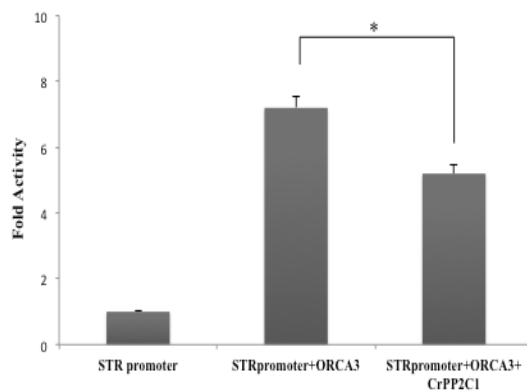


Figure 5.2. CrPP2C1 significantly reduces the transactivation potential of ORCA3 on *STR* promoter in a protoplast assay.

Future direction 2: The posttranscriptional modification also gives another layer of complexity to the regulation of secondary metabolism, including TIAs biosynthesis in *C. roseus* (Patra et al., 2013a; Shen et al., 2017). miRNAs, 21 to 24 nucleotides long, are a major class of endogenous non-coding small regulatory RNAs that are ubiquitous to plants (Bartel, 2009). In plants, miRNAs are involved in a number of biological processes, such as hormone homeostasis (Meng et al., 2009), embryogenesis (Lin and Lai, 2013; Zhai et al., 2014), root (Meng et al., 2010; Khan et al., 2011), and flower development (Wang et al., 2012). However, the role of miRNA in the regulation of secondary metabolism is less studied (Gou et al., 2011; Jia et al., 2015; Sharma et al., 2016). Moreover, the regulatory roles miRNAs in TIAs biosynthesis in *C. roseus* is relatively unexplored. In our previous studies, we identified 181 conserved and 173 novel cro-miRNAs in *C. roseus* seedlings by using small RNA-sequencing. Interestingly, genome-wide expression analysis and qRT-PCR validation identified a set of cro-miRNAs (cro-miR393d, cro-miR164a, cro-miR164b, cro-miR168a etc.), which are differentially regulated in response to MeJA, the key inducer of TIA biosynthesis (Shen et al., 2017). Identification and functional characterization of candidate cro-miRNA(s), which directly regulate(s) the structural and/or regulatory genes in TIA pathway will be promising future aspects in TIA regulation studies.

Copyright © Priyanka Paul 2017

APPENDIX-A

LIST OF ABBREVIATIONS

Acronym/abbreviation	Expansion
ACC	1-aminocyclopropane-1-carboxylic acid
cDNA	Complementary DNA
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribo nucleic triphosphate
DTT	Dithiothreitol
DW	Dry weight
EDTA	Ethylene diamine tetra acetic acid
EF1 ∞	Elongation Factor 1 ∞
eGFP	Enhanced green fluorescent protein
EMSA	Electrophoretic mobility shift assays
g/mg/ μ g/ng	Gram/ milligram/ microgram/ nanogram
GFP	Green fluorescent protein
GUS	β -glucuronidase
h/min/sec	Hours/minutes/seconds
HMM	Hidden Markov Models
HPLC	High performance liquid chromatography
JA	Jasmonic acid
kDa	Kilo Dalton
L/mL/ μ L	Liter/ milliliter/ microliter
LC-ESI-MS/MS	Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometric
M/mM/ μ M	Molar/millimolar/ micromolar
MeJA	Methyl jasmonic acid
MS	Murashige and skoog
MS media	Murashige & Skoog media
NaCl	Sodium chloride
NaOH	Sodium hydroxide
$^{\circ}$ C	Degree centigrade
PBS	Phosphate buffered saline
pBS	pBlueScript
PCR	Polymerase chain reaction
qRT-PCR	Quantitative real-time PCR
RNA	Ribonucleic acid
RPKM	Reads per kilobase of transcript per million mapped reads
RPS9	40S Ribosomal Protein S9
SD	Standard deviation
SE	Standard error
SRA	Sequence read archive
TBE	Tris- borate/ EDTA electrophoresis buffer
TE	TRIS-EDTA

TRIS	Hydroxymethyl Aminomethane
var.	Variety
Wt	Wild-type
μM	Micron meter

References

- Agrawal, G.K., Agrawal, S.K., Shibato, J., Iwahashi, H., and Rakwal, R.** (2003). Novel rice MAP kinases OsMSRMK3 and OsWJUMK1 involved in encountering diverse environmental stresses and developmental regulation. *Biochemical and biophysical research communications* **300**, 775-783.
- Alves, M.S., Dadalto, S.P., Gonçalves, A.B., De Souza, G.B., Barros, V.A., and Fietto, L.G.** (2013). Plant bZIP Transcription Factors Responsive to Pathogens: A Review. *International Journal of Molecular Sciences* **14**, 7815-7828.
- An, J.-P., Qu, F.-J., Yao, J.-F., Wang, X.-N., You, C.-X., Wang, X.-F., and Hao, Y.-J.** (2017). The bZIP transcription factor MdHY5 regulates anthocyanin accumulation and nitrate assimilation in apple **4**, 17023.
- Andreasson, E., and Ellis, B.** (2010). Convergence and specificity in the Arabidopsis MAPK nexus. *Trends in plant science* **15**, 106-113.
- Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H., and De Luca, V.** (2013). A 7-deoxyloganetic acid glucosyltransferase contributes a key step in Secologanin Biosynthesis in Madagascar periwinkle. *The Plant Cell Online* **25**, 4123-4134.
- Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.-L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., and Sheen, J.** (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature* **415**, 977-983.
- Atchley, W.R., and Fitch, W.M.** (1997). A natural classification of the basic helix-loop-helix class of transcription factors. *Proceedings of the National Academy of Sciences* **94**, 5172-5176.
- Bailey, T.L., Williams, N., Misleh, C., and Li, W.W.** (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* **34**.
- Bartel, D.P.** (2009). MicroRNAs: target recognition and regulatory functions. *cell* **136**, 215-233.
- Berriri, S., Garcia, A.V., dit Frey, N.F., Rozhon, W., Pateyron, S., Leonhardt, N., Montillet, J.-L., Leung, J., Hirt, H., and Colcombet, J.** (2012). Constitutively active mitogen-activated protein kinase versions reveal functions of Arabidopsis MPK4 in pathogen defense signaling. *The Plant cell* **24**, 4281-4293.
- Besseau, S., Kellner, F., Lanoue, A., Thamm, A.M., Salim, V., Schneider, B., Geu-Flores, F., Höfer, R., Guirimand, G., and Guihur, A.** (2013). A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. *Plant physiology* **163**, 1792-1803.
- Blom, T.J.M., Sierra, M., van Vliet, T.B., Franke-van Dijk, M.E.I., de Koning, P., van Iren, F., Verpoorte, R., and Libbenga, K.R.** (1991). Uptake and accumulation of ajmalicine into isolated vacuoles of cultured cells of *Catharanthus roseus* (L.) G. Don. and its conversion into serpentine. *Planta* **183**, 170-177.
- Brown, S., Clastre, M., Courdavault, V., and O'Connor, S.E.** (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. *Proceedings of the National Academy of Sciences* **112**, 3205-3210.
- Burlat, V., Oudin, A., Courtois, M., Rideau, M., and St - Pierre, B.** (2004). Co - expression of three MEP pathway genes and geraniol 10 - hydroxylase in internal

- phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid - derived primary metabolites. *The Plant Journal* **38**, 131-141.
- Çakır, B., and Kılıçkaya, O.** (2015). Mitogen-activated protein kinase cascades in *Vitis vinifera*. *Frontiers in Plant Science* **6**, 556.
- Cárdenas, P.D., Sonawane, P.D., Pollier, J., Bossche, R.V., Dewangan, V., Weithorn, E., Tal, L., Meir, S., Rogachev, I., and Malitsky, S.** (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nature communications* **7**.
- Cardinale, F., Meskiene, I., Ouaked, F., and Hirt, H.** (2002). Convergence and divergence of stress-induced mitogen-activated protein kinase signaling pathways at the level of two distinct mitogen-activated protein kinase kinases. *The Plant cell* **14**, 703-711.
- Cardoso, M.I.L., Meijer, A.H., Rueb, S., Machado, J.Q., Memelink, J., and Hoge, J.H.C.** (1997). A promoter region that controls basal and elicitor-inducible expression levels of the NADPH:cytochrome P450 reductase gene (Cpr) from *Catharanthus roseus* binds nuclear factor GT-1. *Molecular and General Genetics MGG* **256**, 674-681.
- Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martínez-García, J.F., Bilbao-Castro, J.R., and Robertson, D.L.** (2010). Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in *Arabidopsis*, poplar, rice, moss, and algae. *Plant physiology* **153**, 1398-1412.
- Chahed, K., Oudin, A., Guivarc'h, N., Hamdi, S., Chénieux, J.-C., Rideau, M., and Clastre, M.** (2000). 1-Deoxy-D-xylulose 5-phosphate synthase from periwinkle: cDNA identification and induced gene expression in terpenoid indole alkaloid-producing cells. *Plant Physiology and Biochemistry* **38**, 559-566.
- Champion, A., Picaud, A., and Henry, Y.** (2004). Reassessing the MAP3K and MAP4K relationships. *Trends in plant science* **9**, 123-129.
- Chávez Suárez, L., and Ramírez Fernández, R.** (2010). Signalling pathway in plants affected by salinity and drought. *ITEA* **106**, 157-169.
- Chen, L., Hu, W., Tan, S., Wang, M., Ma, Z., Zhou, S., Deng, X., Zhang, Y., Huang, C., and Yang, G.** (2012). Genome-wide identification and analysis of MAPK and MAPKK gene families in *Brachypodium distachyon*. *PLoS One* **7**, e46744.
- Cohen, P.T.** (1997). Novel protein serine/threonine phosphatases: variety is the spice of life. *Trends in biochemical sciences* **22**, 245-251.
- Collu, G., Unver, N., Peltenburg-Looman, A.M., van der Heijden, R., Verpoorte, R., and Memelink, J.** (2001). Geraniol 10 - hydroxylase1, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS letters* **508**, 215-220.
- Contin, A., van der Heijden, R., Lefeber, A.W., and Verpoorte, R.** (1998). The iridoid glucoside secologanin is derived from the novel triose phosphate/pyruvate pathway in a *Catharanthus roseus* cell culture. *FEBS letters* **434**, 413-416.
- Costa, M.M.R., Hilliou, F., Duarte, P., Pereira, L.G., Almeida, I., Leech, M., Memelink, J., Barceló, A.R., and Sottomayor, M.** (2008). Molecular cloning and characterization of a vacuolar class III peroxidase involved in the metabolism of anticancer alkaloids in *Catharanthus roseus*. *Plant Physiology* **146**, 403-417.

- Cristina, M.S., Petersen, M., and Mundy, J.** (2010). Mitogen-activated protein kinase signaling in plants. *Annual review of plant biology* **61**, 621-649.
- De Boer, K., Tilleman, S., Pauwels, L., Vanden Bossche, R., De Sutter, V., Vanderhaeghen, R., Hilson, P., Hamill, J.D., and Goossens, A.** (2011). APETALA2/ETHYLENE RESPONSE FACTOR and basic helix-loop-helix tobacco transcription factors cooperatively mediate jasmonate - elicited nicotine biosynthesis. *The Plant Journal* **66**, 1053-1065.
- De Carolis, E., and De Luca, V.** (1993). Purification, characterization, and kinetic analysis of a 2-oxoglutarate- dependent dioxygenase involved in vindoline biosynthesis from *Catharanthus roseus*. *Journal of Biological Chemistry* **268**, 5504-5511.
- De Carolis, E., Chan, F., Balsevich, J., and De Luca, V.** (1990). Isolation and characterization of a 2-oxoglutarate dependent dioxygenase involved in the second-to-last step in vindoline biosynthesis. *Plant physiology* **94**, 1323-1329.
- De Luca, V., and Cutler, A.J.** (1987). Subcellular localization of enzymes involved in indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant physiology* **85**, 1099-1102.
- De Luca, V., Marineau, C., and Brisson, N.** (1989). Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases. *Proceedings of the National Academy of Sciences* **86**, 2582-2586.
- Denu, J.M., Stuckey, J.A., Saper, M.A., and Dixon, J.E.** (1996). Form and Function in Protein Dephosphorylation. *Cell* **87**, 361-364.
- Ding, X., Richter, T., Chen, M., Fujii, H., Seo, Y.S., Xie, M., Zheng, X., Kanrar, S., Stevenson, R.A., and Dardick, C.** (2009). A rice kinase-protein interaction map. *Plant physiology* **149**, 1478-1492.
- Duan, P.G., Rao, Y.C., Zeng, D.L., Yang, Y.L., Xu, R., and Zhang, B.L.** (2014). SMALL GRAIN 1, which encodes a mitogen-activated protein kinase kinase 4, influences grain size in rice. *The Plant journal : for cell and molecular biology* **77**.
- Duarte, J.M., Cui, L., Wall, P.K., Zhang, Q., Zhang, X., Leebens-Mack, J., Ma, H., and Altman, N.** (2006). Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular biology and evolution* **23**, 469-478.
- Eckardt, N.A.** (2007). A complete MAPK signaling cascade that functions in stomatal development and patterning in *Arabidopsis* (*Am Soc Plant Biol*).
- El-Sayed, M., and Verpoorte, R.** (2007). *Catharanthus* terpenoid indole alkaloids: biosynthesis and regulation. *Phytochemistry Reviews* **6**, 277-305.
- Ellenberger, T.E., Brandl, C.J., Struhl, K., and Harrison, S.C.** The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: Crystal structure of the protein-DNA complex. *Cell* **71**, 1223-1237.
- Facchini, P.J., and De Luca, V.** (2008). Opium poppy and Madagascar periwinkle: model non - model systems to investigate alkaloid biosynthesis in plants. *The Plant Journal* **54**, 763-784.
- Fahn, W., Laußermair, E., Deus-Neumann, B., and Stöckigt, J.** (1985). Late enzymes of vindoline biosynthesis. *Plant Cell Reports* **4**, 337-340.
- Fanger, G.R., Johnson, N.L., and Johnson, G.L.** (1997). MEK kinases are regulated by EGF and selectively interact with Rac/Cdc42. *The EMBO Journal* **16**, 4961-4972.

- Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.-M., Gimenez-Ibanez, S., Geerinck, J., Eeckhout, D., Schweizer, F., Godoy, M., and Franco-Zorrilla, J.M.** (2011). The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *The Plant cell* **23**, 701-715.
- Field, B., Fiston-Lavier, A.S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A.E.** (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proceedings of the National Academy of Sciences USA* **108**, 16116-16121.
- Finn, R.D., Clements, J., and Eddy, S.R.** (2011). HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39**, W29-37.
- Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., and Bateman, A.** (2010). The Pfam protein families database. *Nucleic acids research* **38**, D211-222.
- Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grun, S., Winklmaier, A., Eisenreich, W., Bacher, A., Meeley, R.B., Briggs, S.P., Simcox, K., and Gierl, A.** (1997). Analysis of a chemical plant defense mechanism in grasses. *Science* **277**, 696-699.
- Frye, C.A., and Innes, R.W.** (1998). An Arabidopsis mutant with enhanced resistance to powdery mildew. *The Plant cell* **10**, 947-956.
- Frye, C.A., Tang, D., and Innes, R.W.** (2001). Negative regulation of defense responses in plants by a conserved MAPKK kinase. *Proceedings of the National Academy of Sciences* **98**, 373-378.
- Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M.** (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *The Plant Cell Online* **12**, 393-404.
- Galletti, R., Ferrari, S., and De Lorenzo, G.** (2011). Arabidopsis MPK3 and MPK6 play different roles in basal and oligogalacturonide-or flagellin-induced resistance against *Botrytis cinerea*. *Plant Physiology* **157**, 804-814.
- Geu-Flores, F., Sherden, N.H., Courdavault, V., Burlat, V., Glenn, W.S., Wu, C., Nims, E., Cui, Y., and O'Connor, S.E.** (2012). An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138-142.
- Giddings, L.-A., Liscombe, D.K., Hamilton, J.P., Childs, K.L., DellaPenna, D., Buell, C.R., and O'Connor, S.E.** (2011). A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus*. *Journal of Biological Chemistry* **286**, 16751-16757.
- Góngora-Castillo, E., Childs, K.L., Fedewa, G., Hamilton, J.P., Liscombe, D.K., Magallanes-Lundback, M., Mandadi, K.K., Nims, E., Runguphan, W., and Vaillancourt, B.** (2012). Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS One* **7**, e52506.
- Gou, J.-Y., Felippes, F.F., Liu, C.-J., Weigel, D., and Wang, J.-W.** (2011). Negative regulation of anthocyanin biosynthesis in Arabidopsis by a miR156-targeted SPL transcription factor. *The Plant cell* **23**, 1512-1522.

- Hallahan, D.L., West, J.M., Wallsgrove, R.M., Smiley, D.W., Dawson, G.W., Pickett, J.A., and Hamilton, J.G.** (1995). Purification and characterization of an acyclic monoterpene primary alcohol: NADP⁺ oxidoreductase from catmint (*Nepeta racemosa*). *Archives of biochemistry and biophysics* **318**, 105-112.
- Hamel, L., Nicole, M., Sritubtim, S., Morency, M., Ellis, M., Ehltung, J., Beaudoin, N., Barbazuk, B., Klessig, D., Lee, J., Martin, G., Mundy, J., Ohashi, Y., Scheel, D., Sheen, J., Xing, T., Zhang, S., Seguin, A., and Ellis, B.** (2006). Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci* **11**, 192 - 198.
- Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., and Shiu, S.-H.** (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant physiology* **148**, 993-1003.
- Hanks, S.K., and Hunter, T.** (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *The FASEB journal* **9**, 576-596.
- Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C.** (2003). The Basic Helix–Loop–Helix Transcription Factor Family in Plants: A Genome-Wide Study of Protein Structure and Functional Diversity. *Molecular Biology and Evolution* **20**, 735-747.
- Heinrich, M., Baldwin, I.T., and Wu, J.** (2011). Two mitogen-activated protein kinase kinases, MKK1 and MEK2, are involved in wounding-and specialist lepidopteran herbivore *Manduca sexta*-induced responses in *Nicotiana attenuata*. *Journal of experimental botany*, err162.
- Hemscheidt, T., and Zenk, M.** (1985). Partial purification and characterization of a NADPH dependent tetrahydroalstonine synthase from *Catharanthus roseus* cell suspension cultures. *Plant cell reports* **4**, 216-219.
- Hiratsu, K., Matsui, K., Koyama, T., and Ohme - Takagi, M.** (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *The Plant Journal* **34**, 733-739.
- Hord, C.L.H., Suna, Y.J., Pillitteri, L.J., Torii, K.U., Wang, H.C., and Zhang, S.Q.** (2008). Regulation of *Arabidopsis* early anther development by the mitogen-activated protein kinases, MPK3 and MPK6, and the *ERECTA* and related receptor-like kinases. *Mol Plant* **1**.
- Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., Ren, Y., Zhu, H., Li, J., Lin, K., Jin, W., Fei, Z., Li, G., Staub, J., Kilian, A., van der Vossen, E.A., Wu, Y., Guo, J., He, J., Jia, Z., Ren, Y., Tian, G., Lu, Y., Ruan, J., Qian, W., Wang, M., Huang, Q., Li, B., Xuan, Z., Cao, J., Asan, Wu, Z., Zhang, J., Cai, Q., Bai, Y., Zhao, B., Han, Y., Li, Y., Li, X., Wang, S., Shi, Q., Liu, S., Cho, W.K., Kim, J.Y., Xu, Y., Heller-Uszynska, K., Miao, H., Cheng, Z., Zhang, S., Wu, J., Yang, Y., Kang, H., Li, M., Liang, H., Ren, X., Shi, Z., Wen, M., Jian, M., Yang, H., Zhang, G., Yang, Z., Chen, R., Liu, S., Li, J., Ma, L., Liu, H., Zhou, Y., Zhao, J., Fang, X., Li, G., Fang, L., Li, Y., Liu, D., Zheng, H., Zhang, Y., Qin, N., Li, Z., Yang, G., Yang, S., Bolund, L., Kristiansen, K., Zheng, H., Li, S., Zhang, X., Yang, H., Wang, J., Sun, R., Zhang, B., Jiang, S., Wang, J., Du, Y., and Li, S.** (2009). The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* **41**, 1275-1281.

- Huang, Y., Li, H., Hutchison, C.E., Laskey, J., and Kieber, J.J.** (2003). Biochemical and functional analysis of CTR1, a protein kinase that negatively regulates ethylene signaling in Arabidopsis. *The Plant Journal* **33**, 221-233.
- Huang, Y., Li, H., Gupta, R., Morris, P.C., Luan, S., and Kieber, J.J.** (2000). ATMPK4, an Arabidopsis homolog of mitogen-activated protein kinase, is activated in vitro by AtMEK1 through threonine phosphorylation. *Plant Physiology* **122**, 1301-1310.
- Ichimura, K., Mizoguchi, T., Irie, K., Morris, P., Giraudat, J., Matsumoto, K., and Shinozaki, K.** (1998). Isolation of ATMEKK1 (a MAP Kinase Kinase Kinase)-interacting proteins and analysis of a MAP kinase cascade in Arabidopsis. *Biochemical and biophysical research communications* **253**, 532-543.
- Ichimura, K., Shinozaki, K., Tena, G., Sheen, J., Henry, Y., Champion, A., Kreis, M., Zhang, S., Hirt, H., and Wilson, C.** (2002). Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in plant science* **7**, 301-308.
- Ikeda, H., Esaki, N., Nakai, S., Hashimoto, K., Uesato, S., Soda, K., and Fujita, T.** (1991). Acyclic monoterpene primary alcohol: NADP⁺ oxidoreductase of *Rauwolfia serpentina* cells: the key enzyme in biosynthesis of monoterpene alcohols.
- Irmiler, S., Schröder, G., St - Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J.** (2000). Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *The Plant Journal* **24**, 797-804.
- Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R., Tikunov, Y., Bovy, A., Chikate, Y., Singh, P., Rogachev, I., Beekwilder, J., Giri, A.P., and Aharoni, A.** (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* **341**, 175-179.
- Iyer, G.H., Moore, M.J., and Taylor, S.S.** (2005). Consequences of lysine 72 mutation on the phosphorylation and activation state of cAMP-dependent kinase. *Journal of Biological Chemistry* **280**, 8800-8807.
- Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F.** (2002). bZIP transcription factors in Arabidopsis. *Trends in Plant Science* **7**, 106-111.
- Jalmi, S.K., and Sinha, A.K.** (2016). Functional involvement of a mitogen activated protein kinase module, OsMKK3-OsMPK7-OsWRK30 in mediating resistance against *Xanthomonas oryzae* in rice. *Scientific reports* **6**.
- Jander, G., and Barth, C.** (2007). Tandem gene arrays: a challenge for functional genomics. *Trends in Plant Science* **12**, 203-210.
- Jia, X., Shen, J., Liu, H., Li, F., Ding, N., Gao, C., Pattanaik, S., Patra, B., Li, R., and Yuan, L.** (2015). Small tandem target mimic-mediated blockage of microRNA858 induces anthocyanin accumulation in tomato. *Planta* **242**, 283-293.
- Jiménez, C., Cossío, B.R., Rivard, C.J., Berl, T., and Capasso, J.M.** (2007). Cell division in the unicellular microalga *Dunaliella viridis* depends on phosphorylation of extracellular signal-regulated kinases (ERKs). *Journal of experimental botany* **58**, 1001-1011.

- Jin, H., Liu, Y., Yang, K.Y., Kim, C.Y., Baker, B., and Zhang, S.** (2003). Function of a mitogen-activated protein kinase pathway in N gene-mediated resistance in tobacco. *Plant Journal* **33**, 719-731.
- Jonak, C., Ökrész, L., Bögre, L., and Hirt, H.** (2002). Complexity, cross talk and integration of plant MAP kinase signalling. *Current opinion in plant biology* **5**, 415-424.
- Jouannic, S., Hamal, A., Leprince, A.-S., Tregear, J.W., Kreis, M., and Henry, Y.** (1999). Characterisation of novel plant genes encoding MEKK/STE11 and RAF-related protein kinases. *Gene* **229**, 171-181.
- Jun, J.H., Ha, C.M., and Nam, H.G.** (2002). Involvement of the VEP1 gene in vascular strand development in *Arabidopsis thaliana*. *Plant and cell physiology* **43**, 323-330.
- Kato, N., Dubouzet, E., Kokabu, Y., Yoshida, S., Taniguchi, Y., Dubouzet, J.G., Yazaki, K., and Sato, F.** (2007). Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant and cell physiology* **48**, 8-18.
- Keat, H., Elizabeth, G., and McKnight, T.** (2000). Characterization and cloning of 10-hydroxygeraniol oxidoreductase. *Plant Biology*.
- Kellner, F., Kim, J., Clavijo, B.J., Hamilton, J.P., Childs, K.L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., and Clissold, L.** (2015). Genome - guided investigation of plant natural product biosynthesis. *The Plant Journal* **82**, 680-692.
- Kerk, D., Bulgrien, J., Smith, D.W., Barsam, B., Veretnik, S., and Gribskov, M.** (2002). The Complement of Protein Phosphatase Catalytic Subunits Encoded in the Genome of *Arabidopsis*. *Plant Physiology* **129**, 908.
- Khan, G.A., Declerck, M., Sorin, C., Hartmann, C., Crespi, M., and Lelandais-Brière, C.** (2011). MicroRNAs as regulators of root development and architecture. *Plant molecular biology* **77**, 47-58.
- Kieber, J.J., Rothenberg, M., Roman, G., Feldmann, K.A., and Ecker, J.R.** (1993). CTR1, a negative regulator of the ethylene response pathway in *Arabidopsis*, encodes a member of the raf family of protein kinases. *Cell* **72**, 427-441.
- Kiegerl, S., Cardinale, F., Siligan, C., Gross, A., Baudouin, E., Liwosz, A., Eklof, S., Till, S., Bogre, L., Hirt, H., and Meskiene, I.** (2000). SIMKK, a mitogen-activated protein kinase (MAPK) kinase, is a specific activator of the salt stress-induced MAPK, SIMK. *The Plant cell* **12**, 2247-2258.
- Kolch, W.** (2000). Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem. J* **351**, 289-305.
- Kong, F., Wang, J., Cheng, L., Liu, S., Wu, J., Peng, Z., and Lu, G.** (2012). Genome-wide analysis of the mitogen-activated protein kinase gene family in *Solanum lycopersicum*. *Gene* **499**, 108-120.
- Kong, X., Lv, W., Zhang, D., Jiang, S., Zhang, S., and Li, D.** (2013a). Genome-wide identification and analysis of expression profiles of maize mitogen-activated protein kinase kinase kinase. *PloS one* **8**, e57714.
- Kong, X., Pan, J., Zhang, D., Jiang, S., Cai, G., Wang, L., and Li, D.** (2013b). Identification of mitogen-activated protein kinase kinase gene family and MKK-

- MAPK interaction network in maize. *Biochemical and biophysical research communications* **441**, 964-969.
- Kosetsu, K., Matsunaga, S., Nakagami, H., Colcombet, J., Sasabe, M., Soyano, T., Takahashi, Y., Hirt, H., and Machida, Y.** (2010). The MAP kinase MPK4 is required for cytokinesis in *Arabidopsis thaliana*. *The Plant cell* **22**, 3778-3790.
- Krithika, R., Srivastava, P.L., Rani, B., Kolet, S.P., Chopade, M., Soniya, M., and Thulasiram, H.V.** (2015). Characterization of 10-Hydroxygeraniol Dehydrogenase from *Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Scientific reports* **5**.
- Kumar, K., Rao, K.P., Biswas, D.K., and Sinha, A.K.** (2011). Rice WNK1 is regulated by abiotic stress and involved in internal circadian rhythm. *Plant Signal Behav* **6**.
- Laflamme, P., St-Pierre, B., and De Luca, V.** (2001). Molecular and biochemical analysis of a Madagascar periwinkle root-specific minovincinine-19-hydroxy-O-acetyltransferase. *Plant physiology* **125**, 189-198.
- Larkin, M.A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., and Lopez, R.** (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
- Larsen, B., Fuller, V.L., Pollier, J., Van Moerkercke, A., Schweizer, F., Payne, R., Colinas, M., O'Connor, S.E., Goossens, A., and Halkier, B.A.** (2017). Identification of Iridoid Glucoside Transporters in *Catharanthus roseus*. *Plant and Cell Physiology*.
- Lee, J.S., Huh, K.W., Bhargava, A., and Ellis, B.E.** (2008). Comprehensive analysis of protein-protein interactions between *Arabidopsis* MAPKs and MAPK kinases helps define potential MAPK signalling modules. *Plant signaling & behavior* **3**, 1037-1041.
- Lehti-Shiu, M.D., and Shiu, S.-H.** (2012). Diversity, classification and function of the plant protein kinase superfamily. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **367**, 2619-2639.
- Lehti-Shiu, M.D., Zou, C., Hanada, K., and Shiu, S.-H.** (2009). Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiology* **150**, 12-26.
- Levac, D., Murata, J., Kim, W.S., and De Luca, V.** (2008). Application of carborundum abrasion for investigating the leaf epidermis: Molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-O-methyltransferase. *Plant Journal* **53**, 225-236.
- Li, F., Li, M., Wang, P., Cox, K.L., Duan, L., Dever, J.K., Shan, L., Li, Z., and He, P.** (2017). Regulation of cotton (*Gossypium hirsutum*) drought responses by mitogen-activated protein (MAP) kinase cascade-mediated phosphorylation of GhWRKY59. *New Phytologist* **215**, 1462-1475.
- Li, M., and Sack, F.D.** (2014). Myrosin Idioblast Cell Fate and Development Are Regulated by the Arabidopsis Transcription Factor FAMA, the Auxin Pathway, and Vesicular Trafficking. *The Plant cell* **26**, 4053.
- Liang, W., Yang, B., Yu, B.-J., Zhou, Z., Li, C., Jia, M., Sun, Y., Zhang, Y., Wu, F., and Zhang, H.** (2013). Identification and analysis of MKK and MPK gene families in canola (*Brassica napus* L.). *BMC genomics* **14**, 392.

- Lichtenthaler, H.K.** (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual review of plant biology* **50**, 47-65.
- Lin, Y., and Lai, Z.** (2013). Comparative analysis reveals dynamic changes in miRNAs and their targets and expression during somatic embryogenesis in longan (*Dimocarpus longan* Lour.). *PLoS One* **8**, e60337.
- Liscombe, D.K., Usera, A.R., and O'Connor, S.E.** (2010). Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis. *Proceedings of the National Academy of Sciences USA* **107**, 18793-18798.
- Liu, D.H., Jin, H.B., Chen, Y.H., Cui, L.J., Ren, W.W., Gong, Y.F., and Tang, K.X.** (2007). Terpenoid indole alkaloids biosynthesis and metabolic engineering in *Catharanthus roseus*. *Journal of integrative plant biology* **49**, 961-974.
- Liu, H., Wang, Y., Xu, J., Su, T., Liu, G., and Ren, D.** (2008). Ethylene signaling is required for the acceleration of cell death induced by the activation of AtMEK5 in *Arabidopsis*. *Cell research* **18**, 422.
- Liu, J.Z., Horstman, H.D., Braun, E., Graham, M.A., Zhang, C.Q., and Navarre, D.** (2011). Soybean homologs of MPK4 negatively regulate defense responses and positively regulate growth and development. *Plant Physiol* **157**.
- Liu, Y., and Zhang, S.** (2004). Phosphorylation of 1-aminocyclopropane-1-carboxylic acid synthase by MPK6, a stress-responsive mitogen-activated protein kinase, induces ethylene biosynthesis in *Arabidopsis*. *The Plant cell* **16**, 3386 - 3399.
- Liu, Y., Zhang, D., Wang, L., and Li, D.** (2013). Genome-wide analysis of mitogen-activated protein kinase gene family in maize. *Plant molecular biology reporter* **31**, 1446-1460.
- Liu, Z., Shi, L., Liu, Y., Tang, Q., Shen, L., Yang, S., Cai, J., Yu, H., Wang, R., and Wen, J.** (2014). Genome-wide identification and transcriptional expression analysis of mitogen-activated protein kinase and mitogen-activated protein kinase kinase genes in *Capsicum annuum*. *Frontiers in plant science* **6**, 780-780.
- Lu, K., Guo, W., Lu, J., Yu, H., Qu, C., Tang, Z., Li, J., Chai, Y., and Liang, Y.** (2015). Genome-wide survey and expression profile analysis of the mitogen-activated protein kinase (MAPK) gene family in *Brassica rapa*. *PloS one* **10**, e0132051.
- Ma, D., Pu, G., Lei, C., Ma, L., Wang, H., Guo, Y., Chen, J., Du, Z., Wang, H., and Li, G.** (2009). Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription factor that regulates the amorpha-4, 11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant and Cell Physiology* **50**, 2146-2161.
- MacAlister, C.A., and Bergmann, D.C.** (2011). Sequence and function of basic helix-loop-helix proteins required for stomatal development in *Arabidopsis* are deeply conserved in land plants. *Evolution & Development* **13**, 182-192.
- Maier, A., Schrader, A., Kokkelink, L., Falke, C., Welter, B., Iniesto, E., Rubio, V., Uhrig, J.F., Hülkamp, M., and Hoecker, U.** (2013). Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in *Arabidopsis*. *The Plant Journal* **74**, 638-651.
- Major, G., Daigle, C., Stafford-Richard, T., Tebbji, F., Lafleur, É., Caron, S., and Matton, D.P.** (2009). Characterization of ScMAP4K1, a MAP kinase kinase

- kinase kinase involved in ovule, seed and fruit development in *Solanum chacoense* Bitt. *Plant Biology* **10**.
- Mao, G., Meng, X., Liu, Y., Zheng, Z., Chen, Z., and Zhang, S.** (2011). Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in *Arabidopsis*. *The Plant cell* **23**, 1639-1653.
- McGrath, K.C., Dombrecht, B., Manners, J.M., Schenk, P.M., Edgar, C.I., Maclean, D.J., Scheible, W.-R., Udvardi, M.K., and Kazan, K.** (2005). Repressor-and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiology* **139**, 949-959.
- McKnight, T., Roessner, C., Devagupta, R., Scott, A., and Nessler, C.** (1990). Nucleotide sequence of a cDNA encoding the vacuolar protein strictosidine synthase from *Catharanthus roseus*. *Nucleic acids research* **18**, 4939-4939.
- McKnight, T.D., Bergey, D.R., Burnett, R.J., and Nessler, C.L.** (1991). Expression of enzymatically active and correctly targeted strictosidine synthase in transgenic tobacco plants. *Planta* **185**, 148-152.
- Meijer, A.H., Cardoso, M., Voskuilen, J.T., Waal, A., Verpoorte, R., and Hoge, J.H.C.** (1993). Isolation and characterization of a cDNA clone from *Catharanthus roseus* encoding NADPH: cytochrome P - 450 reductase, an enzyme essential for reactions catalysed by cytochrome P - 450 mono - oxygenases in plants. *The Plant Journal* **4**, 47-60.
- Meng, X., Xu, J., He, Y., Yang, K.-Y., Mordorski, B., Liu, Y., and Zhang, S.** (2013). Phosphorylation of an ERF transcription factor by *Arabidopsis* MPK3/MPK6 regulates plant defense gene induction and fungal resistance. *The Plant cell* **25**, 1126-1142.
- Meng, Y., Ma, X., Chen, D., Wu, P., and Chen, M.** (2010). MicroRNA-mediated signaling involved in plant root development. *Biochemical and biophysical research communications* **393**, 345-349.
- Meng, Y., Huang, F., Shi, Q., Cao, J., Chen, D., Zhang, J., Ni, J., Wu, P., and Chen, M.** (2009). Genome-wide survey of rice microRNAs and microRNA–target pairs in the root of a novel auxin-resistant mutant. *Planta* **230**, 883-898.
- Menke, F., Stefanie, P., Mueller, M., Kijne, J., and Memelink, J.** (1999a). Involvement of the Octadecanoid pathway and protein phosphorylation in fungal elicitor induced expression of Terpenoid Indole Alkaloid biosynthetic genes in *Catharanthus roseus*. *Plant Physiol* **119**, 1289 - 1296.
- Menke, F.L., Champion, A., Kijne, J.W., and Memelink, J.** (1999b). A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO Journal* **18**, 4455-4463.
- Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., Van Der Krol, S., Lugan, R., and Ilc, T.** (2014). The seco-iridoid pathway from *Catharanthus roseus*. *Nature communications* **5**.
- Mirjalili, N., and Linden, J.C.** (1996). Methyl jasmonate induced production of taxol in suspension cultures of *Taxus cuspidata*: ethylene interaction and induction models. *Biotechnology progress* **12**, 110-118.

- Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K.** (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1819**, 86-96.
- Munkert, J., Pollier, J., Miettinen, K., Van Moerkercke, A., Payne, R., Müller-Uri, F., Burlat, V., O'Connor, S.E., Memelink, J., and Kreis, W.** (2015). Iridoid synthase activity is common among the plant progesterone 5 β -reductase family. *Molecular plant* **8**, 136-152.
- Murakami-Kojima, M., Nakamichi, N., Yamashino, T., and Mizuno, T.** (2002). The APRR3 component of the clock-associated APRR1/TOC1 quintet is phosphorylated by a novel protein kinase belonging to the WNK family, the gene for which is also transcribed rhythmically in *Arabidopsis thaliana*. *Plant Cell Physiol* **43**.
- Murata, J., Roepke, J., Gordon, H., and De Luca, V.** (2008). The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *The Plant cell* **20**, 524-542.
- Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H.** (2006). Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant physiology* **140**, 411-432.
- Neupane, A., Nepal, M.P., Piya, S., Subramanian, S., Rohila, J.S., Reese, R.N., and Benson, B.V.** (2013). Identification, nomenclature, and evolutionary relationships of mitogen-activated protein kinase (MAPK) genes in soybean. *Evolutionary Bioinformatics* **9**, 363.
- Nutzmann, H.W., Huang, A., and Osbourn, A.** (2016). Plant metabolic clusters - from genetics to genomics. *New Phytologist*.
- Ohme-Takagi, M., and Shinshi, H.** (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *The Plant Cell Online* **7**, 173-182.
- Oudin, A., Courtois, M., Rideau, M., and Clastre, M.** (2007). The iridoid pathway in *Catharanthus roseus* alkaloid biosynthesis. *Phytochemistry Reviews* **6**, 259-276.
- Ouwerkerk, P.B., and Memelink, J.** (1999). Elicitor-responsive promoter regions in the tryptophan decarboxylase gene from *Catharanthus roseus*. *Plant molecular biology* **39**, 129-136.
- Parages, M.L., Heinrich, S., Wiencke, C., and Jiménez, C.** (2013). Rapid phosphorylation of MAP kinase-like proteins in two species of Arctic kelps in response to temperature and UV radiation stress. *Environmental and Experimental Botany* **91**, 30-37.
- Patra, B., Pattanaik, S., and Yuan, L.** (2013a). Ubiquitin protein ligase 3 mediates the proteasomal degradation of GLABROUS 3 and ENHANCER OF GLABROUS 3, regulators of trichome development and flavonoid biosynthesis in *Arabidopsis*. *The Plant journal : for cell and molecular biology* **74**, 435-447.
- Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S., and Yuan, L.** (2013b). Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1829**, 1236-1247.
- Pattanaik, S., Werkman, J.R., Kong, Q., and Yuan, L.** (2010a). Site-Directed Mutagenesis and Saturation Mutagenesis for the Functional Study of Transcription Factors Involved in Plant Secondary Metabolite Biosynthesis. In

- Plant Secondary Metabolism Engineering: Methods and Applications, A.G. Fett-Neto, ed (Totowa, NJ: Humana Press), pp. 47-57.
- Pattanaik, S., Kong, Q., Zaitlin, D., Werkman, J., Xie, C., Patra, B., and Yuan, L.** (2010b). Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco. *Planta* **231**, 1061-1076.
- Paul, P., Singh, S.K., Patra, B., Sui, X., Pattanaik, S., and Yuan, L.** (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *New Phytologist* **213**, 1107-1123.
- Pauw, B., Hilliou, F., Sandonis Martin, V., Chatel, G., de Wolf, C., Champion, A., Pre, M., Van Duijn, B., Kijne, J., van der Fits, L., and Memelink, J.** (2004). Zinc finger proteins act as transcriptional repressors of alkaloid biosynthesis genes in *Catharanthus roseus*. *J Biol Chem* **279**, 52940 - 52948.
- Payne, R.M.E., Xu, D., Foureau, E., Teto Carqueijeiro, M.I.S., Oudin, A., Bernonville, T.D.d., Novak, V., Burow, M., Olsen, C.-E., Jones, D.M., Tatsis, E.C., Pendle, A., Ann Halkier, B., Geu-Flores, F., Courdavault, V., Nour-Eldin, H.H., and O'Connor, S.E.** (2017). An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nature Plants* **3**, 16208.
- Peebles, C.A., Hughes, E.H., Shanks, J.V., and San, K.Y.** (2009). Transcriptional response of the terpenoid indole alkaloid pathway to the overexpression of ORCA3 along with jasmonic acid elicitation of *Catharanthus roseus* hairy roots over time. *Metabolic Engineering* **11**, 76-86.
- Petersen, M., Brodersen, P., Naested, H., Andreasson, E., Lindhart, U., Johansen, B., Nielsen, H.B., Lacy, M., Austin, M.J., and Parker, J.E.** (2000). Arabidopsis MAP kinase 4 negatively regulates systemic acquired resistance. *Cell* **103**, 1111-1120.
- Pitzschke, A., Schikora, A., and Hirt, H.** (2009). MAPK cascade signalling networks in plant defence. *Current opinion in plant biology* **12**, 421-426.
- Posas, F., and Saito, H.** (1997). Osmotic activation of the HOG MAPK pathway via Ste11p MAPKKK: scaffold role of Pbs2p MAPKK. *Science* **276**, 1702-1705.
- Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A.** (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc Natl Acad Sci U S A* **101**, 8233-8238.
- Qiu, J.-L., Zhou, L., Yun, B.-W., Nielsen, H.B., Fiil, B.K., Petersen, K., MacKinlay, J., Loake, G.J., Mundy, J., and Morris, P.C.** (2008). Arabidopsis mitogen-activated protein kinase kinases MKK1 and MKK2 have overlapping functions in defense signaling mediated by MEKK1, MPK4, and MKS1. *Plant physiology* **148**, 212-222.
- Qu, Y., Easson, M.L., Froese, J., Simionescu, R., Hudlicky, T., and De Luca, V.** (2015). Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences* **112**, 6224-6229.
- Quimby, B.B., Wilson, C.A., and Corbett, A.H.** (2000). The interaction between Ran and NTF2 is required for cell cycle progression. *Molecular biology of the cell* **11**, 2617-2629.

- Qureshi, A., and Scott, A.** (1968). Interconversion of Corynanthe, Aspidosperma, and Iboga alkaloids. A model for indole alkaloid biosynthesis. *Chemical Communications* (London), 945-946.
- Raina, S., Wankhede, D., Jaggi, M., Singh, P., Jalmi, S., Raghuram, B., Sheikh, A., and Sinha, A.** (2012). CrMPK3, a mitogen activated protein kinase from *Catharanthus roseus* and its possible role in stress induced biosynthesis of monoterpenoid indole alkaloids. *BMC plant biology* **12**, 134.
- Rao, K., Vani, G., Kumar, K., and Sinha, A.** (2009). Rhythmic expression of Mitogen Activated Protein Kinase activity in rice. *Mol Cells* **28**, 417 - 422.
- Rao, K.P., Richa, T., Kumar, K., Raghuram, B., and Sinha, A.K.** (2010). In silico analysis reveals 75 members of mitogen-activated protein kinase kinase kinase gene family in rice. *DNA research* **17**, 139-153.
- Ren, D., Yang, H., and Zhang, S.** (2002). Cell death mediated by MAPK is associated with hydrogen peroxide production in *Arabidopsis*. *Journal of Biological Chemistry* **277**, 559-565.
- Rizvi, N.F., Weaver, J.D., Cram, E.J., and Lee-Parsons, C.W.** (2016). Silencing the transcriptional repressor, ZCT1, illustrates the tight regulation of terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* hairy roots. *PloS one* **11**, e0159712.
- Rizzon, C., Ponger, L., and Gaut, B.S.** (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS computational biology* **2**, e115.
- Rodriguez, S., Compagnon, V., Crouch, N.P., St-Pierre, B., and De Luca, V.** (2003). Jasmonate-induced epoxidation of tabersonine by a cytochrome P-450 in hairy root cultures of *Catharanthus roseus*. *Phytochemistry* **64**, 401-409.
- Roepke, J., Salim, V., Wu, M., Thamm, A.M., Murata, J., Ploss, K., Boland, W., and De Luca, V.** (2010). Vinca drug components accumulate exclusively in leaf exudates of Madagascar periwinkle. *Proceedings of the National Academy of Sciences* **107**, 15287-15292.
- Roytrakul, S., and Verpoorte, R.** (2007). Role of vacuolar transporter proteins in plant secondary metabolism: *Catharanthus roseus* cell culture. *Phytochemistry Reviews* **6**, 383-396.
- Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K.** (2002). DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. *Biochemical and biophysical research communications* **290**, 998-1009.
- Salim, V., Yu, F., Altarejos, J., and Luca, V.** (2013). Virus - induced gene silencing identifies *Catharanthus roseus* 7 - deoxyloganic acid - 7 - hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *The Plant Journal* **76**, 754-765.
- Salim, V., Wiens, B., Masada-Atsumi, S., Yu, F., and De Luca, V.** (2014). 7-deoxyloganic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganic acid in *Catharanthus roseus* iridoid biosynthesis. *Phytochemistry* **101**, 23-31.
- Samuel, M.A., Miles, G.P., and Ellis, B.E.** (2000). Ozone treatment rapidly activates MAP kinase signalling in plants. *The Plant Journal* **22**, 367-376.

- Saraste, M., Sibbald, P.R., and Wittinghofer, A.** (1990). The P-loop—a common motif in ATP-and GTP-binding proteins. *Trends in biochemical sciences* **15**, 430-434.
- Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., and Wisman, E.** (2001). Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *The Plant cell* **13**, 113-123.
- Schluttenhofer, C., and Yuan, L.** (2015). Regulation of Specialized Metabolism by WRKY Transcription Factors. *Plant Physiology* **167**, 295.
- Schluttenhofer, C., Pattanaik, S., Patra, B., and Yuan, L.** (2014). Analyses of *Catharanthus roseus* and *Arabidopsis thaliana* WRKY transcription factors reveal involvement in jasmonate signaling. *BMC Genomics* **15**, 502.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., and Jackson, S.A.** (2010). Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178-183.
- Schröder, G., Unterbusch, E., Kaltenbach, M., Schmidt, J., Strack, D., De Luca, V., and Schröder, J.** (1999). Light - induced cytochrome P450 - dependent enzyme in indole alkaloid biosynthesis: tabersonine 16 - hydroxylase. *FEBS letters* **458**, 97-102.
- Schweighofer, A., Hirt, H., and Meskiene, I.** (2004). Plant PP2C phosphatases: emerging functions in stress signaling. *Trends in Plant Science* **9**, 236-243.
- Schweighofer, A., Kazanaviciute, V., Scheickl, E., Teige, M., Doczi, R., Hirt, H., Schwanninger, M., Kant, M., Schuurink, R., Mauch, F., Buchala, A., Cardinale, F., and Meskiene, I.** (2007). The PP2C-Type Phosphatase AP2C1, Which Negatively Regulates MPK4 and MPK6, Modulates Innate Immunity, Jasmonic Acid, and Ethylene Levels in *Arabidopsis*. *The Plant cell* **19**, 2213-2224.
- Sells, M.A., Knaus, U.G., Bagrodia, S., Ambrose, D.M., Bokoch, G.M., and Chernoff, J.** (1997). Human p21-activated kinase (Pak1) regulates actin organization in mammalian cells. *Current Biology* **7**, 202-210.
- Seo, S., Okamoto, M., Seto, H., Ishizuka, K., Sano, H., and Ohashi, Y.** (1995). Tobacco MAP kinase: a possible mediator in wound signal transduction pathways. *SCIENCE-NEW YORK THEN WASHINGTON-*, 1988-1991.
- Sharma, D., Tiwari, M., Pandey, A., Bhatia, C., Sharma, A., and Trivedi, P.K.** (2016). MicroRNA858 is a potential regulator of phenylpropanoid pathway and plant development in *Arabidopsis*. *Plant physiology*, pp. 01831.02015.
- Sharrocks, A.D., Yang, S.-H., and Galanis, A.** (2000). Docking domains and substrate-specificity determination for MAP kinases. *Trends in biochemical sciences* **25**, 448-453.
- Sheikh, A.H., Raghuram, B., Jalmi, S.K., Wankhede, D.P., Singh, P., and Sinha, A.K.** (2013). Interaction between two rice mitogen activated protein kinases and its possible role in plant defense. *BMC plant biology* **13**, 121.

- Shen, E.M., Singh, S.K., Ghosh, J.S., Patra, B., Paul, P., Yuan, L., and Pattanaik, S.** (2017). The miRNAome of *Catharanthus roseus*: identification, expression analysis, and potential roles of microRNAs in regulation of terpenoid indole alkaloid biosynthesis. *Sci Rep* **7**.
- Shen, Q., Lu, X., Yan, T., Fu, X., Lv, Z., Zhang, F., Pan, Q., Wang, G., Sun, X., and Tang, K.** (2016). The jasmonate-responsive AaMYC2 transcription factor positively regulates artemisinin biosynthesis in *Artemisia annua*. *The New phytologist* **210**, 1269-1281.
- Shi, J., An, H.L., Zhang, L.A., Gao, Z., and Guo, X.Q.** (2010). GhMPK7, a novel multiple stress-responsive cotton group C MAPK gene, has a role in broad spectrum disease resistance and plant development. *Plant Mol Biol* **74**.
- Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N., Koga, J., Nojiri, H., and Yamane, H.** (2007). Identification of a biosynthetic gene cluster in rice for momilactones. *Journal of Biological Chemistry* **282**, 34013-34018.
- Shiu, S.H., Shih, M.C., and Li, W.H.** (2005). Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol* **139**, 18-26.
- Shoji, T., and Hashimoto, T.** (2011). Tobacco MYC2 regulates jasmonate-inducible nicotine biosynthesis genes directly and by way of the NIC2-locus ERF genes. *Plant and cell physiology* **52**, 1117-1130.
- Shoji, T., Nakajima, K., and Hashimoto, T.** (2000). Ethylene Suppresses Jasmonate-Induced Gene Expression in Nicotine Biosynthesis. *Plant and Cell Physiology* **41**, 1072-1076.
- Shoji, T., Kajikawa, M., and Hashimoto, T.** (2010). Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *The Plant cell* **22**, 3390-3409.
- Shoji, T., Mishima, M., and Hashimoto, T.** (2013). Divergent DNA-binding specificities of a group of ETHYLENE RESPONSE FACTOR transcription factors involved in plant defense. *Plant physiology* **162**, 977-990.
- Sibéril, Y., Benhamron, S., Memelink, J., Giglioli-Guivarc'h, N., Thiersault, M., Boisson, B., Doireau, P., and Gantet, P.** (2001). *Catharanthus roseus* G-box binding factors 1 and 2 act as repressors of strictosidine synthase gene expression in cell cultures. *Plant molecular biology* **45**, 477-488.
- Simkin, A.J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., and St-Pierre, B.** (2013). Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* **85**, 36-43.
- Singh, A.K., Kumar, S.R., Dwivedi, V., Rai, A., Pal, S., Shasany, A.K., and Nagegowda, D.A.** (2017). A WRKY transcription factor from *Withania somnifera* regulates triterpenoid withanolide accumulation and biotic stress tolerance through modulation of phytosterol and defense pathways. *New Phytologist*.
- Singh, S.K., Wu, Y., Ghosh, J.S., Pattanaik, S., Fisher, C., Wang, Y., Lawson, D., and Yuan, L.** (2015). RNA-sequencing Reveals Global Transcriptomic Changes in *Nicotiana tabacum* Responding to Topping and Treatment of Axillary-shoot Control Chemicals. *Scientific Reports* **5**, 18148.

- Smékalová, V., Doskočilová, A., Komis, G., and Šamaj, J.** (2014). Crosstalk between secondary messengers, hormones and MAPK modules during abiotic stress signalling in plants. *Biotechnology advances* **32**, 2-11.
- Song, S., Qi, T., Fan, M., Zhang, X., Gao, H., Huang, H., Wu, D., Guo, H., and Xie, D.** (2013). The bHLH Subgroup IIIId Factors Negatively Regulate Jasmonate-Mediated Plant Defense and Development. *PLoS Genetics* **9**, e1003653.
- Sottomayor, M., and Barceló, A.R.** (2003). Peroxidase from *Catharanthus roseus* (L.) G. Don and the biosynthesis of α -3', 4'-anhydrovinblastine: a specific role for a multifunctional enzyme. *Protoplasma* **222**, 97-105.
- Sottomayor, M., Lopez-Serrano, M., DiCosmo, F., and Barceló, A.R.** (1998). Purification and characterization of α -3', 4'-anhydrovinblastine synthase (peroxidase-like) from *Catharanthus roseus* (L.) G. Don. *FEBS letters* **428**, 299-303.
- Sottomayor, M., Cardoso, I.L., Pereira, L., and Barceló, A.R.** (2004). Peroxidase and the biosynthesis of terpenoid indole alkaloids in the medicinal plant *Catharanthus roseus* (L.) G. Don. *Phytochemistry Reviews* **3**, 159-171.
- Soyano, T., Nishihama, R., Morikiyo, K., Ishikawa, M., and Machida, Y.** (2003). NQK1/NtMEK1 is a MAPKK that acts in the NPK1 MAPKKK-mediated MAPK cascade and is required for plant cytokinesis. *Gene Dev* **17**.
- St-Pierre, B., and De Luca, V.** (1995). A cytochrome P-450 monooxygenase catalyzes the first step in the conversion of tabersonine to vindoline in *Catharanthus roseus*. *Plant physiology* **109**, 131-139.
- 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. *The Plant cell* **11**, 887-900.
- St-Pierre, B., Besseau, S., Clastre, M., Courdavault, V., Courtois, M., Crèche, J., Ducos, E., de Bernonville, T.D., Dutilleul, C., and Glévarec, G.** (2013). Deciphering the evolution, cell biology and regulation of monoterpene indole alkaloids. *Adv. Bot. Res* **68**, 73-109.
- St - Pierre, B., Laflamme, P., Alarco, A.M., and Luca, E.** (1998). The terminal O - acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A - dependent acyl transfer. *The Plant Journal* **14**, 703-713.
- Sugden, P.H., and Clerk, A.** (1997). Regulation of the ERK subgroup of MAP kinase cascades through G protein-coupled receptors. *Cellular signalling* **9**, 337-351.
- Sun, Y., Wang, C., Yang, B., Wu, F., Hao, X., Liang, W., Niu, F., Yan, J., Zhang, H., and Wang, B.** (2014). Identification and functional analysis of mitogen-activated protein kinase kinase kinase (MAPKKK) genes in canola (*Brassica napus* L.). *Journal of experimental botany* **65**, 2171-2188.
- Suttipanta, N., Pattanaik, S., Gunjan, S., Xie, C.H., Littleton, J., and Yuan, L.** (2007). Promoter analysis of the *Catharanthus roseus* geraniol 10-hydroxylase gene involved in terpenoid indole alkaloid biosynthesis. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1769**, 139-148.
- Suttipanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S.K., and Yuan, L.** (2011). The Transcription Factor CrWRKY1 Positively Regulates the Terpenoid

- Indole Alkaloid Biosynthesis in *Catharanthus roseus*. *Plant physiology* **157**, 2081-2093.
- Szabó, L.F.** (2008). Molecular evolutionary lines in the formation of indole alkaloids derived from secologanin. *Arkivoc* **2008**, 167-181.
- Takahashi, Y., Soyano, T., Kosetsu, K., Sasabe, M., and Machida, Y.** (2010). HINKEL kinesin, ANP MAPKKs and MKK6/ANQ MAPKK, which phosphorylates and activates MPK4 MAPK, constitute a pathway that is required for cytokinesis in *Arabidopsis thaliana*. *Plant and cell physiology* **51**, 1766-1776.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S.** (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731-2739.
- Tanoue, T., Adachi, M., Moriguchi, T., and Nishida, E.** (2000). A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nature cell biology* **2**, 110-116.
- Tena, G., Asai, T., Chiu, W.-L., and Sheen, J.** (2001). Plant mitogen-activated protein kinase signaling cascades. *Current opinion in plant biology* **4**, 392-400.
- Thagun, C., Imanishi, S., Kudo, T., Nakabayashi, R., Ohyama, K., Mori, T., Kawamoto, K., Nakamura, Y., Katayama, M., Nonaka, S., Matsukura, C., Yano, K., Ezura, H., Saito, K., Hashimoto, T., and Shoji, T.** (2016). Jasmonate-Responsive ERF Transcription Factors Regulate Steroidal Glycoalkaloid Biosynthesis in Tomato. *Plant and Cell Physiology* **57**, 961-975.
- Thamm, A.M., Qu, Y., and De Luca, V.** (2016). Discovery and metabolic engineering of iridoid/secoiridoid and monoterpene indole alkaloid biosynthesis. *Phytochemistry Reviews* **15**, 339-361.
- van der Fits, L., and Memelink, J.** (2000). ORCA3, a Jasmonate-Responsive Transcriptional Regulator of Plant Primary and Secondary Metabolism. *Science* **289**, 295-297.
- Van Der Fits, L., and Memelink, J.** (2001). The jasmonate - inducible AP2/ERF - domain transcription factor ORCA3 activates gene expression via interaction with a jasmonate - responsive promoter element. *The Plant Journal* **25**, 43-53.
- van der Fits, L., Zhang, H., Menke, F.L., Deneka, M., and Memelink, J.** (2000). A *Catharanthus roseus* BPF-1 homologue interacts with an elicitor-responsive region of the secondary metabolite biosynthetic gene Str and is induced by elicitor via a JA-independent signal transduction pathway. *Plant molecular biology* **44**, 675-685.
- van der Heijden, R., Schripsema, J., and Verpoorte, R.** (2004). The *Catharanthus* alkaloids: pharmacognosy and biotechnology. *Curr Med Chem* **11**, 607-628.
- Van Moerkercke, A., Steensma, P., Schweizer, F., Pollier, J., Gariboldi, I., Payne, R., Bossche, R.V., Miettinen, K., Espoz, J., and Purnama, P.C.** (2015). The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpene indole alkaloid pathway in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences* **112**, 8130-8135.
- Van Moerkercke, A., Steensma, P., Gariboldi, I., Espoz, J., Purnama, P.C., Schweizer, F., Miettinen, K., Vanden Bossche, R., De Clercq, R., Memelink, J., and Goossens, A.** (2016). The basic helix-loop-helix transcription factor BIS2

- is essential for monoterpene indole alkaloid production in the medicinal plant *Catharanthus roseus*. *Plant Journal*.
- Vazquez-Flota, F., De Carolis, E., Alarco, A.-M., and De Luca, V.** (1997). Molecular cloning and characterization of desacetylvindoline-4-hydroxylase, a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vindoline in *Catharanthus roseus* (L.) G. Don. *Plant Molecular Biology* **34**, 935-948.
- Veau, B., Courtois, M., Oudin, A., Chénieux, J.-C., Rideau, M., and Clastre, M.** (2000). Cloning and expression of cDNAs encoding two enzymes of the MEP pathway in *Catharanthus roseus*. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1517**, 159-163.
- Vinson, C., Acharya, A., and Taparowsky, E.J.** (2006). Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1759**, 4-12.
- Vinson, C.R., Sigler, P.B., and McKnight, S.L.** (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246**, 911.
- Vogel, M.O., Moore, M., König, K., Pecher, P., Alsharafa, K., Lee, J., and Dietz, K.-J.** (2014). Fast retrograde signaling in response to high light involves metabolite export, MITOGEN-ACTIVATED PROTEIN KINASE6, and AP2/ERF transcription factors in *Arabidopsis*. *The Plant cell* **26**, 1151-1165.
- Vom Endt, D., Soares e Silva, M., Kijne, J.W., Pasquali, G., and Memelink, J.** (2007). Identification of a Bipartite Jasmonate-Responsive Promoter Element in the *Catharanthus roseus* ORCA3 Transcription Factor Gene That Interacts Specifically with AT-Hook DNA-Binding Proteins. *Plant physiology* **144**, 1680-1689.
- Wang, G., Lovato, A., Polverari, A., Wang, M., Liang, Y.-H., Ma, Y.-C., and Cheng, Z.-M.** (2014a). Genome-wide identification and analysis of mitogen activated protein kinase kinase kinase gene family in grapevine (*Vitis vinifera*). *BMC plant biology* **14**, 219.
- Wang, J., Pan, C., Wang, Y., Ye, L., Wu, J., Chen, L., Zou, T., and Lu, G.** (2015). Genome-wide identification of MAPK, MAPKK, and MAPKKK gene families and transcriptional profiling analysis during development and stress response in cucumber. *BMC genomics* **16**, 386.
- Wang, L., Liu, Y., Cai, G., Jiang, S., Pan, J., and Li, D.** (2014b). Ectopic expression of *ZmSIMK1* leads to improved drought tolerance and activation of systematic acquired resistance in transgenic tobacco. *Journal of biotechnology* **172**, 18-29.
- Wang, M., Yue, H., Feng, K., Deng, P., Song, W., and Nie, X.** (2016). Genome-wide identification, phylogeny and expressional profiles of mitogen activated protein kinase kinase kinase (MAPKKK) gene family in bread wheat (*Triticum aestivum* L.). *BMC genomics* **17**, 668.
- Wang, Y., Liu, K., Liao, H., Zhuang, C., Ma, H., and Yan, X.** (2008). The plant WNK gene family and regulation of flowering time in *Arabidopsis*. *Plant Biol* **10**.
- Wang, Z.J., Huang, J.Q., Huang, Y.J., Li, Z., and Zheng, B.S.** (2012). Discovery and profiling of novel and conserved microRNAs during flower development in *Carya cathayensis* via deep sequencing. *Planta* **236**, 613-621.

- Wankhede, D.P., Misra, M., Singh, P., and Sinha, A.K.** (2013). Rice mitogen activated protein kinase kinase and mitogen activated protein kinase interaction network revealed by in-silico docking and yeast two-hybrid approaches. *PLoS One* **8**.
- Weigel, D., and Glazebrook, J.** (2006). Transformation of agrobacterium using the freeze-thaw method. *CSH protocols* **2006**, 1031-1036.
- Wickham, H.** (2009). *ggplot2: elegant graphics for data analysis.* (Springer Science & Business Media).
- Widmann, C., Gibson, S., Jarpe, M.B., and Johnson, G.L.** (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol Rev* **79**, 143-180.
- 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**, 1704-1708.
- Wrzaczek, M., and Hirt, H.** (2001). Plant MAP kinase pathways: how many and what for? *Biology of the Cell* **93**, 81-87.
- Wu, J., Wang, J., Pan, C., Guan, X., Wang, Y., Liu, S., He, Y., Chen, J., Chen, L., and Lu, G.** (2014). Genome-wide identification of MAPKK and MAPKKK gene families in tomato and transcriptional profiling analysis during development and stress response. *PLoS One* **9**, e103032.
- Xu, B.-e., Byung-Hoon, L., Xiaoshan, M., Lenertz, L., Heise, C.J., Stippec, S., Goldsmith, E.J., and Melanie, H.C.** (2005). WNK1: analysis of protein kinase structure, downstream targets, and potential roles in hypertension. *Cell research* **15**, 6-10.
- Xu, J., and Zhang, S.** (2015). Mitogen-activated protein kinase cascades in signaling plant growth and development. *Trends in plant science* **20**, 56-64.
- Xu, J., Li, Y., Wang, Y., Liu, H., Lei, L., Yang, H., Liu, G., and Ren, D.** (2008). Activation of MAP kinase kinase 9 induces ethylene and camalexin biosynthesis and enhances sensitivity to salt stress in *Arabidopsis*. *J Biol Chem* **283**, 26996 - 27006.
- Yamamoto, S., Nakano, T., Suzuki, K., and Shinshi, H.** (2004). Elicitor-induced activation of transcription via W box-related cis-acting elements from a basic chitinase gene by WRKY transcription factors in tobacco. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* **1679**, 279-287.
- Yang, K., Moon, Y., Choi, K., Kim, Y., Eun, M., Guh, J., Kim, K., and Cho, B.** (1997). Structure and expression of the AWI 31 gene specifically induced by wounding in *Arabidopsis thaliana*. *Molecules and cells* **7**, 131-135.
- Yang, Z., Patra, B., Li, R., Pattanaik, S., and Yuan, L.** (2013). Promoter analysis reveals cis-regulatory motifs associated with the expression of the WRKY transcription factor CrWRKY1 in *Catharanthus roseus*. *Planta* **238**, 1039-1049.
- Yin, Z., Wang, J., Wang, D., Fan, W., Wang, S., and Ye, W.** (2013). The MAPKKK gene family in *Gossypium raimondii*: genome-wide identification, classification and expression analysis. *International journal of molecular sciences* **14**, 18740-18757.
- Yoo, S.-D., Cho, Y.-H., Tena, G., Xiong, Y., and Sheen, J.** (2008). Dual control of nuclear EIN3 by bifurcate MAPK cascades in C2H4 signalling. *Nature* **451**, 789.

- Yoo, S.J., Kim, S.-H., Kim, M.-J., Ryu, C.-M., Kim, Y.C., Cho, B.H., and Yang, K.-Y.** (2014). Involvement of the OsMKK4-OsMPK1 cascade and its downstream transcription factor OsWRKY53 in the wounding response in rice. *The plant pathology journal* **30**, 168.
- Yu, F., and De Luca, V.** (2013). ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proceedings of the National Academy of Sciences* **110**, 15830-15835.
- Yu, Z.-X., Li, J.-X., Yang, C.-Q., Hu, W.-L., Wang, L.-J., and Chen, X.-Y.** (2012). The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Molecular Plant* **5**, 353-365.
- Zhai, L., Xu, L., Wang, Y., Huang, D., Yu, R., Limera, C., Gong, Y., and Liu, L.** (2014). Genome-wide identification of embryogenesis-associated microRNAs in radish (*Raphanus sativus* L.) by high-throughput sequencing. *Plant molecular biology reporter* **32**, 900-915.
- Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, 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**, 61-71.
- Zhang, S., and Klessig, D.F.** (2001). MAPK cascades in plant defense signaling. *Trends in plant science* **6**, 520-527.
- Zhang, S., and Liu, Y.** (2001). Activation of salicylic acid-induced protein kinase, a mitogen-activated protein kinase, induces multiple defense responses in tobacco. *The Plant Cell Online* **13**, 1877-1889.
- Zheng, L., Baumann, U., and Reymond, J.L.** (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research* **32**, e115.
- Zhou, C., Cai, Z., Guo, Y., and Gan, S.** (2009). An Arabidopsis mitogen-activated protein kinase cascade, MKK9-MPK6, plays a role in leaf senescence. *Plant physiology* **150**, 167-177.
- Zhou, M., and Memelink, J.** (2016). Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnology Advances* **34**, 441-449.
- Zhu, J.-P., Guggisberg, A., Kalt-Hadamowsky, M., and Hesse, M.** (1990). Chemotaxonomic study of the genus *Tabernaemontana* (Apocynaceae) based on their indole alkaloid content. *Plant Systematics and Evolution* **172**, 13-34.
- Ziegler, J., and Facchini, P.J.** (2008). Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.* **59**, 735-769.

Vita

Name

Priyanka Paul

Education

University of Calcutta, Bethune College (2006-2008)

- Masters in Plant Physiology, Biochemistry and Molecular Biology
- 1st Class (74.4%)

University of Calcutta, Rammohan College (2003-2006)

- Bachelor of Science in Botany (Honours) with Chemistry and Zoology (Minor)
- 1st Class (60.1%)

Fellowships Received

- Obtained National Scholarship on merit in Secondary Examination from State Government of India, 2001
- Obtained National Scholarship on merit cum means in B.Sc. Examination from State Government of India, 2006
- Qualified the Graduate Aptitude Test in Engineering (GATE) in Life Sciences with 93.27 percentile score (February 2009) and obtained junior research fellowship.
- Qualified the National Eligibility Test (NET), jointly conducted by Council of Scientific and Industrial Research (CSIR) and University Grants Commission (UGC) of India for junior research fellowship and eligibility for lectureship (December 2009).

Membership

- American Society of Plant Biologists (ASPB)

Meetings, presentations and workshops

- Participated in KBRIN Next Generation Sequencing Workshop, 2017
- Presented poster entitled “Transcriptional and post-translational regulation of terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* (Madagascar periwinkle)”, Priyanka Paul, Sanjay K. Singh, Barunava Patra, Xueyi Sui, Sitakanta Pattanaik, and Ling Yuan, in American Society of Plant Biologists meeting, Hawaii, 2017.
- Presented poster entitled “Gene regulation of complex alkaloid biosynthesis”, Priyanka Paul, Sanjay K. Singh, Barunava Patra, Xueyi Sui, Sitakanta Pattanaik, and Ling Yuan, in Super Collider 2017, Organized by Kentucky NSF EPSCoR, in Lexington, February 24, 2017.

- Oral presentation entitled “Jasmonate-Inducible Mitogen-Activated Protein Kinase Kinase 1 (MAPKK1) Modulates Terpenoid Indole Alkaloid Biosynthesis in *Catharanthus roseus*” at the IPSS Graduate Student Mini symposium in University of Kentucky, August 2016.
- Oral presentation entitled “Potential role of protein phosphorylation in TIA pathway gene regulation in *Catharanthus roseus*” at the IPSS Graduate Student Mini symposium in the University of Kentucky, August 2014.
- Attended the Southern Section of the American Society of Plant Biologists (SS-ASPB), Lexington 2014 meeting.
- Presented poster entitled “Regulation of gene expression during fruit ripening”, Sudip K. Sinha, Priyanka Paul, Saptadwipa Sanyal, Swarup Roy Choudhury and Dibyendu Narayan Sengupta, in XXXIV All India Cell Biology Conference and Symposium on Quantitative Biology: From Molecules to cells, Bose Institute, India, December, 2010.
- Presented poster entitled “Expression of the *rin* gene involved in developmental programme of tomato fruit ripening”, Sudip K. Sinha, Priyanka Paul, and Dibyendu Narayan Sengupta, in National Symposium on Plant Cell Tissue and Organ Culture: The Present Scenario University of Calcutta, India, March 2010.

Publications

- **Paul, P.**, Singh, S. K., Patra, B., Sui, X., Pattanaik, S., & Yuan, L. (2017). A differentially regulated AP2/ERF transcription factor gene cluster acts downstream of a MAP kinase cascade to modulate terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. **New Phytologist**, 213(3), 1107-1123.
- Shen, E. M., Singh, S. K., Ghosh, J. S., Patra, B., **Paul, P.**, Yuan, L., & Pattanaik, S. (2017). The miRNAome of *Catharanthus roseus*: identification, expression analysis, and potential roles of microRNAs in regulation of terpenoid indole alkaloid biosynthesis. **Scientific Reports**, 7.

Awards received

- Outstanding graduate student scholarship (Sashi Sathaye Memorial Scholarship 2017) given by The Bluegrass Indo-American Civic Society (BIACS), USA.
- UK Graduate School Travel Award for attending American Society of Plant Biologists (ASPB) conference 2017.
- Travel award by Department of Plant and Soil Sciences (PSS) for attending American Society of Plant Biologists (ASPB) conference 2017.

Priyanka Paul