EPIGENETIC BARRIERS TO X CHROMOSOME REACTIVATION DURING REPROGRAMMING TO INDUCED PLURIPOTENCY

&

MOLECULAR EVOLUTION OF MYB GENES IN HAWAIIAN SILVERSWORDS AND CALIFORNIA TARWEEDS

A thesis submitted in partial fulfillment of the requirements

For the degree of Master of Science in

Biology

By

Anupama Reddy Sadhu Dimashkie

May 2013
The thesis of Anupama Reddy Sadhu Dimashkie is approved:

Dr. Maurie Beck
Date

Dr. Cindy Malone
Date

Dr. Ray Hong
Date

Dr. Virginia Oberholzer Vandergon, Chair
Date

California State University, Northridge
ACKNOWLEDGEMENTS

I would like to give a special recognition to Dr. Virginia Vandergon, my graduate advisor, who accepted me into the graduate program and provided me with the guidance and encouragement to be successful throughout my graduate studies in both my work at CSU, Northridge and at UCLA. I would also like to thank Dr. Cindy Malone, my faculty mentor, who gave me an amazing opportunity to continue my graduate studies at UCLA through the CIRM CSUN-UCLA Bridges to stem cell research program. I was privileged enough to have two excellent mentors during my research program at UCLA, Dr. Vincent Pasque and Dr. Kathrin Plath. Both gave me the opportunity to continue my research in a cutting edge field of biology, stem cell research. A special thanks for the two other members of my thesis committee, Dr. Ray Hong and Dr. Maurie Beck.

I would like to thank my family and friends for their continued support throughout my graduate degree. Lastly, I would like to give a special acknowledgement to my husband who has been supportive, encouraging and patient for the past few years while I achieved this graduate degree.
TABLE OF CONTENTS
SIGNATURE PAGE ........................................................................................................... ii
ACKNOWLEDGEMENTS ................................................................................................. iii
LIST OF FIGURES ........................................................................................................... vi
ABSTRACTS .................................................................................................................... vii
SECTION 1: EPIGENETIC BARRIERS TO X CHROMOSOME REACTIVATION
DURING REPROGRAMMING TO INDUCED PLURIPOTENCY ..................................... 1
INTRODUCTION .............................................................................................................. 1
   Nuclear Reprogramming ............................................................................................. 1
   X chromosome inactivation ....................................................................................... 2
   X chromosome reactivation in iPSCs ........................................................................... 3
CHAPTER 1: MATERIALS & METHODS ........................................................................ 5
   Cell culture, cell lines, reprogramming, and isolation of partially reprogrammed cells ...................................................................................................................... 5
   Immunostaining and antibodies ............................................................................... 5
   Bisulfite Sequencing ................................................................................................. 5
   siRNA transfection and chemical treatments ......................................................... 6
CHAPTER 2: RESULTS .................................................................................................... 7
   The role of DNA methylation and Xist RNA in maintaining the Xi in pre-iPs cells .............................................................................................................................. 7
   DNA demethylation treatments reduce DNA methylation of X-linked genes in pre-iPS cells .................................................................................................................... 7
   Isolation of reprogramming intermediates from a reprogramming culture ........... 8
   DNA methylation status of pluripotency and Xi-linked genes in SSEA-1 sorted reprogramming intermediates .................................................................................. 8
CHAPTER 3: DISCUSSION ............................................................................................. 10
SECTION 2: MOLECULAR EVOLUTION OF MYB GENES IN HAWAIIAN SILVERWORDS AND CALIFORNIA TARWEEDS ................................................. 12
INTRODUCTION ........................................................................................................... 12
   Myb genes ............................................................................................................... 12
   MYB gene in plants ................................................................................................ 12
   Subcategories of the myb gene family .................................................................. 13
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene Structure of MYB</td>
<td>14</td>
</tr>
<tr>
<td>Myb complex with myc and wd40</td>
<td>14</td>
</tr>
<tr>
<td>Anthocyanin pathway</td>
<td>15</td>
</tr>
<tr>
<td>Evolution of anthocyanin pathway genes</td>
<td>15</td>
</tr>
<tr>
<td>Regulation of the Anthocyanin pathway</td>
<td>16</td>
</tr>
<tr>
<td><strong>CHAPTER 4: MATERIALS AND METHODS</strong></td>
<td>17</td>
</tr>
<tr>
<td>Primer design</td>
<td>17</td>
</tr>
<tr>
<td>PCR and subcloning</td>
<td>17</td>
</tr>
<tr>
<td>Plasmid and sequence isolation</td>
<td>18</td>
</tr>
<tr>
<td>Sequencing and editing</td>
<td>18</td>
</tr>
<tr>
<td>Evolutionary analysis</td>
<td>18</td>
</tr>
<tr>
<td><strong>CHAPTER 5: RESULTS</strong></td>
<td>19</td>
</tr>
<tr>
<td>Isolation of myb gene</td>
<td>19</td>
</tr>
<tr>
<td>Phylogenetic Analysis</td>
<td>20</td>
</tr>
<tr>
<td>Evolutionary rates of myb genes</td>
<td>20</td>
</tr>
<tr>
<td><strong>CHAPTER 6: DISCUSSION</strong></td>
<td>22</td>
</tr>
<tr>
<td>Myb gene sequences</td>
<td>22</td>
</tr>
<tr>
<td>Evolution of myb genes</td>
<td>22</td>
</tr>
<tr>
<td>Rates of evolution</td>
<td>23</td>
</tr>
<tr>
<td><strong>CHAPTER 7: CONCLUSION</strong></td>
<td>24</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>25</td>
</tr>
<tr>
<td>APPENDIX A:</td>
<td>32</td>
</tr>
<tr>
<td>APPENDIX B:</td>
<td>43</td>
</tr>
</tbody>
</table>
LIST OF FIGURES:

APPENDIX A:

Figure 1. Epigenetic changes leading to XCI. .......................................................... 32
Figure 2. Schematic of X and pluripotent features during reprogramming. .................. 32
Figure 3. Experimental design of Xi-linked GFP reactivation. .................................. 33
Figure 4. Visual illustrations of Xi-linked GFP pre-iPSC cells with treatments. ......... 34
Figure 5. Graphical representation of the percent of Xi-linked GFP reactivation in treated cells. ................................................................................................................... 35
Figure 6. Bisulfite Sequencing ...................................................................................... 36
Figure 7. Proposed model of effects of treatments. ..................................................... 37
Figure 8. Images of reprogrammed cells on days 1,3,5,7,9 and 12. ............................. 38
Figure 9. Experimental design for isolating reprogramming intermediates. ............... 38
Figure 10. Cells were sorted for SSEA1+/− with FACS ............................................ 39
Figure 11. Early and late immunofluorescence of Nanog and Ezh2 ......................... 39
Figure 12. Isolated reprogramming intermediates from a reprogramming culture ...... 40
Figure 13. Bisulfite sequencing analysis of pluripotent genes ................................. 41
Figure 14. Bisulfite sequencing analysis of X-linked genes ........................................ 42

APPENDIX B:

Figure 1. Diagram of enzymatic genes anthocyanin biochemical pathway ............. 43
Figure 2. Schematic of R2R3 myb gene structure ...................................................... 44
Figure 3. Table of sequences downloaded from the NCBI database. .......................... 44
Figure 4. DNA sequence obtained from gDNA from Hawaiian Silverswords and California tarweeds. .......................................................... 44
Figure 5. Multiple sequence alignment made of protein coding sequence obtained from Hawaiian Silversword and California tarweed ......................................... 44
Figure 6. Maximum likelihood tree .......................................................................... 44
Figure 7. Bayesian analysis tree .................................................................................. 44
Figure 8. Ka/Ks ratios for structural and regulatory genes in tarweed and silverswords.51
ABSTRACTS

EPIGENETIC BARRIERS TO X CHROMOSOME REACTIVATION DURING REPROGRAMMING TO INDUCED PLURIPOTENCY

&

MOLECULAR EVOLUTION OF THE MYB GENE FAMILY IN HAWAIIAN SILVERSWORDS AND CALIFORNIA TARWEEDS

By

Anupama Reddy Sadhu Dimashkie

Masters of Science in Biology

The differentiated state of somatic cells can be reprogrammed to induced pluripotent stem cells (iPSCs), leading to the reactivation of the inactive X chromosome (Xi) of female differentiated cells in the mouse. In somatic cells, the Xi is very stable however, when reprogramming mouse embryonic fibroblasts, the Xi reactivates upon completion to a pluripotent state. How somatic character is reversed as well as the sequence of epigenetic events leading to X chromosome reactivation during reprogramming to iPSCs, is not well understood. We found that, the non-coding RNA $Xist$ and DNA methylation are present on the Xi in clonal late reprogramming intermediates, pre-iPSCs, suggesting that these are reversed late during reprogramming. To test whether these function in maintaining Xi repression at this stage of reprogramming we functionally interfered with $Xist$ and DNA methylation using pre-iPSCs genetically deleted for $Xist$ and/or treated with siDNMT1 (DNA methyltransferase 1) and DNA methyltransferase inhibitor 5AzadC (5-aza-2′-deoxycytidine). Reactivation of the X chromosome occurs upon deletion of $Xist$ and both treatments. We used bisulfite sequencing to confirm the loss of DNA methylation of X-linked genes in pre-iPSC and found the loss of DNA methylation on the Xi occurs in the presence or absence of $Xist$ with both 5AzadC and siDNMT1. This suggests DNA demethylation is a late event in Xi reactivation and occurs very late in reprogramming. To further test this, we purified the SSEA1+ population directly from a reprogramming culture that represents a state where the pluripotency marker Nanog is already active but the Xi is not yet reactivated. We show from these intermediates that the promoter region of Nanog is demethylated. However, promoter regions of Xi-linked genes from these intermediates are still completely methylated. Together, our results identify new late stages of reprogramming to iPSCs.
Myb genes are part of an ancient, large, super family of transcription factors and play important roles in different physiological and biochemical processes, for example, the anthocyanin pathway. Anthocyanin’s give specific color phenotype and play key roles in pollination and UV protection for flower specific plants. They also have significant benefits in human health, including fighting against cancer. Many of the structural genes in the anthocyanin pathway, such as chs, dfr and ans, are regulated by MYB genes. Isolation of these genes that are involved in the anthocyanin pathway and bioinformatic analysis will provide a better understanding of how this gene family and its linked connection to the anthocyanin pathway evolved in Hawaiian Silversword and California Tarweed. Using designed degenerate primers in conserved regions of the myb gene, PCR, and subcloning were performed and indicated that there are many myb genes in both Hawaiian Silversword and California Tarweed as expected since myb is a large gene family. Sequences from two species of California tarweed (Madia elegans, Madia gracilis sp.) and two species of Hawaiian silversword (Dubautia linearis, Wilkesia gymnoxiphium sp.) have been isolated using genomic DNA. Gene sequences were spliced and edited with BioEdit, and were aligned with ClustalW 1.8. A maximum likelihood tree was generated with MEGA 5 and a Bayesian tree was generated using Mr.Bayes 3.1.2. The Kimura-2 model was used to calculate distances in tree branches based on the results from a best-fit model test. Phylogenetic analysis shows the diversity of myb genes and how selective pressures at the gene level have played a role in the plants adaptive radiation. The analysis groups one of the tarweeds and one of the silverswords within the same clade with high support on the branching suggesting these myb genes are related. Analyses of the Ka/Ks (non-synonymous/synonymous changes) values detect purifying selection in the silverswords which is consistent with previous studies on the structural genes such as, ANS, in the anthocyanin pathway. Currently, we are working on obtaining intronic regions of the gene which will give us a better understanding of the genetic mechanisms behind adaptive radiation.
INTRODUCTION

Nuclear Reprogramming

The differentiated state of somatic cells is very stable. Once differentiated, cells generally do not change from one cell type to another. However, somatic cells can be reprogrammed by a number of experimental procedures including nuclear transfer, cell fusion and transcription factor overexpression (Gurdon et al. 1958, Blau et al. 1983, Takahashi and Yamanaka 2006, Plath et al., 2011). Reprogrammed cells may become pluripotent, meaning that they can give rise to all the cell types in the body and are then referred to as induced pluripotent stem cells (iPSCs). In this thesis, I will reprogram mouse embryonic fibroblasts (MEFs) by transcription factor overexpression to pluripotency.

The discovery of iPSCs can be used as an important tool to model diseases and to apply in clinical therapeutics (Bock et al. 2011). iPSCs are similar to embryonic stem cells (ESCs) based on function and molecular genetic profile (Sridharan et al., 2009 and Chin et al. 2009). ESCs are pluripotent cells derived from the inner cell mass (ICM) of early mouse or human blastocysts. Similar to ESCs, iPSCs can also form all three germ layers: ectoderm, mesoderm and endoderm (Maherali et al., 2007). iPSCs create the possibility of reprogramming patient specific somatic cells to any cell type, making them an excellent tool for potential cell replacement therapies. iPSCs are generated from somatic cells by overexpressing four transcription factors: Oct4, Sox2, Klf4 and c-Myc (Takahashi and Yamanaka 2006). These transcription factors play key roles in reprogramming the regulation of the pluripotent network. An interesting feature of iPSCs is that only a small percentage of the differentiated cells fully reprogram (Hanna et al., 2009). Less than 0.1% of the somatic cells reprogram to a pluripotent state implying that reprogramming is very inefficient (Wering et al., 2007). However inefficient, some somatic cells do reprogram to pluripotency. Why this process is so inefficient is not well understood.

The stages of early embryonic development have been extensively studied and are well understood. Reprogramming to iPSCs may also be comprised of similar stages in an inverse order but these are unknown. Recent studies have identified at least two waves during reprogramming to pluripotency (Buganim et al., 2012,Golipour et al., 2012, Polo et al., 2012). The first is characterized by a mesenchymal to epithelial transition (MET) and reactivation of the cell adhesion molecule, E-cadherin (CDH1) (Li et al., 2010). The second is marked by reactivation of endogenous pluripotency genes such as Nanog. Cells in the second wave can also be identified by stage specific embryo antigen- 1 (SSEA-1) (Polo et al., 2012, Silva et al., 2009). To determine how the somatic character of a cell can be reversed to a pluripotent state we studied the epigenetic changes that occur during reprogramming to iPSCs. The epigenetic chromatin changes acquired during development may act as epigenetic barriers and cause at least part of the inefficiency of
With chromatin states being stable in somatic cells, how these stable states are reversed during reprogramming is still highly debated and largely unknown. One way to study epigenetic changes that occur during reprogramming is to use partially reprogrammed iPSCs (pre-iPSC) (Sridharan et al., 2009). Pre-iPSC are picked and clonally expanded cells from a reprogramming culture. Although pre-iPSCs share similar characteristics to iPSCs, e.g. somatic genes are silenced; they do not express endogenous pluripotency markers such as Oct4 and Nanog (Silva et al., 2008 and Sridharan et al., 2009). However, pre-iPSCs can still be induced to become fully reprogrammed by small molecule treatment and these cells can be used to examine mechanisms late in reprogramming. It is also suggested that the progression to pluripotency is a stochastic process, where cells are undergoing reprogramming randomly (Hanna et al., 2009). By looking at chromatin changes it may be possible to achieve a better understanding of the steps to reprogramming and address that chromatin changes occur in a stochastic manner.

X chromosome inactivation

During female mammalian development, X chromosome inactivation (XCI) acts as a means for gene dosage compensation between males (XY) and females (XX) (Lyon, 1961). Females have one active X chromosome (Xa) and one inactive X chromosome (Xi), to match the gene dosage of males. While this is a naturally occurring process, the complete molecular mechanisms are not clear. It is known that Xi during development occurs in two forms in female mice: imprinted inactivation and random inactivation (Augui et al., 2011). First, in the 4 cell stage, the paternal X (Xp) is preferentially inactivated in the extra-embryonic tissues (trophectoderm and placenta) by imprinting (Augui et al., 2011). Second, the inactivation of Xp is maintained until the blastocyst stage development when the Xp is reactivated in the ICM but remains inactive in the extra-embryonic tissue (Augui et al., 2011). The cells in the ICM then undergo random XCI upon differentiation to specific cell lineages (Gartler et al., 1976). The silenced stage of the Xi is then maintained throughout life and subsequent cell divisions. Inactivation of the X chromosome involves a series of epigenetic changes and occurs in two phases: initiation of Xi followed by maintenance of Xi. Initiation begins with expression of the Xist gene, a long non-coding RNA, which coats the Xi in cis and initiates a cascade of events leading to chromosome wide gene silencing, including silencing of X-linked genes (Plath et al., 2003, Silva et al., 2003). After the initiation by the expression of Xist RNA, Polycomb repressive complex 2 (PRC2) is recruited, proceeded by the maintenance phase with incorporation of repressive element histone variant macroH2A and DNA methylation (Figure 1).

Polycomb group proteins have known function in chromatin organization and are crucial in the epigenetic regulation of gene expression during development (Surface et al., 2010, Posfai et al., 2012). The primary role of PRC2 is in post-transcriptional modifications of histones. Previous studies have shown PRC2 proteins are localized to the Xi sparking interest in the role of PRC2 and its relationship to the Xi (Mak et al., 2002). PRC2 contains several proteins: EED, SUZ12 and EZH2. More specifically, EZH2 is the catalytic subunit of this complex and trimethylates histone 3 lysine 27.
(H3K27me3) on the Xi. EZH2 and SUZ12 are recruited by Xist RNA to the Xi and can be seen enriched on the Xi of cells induced to the differentiated state (Plath et al., 2003, Silva et al., 2003). Therefore, EZH2 and SUZ12 can be used as markers for Xi during development, as they are transiently expressed (Figure 1). These factors are attributed to the maintenance phase of the Xi state in differentiated cells.

PRC2 recruitment is followed by co-localization of histone variant macroH2A with the Xi. MacroH2A is a variant of histone 2A and contains a non-histone dense region that may facilitate the binding of RNA (Pehrson 1998). For example, this region could mediate an interaction with Xist RNA, although Xist RNA coating proceeds macroH2A recruitment (Coztansi et. al, 1999). After Xist upregulation and recruitment of PRC2, DNA methylation is active which is associated with long term maintenance of the Xi (de Napoles et al., 2004).

DNA methylation is a classic epigenetic mark that is largely associated with gene silencing (Holliday et al., 1975, Riggs 1975). Most previous work in DNA methylation has concentrated on the 5’-methylation of the nucleotide cytosine in the CpG dinucleotide context. In mammals, CpG sites are associated with dense regions known as CpG islands (CGI) and are usually found in promoter regions of genes (Jones, 2012). A methylated CGI promoter region generally correlates with transcriptional inactivity, acting as a repressor. Previous work has shown that DNA methylation on the Xi is maintained by DNA methyltransferase-1 (DNMT1) (Beard et al., 1995). Thus, the removal of DNMT1 by small interfering molecules (siDNMT1) or DNMT1 deficient mice causes the loss of DNA methylation (Sado et al., 2000). It has also long been known that 5-aza-deoxycytidine (5AzadC) is a chemical responsible for inhibiting DNA methyltransferases which reduces DNA methylation levels (Jones et al., 1983). Other studies have also shown chemical 5-aza-deoxycytidine (5AzadC) induces the loss of DNA methylation in mouse embryonic fibroblasts and iPSCs (Csankovszki et al., 2001 and Kim et al., 2010). For these reasons, methylation is important in maintaining the silenced state of the Xi.

Together, non-coding RNAs, Polycomb proteins, histone variants and DNA methylation all play an important role in inactivating and maintaining the Xi.

**X chromosome reactivation in iPSCs**

The Xi in female mouse cells is a good tool to look at epigenetic changes during reprogramming (Eggan et al., 2000). It is well known that in female somatic cells one X is active and the other is inactive whereas in embryonic stem cells both X chromosomes are active. XCI is induced during differentiation. Remarkably, while the Xi is maintained in all somatic cells, it is fully reactivated during reprogramming to iPSCs in the mouse system (Maherali et al., 2007). The process of reactivation is poorly understood. Based on our understanding of XCI during development, we can ask in what order are Xi chromatin features reversed during reprogramming. Our preliminary model during reprogramming reveals that some features of Xi are in the reverse order of developmental XCI while other features deviate from this order. For example, loss of Xist RNA, macroH2A and EZH2/SUZ12 are reversed sequentially while biallelic expression of X-linked is reversed late in reprogramming (Figure 2).
In this study, I first asked how \textit{Xist} RNA and DNA methylation contribute to the maintenance of the Xi in pre-iPSCs. In order to do this I characterized the effect of \textit{Xist} RNA deletion, DNMT1 knockdown and 5AzadC on DNA methylation in pre-iPSCs. Second, I asked when during reprogramming to iPSCs does DNA demethylation of the Xi occur. Using a cell sorting strategy to enrich reprogramming intermediates, I found that DNA demethylation of the Xi occurs after DNA demethylation of the pluripotency gene \textit{Nanog}, which is very late during reprogramming. Together, this work helps to understand the stages and mechanisms of epigenetic reprogramming of the Xi. Our results first conclude Xi-reactivation is a step-wise process with many epigenetic stages (Figure 2). Second, I also determined DNA methylation is important for maintaining the silent state of the Xi until the end of reprogramming. Lastly, Xi reactivation occurs after the activation of pluripotency genes.
CHAPTER 1: MATERIALS & METHODS

Cell culture, cell lines, reprogramming, and isolation of partially reprogrammed cells

For reprogramming, MEFs (mouse embryonic fibroblasts) at passages 1–3 were infected at 50% confluency overnight with pooled viral supernatant generated by transfection of PlatE cells with individual pMX vectors encoding Oct4, Sox2 and Klf4 as described previously (Sridharan et al., 2009).

Isolation of pre-iPSCs was performed as previously described (Sridharan et al., 2009). Briefly, MEFs (Xi2loxXist/GFP XaΔ(1lox)Xist) were reprogrammed as stated above, except that a retrovirus encoding cMyc was included in the reprogramming cocktail and reprogramming was performed in complete ESC (embryonic stem cell) media containing 10% FBS. GFP negative ESC-like colonies were picked at day 20, clonally expanded, and characterized as described previously (Sridharan et al., 2009).

Immunostaining and antibodies

Cells were grown on 0.05% gelatinized glass coverslips, washed with 1X PBS, then fixed in PBS containing 4% paraformaldehyde for 10 min at room temperature, permeabilized with PBS containing 0.5% Triton-X100 for 5 minutes and washed with PBS containing 0.2% Tween20 (PBST) for 5 minutes. Cells were blocked for 30 min with 10% goat or donkey serum in 1X PBS containing 0.1% Triton-X and 0.1% fish skin gelatin. After incubation with primary antibodies for 1-3 hours to overnight in blocking solution, cells were washed 3 times with PBST and incubated with appropriate fluorophor-labeled secondary antibodies (Invitrogen) in blocking solution for 30 minutes in the dark, washed 3 times with PBST, stained with DAPI and mounted in AquapolyMount (Polysciences Inc.). Primary antibodies used for immunostaining were: NANOG (ab80892) and EZH2 (BD 612666).

Bisulfite Sequencing

Bisulfite treatment of 200ng-1ug gDNA was performed with the EpiTect Bisulfite Kit (Qiagen) according to manufacturer’s instructions. Bisulfite converted primers for X-linked genes Atrx (5’-TTGGTTTGGAATTTTTTGAT-3’ and 5’-AACAACA ACAACAACAACAACACAC-3’), Rnf12 (5’-TTGGTTTGGAATTTTTTGAT-3’ and 5’-TTAATAACTCCTCCAAAATTTCCTC-3’) and pluripotency genes Nanog (5’-GATTITTTGAGGATTTTTATCAGAT-3’ and 5’-ACCCAAAA AACCCACACTCATACTCAATA-3’), Dppa3 (5’-AATATAAAAAGAGATTTC GATAAAATTGGATTTTG-3’ and 5’- TACCCCCAACTCCTTTTATATATATAAA-3’) were designed using MethPrimer and HotStar Taq (Qiagen) was used to amplify promoter specific CpG islands by PCR. Each 25ul PCR reaction included 1.25ul of 10uM gene specific forward primer (IDT), 1.25ul of 10uM of gene specific reverse primer (IDT), 0.625ul of 10mM dNTPs, 2.5ul of 10X Buffer with MgCl2 (Qiagen), 1.5ul of
50mM MgCl2, 5ul of 5X Q solution (Qiagen), 0.5ul HotStar Taq polymerase (Qiagen),
2ul of extracted gDNA and 10.37ul of pico pure H2O. PCR cycles for Atrx were, 1 cycle
95° C for 15 minutes, 40 cycles of 95° C for 30 s then 60° C for 30 s, 1 cycle 72° C for 1
min then 72° C for 2 mins. PCR cycles for Rnf12 were, 1 cycle 95° C for 15 minutes, 40
cycles of 95° C for 30 s then 62° C for 30 s, 1 cycle 72° C for 1 min then 72° C for 2
mins. PCR cycles for Nanog were, 1 cycle 95° C for 15 minutes, 40 cycles of 95° C for
30 s then 52° C for 30 s, 1 cycle 72° C for 1 min then 72° C for 2 mins. PCR cycles for
Dppa3 were, 1 cycle 95° C for 15 minutes, 40 cycles of 95° C for 30 s then 61° C for 30
s, 1 cycle 72° C for 1 min then 72° C for 2 mins. Amplified products were run on a 1%
agarose gel and gel extracted using a razor blade. Extracted products were purified using
Macherey nagel Nucleospin PCR purification kit 740609.50, cloned into the pJET-vector
(Fermentas), transformed using MACH T1 chemically competent cells (Invitrogen) and
plated on carbenicillin agar plates grown overnight in a 37°C incubator. Random clones
were picked for colony PCR. Each 50ul PCR reaction included one picked colony, 0.1ul
of 100uM pJet forward primer (Fermentas), 0.1ul of 100uM of pJet reverse primer
(Fermentas), 1ul of 10mM DNTPs, 5ul of 10X Buffer with MgCl2 (Qiagen), 0.5ul Taq
polymerase (Qiagen), 1ul formamide (ultrapure) and 42.30ul of pico pure H2O. Colony
PCR cycle was 1 cycle 95° C for 3 minutes, 30 cycles of 94° C for 30 s then 60° C for 30
s, 1 cycle 72° C for 1 min then 72° C for 5 mins. sequenced using a DNA sequencer
(Applied Biosystems 3730xl DNA Analyzer, Retrogen). Sequenced products were
aligned using Bioedit 7.1.3 with a ClustalW a gap penalty of 100 and gap extension
penalty of 10 and compared CpG sites of each bisulfite converted sequences.

siRNA transfection and chemical treatments

siRNAs targeting Dnmt1 were obtained from Dharmacon (Dnmt1-1:056796-09 and
Dnmt1-3:056796-11). 20nM of siRNA were transfected into pre-iPSCs or
reprogramming cultures with Lipofectamine RNAiMax (Invitrogen) as described by the
manufacturer. Confirmation of knockdown efficiency of Dnmt1 was determined by
protein level. Primary antibodies were:DNMT1 (CosmoBio: BAM-70-201-EX) and
GAPDH (Fitzgerald: 10R-1178). 5AzadC (Sigma) was used at a concentration of 0.3uM,
doxycycline at 2ug/ml, and tamoxifen at 5mM.
CHAPTER 2: RESULTS

The role of DNA methylation and Xist RNA in maintaining the Xi in pre-iPS cells

Under normal conditions, the X inactivation is initiated by Xist RNA and maintained by epigenetic features such as DNA methylation (Beard et al., 1995). Previous experiments showed that a conditional knockout of Xist along with DNA demethylation had the most effect in reactivation of the Xi in somatic cells (Csankovszki et al., 2001). These data suggest that in somatic cells, Xist and DNA methylation work together synergistically to maintain the Xi (Csankovszki et al., 2001). In order to test whether this is true during reprogramming we used a similar strategy. Upon the deletion of Xist and treatment with DNA demethylation agents along with using a Xi-linked-GFP reporter we measured the Xi reactivation after the treatments by measuring the population of GFP positive cells (Figure 3). The three day treatments included deleting Xist using dox-inducible Cre 2lox/lox Xist female MEFs and/or 5-aza-2’-deoxycytidine (DNA methyltransferase inhibitor) and/or siDNMT1 (small interfering DNA methyltransferase 1) to reduce DNA methylation (Figure 3).

We noticed the greatest effect in conditions where Xist was deleted and DNA demethylation treatments were used. Figure 4, visually illustrates GFP fluorescence under the various conditions, in the presence or absence of Xist and with or without 5AzadC and/or siDNMT1. The morphology and growth of pre-iPSCs in both DNA demethylation treatments and in the absence or presence of Xist look similar. Xi-GFP begins to reactivate in cells with 5AzadC or siDNMT1 and in the presence of Xist however few cells reactivate the Xi. Cells given both DNA demethylation treatments, 5AzadC and siDNMT1, with deletion of Xist have the most Xi-GFP reactivation compared to both DNA demethylation treatments in the presence of Xist (Figure 5). Cells in the absence or presence of Xist and with either 5AzadC or siDNMT1 had less than 4% Xi-GFP reactivation. Where cells in the presence of Xist and both 5AzadC and siDNMT1 had between 12-15% Xi-GFP reactivation. Cells in the absence of Xist with both DNA demethylation treatments had the greatest effect, with a 2-fold increase of Xi-GFP reactivation (Figure 6). These data indicate that Xist and DNA methylation synergistically maintain the inactive state of the X chromosome in female pre-iPSCs (Figure 7). Our findings also indicate Xist is independent of DNA demethylation in pre-iPSCs.

DNA demethylation treatments reduce DNA methylation of X-linked genes in pre-iPS cells

To look at the DNA methylation pattern in pre-iPSCs, I used bisulfite converted genomic DNA, from the conditions previously described above. Primers were designed around highly dense CGI regions of X-linked gene Atrx. We compared the DNA methylation of each condition with the control. In the control samples where DNA demethylation treatments were not used and in the presence or absence of Xist we expect to see 50% of the clones methylated and 50% of the clones unmethylated. DNA sequences obtained come from clones where each sequence represents alleles derived from either the Xi (methylated) or Xa (unmethylated) (Figure 7). In cells with the
deletion of Xist and 5AzadC or siDNMT1 treatments have a similar pattern to the control. These data show the Xi is still methylated. When cells are deleted for Xist and both 5AzadC and siDNMT1 treatment were used this led to a substantial decrease in methylation of the Xi (Figure 7). The treatments are validated only when they are combined. Hence, the deletion of Xist does not directly affect the methylation pattern of X-linked genes (Figure 7). Our findings suggest DNA methylation is crucial in maintaining Xi-linked genes and DNA demethylation in pre-iPSCs is only efficiently erased by combined 5AzadC and siDNMT1 treatments.

**Isolation of reprogramming intermediates from a reprogramming culture**

To understand the intermediate stages in reprogramming we chose to isolate intermediates from a reprogramming culture. We designed a strategy to isolate such intermediates by using female stem cell cassette containing Oct4, Sox2, Klf4 and c-Myc (stemmca). MEFs which contain a tetO inducible promoter that is doxycycline (dox) inducible by expression of rtTA from the Rosa 26 locus. These cells can be induced to reprogram by simple addition of dox to the culture medium. We cultured these cells in feeder-free conditions in the presence of dox to induce reprogramming. At day 9, reprogramming cultures were disassociated to single cells and stained for SSEA-1. Cells were sorted by fluorescence activated cell sorting (FACS) (Figure 8). This procedure enriches for reprogramming intermediates (Polo et al., 2012). Images of reprogramming days, 1, 3, 5, 7, 9 and 12 show the progression of iPSC colonies forming (Figure 9). The first colonies appear on day 7 and by day 12 there are many more and expanding iPSC colonies (Figure 9). The growth is typical for a reprogramming from stemmca MEFs. By cell sorting at day 9 we obtained a population of cells that are SSEA-1 positive and these were NANOG positive and poised to reprogramming while still maintaining the Xi. Only a small portion of the cells from a reprogramming culture were SSEA-1 positive (0.064%), however, there were enough cells to fully reprogram (Figure 10).

To ensure cells were true reprogramming intermediates while maintaining the Xi we assayed the re-plated cells by immunostaining for NANOG and EZH2 Xi enrichment. Early after re-plating NANOG positive colonies were seen with EZH2 Xi enrichment. Late after re-plating, NANOG positive cells with no EZH2 Xi enrichment were seen (Figure 11). The number of NANOG positive colonies increases significantly from day 7 to day 18 as reprogramming colonies (Figure 12a). The question then is whether or not the Xi is reactivated in these NANOG positive colonies. Colonies on day 9 after re-plating are 100% NANOG positive and EZH2 Xi enriched. By day 15, 50% of the NANOG positive colonies have now lost the enrichment on the Xi and by day 18 all of the colonies have lost EZH2 Xi enrichment (Figure 12b). These results suggest that sorting for cells at day 9 for SSEA-1 from a reprogramming culture enriches for true reprogramming.

**DNA methylation status of pluripotency and Xi-linked genes in SSEA-1 sorted reprogramming intermediates**
We sought to determine the DNA methylation status of pluripotency and X-linked genes in SSEA-1 FACS sorted intermediates. We previously showed that Nanog was positive in all re-plated reprogramming intermediate colonies (Figure 12b). In order to determine the methylation state of genes, we bisulfite converted genomic DNA from the sorted samples and then amplified, cloned and sequenced clones using specific primers designed in promoter regions with high CGIs of each gene. The methylation state of Nanog in this intermediate stage shows the promoter region of Nanog in SSEA-1 negative cells is fully methylated and had the same pattern as in MEFs. The SSEA-1 positive cells are fully demethylated for the Nanog promoter region, which is consistent with the fact that all the re-plated cells are Nanog positive (Figure 13a). The methylation state of Nanog in SSEA-1 positive cells fully demethylated and is comparable to the methylation state in embryonic cells confirming in terms of DNA methylation pattern that these cells are on their way to pluripotency. We found the methylation state of Dppa3/Stella to also confirm the pluripotency of these sorted cells. The methylation in the promoter region of Dppa3 shows SSEA-1 negative cells are unmistakably fully methylated, as also seen in the MEF control, while there is a significant loss of methylation in SSEA-1 positive cells (Figure 13b). We conclude, that these cells may be transitioning to a demethylated state again consistent with that there are reprogramming intermediates.

I specifically looked at two X-linked genes to determine the methylation state of the Xi. In the sorted population of cells for SSEA-1 positive we expected that the Xi was still methylated in CGI. We first examined a dense CGI region about 300 nucleotide base pairs downstream of the transcriptional start site (TSS) of X-linked gene Atrx. We found that the Xi was still methylated in the sorted SSEA-1 positive cells and the SSEA-1 negative cells. The DNA methylation was maintained until late stages of reprogramming in which pluripotency gene reactivation and DNA demethylation already occurred (Figure 14). A similar result is found in the X-linked gene Rnf12. The Xi was still methylated in Rnf12 and bisulfite sequencing showed 50% of the clones were methylated and 50% of the clones were unmethylated (Figure 14). These data uncovered that one of the key stages in reprogramming to pluripotency is DNA demethylation of the inactive X chromosome and occurs late in reprogramming.
CHAPTER 3: DISCUSSION

The molecular mechanisms behind the intermediate stages of reprogramming are still unknown. Although much is known of intermediate stages to differentiation we cannot assume that the stages are merely reversed during reprogramming. In order to understand the dynamics behind reprogramming we looked at the intermediate stages cultures that were poised to fully reprogram. There are many components to reprogramming. This study focused on identifying one of the major epigenetic components in reactivation of the X chromosome during nuclear reprogramming to iPSC, DNA demethylation. We find DNA methylation of Xi is removed very late in this process. DNA methylation in female differentiated cells is known to maintain the Xi (Csankovszki et al., 2001). We can use these previous findings as a model to address DNA demethylation of the Xi during reprogramming.

First, we looked at pre-iPSCs to address the question of DNA methylation in intermediate stages of reprogramming. We used pre-iPSCs that were isolated, expanded and clonally derived from a reprogramming culture. Pre-iPSCs may eventually fully reprogram or they will remain in a pre-iPS state. From these cells, we were able to treat them with DNA demethylation agents and measure X reactivation. X reactivation was most successful when Xist was deleted in combination with both DNA demethylation agents, siDNMT1 and 5Azadc (Figure 4). The two fold increase or Xi reactivation that we observed means together DNA methylation along with Xist maintain the Xi in pre-iPSCs. When we look at the actual DNA methylation pattern for X-linked genes in pre-iPSCs we showed X-linked genes on the Xi are still fully methylated. Whereas, in cells with or without Xist and both DNA demethylation treatments there was a significant reduction in DNA methylation, about 50%, in the Xi (Figure 5). Although there was no notable difference in samples with or without Xist, this means DNA demethylation was not dependent on Xist. The main conclusion from this data was that DNA demethylation of the Xi in pre-iPSCs occurs late in reprogramming.

However, the epigenetic state of expanded pre-iPSCs may differ from true reprogramming intermediates. The data above shows that we can isolate reprogramming intermediates from a reprogramming culture and these intermediate cells were poised to reprogram. Although few cells were isolated after cell sorting these cells can be cultured and used as an excellent tool for study. Early in reprogramming NANOG is initiated and EZH2 Xi enrichment was still on the Xi. Late in reprogramming NANOG is still upregulated and we no longer see EZH2 Xi enrichment. We see the same cascade of events in this reprogramming culture obtained from re-plating SSEA-1 positive sorted cells, where there was progression from an EZH2 Xi enriched state to X reactivation and loss of Xi PRC2 enrichment (Figure 9). On day 7 of reprogramming there are only NANOG positive colonies which are EZH2 Xi enriched, however as reprogramming continues there was a progressive loss of EZH2 Xi enrichment and all colonies maintain NANOG upregulation (Figure 10). This shows evidence that colonies progress from a population having EZH2 Xi enrichment to no longer having this enrichment while still maintaining NANOG expression.
Having established that these isolated cells are true reprogramming intermediates, I looked at DNA methylation of pluripotent genes. The low percentage of Nanog methylation is akin to the methylation state of Nanog in embryonic stem cells for SSEA-1 sorted cells. I also found the DNA methylation pattern for Nanog in SSEA-1 negative sorted cells was very high (Figure 11). The pattern was similar to mouse embryonic fibroblasts, therefore I conclude these SSEA-1 negative cells lack DNA demethylation and may be at an earlier stage of reprogramming than the SSEA-1 positive cells. The demethylation of the Nanog promoter establishes that mechanisms behind DNA demethylation are initiated at this early stage of reprogramming however in a specific manner. Immunofluorescence of NANOGL in sorted cells confirms protein expression. Thus DNA demethylation was possibly occurring in a step-wise manner. The results we found in Nanog agree with the hypothesis that Nanog is transcriptionally activated. The proposed model of stages of reprogramming suggests that Nanog is activated first followed by Dppa3. Analyzing the DNA methylation results from bisulfite sequencing I find an interesting result. Dppa3 was significantly demethylated in SSEA-1 positive cells however, not fully demethylated. This partial demethylation suggests an intermediate stage of DNA demethylation and the pattern may gradually become more similar to that of embryonic stem cells as reprogramming progresses (Figure 11).

It is important to determine the DNA methylation state of pluripotency genes and whether or not the SSEA-1 positive cells are reprogramming intermediates. We used these results as controls when investigating the DNA methylation of X-linked genes in SSEA-1 sorted cells. The lack of DNA demethylation in X-linked genes of reprogramming intermediates as previously described, suggests that the Xi is methylated and DNA demethylation occurs late in reprogramming. These data also suggest DNA demethylation of the Xi may not occur in an inverse order to development as one would expect. Whether DNA demethylation is active or passive is not addressed in this study. Active DNA demethylation refers to the enzymatic removal of a methyl group and passive DNA demethylation implies inhibition of DNMT1 during DNA replication. One way to address whether DNA demethylation is active or passive may be to induce cell cycle arrest and knockdown candidate enzymes that may be required for active DNA demethylation, then ask what the methylation state of X-linked genes is. For the purposes of our study, the findings ultimately show that DNA demethylation of the Xi occurs late in reprogramming downstream of Nanog promoter demethylation and reveals that DNA demethylation is therefore a staged process.

When comparing the experiments done in pre-iPSCs and in SSEA-1 sorted cells, both uncover that DNA demethylation occurs late in reprogramming. Both experiments also confirm Nanog occurs early in reprogramming and DNA demethylation is late in reprogramming. It will be of high interest to further dissect the stages of reprogramming to iPSCs and this will greatly aid to unravel the mechanisms that regulate reprogramming and maintain cellular identity.
SECTION 2:
MOLECULAR EVOLUTION OF MYB GENES IN HAWAIIAN SILVERSWORDS AND CALIFORNIA TARWEEDS

INTRODUCTION

Myb genes

Myb genes are important transcription factors that regulate many functions in organisms, from cell proliferation to apoptosis (Lipsick, 1996). In most cases, MYB is seen as a gene activator to turn on target genes in pathways, rather than acting as a repressor. The gene was first identified as a proto-oncogene in mammals because of its ability to control cell proliferation (Lipsick, 1996). Myb genes are ancient genes and have roots across taxa from slime molds and yeast to plants and animals (Jin et al., 1999). An example of the functional diversity of myb is, the three v-myb gene which has been identified in vertebrates functioning in cell proliferation, cell differentiation and apoptosis (Klempnauer et al. 1982; Yanhui C. et al. 2006). Since this gene family not only has various functions but exists among various organisms mybs evolutionary history has been extensively studied (Lipsick JS. 1996).

The genes’ main role is to regulate transcription of enzymatic structural genes in various biochemical pathways (Urao et al. 1993; Daniel et al. 1999; Sugimoto et al. 2000; Hemm et al. 2001; Stockinger et al. 2001; Vailleau et al. 2002; Abe et al. 2003; Denekamp and Smeekens, 2003; Nagaoka and Takano, 2003; Yanhui C. et al. 2006). Some of these pathways are critical for most biological functions in organisms and particularly significant in development and metabolism (Ito et al., 2001; Araki et al., 2004). For example, in rice myb genes transcriptionally regulate stress induced pathways that maintain a plant’s ability to survive conditions such as cold and dehydration (Yang et al. 2012). In this way, examining myb genes can also lead to a better understanding of gene regulation and pathway evolution. In this study we use plants as a model system to understand pathway evolution of gene families. It is interesting that myb genes play many roles in regulation and therefore by isolating sequences and comparing evolution rates we will be able to obtain evolutionary analysis of this gene family.

MYB gene in plants

The widespread interest in myb encouraged researchers to look at these genes in plants to identify the function and role of myb genes. What was found was myb genes in plants are part of a large family of transcription factors and contain multiple subcategories that vary vastly in function. The C1 gene was the first myb to be isolated from maize in the late 1980’s which showed similarities at the DNA sequence level between mybs in plants and animals (Paz-Ares et al. 1987). Further investigation proved that there are conserved regions of the gene among species and in some plants myb genes have undergone gene duplication events. (Braun and Grotewold 1999; Rabinowicz et al. 1999). An example of a gene duplication event is when an organism changes from
diploid to tetraploid resulting in all genes duplicating. These polyploidy events are specific to plants whereas in mammals’ chromosome duplications do not usually lead to viable offspring. In some cases in plants, gene duplication events can lead to novel functions of genes as well as pseudogenes. For instance, myb genes not only control primary metabolic pathways that are responsible for growth, development and stress response (Urao et al., 1993; Daniel et al., 1999; Sugimoto et al., 2000;) but also secondary metabolic pathways that mediate important ecological processes such as pollination. In plants, secondary metabolic pathways are often important for plant survival when under environmental stress.

There are hundreds of mybs found in plants, many of which overlap in function. A majority of myb genes in Arabidopsis have been isolated and it was discovered that there are 198 myb genes are in the superfamily of which 126 are R2R3 myb, 5 are R1R2R3 myb and 64 are myb-related genes (Yanhui C. et al. 2006). Myb genes have also been isolated in other model plants including maize, perilla, snapdragons, potato and tobacco (Pattnaik et al. 2010). Many of which are part of subcategory, R2R3 myb transcription factors known to regulate the biochemical pathways (Mol et al. 1998; Koes et al. 200; Pattnaik et al. 2010). In addition to the work done in plants, myb genes have also been isolated in animals where they have been extensively studied to isolate and characterize the role of MYB genes in cancer (Ganter and Lipsick 1999).

Subcategories of the myb gene family

The myb gene family is divided into three sub-families, which include, R2R3, R1R2R3 and MYB-related genes (Rosinski, J.A. and Atchley 1998; Jin and Martin 1999; Stracke et al., 2001) They are divided in these groups based on a highly conserved repeat region found in the N-terminus of the gene. R1R2R3 myb genes are found to have general influences in regulation of pathways involved in cell cycle as well as stress response systems for plants (Urao et al., 1993; Daniel et al., 1999; Sugimoto et al., 2000). For example, by increasing tolerance to drought and freezing conditions (Xiaoyan et al. 2007). As seen in Arabidopsis, the number of R1R2R3 is small compared to the other two subcategories of mybs (Yanhui C. et al. 2006). Myb-related genes refer to genes that have at least one of the highly conserved repeats. These genes again can be involved in many different functions and are sometimes homologous to functions of R2R3 mybs. The most significant example is, C1 R2R3 myb in Maize for example, was found to regulated the anthocyanin pathway, much like many of the R2R3 mybs (Paz-Ares et al., 1987).

In plants, R2R3 mybs only have two repeats and most likely lost the first repeat. The reason for the loss of repeat one is unknown; nevertheless R2R3 myb is the largest subcategory with the greatest number of functions. This subfamily, act as a transcription factor for regulating genes which are involved in the control of cell morphogenesis, regulation of secondary metabolism, control of cell cycle, light and hormone signaling pathways and the regulation of meristem formation and floral seed development (Yanhui C. et al. 2006). In addition, the R2R3 myb genes regulate the enzymatic structural genes in the anthocyanin pathway, which is important for flower color, fruit color and UV protection of leaves in plants (Yanhui C. et al. 2006).
Gene Structure of MYB

The gene structure of R2R3 \textit{myb} consists of the repeat regions in the N-terminus end of the gene. The repeat regions are highly conserved motifs which constitute the DNA binding domains. In R2R3 \textit{myb}, the DNA binding domain can be found in repeat three and is part of the helix-turn-helix structure of the protein (Jia et al. 2004). Each repeat is comprised of about 52 amino acids which have highly conserved regions and regions of similarity. Within the repeat there are evenly spaced tryptophans, which are hydrophobic, and establish the helix-turn-helix structure of the protein (Ogata et al. 1994; Lipsick 1996; Kranz et al. 1998). The basic-helix-turn-helix structure is the most conserved region of the gene and can be found as the last two helices in each of the repeats (Ogata et al., 1994). In the R2R3 gene, an amino acid signature of [DE]Lx2[RK]x3Lx6Lx3R, can be found in the third repeat. This signature in \textit{Arabidopsis} is where the basic-helix-loop-helix (\textit{myc} transcription factor) binds (Pattnai et al. 2010). The C-terminus end of the gene has few small domains of homology with percent similarity being low compared to the N-terminus. Functionality of the protein may be derived from the C-terminus end.

R2R3 \textit{myb} is found to have three exons and two introns with an average gene size totaling approximately 1500bp. The size of exon one and exon two range from 150 to 200 bp. Exon three is usually double the size of the first two exons (Lin Wang et al. 2010). Part of R2 is found in exon one and exon two is the remaining part of the R2 repeat and the first part of the R3 repeat. The last exon consists of the last part of R3 repeat and the C-terminus end of the gene. The introns are known to have varying size. Intron one is small with up to 100 bp where intron two can range anywhere from 80 to 2000 bp. The varying intron size has little to no link to functionality of the gene when using phylogenetic analysis to compare (Lin Wang et al. 2010).

\textit{Myb} complex with \textit{myc} and \textit{wd40}

Another interesting aspect of the R2R3 \textit{myb} is that the gene forms a complex with \textit{myc}, a basic-helix-loop-helix transcription factor, and the \textit{wd40} transcription factor (Hirichi et al. 2011). When bound together this complex binds to other genes and regulates transcription in plants. This however is not the same in mammals, where \textit{mybs} can work alone in regulation. The \textit{myc} gene transcription factor is much larger in size compared to \textit{myb} and has highly conserved motifs much like \textit{myb}. \textit{Wd40} is a smaller gene in size and consists of a highly conserved tryptophan and aspartic acid repeat 40 times. WD40 serves as a docking platform for other transcription factors (Smith et al. 1999; Van Nocker and Ludwig 2003). Together they form a tertiary protein conformation which is important to binding and activating genes.

The tertiary complex itself is what regulates genes in a specific pathway. However, the accumulation of each of these transcription factors relies on the recruitment
and interaction of the complex. Knocking out one of these genes effectively renders the complex nonfunctional and eliminates the transcription of target genes (Hirichi et al. 2011). The complex needs to be intact for there to be functionality of transcribing genes in a specific pathway, although in some cases MYB seems to be the determining factor of whether or not binding to the target gene will occur.

**Anthocyanin pathway**

The anthocyanin pathway is a secondary metabolic pathway found in flowering and fruit plants. The pathway is part of the larger phenylpropanoid biochemical pathway. This pathway has been well documented and studied over a long period of time (Grotewold, 2006; Winkel-Shirley, 2001). Various studies of flower coloration date back to the early 1900’s with the discovery of anthocyanins which subsequently lead to the discovery of more enzymes involved in this pathway (Winkel 2006). By the late 1990’s most of the enzymes’ structures and functions were known and analysis began on the regulatory aspects of this pathway (Saito et al. 1999). The pathway is responsible for producing the red, blue and purplish color pigment found in flowers and fruits, as well as, having health benefits for humans such as acting as an anticancer agent and limiting levels of cholesterol (Winkel 2006). The pathway is divided into two main categories, those genes involved in the enzymatic catalytic reaction and those involved in regulating the pathway genes.

The pathway is divided into two main categories; those genes involved in the enzymatic catalytic reaction and those involved in regulating the pathway genes. There are seven enzymatic genes involved in the direct production of anthocyanin. The pathway begins with pre-pathway and early pathway genes known as, phenylalanine ammonia lyase (PAL), chalcone synthase (CHS), chalcone isomerise (CHI) and flavonone 3-hydroxylase (F3H). Late pathway genes are considered, diflavaonal 4-reductase (DFR), leucoanthocyanidin dioxygenase (LDOX), UDP glucose-flavonoid 3-o-glucosyl transferase (UFGT), and anthocyanidin synthase (ANS) (Boss et al. 1996). The actual production of anthocyanin comes from the glycosylation of anthocyanidin, which is unstable, to the more stable compound anthocyanin (Boss et al. 1996). The pre and early pathway genes can be involved in several branches that occur to produce other known enzymes in the larger flavonoid biochemical pathway (Figure 1). Once the products of the late pathway genes are produced there is no other branching and the pathway is committed to making anthocyanins.

**Evolution of anthocyanin pathway genes**

Many studies have been done on what are known as the enzymatic structural genes of the anthocyanin pathway. They have determined structure and function and used this information to further examine the molecular evolution of these genes. A common theme in these enzymatic genes are similar in function even though at the sequence level there may be large differences. For example, in *Ipomea, chs* was described as having large nonsynonymous changes; although there are evolutionary factors that play a role in the production of these genes (Rauscher et al. 1999). Some evolutionary
factors include environmental stresses and selecting for certain pollinators. These factors have a major impact on the genes in the pathway.

Another interesting aspect is comparing the early pathway genes to the late pathway genes. Again, as described, the later in the pathway the more committed the production of anthocyanin becomes. When comparing the nonsynonymous and synonymous changes between the early and late pathway genes, what was found in the upstream genes, such as CHS, was less nonsynonymous changes than later pathway genes, such as \textit{ans} (Lu and Rauscher 2003). The expectation would be that there would be little difference in the number of nonsynonymous changes in early pathway genes like CHS compared to late pathway genes like ANS. However, this was not the case when they examined the genes in \textit{Ipomoea}. Here they found more nonsynonymous changes in ANS, a late pathway gene (Lu and Rauscher 2003). This leads to a possibility that other factors play a significant role in the evolution of this pathway, in particular the regulatory elements on this pathway (Oberholzer et al. 1999).

\textbf{Regulation of the Anthocyanin pathway}

The regulation of the anthocyanin pathway may be the link to determining its pathway evolution. The three transcription factors involved in regulating the pathway are R2R3 MYB, MYC, and WD40 (Figure 1). These regulatory genes are responsible for turning on the genes in the pathway. Evolutionarily the R2R3 MYBs remain conserved in the N-terminus region of the genes, primarily due to the DNA binding domain within the R2 and R3 repeats (Hirichi et al. 2011). In this study, we use the Hawaiian Silversword Alliance (HSA) and California tarweeds to examine the anthocyanin pathway evolution of the \textit{myb} gene family.

HSA are comprised of 32 species and were derived from the California tarweeds over 5 million years (Baldwin 1998). This plant group consists of three genera \textit{Argyroxyphium}, \textit{Dubautia} and \textit{Wilkesia} (Carr 1985, Robichaux et al. 1990). Adaptive radiation has allowed HSA to diversify in new climates and terrain. As a result, HSA’s morphology varies from small shrubs grown in the low lands to trees and rosettes with Silversword shaped leaves, which grow in volcanic regions of the islands (Carr 1985, Robichaux et al. 1990). The flowers on these plants can also vary from white, yellow and purple. All colors are derived from the anthocyanin pathway and the variation between the genera makes HSA ideal to study its molecular evolution. Using different bioinformatics techniques we can ask the question of what selective pressures are acting on transcriptions factors like \textit{myb} in the anthocyanin pathway of HSA and California tarweeds. Based on previous studies we expect purifying selection to be influencing \textit{myb} genes in the anthocyanin pathway.
CHAPTER 4: MATERIALS AND METHODS

Primer design

Over 20 known MYB gDNA or mRNA sequences downloaded from GenBank (NCBI) were aligned with ClustalW 1.8 available on Biology Workbench (SDSC) using gap penalty of 10 and gap extension penalty of 6.66. Degenerate nested primers were designed in conserved regions (Figure 2). Myb primers are as follows: F61 (5’-TGARRAAAGGKBCWTGGA-3’ and R376 5’-KMARDTGDGTGTCCAGTA-3’), F187 (5’-AARAGYTGYAGAYTRAGRTGG-3’ and R324 5’-TCTMCCVGCRTMARYGACCA-3’) and published primers F1 (5’-AARAGYTGYAGATTTAGTGG-3’ and R2 5’-CCARTAGTTTTSACATCGTT-3’)

PCR and subcloning

PCR was performed with degenerate primers based on sequences specified above. PCR cycles for F61 and R376 primers were, 1 cycle 94° C for 2 minutes, 35 cycles of 95° C for 20s then 45° C for 20 s, 1 cycle 72° C for 40 min then 72° C for 10 mins. PCR cycles for F187 and R324 were, 1 cycle 94° C for 2 minutes, 35 cycles of 94° C for 20 s then 47° C for 20 s, 1 cycle 72° C for 40s then 72° C for 10 mins. PCR cycles for F1R2 published primers were, 1 cycle 94° C for 2 minutes, 35 cycles of 94° C for 20 s then 50° C for 20 s, 1 cycle 72° C for 40s then 72° C for 10 mins. Eppendorf thermal cyclers, either a Mastercycler®, or a Mastercycler® gradient were used in the early stages of the experiment, and a BioRad thermal cycler was used later. PCR products were stored at 4°C for up to 2 days before screening and cloning. Screening was done on 1% TAE (Tris base, glacial acetic acid, EDTA) agarose gel, with 100bp DNA ladder (Fisher, BP2573100). Six to eight microliters of PCR product were loaded into a well, and 1: 1 of loading dye unless the loading dye was already present in the PCR buffer. Gels were stained in a 0.5% ethidium bromide solution for 5-10 minutes and screened on a Fisher FBTIV-816 UV transilluminator. PCR products were extracted using a gel extraction kit (Qiagen, 28704).

PCR products or gel extracted bands were cloned into a TOPO-TA vector (Invitrogen, 45-0641) as follows: 0.5 µl TOPO™ vector, 0.5µl salt solution, 1-2 µl PCR product, 0-1 µl water, for a total of 3 µl. Reactions were incubated at room temperature for 10-20 minutes and either used immediately in transformation or kept at -20°C overnight (15-20 hours). Two or three microliters of the cloning reaction was transferred to a vial of OneShot™ TOP10 competent cells (Invitrogen), previously thawed on ice for 10-15 minutes. Vials were incubated on ice for 10-15 minutes, heat-shocked by placing in a 42°C water bath for 30 seconds, and placed immediately on ice. Each vial had 250ml of the kit’s SOC medium added to it, and was placed in a shaking incubator at 37°C, 200rpm. Forty microliters of X-gal and 50-100ml of cells were antiseptically added to LB agar plates with 50mg/ml kanamycin; usually two plates with different volumes of cells were done to ensure proper spacing between colonies. Plates were kept at 37°C for 14-18 hours or longer if colonies were too small. Between 10 and 20 white colonies were
picked from each original PCR product, and each colony was grown in 4ml of LB broth with 50mg/ml kanamycin at 37°C, 200rpm, for 16-20 hours.

**Plasmid and sequence isolation**

Plasmid isolation was done using Promega SV Miniprep Wizard® kit (Promega A1220) and manufacture’s standard protocol. A restriction enzyme digest was used to test for the presence of PCR insert in the plasmids. Five microliters of reaction was mixed with 10ml of digestion master mix (1.5μL 1mg/ml bovine serum albumin, 1.5μL 10X Buffer H, 0.5μL of 5U/μL Fast Digest EcoRI and 6.5μL nanopure water). The digestion was allowed to proceed for 15minns at 37°C water bath. The products were run on 1% agarose gels as described above. Plasmids that were positive for PCR inserts were tested for concentration with a NanoDrop spectrophotometer and sent to the CSUN Sequencing Facility for sequencing.

**Sequencing and editing**

Sequencing was performed at the CSUN DNA Sequencing Facility using an ABI Prism 377 DNA sequencer. Primers used were forward and reverse M13 primers, which were suitable for the plasmids employed by our laboratory. The plasmid sequences were omitted from each clone sequence before performing a search through BLAST. Sequences that matched known myb genes were saved and then aligned with ClustalW on either Biology Workbench or BioEdit. If necessary, sequences were inverted using reverse complement.

**Evolutionary analysis**

Known myb sequences from over 60 gymnosperms, dicots and monocots were obtained from the NCBI GenBank® database. A multiple sequence alignment of these was generated with ClustalW (Figure 3). The alignment was done with gap open/extension penalties of 100/6.6 for both the pairwise and multiple alignments. A model test was run on Jmodeltest (Posada, 2008), which estimated the GTR+G+I (general time-reversible, gamma distributed, with invariant sites) method to be the best model for a maximum likelihood (ML) analysis. A ML tree was constructed with MEGA 5 (Tamura et al., 2007) with the suggested GTR+G+I model of substitution, and 5000 bootstraps to test the phylogeny. A Bayesian tree was generated with MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001) with GTR+G+I for 3,000,000 generations and sampling every 100 generations. The first 25% of the samples were discarded and not included in the automatic analysis.

Nonsynonymous (Ka) and synonymous (Ks) substitutions were determined with MEGA 5, respectively (Tamura et al., 2007). The Ka/Ks ratios were calculated using Microsoft® Excel. Tests of selection (neutral, positive or purifying) were done using MEGA 5.
Isolation of myb gene

First to address whether these genes could be isolated, we designed degenerate primers in the highly conserved DNA binding domain of R2R3 myb. We reasoned over time the DNA binding domain would have fewer changes and be more evolutionarily constrained. Genomic DNA (gDNA) was extracted from leaf tissues from *M. gracilis* and *M. elegans* (California tarweeds) yielded products with concentrations between 250 ng/µl and 500 ng/µl. gDNA was previously extracted in the lab from *Dubautia linearis* and *Wilkesia gymnoxiphium* (HSA).

Myb genes are on average 1400bps long and comprised of three exons and two introns (Figure 2). Our goal was to isolate both exonic and intronic regions of the gene. The exons or coding sequence could be used to further evaluate the evolutionary basis of changes at the molecular level to this gene family while intronic regions can also show repeats in DNA sequence that may have evolved for stability of the plant types. Outside primers were designed to cover exon 1, intron 1 and exon 2 and inside primers were designed to cover only exon 2. Since exon 1 and 2 contain the highly conserved DNA binding domain, we were more easily able to design primers here (Figure 2). Figure 2 shows segments A, B and C which were obtained from HSA and California tarweeds (Figure 2).

Degenerate primers were designed based on known myb genes downloaded from the National Center for Biotechnology Information (NCBI) and a multiple sequence alignment was generated using software programs BioWorkbench and BioEdit Sequence Alignment Editor (Figure 3). Since myb genes had never been isolated from either HSA or California tarweeds before we used many different plant species, such as maize, sweet potato, and raspberry and gingko biloba from only R2R3 myb genes to complete the MSA (Figure 3). The degenerate primers were used to PCR amplify specific R2R3 myb genes from the gDNA of HSA and California tarweeds. PCR products were purified and cloned. Clones were grown and lysed for isolation of plasmid with the PCR insert. These products were sequenced and the results were partial myb DNA sequences from both Hawaiian Silversword and California tarweed. To initially confirm the products, obtained DNA sequences were entered into NCBI’s basic local alignment search tool (BLAST). All sequences matched to known R2R3 myb genes with a minimum of 60% max identity. These sequences were then entered into the multiple sequence alignment along with more known myb sequences (Figure 4). Sequences had a low percentage of similarity at the DNA level. However, when the sequences were translated as expected there was high sequence conservation of amino acids in the R2R3 DNA binding domain region. The highlighted region shows 70% conservation between amino acids of the R2R3 DNA binding domain of sequences obtained from HSA, California tarweeds and other plant species. These results initially suggest the sequences are from myb genes and the variation in sequences could partially explain how these plants evolved at the molecular level. Thus, using these sequences we can start analyzing the molecular evolution of the myb gene family involved in the regulation of anthocyanin genes.
Phylogenetic Analysis

To further our understanding of the molecular evolution between California tarweeds and HSAs we analyzed the myb DNA sequences found in both plant types by generating phylogenetic trees. For phylogenetic analysis we collected 58 myb complete coding sequences from monocots and dicots from the NCBI database along with 6 discovered sequences in HSA and California tarweeds. Based on these sequences two trees were constructed. The first tree was a maximum likelihood (ML) tree using bioinformatics program MEGA 5 (Tamura et al., 2007). One observation of the ML tree is that 3 sequence, 2 California tarweeds (Madia gracillis) and 1 HAS Dubautia linieras) cluster together with high bootstrap values. The Asteraceae clade itself with Gynua and Dahlia also group together with one sequence from Wilkesia, although this branch has very low support and my not be statistically significant. Most of the other clades in this tree group as expected.

The ML tree did not have enough reliable clades with high bootstrap values therefore we ran a Bayesian analysis with MrBayes 3.1.2 of the sequences to see if this would yield any other results (Huelsenbeck et al., 2001). The support for most of the branching of the Bayesian analysis was much higher than we saw in the ML tree. Most of the branches have over 80 percent statistical significance (Figure 7). The Bayesian tree confirms the clade with Madia gracillis and Dubautia linieras with very high support (Figure 7). The tree also reveals a small clade with one Wilkesia sequence and one Gerbera sequence, both of which are part of the Asteraceae family. This clade again shows high support (Figure 7). However, in the Bayesian tree we still find the other Asteraceae clade that does not group with either the HSA or California tarweed. Essentially both groups show similar results but the Bayesian tree has the statistically significance to back up this data where the ML tree did not.

Evolutionary rates of myb genes

To test our original hypothesis of whether there is positive or negative selection taking place on HSA and California tarweeds we examined the ratio of nonsynonymous (Ka) and synonymous (Ks) substitutions among coding sequences of different species. A nonsynonymous substitution is a change in the nucleotide of an amino acid resulting in an amino acid change for a given reading frame. A synonymous substitution is also a change in the nucleotide that results in the same amino acid, therefore no change in protein. Equal ratios between Ka and Ks suggest no selection. Ratios larger than one imply positive selection and ratios lower than one suggest negative or purifying selection.

Nonsynonymous to synonymous substitution ratios (Ka/Ks) were calculated for all myb sequences obtained from HSA and California tarweeds (Figure 8). Previous studies showed Ka/Ks values from enzymatic structural genes CHS and ANS. Both of which are transcriptionally regulated by MYB. Both enzymatic genes ANS and CHS show purifying selection in both HSA and California tarweeds. We expected that positive selection may be acting on the myb gene family which could be an explanation for the varying phenotype we see amongst the HSA. However our data suggest purifying
selection in HSA and tarweeds. Although the Ka/Ks values between MYB in HSA (0.482) and tarweeds (0.092) vary greatly (Figure 8). One issue with the MYB sequences used for this analysis is the full coding sequence of exon three has not yet been isolated from either HSA or California tarweed (Figure 2). Therefore, with more sequence we would be able to have a more complete Ka/Ks analysis.
CHAPTER 6: DISCUSSION

Myb gene sequences

In order to understand the pathway evolution of myb genes in the anthocyanin pathway in HSA and California tarweeds we first needed to isolated myb sequences from these plant types. Many studies have shown these genes can be isolated, however in most cases the genomic sequence of the species is known. In our case both HSA and California tarweeds have no reference genomic sequence therefore we needed to first isolate these sequences. Previous research has shown myb genes can be isolated using degenerate primers in highly conserved regions of the gene (Pattnai et al., 2010). Based on these finding we also designed degenerate primers. To ensure specific binding we designed the primers in the highly conserved R2R3 DNA binding domain. Sequences were obtained using this method and clones were aligned to other known myb sequences using a multiple sequences alignment. Interestingly, the homology between the sequences can mostly be seen at the amino acid level and when comparing DNA sequences there is much more variation. One explanation for this is that single nucleotide polymorphisms (SNPs) are found throughout gene sequences and these SNPs do not necessarily lead to an amino acid change and therefore may not interfere with protein structure and function.

A total of six R2R3 myb gDNA sequences were obtained using the methods previously described. Four sequences were from two different HSA and two sequences were from two different California tarweeds (Figure 4). Initial BLAST results revealed that these sequences were very similar to other R2R3 myb genes found in the NCBI database. When we tested for regions of homology using a MSA we found the highly conserved DNA binding domain as expected. Together both results from BLAST and the MSA suggest the genes isolated are part of the R2R3 myb gene family.

One aspect that was not addressed in this study was gene copies. Since the myb gene family is very large and there are many myb genes with different functions one way to address gene copies is to analyze intronic regions. However we were not able to obtain enough myb sequences in order to make any conclusions. Future work will be performed to obtain more myb sequences and comparing intronic regions.

Evolution of myb genes

Of the six myb sequences obtained we were unable to isolate the complete gDNA sequence. However, we did isolate the DNA binding domain from each of the sequences obtained which directly related to functionality at the protein level and facilitates transcription of key enzymatic genes in the pathway such as, CHS and ANS, though we can’t be sure we isolated anthocyanin myb genes. Using the isolated sequence an MSA was generated with 58 known R2R3 myb sequences. Special attention was given to splicing out intronic regions from the myb genomic DNA sequences. We found the isolated myb sequences have the canonical R2R3 repeat domain (Figure 3). These data imply that these sequences are myb sequences however functional studies need to confirm
the expression of these genes and more specific whether these genes are involved in anthocyanin production.

The next question was which phylogenetic tree would give the best representation of the molecular evolution of myb genes in HSA and California tarweeds. To address this question two phylogenetic trees were generated based on the MSA of isolated myb sequences from this study and previously published myb sequences from NCBI (Figure 3). As expected in the maximum likelihood tree myb sequences from HSA and California tarweeds form their own clade (Figure 6). These sequences actually group with Glycine max (soybean). Surprisingly, none of the myb sequences cluster with other known species from the Asteraceae family. One explanation for these results is the maximum likelihood tree was generated using the Kimura parameters (Kimura et al., 1980). Kimura parameters take into account transitions and transversions between the sequences. Even though the amino acid sequence is highly conserved the DNA sequence between HSA and tarweeds varies from other myb genes from the Asteracaea family. Another possible explain for this is all the sequences isolated are only fragments therefore when the maximum likelihood was generated it only took into account the DNA binding domain. However we argue that the DNA binding domain is the most important region for analysis since functionality of myb genes is characterized by this region. The myb family is also an ancient gene and the region isolated is known to be conserved over millions of years (Oberholzer et al., 2001) Either way the complete coding sequence needs to be isolated for further analysis. Lastly, most of the branches in the maximum likelihood tree have very low support except for the clade with myb sequences from HSA and California tarweeds. The conclusion from these results suggests that the isolated myb sequences share homology with other known R2R3 mybs and that HSA and California tarweed sequences are related.

To further examine the evolution of myb genes that were isolated we constructed a tree using Bayesian analysis. One of the differences between Bayesian analysis and is trees are calculated using the General time reversible model (GTR). GTR gives values to prior probabilities where likelihood estimations do not. Like the maximum likelihood tree, the Bayesian tree are also calculated prior probabilities (Huelsenbeck et al., 2001). Branch support values for the Bayesian tree are much higher therefore may depict an accurate representation of gene speciation and divergence. The Bayesian tree generated from the same data set as the maximum likelihood is more statistically significant with most branches having over 80% likelihood (Figure 7). Also, the clades in the Bayesian tree seem more accurate. For example, the HSA and California tarweeds group together but one Wilkesia sequence groups with Gerbera, both species being from the Asteracaea family. The Bayesian tree fits our data set the best. Both trees show that the sequences isolated from HSA and California tarweeds are related and may also be functionally similar.

Rates of evolution

To determine whether positive selection or purifying selection is acting on the myb gene family in HSA and California tarweeds we used nonsynonymous and
synonymous substitution ratios (Ka/Ks). Over time speciation can lead to a duplication of genes and divergence of sequences which can lead to redundancy of genes or genes with new functions. Divergence in sequence can be due to evolutionary constraints at the DNA level and can be attributed to nonsynonymous and synonymous substitutions. These changes can be very beneficial for species to rapidly adapt to their environment and in time will also give rise to morphological changes. By analyzing the Ka/Ks ratios we can better understand the evolutionary pressures in HSA. We expect to see purifying selection in HSA, a Ka/Ks ratio of less than 1, because higher Ka/Ks values suggest more nonsynonymous changes which are usually associated with deleterious effects. Also, previous results have shown both chs and ans gene families are undergoing purifying selection (Kovacheva, 2011) (Figure 8). We indeed see purifying selection on isolated myb sequences in HSA and California tarweeds however HSA has a higher rate than the tarweeds (Figure 8). These data suggest myb genes potentially regulating the anthocyanin pathway in HSA and California tarweeds are undergoing purifying selection.

CHAPTER 7: CONCLUSION

The data shown in this thesis show that based on DNA sequences of the anthocyanin pathway we can explore how pathways evolve. By using a group of species that has undergone adaptive radiation we show purifying selection to be the determining factor in myb evolution in HSA and California tarweeds. However, the complete coding sequence for the isolated myb genes is necessary to make a better conclusion on the selective pressures of this gene family.
REFERENCES

SECTION 1:


SECTION 2:


**APPENDIX A:**

**SECTION 1:**

---

**Figure 1.** Epigenetic changes leading to XCI. Representation of key Xi marks that are accumulated during differentiation. *Xist* RNA is responsible for the initiation phase followed by polycomb recruitment, macroH2A, and DNA methylation which maintain the Xi.

---

**Figure 2.** Schematic of X and pluripotent features during reprogramming. X features include *Xist* RNA coating, macroH2A, EZH2/Suz12, biallelic expression of X-linked genes and DNA methylation. Pluripotent features include E-cadherin(CDH1) and NANOG.
**Figure 3.** Experimental design of Xi-linked GFP reactivation. Pre-induced pluripotent cells with Xi linked GFP were examined for the percentage of GFP reactivation under various treatments for 3 days.
<table>
<thead>
<tr>
<th>siControl</th>
<th>5Azadc + siControl</th>
<th>siDNMT1</th>
<th>5Azadc + siDNMT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xist+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xist-</td>
<td></td>
<td>GFP+</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GFP+</td>
<td>Control</td>
</tr>
</tbody>
</table>

**Figure 4.** Visual illustrations of Xi-linked GFP pre-iPSC cells with treatments. Top (left to right), Xist+ siControl, Xist+ 5Azadc+siControl, Xist+5Azadc+siDNMT1, ΔXist+siControl, ΔXist+siControl+5Azadc, ΔXist+siDNMT, ΔXist+siDNMT+5Azadc. There is an increase in Xi-linked GFP reactivation in cells with the combined treatment of ΔXist+siDNMT+5Azadc.
Figure 5. Graphical representation of the percent of Xi-linked GFP reactivation in treated cells. Cells with ΔXist+siDNMT+5AzadC have a 2-fold increase in Xi-reactivation compared Xist+siDNMT+5AzadC. X-linked GFP+ reactivation occurs synergistically with the deletion of Xist, knockdown of DNMT1 and chemical inhibitor 5AzadC.
Figure 6. (A) Schematic of X chromosome with the location of Atrx. Target sequence is 200bp downstream of the transcription start site.

(B) Bisulfite sequence analysis of Atrx demonstrates significant DNA demethylation with deletion of Xist, knockdown of DNMT1 and 5AzadC treatment. This suggests DNA methylation occurs late in reprogramming.
**Figure 7.** Proposed model of effects of treatments. Treatments include deletion of Xist, knockdown of DNMT1 (DNA methyltransferase) and 5-aza-deoxycytidine.
Figure 8. Experimental design for isolating reprogramming intermediates. Female stemmca mouse (Tet-O dox-inducible promoter) cell cultured until day 9 and sorted in two populations SSEA-1 (stage specific embryonic antigen -1)+/- to isolate intermediates.

Figure 9. Images of reprogrammed cells on days 1,3,5,7,9 and 12. Arrows, highlight initial reprogramming colonies on day 7,9 and 12.
Figure 10. Cells were sorted for SSEA1+/- with FACS (Fluorescence activated cell sorting) at day 7. Unstained control, isotope control and sorted samples are shown (Left to right).

Figure 11. Early and late immunofluorescence of Nanog and Ezh2 on reprogrammed colonies show cells are positive for Nanog and Ezh2 enriched early in reprogramming and Ezh2 Xi enrichment is lost late in reprogramming.
Figure 12. Isolated reprogramming intermediates from a reprogramming culture. a) There is a progressive increase in NANOG+ cells from d7 to d18 b) SSEA-1+/ cells sorted at day 7 and re-plated. Blue indicates NANOG+ and EZH2 Xi+, Red indicates NANOG+ and EZH2-.
Figure 13. Bisulfite sequencing analysis of pluripotent genes Nanog and Dppa3 on chromosome 6. Each circle represents a CpG island and each row is a cloned sequence. Filled circles represent methylated CpG islands and open circles represent demethylated CpG islands. Nanog is fully demethylated in SSEA-1 + while in SSEA-1 - cells Nanog is methylated. Dppa3 is significantly demethylated in SSEA-1+ compared to SSEA-1 -
Figure 14. Bisulfite sequencing analysis of X-linked genes Rnf12 and Atrx. The inactive X chromosome remains methylated in SSEA-1+ sorted cells for both X-linked genes.
Figure 1. Diagram of enzymatic genes anthocyanin biochemical pathway with synthesized products from pathway. NtAn2 (R2R3 myb gene) is known to transcriptionally regulated enzymatic genes chs, chi, dfr and ans. (Pattniak et.al, 2010)
Figure 2. Schematic of R2R3 Myb gene structure with location of degenerate primers. Sequences are labeled A, B, C. Sequences were isolated from Wilkesia gymnoxiphium, Dubautia linearis, Madia elegans and Madia gracillis.
<table>
<thead>
<tr>
<th>Organism Name</th>
<th>Common Name</th>
<th>Gen Bank accession #</th>
<th>gDNA or mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ricinus communis</td>
<td>Castor Oil Plant</td>
<td>NW_002994722</td>
<td>gDNA</td>
</tr>
<tr>
<td>Rubus idaeus</td>
<td>Raspberry</td>
<td>EU155165.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Rosa hybrid cultivar*</td>
<td>Rose</td>
<td>EU155164.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Fragaria vesca*</td>
<td>Strawberry</td>
<td>EU155163.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Prunus salicina*</td>
<td>Chinese Plum</td>
<td>EU155161.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Mespilus germanica*</td>
<td>Medlar</td>
<td>EU153574.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Malus sylvestris*</td>
<td>European apple</td>
<td>EU153573.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Eriobotrya japonica*</td>
<td>Loquat</td>
<td>EU153572.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Cydonia oblonga*</td>
<td>Quince</td>
<td>EU153571.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>Loblolly Pine</td>
<td>AF5536372.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>Common Wheat</td>
<td>DQ353858.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Populus tremula</td>
<td>Aspen</td>
<td>AJ567346.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Dendrobium XMW-2002-10</td>
<td>Orchid</td>
<td>AF485901.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Pyrus pyrifolia*</td>
<td>Pear</td>
<td>EU153576.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Solanum tuberosum</td>
<td>Tomato</td>
<td>AF426174.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Fagus crenata*</td>
<td>Japanese Beech</td>
<td>AB434650.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Picea glauca</td>
<td>White Spruce</td>
<td>EF601076.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Ginkgo biloba*</td>
<td>Ginkgo biloba</td>
<td>GU968735.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Lilium hybrid division I</td>
<td>Lily</td>
<td>AB534587.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Lotus japonicus</td>
<td>Wild Legume</td>
<td>AB300034.2</td>
<td>mRNA</td>
</tr>
<tr>
<td>Gentiana triflora</td>
<td>Gentiana</td>
<td>AB289444.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Humulus lupulus</td>
<td>Common hop</td>
<td>FN646081.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Ostreococcus tauri</td>
<td>Green algae</td>
<td>XM_003083317.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Garcinia mangostana</td>
<td>Purple mangosteen</td>
<td>FJ197137.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>small legume</td>
<td>FJ199998.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Hordeum vulgare</td>
<td>Domesticated Barley</td>
<td>X70876.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Perilla frutescens</td>
<td>Purple perilla</td>
<td>AB024053.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Nicotiana tabacum*</td>
<td>Tobacco</td>
<td>U72762.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Glycine max*</td>
<td>Soybean</td>
<td>JF510467.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Mimulus aurantiacus</td>
<td>Sticky monkey-flower</td>
<td>EU305682.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Ipomoea purpurea</td>
<td>Morning glory</td>
<td>AB258984.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Petunia x hybrida</td>
<td>Petunia</td>
<td>HQ116170.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Solanum tuberosum</td>
<td>Potato</td>
<td>EU310400.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Zea mays*</td>
<td>Maize</td>
<td>NM_001112467.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Antirrhinum majus</td>
<td>Snapdragon</td>
<td>AY661654.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Species/Cultivar</td>
<td>Common Name</td>
<td>Accession Number</td>
<td>Type</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------------------------------</td>
<td>------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Oryza sativa*</td>
<td>Rice</td>
<td>D88619.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Brassica oleracea</td>
<td>Purple Cauliflower</td>
<td>GU219989.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Arabidopsis thaliana*</td>
<td>Arabidopsis</td>
<td>X90383.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>Arabidopsis</td>
<td>D14712.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Gossypium arboreum</td>
<td>Cotton</td>
<td>GU244349</td>
<td>gDNA</td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
<td>Mexican cotton</td>
<td>AY999070.1</td>
<td>gDNA</td>
</tr>
<tr>
<td>Gynura bicolor *</td>
<td>Okinawa spinach</td>
<td>AB550245.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Capsicum annuum</td>
<td>Chili pepper</td>
<td>AJ608992.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Dahlia pinnata*</td>
<td>Dahlia</td>
<td>AB601003.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Ipomoea nil*</td>
<td>Japanese morning glory</td>
<td>AB234211.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Pyrus pyrifolia*</td>
<td>Asian Pear</td>
<td>HM585181.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Vitis vinifera*</td>
<td>Grape</td>
<td>FJ265729.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Morella rubra</td>
<td>Red Bayberry</td>
<td>GQ340767.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Gerbera hybrid cultivar*</td>
<td>Daisy</td>
<td>EU684239.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Ipomoea batatas</td>
<td>Sweet potato</td>
<td>AB258984.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Oncidium Gower Ramsey</td>
<td>Dancing lady orchid</td>
<td>EF570115.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Phalaenopsis schilleriana</td>
<td>Orchid</td>
<td>FJ039860.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Daucus carota</td>
<td>Wild carrot</td>
<td>AB298507.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Fagopyrum tataricum</td>
<td>Buckwheat</td>
<td>JF313345.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Theobroma cacao</td>
<td>Cocoa tree</td>
<td>GU324346.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Vaccinium corymbosum</td>
<td>Highbush blueberry</td>
<td>JQ085966.1</td>
<td>mRNA</td>
</tr>
<tr>
<td>Epimedium sagittatum</td>
<td>rowdy lamb herb</td>
<td>JN426959.1</td>
<td>mRNA</td>
</tr>
</tbody>
</table>

**Figure 3.** Table of sequences downloaded from the NCBI database. These sequences were used to make multiple sequence alignments and for generating phylogenetic trees. Starred sequences were used to design degenerate primers.
Figure 4. DNA sequence obtained from gDNA from Hawaiian Silverswords and California tarweeds. Sequenced had a minimum of 60% similarity to other known R2R3 myb genes in NCBI database.
Figure 5. Multiple sequence alignment made of protein coding sequence obtained from Hawaiian Silversword and California tarweed. Highlighted regions show 70% conserved regions.
Figure 6. Maximum likelihood tree comparing R2R3 myb sequences obtained from Hawaiian Silversword and California tarweeds with other R2R3 myb genes from different plant types. Inset is a blowup of Silverswords and tarweeds
Figure 7. Bayesian analysis tree comparing R2R3 myb sequences obtained from Hawaiian Silversword and California tarweeds with other R2R3 myb genes from different plant types. Boxed region is HSA and California tarweed clade.
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Silversword</th>
<th>Tarweeds</th>
<th>Gene type</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYB</td>
<td>0.482</td>
<td>0.092</td>
<td>Regulatory</td>
</tr>
<tr>
<td>MYC (bHLH)</td>
<td>0.323</td>
<td>N/A</td>
<td>Regulatory</td>
</tr>
<tr>
<td>GAI homologue</td>
<td>0.196</td>
<td>0.140</td>
<td>Regulatory</td>
</tr>
<tr>
<td>ANS</td>
<td>0.405</td>
<td>0.114</td>
<td>Structural</td>
</tr>
<tr>
<td>CHS-Copy 1</td>
<td>0.192</td>
<td>0.140</td>
<td>Structural</td>
</tr>
<tr>
<td>CHS-Copy 2</td>
<td>0.253</td>
<td>0.147</td>
<td>Structural</td>
</tr>
<tr>
<td>ASCAB9</td>
<td>0.210</td>
<td>0.140</td>
<td>Structural</td>
</tr>
</tbody>
</table>

**Figure 8.** Ka/Ks ratios for structural and regulatory genes in tarweed and silversword.
MYB: Myeloblastosis, transcription factor responsible for regulating ANS and CHS (this study)
MYC: basic-Helix-Loop-Helix transcription factor responsible for regulating ANS and CHS (Dimov)
GAI homologue: plant growth regulatory gene (Remington et al., 2002)
ANS: Anthocyanidin synthase, a late gene in the flower color pathway (Kovacheva, personal communication)
CHS: Chalcone synthase, an early gene in the flower color pathway (Rodriguez, personal communication)
ASCAB9: Chlorophyll A/B binding protein 9 (Barrier et al., 2001)