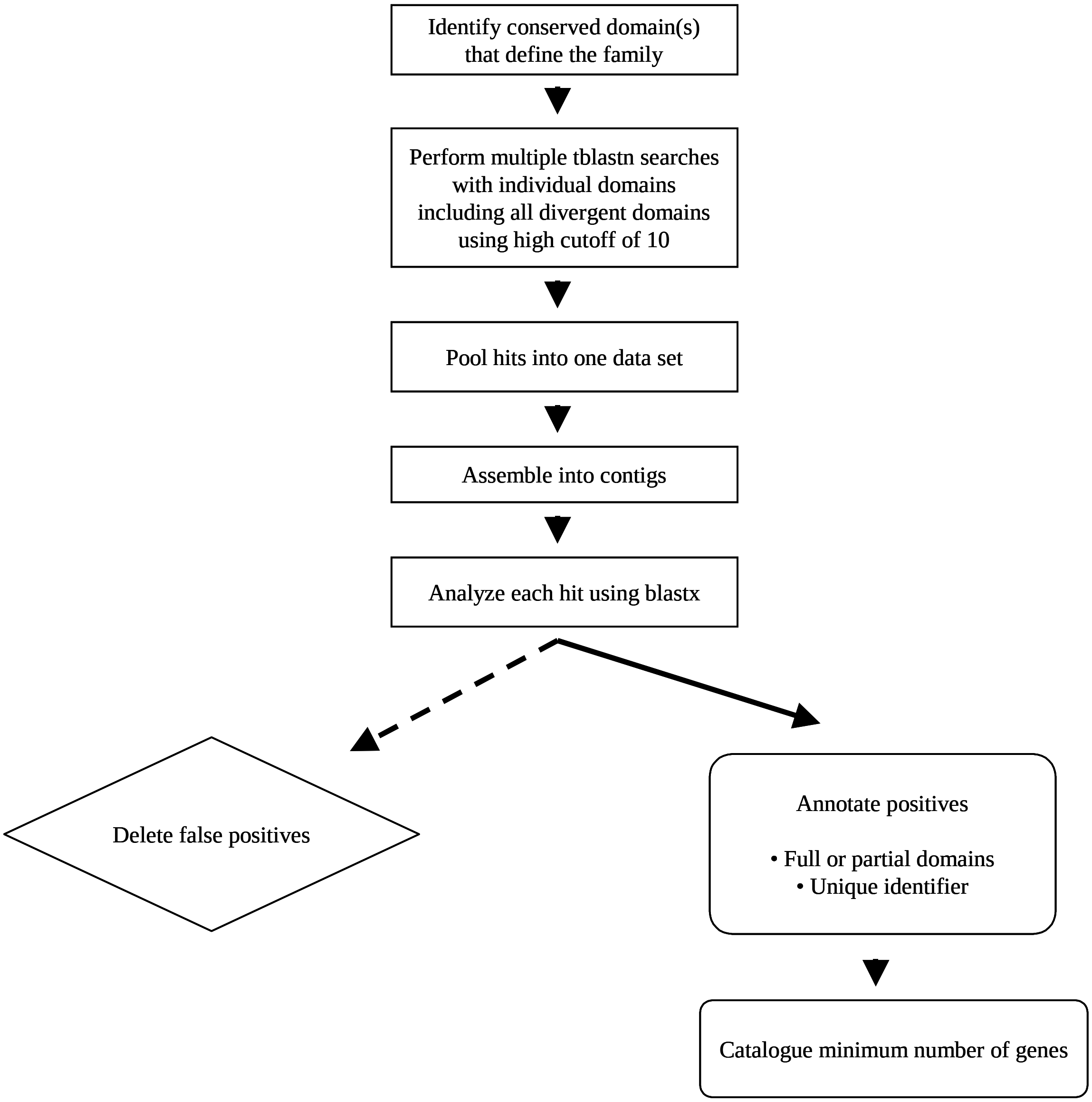
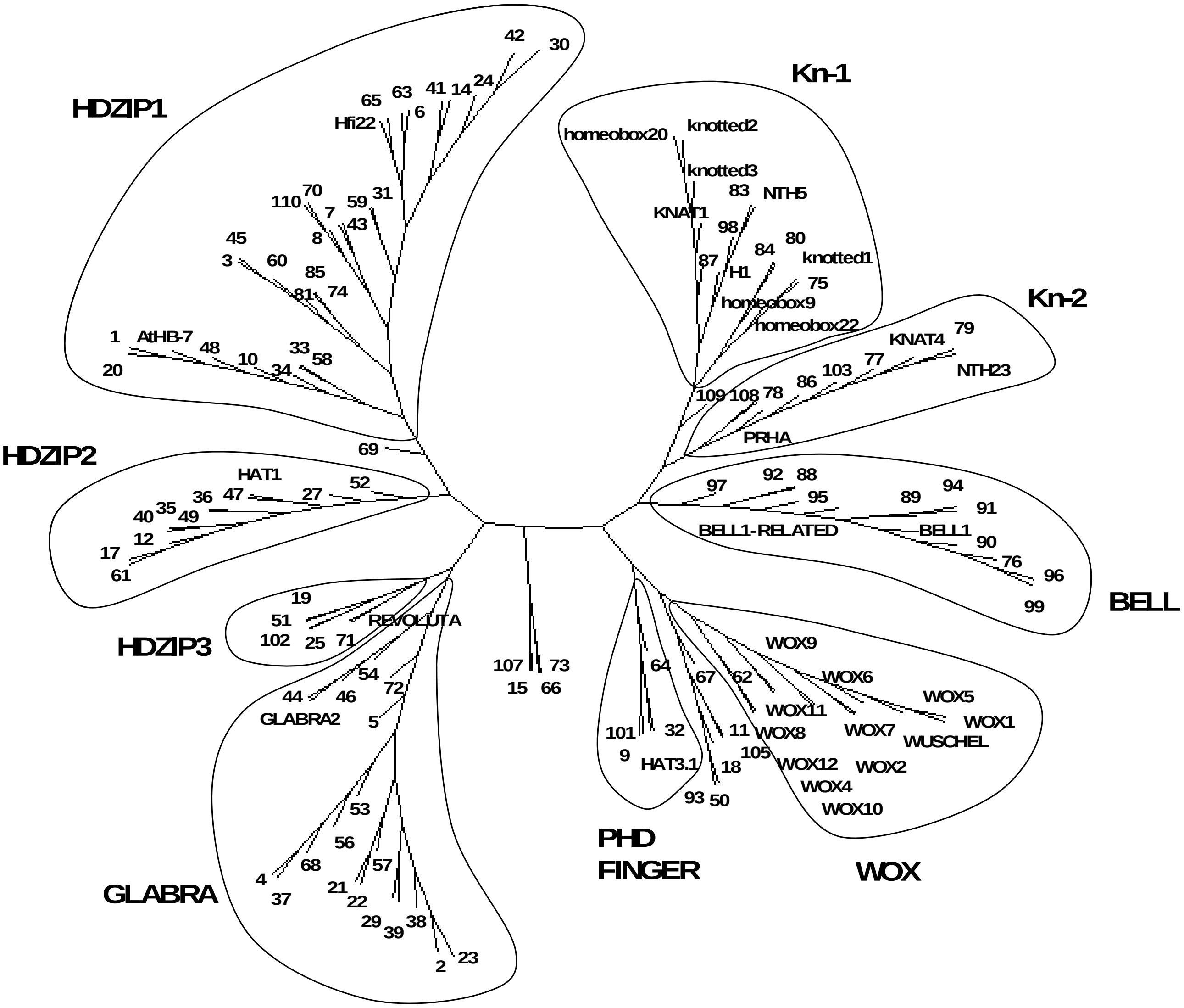
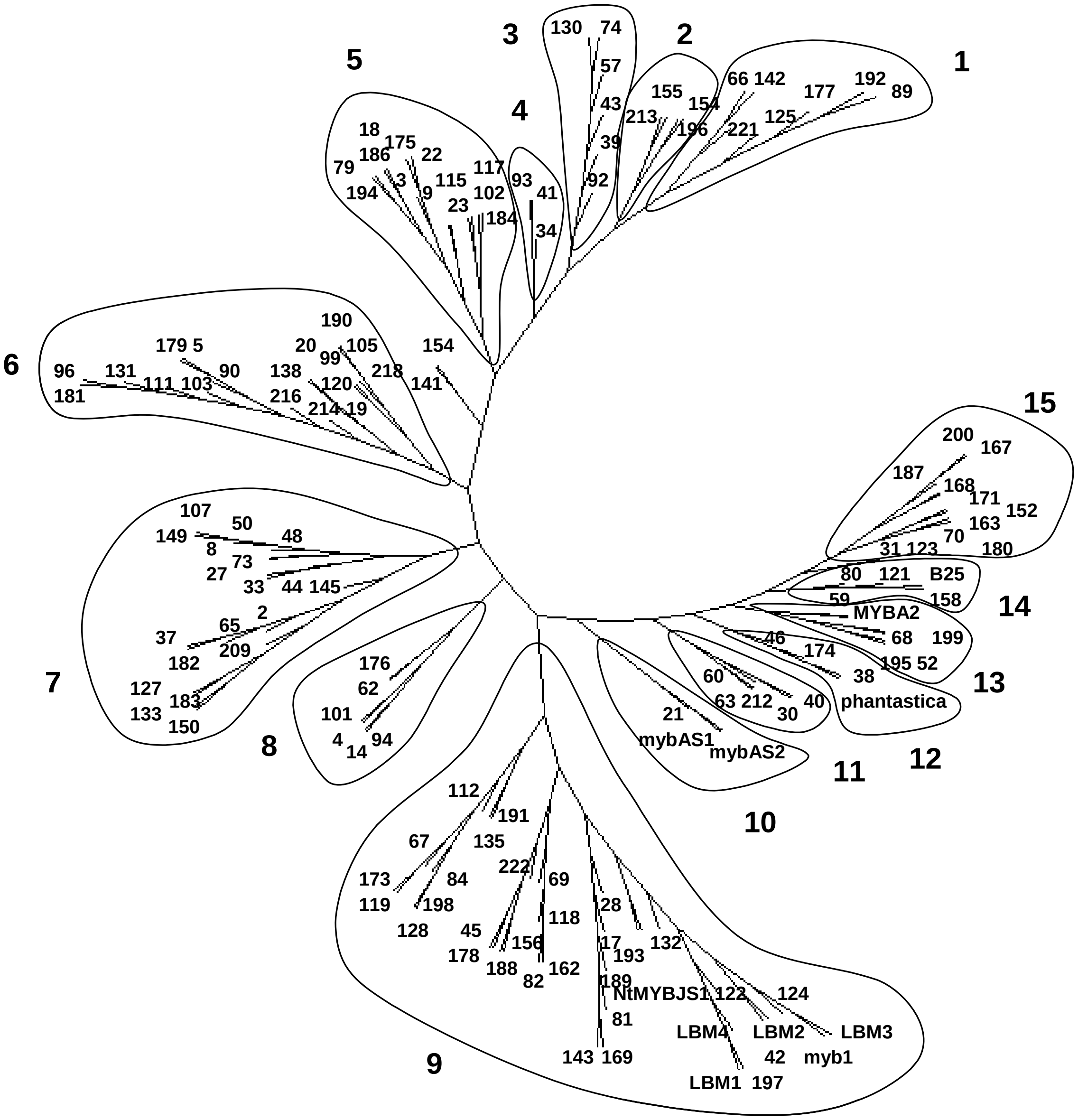
Supplemental Materials



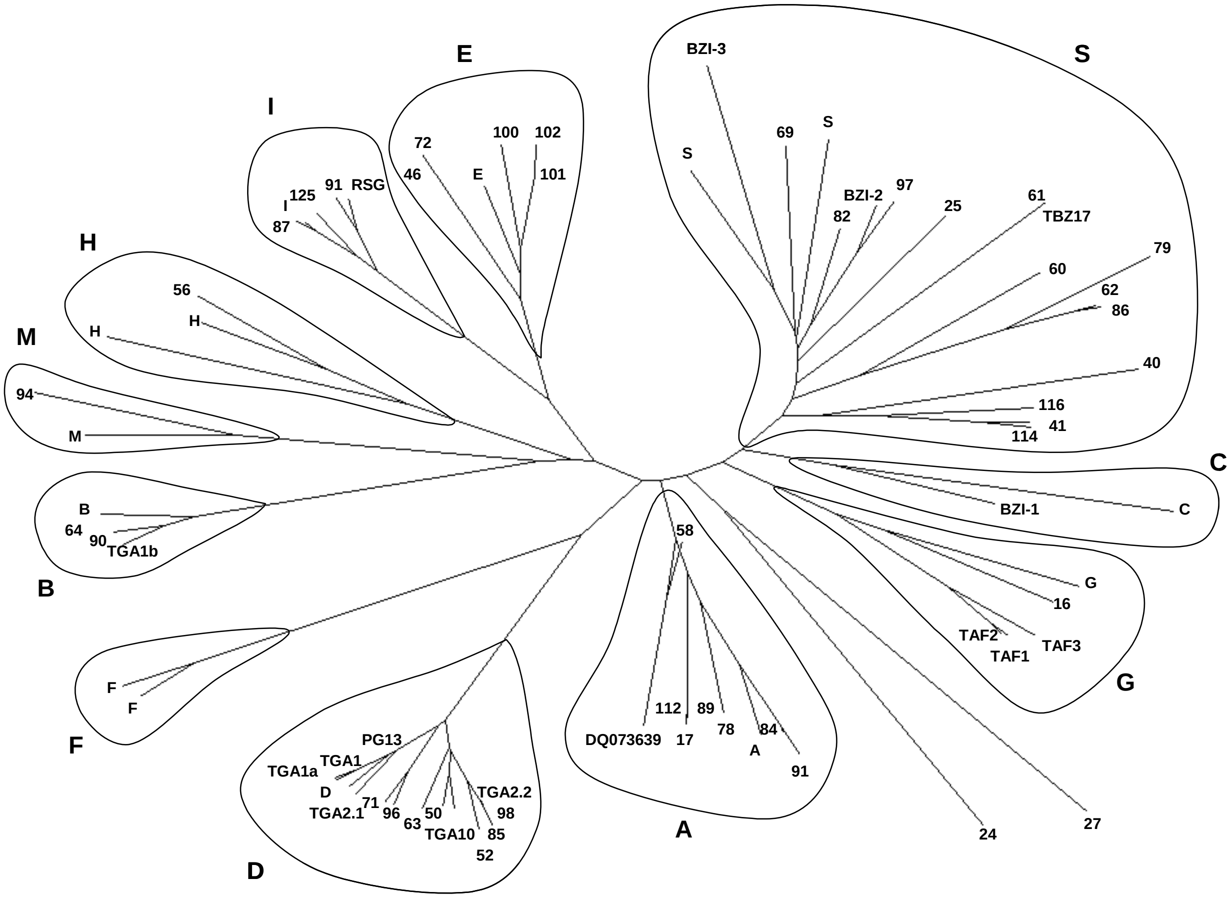
Supplemental Fig. 1. Flow diagram for the identification of GSRs encoding transcription factors in tobacco.



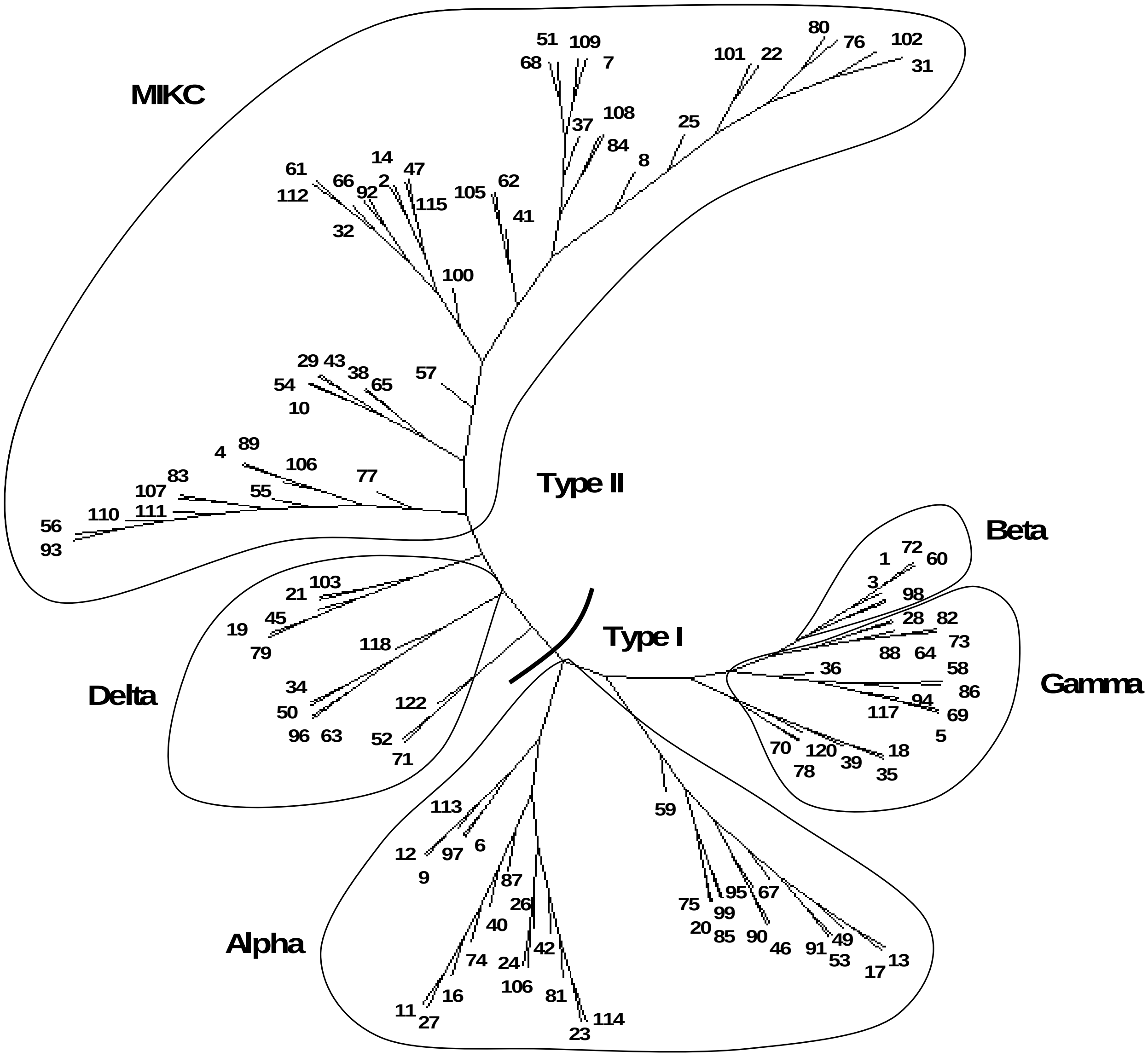
Supplemental Fig. 2. Tobacco homeodomain gene family. Shown is an unrooted phylogenetic tree of the tobacco homeodomain gene family constructed using the Neighbor-Joining method. Each tobacco homeodomain gene is designated by an arbitrary number (e.g., 1, 2, 3). The homeodomain family can be divided into nine major subgroups: Knotted-class 1 (Kn-1), Knotted-class 2 (Kn-2), PHD-finger, Bell, GLABRA, HDZip class 1(HDZIP1), HDZip class 2 (HDZIP2), HDZip class 3 (HDZIP3) and WOX. Homeodomains from Arabidopsis genes were included for group identification purposes and are indicated by their published names.



Supplemental Fig. 3. Tobacco R2R3MYB gene family. Shown is an unrooted phylogenetic tree of the tobacco R2R3MYB gene family constructed using the Neighbor-Joining method. Each R2R3MYB gene is designated by an arbitrary number (e.g., 1, 2, 3). The R2R3MYB gene family can be divided into 15 major clades, numbered 1-15 in the figure. R2R3MYB domains from published Arabidopsis and tobacco genes were included for group identification purposes and are indicated by their published names (e.g., LBM1, NtMYBJS1).



Supplemental Fig. 4. Tobacco bZIP gene family. Shown is an unrooted phylogenetic tree of the tobacco bZIP gene family constructed using the Neighbor-Joining method. Each tobacco bZIP gene identified in the GSR dataset is designated by an arbitrary number (e.g., 1, 2, 3) and published tobacco bZIP family members are represented by their published name. Arabidopsis bZIP genes were included for group identification purposes and are indicated by small letters. The tobacco bZIP gene family forms 10 clades (large letters A-E, G, H, I, M and S); clade F contains only Arabidopsis genes.



Supplemental Fig. 5. The tobacco MADS box gene family. Shown is an unrooted phylogenetic tree of the tobacco MADS box gene family constructed using the Neighbor-Joining method. Each tobacco MADS box gene identified in the GSR dataset is designated by an arbitrary number (e.g., 1, 2, 3). Two types of MADS box genes are recognized (Type I and Type II) and these are further divided into five subgroups: MIKC, Mα, Mβ, Mγ, and Mδ as indicated.

Supplemental Table II. Comparison of predicted TF gene family sizes in tobacco and other sequenced plant genomes.

Shown in the table are the numbers of predicted genes in sequenced plant genomes and the minimum number of genes estimated in the tobacco gene space sequence. Gene families appearing in the upper part of the table are overrepresented in tobacco, whereas those in the lower part of the table are underrepresented.

TF Family	**Arabidopsis**	**Poplar**	**Rice (*O. sativa* ssp.*indica*)**	**Rice (*O. sativa* ssp. *japonica*)**	**Tobacco**
ERF/AP2	146	212	174	182	274
C2H2	134	81	94	113	161
HD	87	106	84	103	129
TCP	23	34	22	24	43
ZF-HD	16	25	14	15	38
GRF	9	9	12	18	23
BES	8	12	7	6	19
SAP	1	1	0	0	4
PcG	34	45	34	34	23
ZIM	18	22	19	29	13
ARF	23	37	24	41	12
CCAAT HAP5	13	19	14	18	6
CPP	8	13	11	16	3

Supplemental Table III. Gene-specific primers used in RT-PCR analysis of methyl jasmonate-induced gene expression.

Gene	**Forward Primer**	**Reverse Primer**
ERF1	GCGAATTTTCCGGGAGAATTTCCGGTG	GTACCACAATTGAATATCCTAAGTTC
ERF115	TCTATAGAAAAGTAAATGAATCCCAA	TTCTAATACAAAAAAATAGAATAC
ERF146	TTGGAGCCTGAAATTTCAGTACAG	AGAACCCTTAAAAGCAGCCCGAGG
ERF165	TCACCCCAATCTTCTTCATGGGATGAG	CTTAAGCTCAATAACTGGCGAACAACTCTC
ERF179	GAATCCAGCTGATAATGCAACCTTCTC	TATGCAGTAATGTTGCTGCATTTGG
ERF192	GTTAACGCTAATGACAATAAGGAATCAAGAGAAG	CCAACAAGTTATCATCCAAGTACTCAAATTC
ERF210	CATTGATAAACCCTACTTTAGTAGGCG	GCTAGTACGACTTAAACTAATTTACA
ERF212	TAGTGCATAATTCAACCAATTTCGTCAAAT	CCATGACTCCTCTTCCTTACACCTA
ERF214	GGAAGTCAAAGTTAAAGTTGAAGCTG	CAGCTGGTGACATTAATCCATCTTTCAA
JAP1	TCCAACATGTTTCCTAATTGCTTGCC	CGGGAAATTAAGCAATGCACGTGACC
ORC1	TCTATAGAAAAGTAAATGAATCCCGC	TTCTAATACAAAGACATAGAATAAA
ACRE1	CAGTGCATAATTCAACCAATTTCATCATGC	TCATGATTCCTCTTCCTTGCACCTC
S25XP1	ATGAAGAGCCAAGCCATGGCAGAAAAATGG	GAAGTTCCTCTAACAACTTATCATCCAAAC