

Appendix A Alignments of the deduced amino acid sequences of the two cotton C4Hs and other known C4Hs

Accession number	Plant species	Identity (%)	
		GhC4H1	GhC4H2
FJ848866	GhC4H1	100	90.89
FJ848867	GhC4H2	90.89	100
DQ485129	BnC4H1	85.15	85.94
DQ485131	BnC4H2	84.95	85.74
DQ355979	SmC4H	89.11	89.70
AF255013	CsC4H1	60.22	59.29
AF255014	CsC4H2	85.16	87.67
DQ211885	PhC4H	88.91	89.70
AAG10197	GaC4H1	100	90.89
AAG10196	GaC4H2	91.29	99.60
ACH56520	GhC4H	90.89	99.21
AAC99993	AtC4H	84.75	85.35
ACR44227	GmC4H	88.74	88.93
ABC69412	NtC4H1	90.89	89.50
ABC69411	NtC4H2	90.89	89.50
AAB42024	ZvC4H	87.92	87.92
CAA78982	HtC4H	87.33	88.12
CAA83552	CrC4H	89.11	90.10
AAC49187	PsC4H	86.73	88.12
AAD11427	McC4H	58.39	58.58
CAA70595	PvC4H	59.04	59.23

Appendix B Analysis of putative *cis*-acting elements of the *GhC4H1* and *GhC4H2* promoter

<i>cis</i> -acting element	Core sequence	<i>GhC4H1</i> promoter	<i>GhC4H2</i> promoter
CBF binding site	RYCGAC	1	0
EEC element	GANTTNC	1	1
E-box	CANNTG	5	4
BS1 binding site	AGCGGG	1	0
RAV1	CAACA	1	5
CAAT-box	CAAT	19	10
ATHB-1	CAATWATTG	1	0
ATHB-6	CAATTATTA	0	1
SORLIP	GAGTGAG	2	0
CACT	YACT	14	14
MYB-I	GGATA	2	0
MYB-II	MACCWAMC	2	3
I-box	GATAA	2	2
DPBF binding sequence	ACACNNG	1	1
ARR1 binding element	NGATT	8	7
GTAC element	GTAC	1	2
GT1-binding sites	GRWAAW	5	13
DOF recognition sequence	AAAG	5	18
TATA-box	TTATTT	9	14
ERE element	AWTTCAAA	1	2
ASF binding site	TGACG	1	1
RHE sequence	KCACGW	1	0
SEBF binding site	YTGTCWC	1	0
GARE	TAACAAR	1	2

GATA-box	GATA	4	6
W-box	TTGAC	1	2
OSE element	CTCTT	3	3
L-box	ACCWWCC	3	2
P-box	ACCWACCNN	5	4
H-box	CCTACCNNNNNNNCT	1	0
SURE element	GAGAC	1	0
C-box	CTCCCAC	1	0
SEF3 binding site	AACCCA	1	0
RBCS sequence	AATCCAA	1	2
L1-box	TAAATGYA	0	1
SRE	TTATCC	1	0
MYC	CANNTG	5	4

Appendix C Summary of primers used in this study

Constructions	Restriction sites	Primer sequences (5'-3') ¹
Primers for full-length cDNA cloning		
C4H1F		CATCATCATGGACCTCCTCTTCCTAG
C4H1R		ACTTGTTAAATCAAAACACCCTTGGCTT
C4H2F		GGCACGAGGAAGAAATAAACTCAAAGC
C4H2R		AGAACAAGTTAAAATTGCCTTGGCTTAGCA
Primes for amplifying promoter sequence		
C4H1SP1		CTTAAACAATCCCATTTGGTGGCC
C4H1SP2		GGTGAAGAAAGGGACGGTCATGAT
C4H1SP3		GAACTCAACTCCTTGGGTGTGCAG
C4H2SP1		GCTGCAATCTCCTCCTCAAAACGAT
C4H2SP2		GTGCTCTCCGTACACCGTGAAAAC
C4H2SP3		GGCCAAATCTGTCAAGTTGCGG
Primers for amplifying genomic sequence		
C4H1iF		ATGGACCTCCTCTTCCTAGAGAAAGTTCTC
C4H1iR		TCAAAACACCCTTGGCTTTGCAAT
C4H2iF		ATGGATCTCCTCTTCCTGGAGAAGG
C4H2iR		AGAACAAGTTAAAATTGCCTTGGCTTAGCA
Primers for transient expression		
pRTL2-C4H1PF	<i>Pst</i> I	ATA <u>CTGCAG</u> ATGCAGCCCCATACGCGATC
pRTL2-C4H1PR	<i>Xho</i> I	TAT <u>CTCGAG</u> GATGATGAATTGTGGGGGTAGGG
pRTL2-C4H2PF	<i>Pst</i> I	ATA <u>CTGCAG</u> GGGTGGAAAGTTAAGTGCCTAGCTC
pRTL2-C4H2PR	<i>Xho</i> I	TAT <u>CTCGAG</u> TTGGCTTTGAGTTTATTTCTTCT
Primers for qRT-PCR analysis		
C4H1RTF		GATGCAAAGCTTGGTGGGTATGAC
C4H1RTFR		ACTTGTTAAATCAAAACACCCTTGGCTT
C4H2RTF		AAAGGTTCTTTGAAGAGGAAGCCAAGG
C4H2RTR		AGAACAAGTTAAAATTGCCTTGGCTTAGCA
ubiF		AAGACCTACACCAAGCCCAAGA
ubiR		CTCTTTCCTCAGCCTCTGAACCT

¹ The underlined sequences are for restriction sites.