Plant Omics Journal POJ

POJ 4(7):391-399 (2011) ISSN:1836-3644

Genomic analysis and gene structure of the two invertase families in the domesticated apple (*Malus* × *domestica* Borkh.)

Tae Kyung Hyun[1], Seung Hee Eom[2], Ju-Sung Kim[3*]

[1]Department of Biochemistry, Gyeongsang National University, Jinju 660-701, Korea
[2]Department of Horticulture, Gyeongnam National University of Science and Technology, Jinju 660-758, Korea
[3]Majors in Plant Resource Sciences and Environment, College of Applied Life Sciences, Jeju National University, Jeju 690-756, Korea

***Corresponding author: Aha2011@jejunu.ac.kr**

Abstract

Plant invertase (β-fructofuranosidase, EC 3.2.1.26) comprises a family of enzymes which plays an important role in the hydrolysis of sucrose to glucose and fructose. Although the specific functions of different invertase families are not clear, they regulate the entry of sucrose into different metabolic pathways in higher plants. Plants contain two unrelated invertase families with different biochemical properties and subcellular localizations. In this study, we identify gene families including vacuolar invertases (designated *MdoVIN1-3*), cell-wall bound invertases (*MdoCIN1-3*) and neutral/alkaline invertases (*MdoNIN1-12*) from the domestic apple (*Malus* × *domestica* Borkh.) genome. Based on phylogeny, most neutral/alkaline invertases could be divided into two subgroups, whereas *MdoNIN12* was found to be a pseudogene. In addition, specific motifs were discovered in neutral/alkaline invertases, which suggested that different motifs are associated with differences in protein function between subgroups α and β. Taken together, our comparative genomic analysis of invertase genes and encoded proteins in the domestic apple provides the first step towards the functional dissection of the role of invertase families in heterotrophic metabolism

Keywords: Comparative genomics analysis, Glycoside hydrolase family 32, *In-silico* analysis, Invertase, *Malus* × *domestica* Borkh.
Abbreviations: CIN- Cell-wall bound invertase; VIN- Vacuole invertase; NIN- Neutral/alkaline invertase.

Introduction

In flowering plants, the non-reducing disaccharide sucrose is the final product of photosynthetic carbon fixation and one of the principal storage forms of carbohydrate together with starch (Winter and Huber, 2000; Tetlow et al., 2004; Charkazi et al., 2010; Di Maro et al., 2011). Sucrose is synthesized in source organs and transported via the phloem to the heterotrophic parts of plant such as the roots, tubers and seeds (sink organs), which are the sites of consumption and storage. The utilization of sucrose as a source of carbon and energy depends on its cleavage by either invertase (β-fructofuranosidase, EC 3.2.1.26) or sucrose synthase (SuSy, EC 2.4.1.13). Invertase uses an irreversible catalytic mechanism to cleave the sucrose molecule into glucose and fructose (Roitsch and Gonzalez, 2004), whereas SuSy catalyses the reversible conversion of sucrose and UDP to UDP-glucose and fructose (Baud et al., 2004). In the model plant *Arabidopsis thaliana*, a total of six putative members in the SuSy gene family have been identified in its genome (Barratt et al., 2001), and the invertase family is composed of two subfamilies that display characteristic pH optima for activity (Roitsch and Gonzalez, 2004). Most plant species contain the acid invertase sub-family (optimum pH 3.5-5.5), which cleaves sucrose on the cell wall (cell-wall bound invertase) or vacuole (vacuole invertase), and neutral/alkaline invertases (optimum pH 6.8-8.0), which are localized to the cytosol (Lee and Sturm, 1996; Roitsch and Gonzalez, 2004; Qi et al., 2007). Acid invertases are glycosylated proteins belonging to glycoside hydrolase family 32 (GH32), while neutral/alkaline invertases are non-glycosylated forms which are classified in the GH100 family (Sturm and Chrispeels, 1990; Bocock et al., 2008; Lammens et al., 2009). The acid invertase sub-family is believed to have originated from respiratory eukaryotes and aerobic bacteria (Sturm and Chrispeels, 1990; Bocock et al., 2008), whereas neutral/alkaline invertases are unique to photosynthetic bacteria and plants (Vargas et al., 2003). Although *Suc2* of yeast is transcribed from two promoters and results in encoding both the cytosolic enzyme and the glycosylated external invertase (Carlson and Botstein, 1982), the acid invertases of bacteria and fungi are periplasmic and extracellular enzymes, respectively (Ehrmann et al., 2003; Ji et al., 2005). On the other hand, plants contain not only extracellular invertase (cell-wall invertase) but also vacuolar invertase, and their amino acid sequences are more closely related to each other than to the neutral/alkaline invertases (Sherson et al., 2003). It is generally believed that the presence of both acid invertases with different subcellular localizations and varying modes of regulation is physiologically advantageous to the plant for optimizing the control of sucrose metabolism, partitioning and storage within different cells (Haouazine-Takvorian et al., 1997). The physiological functions of various invertase isoforms are not fully understood, but the regulation of acid invertases occurs at the level of expression during growth and development by the

concentrations of metabolizable and non-metabolizable sugars, or by environmental stress and phytohormones (Sinha et al., 2002; Roitsch et al., 2003; Roitsch and Gonzalez, 2004; Bonfig et al., 2006; Hyun et al., 2009). This indicates that acid invertases not only mobilize sucrose and/or control sugar composition, but also play a role in heterotrophic metabolism in response to stress related stimuli. Much less is known about the neutral/alkaline invertase sub-family. Following the isolation of neutral/alkaline invertase genes from grass (Gallagher and Pollock, 1998), carrot (Sturm et al., 1999), rice (Murayama and Handa, 2007) and cyanobacteria (Vargas et al., 2003), it has been shown that AtCYT-INV1 (neutral/alkaline invertase in *Arabidopsis*) is involved in sugar/ABA signaling (Qi et al., 2007). In addition, the absence of LjINV1, one of the neutral/alkaline invertases in *Lotus japonicas*, affects both root and aerial parts of plants through an effect on cell proliferation and expansion, indicating that neutral/alkaline invertases also play a crucial role during plant establishment and subsequent development (Welham et al., 2009). The domesticated apple (*Malus* × *domestica* Borkh., family Rosaceae), one of the main fruit crops in temperate regions, is a diploid plant with 17 chromosomes (Kron and Husband, 2009; Farrokh et al., 2011). Recently, two forms of acid apple invertase, namely cell-wall bound invertase (CWI) and vacuole soluble acid invertase (SAI), have been identified and subsequently shown to be specifically activated by ABA via a posttranslational mechanism involving reversible protein phosphorylation (Pan et al., 2005 and 2006). It has also been shown that the induction of extracellular invertase (CIN1; extracellular invertase from *Chenopodium rubrum*) expression by fungal elicitors was inhibited by applying a protein kinase inhibitor (Ehness et al., 1997). Similarly, the activation of the promoter of the tomato extracellular invertase Lin6 by stress-related stimuli requires tomato MAPKs, LpMPK2 and LpMPK3, indicating that protein kinases are involved in the upstream signaling pathway of extracellular invertase expression (Hyun et al., 2009). Although a full list of the apple invertase family is not yet provided in the literature, the different aspects of phosphorylation events may be due to the presence of various invertase isoforms. In fact, tomato cell-wall bound invertases comprise of four enzymes, Lin5, Lin6, Lin7 and Lin8 that are differentially regulated by stress-related stimuli (Godt and Roitsch, 1997). The annotated genome sequences of Golden Delicious (the diploid apple cultivar) have shown that the heterozygous apple genome contains 57,389 putative genes and 31,678 transposable element-related ORFs in 603.9 Mb (Velasco et al., 2010). The completion of the domesticated apple genome has made it possible for the identification of gene families through the analysis of sequence similarity with the *Arabidopsis* genome. In fact, the AP2/ERF family (Zhuang et al., 2011), R2R3-MYB transcription factors (Feng et al., 2010) and FRUITFUL-like genes (Cevik et al., 2010) have already been identified by comparative genomic analysis in this way. Therefore, these findings indicate the possibility of identifying a full list of the apple invertase family using comparative genomic analysis. In this study, we identified apple genes potentially encoding three vacuolar invertases, three cell-wall bound invertases and 12 neutral/alkaline invertases. A phylogenetic tree was constructed to evaluate the evolutionary relationship among invertase amino acid sequences, while gene structure analysis was performed to gain insight into structural divergence among these gene families. Our genomic and bioinformatic analysis will provide the foundation for further functional analysis of the apple invertase family and their role in heterotrophic metabolism.

Results and discussion

Identification of apple invertase families

BLAST was utilized to search the Genebank database, using previously identified invertase genes encoding cell-wall bound, vacuolar or neutral/alkaline invertases from tomato, rice or *Arabidopsis* as queries. Candidate invertase sequences were identified in the genome database of Rosaceae. Subsequently, the redundant sequences were removed according to the chromosome locations, resulting in a total of 18 putative invertase genes (six in the acid invertase sub-family and 12 in the neutral/alkaline invertase sub-family), following the rice and poplar invertase nomenclature (Ji et al., 2005; Bocock et al., 2008). Thus we designated three CIN genes *MdoCIN1-3*, three VIN genes *MdoNIV1-3* and 12 NIN genes *MdoNIN1-12* (Table 1). The acid invertase genes are spread over six different chromosomes, whereas NIN genes were located across chromosomes 1, 2, 3, 4, 5, 8, 11, 12 and 16. Chromosome 12 harbors two *NIN* gene copies but they are separated by more than a hundred thousand base pairs (Table 1).According to subcellular localization, the acid invertases are further divided into two subgroups, cell-wall and vacuolar type. To predict the subcellular localization of apple acid invertases, we employed SignalP, MitoPro, PSORT and TargetP tools. Two CINs were predicted to have the hydrophobic N-terminal signal peptide required for secretory proteins (Table 1). Similarly, most of *Arabidopsis* and rice cell-wall bound invertases have the hydrophobic N-terminal signal peptide, and are involved in the secretory pathway to locate cell-walls (Ji et al., 2005). In addition, analysis of SignalP data predicted that all apple vacuolar invertase sequences form a single anchor containing an initial hydrophilic portion followed by a long hydrophobic region (Table 1). The predicted membrane spanning domain from MdoVIN1, 2 and 3 was located downstream from the N-terminus at residues 30-49, 32-51 and 39-58 respectively (HMMTop prediction). Vacuolar invertases from *Arabidopsis*, rice (Ji et al., 2005), barley and sugarcane (Rae et al., 2011) are also predicted to form membrane anchors. Although MdoVIN1 and 2 and MdoVIN3 appear to be type II membrane proteins and a type III membrane protein, respectively (Table 1), the signal anchor is located near the N-terminus of all MdoVINs. This may indicate that the major part of MdoVIN1 and 2 extends into the lumen of the vacuole, while the 38 amino acid at the N-terminus and 29 amino acid at the C-terminus in MdoVIN3 remains in the cytoplasm. Plant neutral/alkaline invertases are believed to be cytosolic. However, the prediction of *Arabidopsis* and rice neutral/alkaline invertase localization using computational analysis suggests that neutral/alkaline invertase proteins located in cell organelles such as mitochondria and plastids (Ji et al., 2005; Murayama and Handa, 2007). In the case of MdoNINs, MitoProtII predicted that the probability of mitochondrial targeting was higher than 85% for six of the 12 MdoNINs (Table 1). The presence of MdoNIN activity in mitochondria and other organelles suggested that organellar MdoNINs generate glucose as a substrate for organellar hexokinases (Xiang et al., 2011), indicating that they may be involved in signaling function.

Phylogenetic and gene structure analysis of apple invertase families

In order to determine the evolutionary relationships among the invertase gene family in the domestic apple, a phylogenetic tree was constructed by comparing the whole invertase amino acid sequence using the neighbor-joining method. As shown in Fig.

Table 1. Nomenclature and chromosomal location of 18 apple invertase genes

Gene Name	Protein ID	Position (5'-3')	Contig (5'-3')	ORF Length	AA length	Subcellular localization	
						SignalP	MitoPro/ChloroP
MdoCIN1	MDP0000275150	Chr 12:2618511-2623108	MDC002872.298:6961-11518	1734	577	Secretory protein	2.2[b]
MdoCIN2	MDP0000268052	Chr14:2743616-2747685	MDC001056.589:8530-12599	1833	607	Secretory protein	10.6[b]
MdoCIN3	MDP0000561738	Chr13:29305782-29311850	MDC009168.711:11397-17465	2112	703	Non- secretory protein	99.9[b]
MdoVIN1	MDP0000149570	Chr1:23870970-23875107	MDC010572.241:580-4717	1998	665	Signal anchor (Type II-membrane protein)[a]	0.2[b]
MdoVIN2	MDP0000377084	Chr7:22464359-22468029	MDC026391.28:9477-13147	1929	642	Signal anchor (Type II-membrane protein)[a]	0.2[b]
MdoVIN3	MDP0000124776	Chr6:9161660-9166433	MDC009378.171:1409-6182	2037	678	Signal anchor (Type III-membrane protein)[a]	0.2[b]
MdoNIN1	MDP0000163452	Chr12:21150009-21152046	MDC002401.431:8929-10966	1644	547	Non- secretory protein (plasma membrane)[a]	0.5[b]/43.8[c]
MdoNIN2	MDP0000186866	Chr5:22192295-22195888	MDC008639.26:3932-7525	1797	598	Non- secretory protein (endoplasmic reticulum)[a]	0.7[b]/43.8[c]
MdoNIN3	MDP0000146680	Chr8:26763784-26767394	MDC000568.122:15-3625	1532	511	Non- secretory protein (peroxisome)[a]	5.2[b]/43.5[c]
MdoNIN4	MDP0000531557	Chr1:11382547-11385966	MDC012813.224:6210-9629	1983	660	Non- secretory protein (endoplasmic reticulum)[a]	6.5[b]/42.7[c]
MdoNIN5	MDP0000261740	Chr11:12761284-12764435	MDC019410.118:19385-22536	1965	654	Non- secretory protein (nucleus)[a]	97.9[b]/54.8[c]
MdoNIN6	MDP0000315220	Chr3:13319281-13324644	MDC019885.318:3626-8989	2178	725	Non- secretory protein (nucleus)[a]	93.6[b]/51.4[c]
MdoNIN7	MDP0000297851	Chr2:29417131-29423194	MDC012104.405:26068-32131	2181	726	Non- secretory protein (nucleus)[a]	7.9[b]/46.1[c]
MdoNIN8	MDP0000652278	Chr4:18807745-18811112	MDC006283.261:18465-21	1998	665	Non- secretory protein (peroxisome)[a]	93.3[b]/50.2[c]
MdoNIN9	MDP0000095481	Chr12:27472254-27475413	MDC008307.240:31820-34979	2028	676	Non- secretory protein (peroxisome)[a]	97.3[b]/52.5[c]
MdoNIN10	MDP0000133399	Chr16:8043095-8046825	MDC026146.11:18480-22210	2049	682	Non- secretory protein (chloroplast)[a]	96.5[b]/47.3[c]
MdoNIN11	MDP0000319075	Chr8:957823-963768	MDC014108.116:9090-15035	2217	738	Non- secretory protein (peroxisome)[a]	0.2[b]/43[c]
MdoNIN12	MDP0000291284	unanchored:22938456-22940247	MDC009193.345:11873-13664	1098	366	Non- secretory protein (mitochondria)[a]	89.1[b]/51.4[c]

[a]Prediction by PSORT, [b]Probability (%) of targeting to mitochondrion, [c]Probability (%) of targeting to chloroplast

1, the acid invertases can be further subdivided into two well-supported groups: cell-wall and vacuolar invertases. Plant genome analyses have also uncovered putative multigene families encoding neutral/alkaline invertases in the genomes of *Arabidopsis* (Qi et al., 2007), rice (Ji et al., 2005), poplar (Bocock et al., 2008) and grapevine (Nonis et al., 2008). Although the functional implications of sub-division in these families are not clear, plant neutral/alkaline invertases can be subdivided into two groups, α and β (Ji et al., 2005; Bocock et al., 2008). In the case of apple neutral/alkaline invertases, we found that MdoNIN1 to 4 and MdoNIN 5 to 10 fell into subgroup α and β, respectively (Fig. 1). Since the gene organization is highly conserved within a gene family, analysis of intron-exon organization can help to understand its evolution (Chauve et al., 2008). To gain insights into the evolution of apple invertases, we analyzed the pattern of exon-intron junctions (Fig. 1).

The first intron from an apple invertase gene is a phase 0 intron, which refers to the splicing after the first nucleotide of the codon (Fedorov et al., 1992). Except for *MdoCIN3* in the acid invertase family, the second intron is a phase 1 intron that interrupts the codons between the first and second nucleotides (Fedorov et al., 1992). Generally, the distribution of intron phases in the acid invertase family is unequal with a bias in favor of phase 0, which indicates that the ancient introns were dominantly of phase 0. In addition, the last intron of cell-wall bound invertases in the apple is in phase 2, which lies between the second and third nucleotides of joining codons (Fig. 1). The exon-intron structure of apple cell-wall bound invertases indicates that the founding gene of cell-wall bound invertases underwent a duplication to produce *MdoCIN3* and the forerunner of *MdoCIN1* and *2*, and then *MdoCIN1* lost an intron at the 3'end of the gene (Fig. 1). The three apple vacuolar invertase genes are closely related at the amino acid level, but each gene contains different numbers of exons (Fig. 1). We suggest that *MdoVIN1* and *2* originated from a common forerunner by duplication, with a gain of an intron in *MdoVIN2*. The neutral/alkaline invertase genes in the α group are completely dominated by the phase 0 introns that are distributed in relatively highly conserved regions (Rogozin et al., 2003), and 3 out of 4 genes contain four exons (Fig. 1). As shown in Fig. 1, the founding genes of the α group underwent an initial duplication to produce *MdoNIN4* and the forerunner of *MdoNIN1* to *3*, which then underwent a second duplication to produce *MdoNIN3*, with a gain of an intron in *MdoNIN2*. However, neutral/alkaline invertase genes of the β group exhibit a different intron-exon structure and a different number of exons compared with the α group, indicating that the α and β groups arose from different ancestral genes.

Annotation of apple invertases

The 18 reannotated apple invertase genes were translated and subjected to protein motif analyses. As shown in Fig. 2, five amino acid residues in the conserved motifs (A185S, M226I, V333P, S439G and A638G based on amino acid numbering of MdoVIN1) were consistently different between the vacuolar invertases and the cell-wall bound invertases. Although these differences are not fully understood, these five amino acid residues may be required for pH optimum and substrate specificity. In fact, the function of the valine/proline residue in the WEC-V/P-D motif has been addressed by substituting proline of CIN1 (extracellular invertase in *C. rubrum*) with a valine residue using site-directed mutagenesis (Goetz and Roitsch, 1999). This substitution of proline by valine results in a shift of the pH optimum curve and significantly reduces the cleavage rate of substrates, suggesting that the P/V amino acid difference between vacuolar invertases and cell-wall bound invertases reflects distinct enzymatic properties (Goetz and Roitsch, 1999). This difference and three others (V333P, A185S, S439G and A638G) are strictly conserved between the vacuolar invertases and the cell-wall bound invertases in the apple. However, a different irregular substitution was found in MdoVIN1 (I instead of M at 226) and MdoCIN1 (M instead of I at 226). This irregular substitution has also been detected in rice cell-wall bound invertases (Ji et al., 2005), indicating that these invertases may alter catalytic properties.

The NDPN, RDP and EC motifs each contain an acidic residue at an equivalent position in all acid invertases and it has been shown that these three motifs are indispensable for binding and catalysis (Lammens et al., 2009). However, acid invertases in the apple do not contain the NDPN motif, though the other two are conserved (Figure 2). Three amino acids (DPN) in the NDPN motif are encoded by a mini-exon, which is one of the smallest exons known in plants. Although this motif is important for the development of a transfructosylation capability (Schroeven et al., 2008), it has been shown that mini-exon skipping induced by cold stress can occur during expression of invertase genes in leaves and stem of potato (Bournay et al., 1996). In apple acid invertase genes, the lack of the DPN amino acids in the NDPN motif was observed in MdoVIN2 and 3 and MdoCIN1 and 2 due to the missing mini-exon, whereas MdoVIN1 and MdoCIN3 contained extensive sequences in this region (Fig. 1 and 2). Therefore, we hypothesize that the disappearance of the mini-exon in apple acid invertases is the result of alternative splicing events.

The neutral/alkaline invertases fall into two groups (α and β) that differed consistently at 9 amino acid residues in the conserved motifs (C252V, C256S, Y266H, Y268H, V367L, S368Q, R439P, V450T and A500S based on the amino acid numbering of MdoNIN1, Fig. 3). Although MdoNIN11 and 12 did not fall into either subgroup α and β based on amino acid homology and gene structure analysis (Fig. 1), the presence of these amino acid residues in the conserved motifs indicates that MdoNIN11 is a subgroup of α neutral/alkaline invertase. However, MdoNIN12 lacks the conserved motifs (Fig. 3), suggesting that it could not encode an active neutral/alkaline invertase and should be classed as a pseudogene. Irregular substitutions were found in MdoNIN4 (T instead of V at 450) and MdoNIN5, 6 and 7 (A instead of S at 500).

It is difficult to clarify whether subgroup α and β correspond to neutral invertase and alkaline invertase, as no experimental data are available determining the pH optimum in subgroup β neutral/alkaline invertases (Ji et al., 2005). To further identify and examine the conserved motifs in the two invertase gene families, we used the Multiple EM for Motif Elicitation (MEME) program (Baliey et al., 2006). The MEME method allows prediction of repetitive sub-sequences within a set of large sequences. As shown in Table 2, we found a total of 15 conserved motifs with low E values (Supplementary data 1). Among these, 15 motifs were shared by most apple acid invertases except MdoCIN2 (Fig. 4a). In neutral/alkaline invertases, motif 11, 12, 13, 14 and 15 were not observed in all the invertases of subgroup α, whereas motif 11 and 13 were present in all the invertases of subgroup β (Fig. 4b). This finding indicated that these differences represent the evolutionary relationship between subgroup α and β in the neutral/alkaline invertase family, suggesting that the presence of motif 11 and 13 may be used as reliable criteria for classification.

Table 2. Motif distribution in invertase families.

Invertase family	Motif number	Length (aa)	Consensus sequence	E value
Acid invertase	Motif 1	50	LRYDYG[KN][FY]YASKTF[FY]D[PQS][NA]K[EN]RR[IV]L[WL]GW[IA][NG]E[ST]D[ST][EVA]T[DA]D[VL][AKQ]KGW[AS][GS][LIV][QH][TA]IPR	5.0e-340
	Motif 2	28	N[DW][PI]NGP[LM][FY]YKGWYHLFYQYNP[KD][GS]AVWGN	1.3e-199
	Motif 3	50	I[VT]W[AG]H[AS]VS[KT]DLI[HN]WL[AHPY]L[PE][LP]A[IM][VF]P[DS][QK][WP][FY]DING[VC]W[ST]GSATILP[DG][GN]K[PI][VI][IM]LYT	1.0e-317
	Motif 4	50	G[EK][KT]LS[LA]R[SV]L[IV]DHS[IV]VES[FY][AG][QA]GG[RK]T[CV]ITSRVYPT[LE]AIY[GD]AA[HR]L[FY][LVA]FNN[AG]T[EG]	4.0e-294
	Motif 5	35	ND[NS]VQVQN[LY]AYP[AK][ND]LSDP[LY]L[RL][EKD]W[VI]K[PY][DP]GNP[LV]L[VT]P[PD]	1.9e-161
	Motif 6	21	[GH][PV]LHS[VA]PGTGMWEC[PV]DF[YF]PVS	2.3e-133
	Motif 7	41	DTSVN[GN]P[DG]VKHVLK[AV]SLDDT[RK][HY][DE][YH]Y[TA][IV]GTY[DF][IP]E[KNT][DE][KT][WY]VPD[NDK]	5.7e-17
	Motif 8	23	[TS]G[KT][QN]L[LV]QWPVEE[ILV]ETLR[GL][KN][SK][VT]K[FL]	2.7e-106
	Motif 9	15	[MV]L[SQ]W[QH]RT[AG][FY]HFQP[PE]K	5.5e-085
	Motif 10	15	[PS]GI[NG]A[TKS][DQ]FRDPTTAW	2.3e-077
	Motif 11	21	PFGLL[VT]LA[SD][EK][NT]L[ES]E[FY]TPV[YF]F[RY]	3.1e-090
	Motif 12	21	L[FM]C[SA]D[QAE][ST]RSSLA[PN][DE]VxK[QP][VT]YG	2.3e-059
	Motif 13	31	[VL]K[AP]GSVV[PE][LV]D[GV]GTA[AT]Q[LA]D[IV]E[VA]EFE[ILV]DKLA[KL][AE]	5.1e-054
	Motif 14	15	G[KH]WR[IMV][TL][IV]G[SG]K[IR]N[KH][RT]G	5.4e-046
	Motif 15	15	[IL][AS]L[VL]Y[RT][TS]KDFK[TH][WY][EV][KL]	2.0e-030
Neutral/alkaline invertase	Motif 1	50	ER[VP][DE][VC]QTGI[KR][LM]IL[NK]LCL[SA][DE]GFD[MT]FPTLL[VC][TA]DG[SC]CMIDRRMG[IV][HY]G[HY]P[LI]EIQ	5.8e-430
	Motif 2	50	MP[LF]KICYPA[LI]E[GS][HQ]EW[RQ]I[IV]TG[SC]DPKNT[PR]WSYHNGGSWP[TV]LLW[QL][LF]T[ALV]ACIK[TM]	6.9e-415
	Motif 3	50	YRYKTEEYS[HTY][TD]AVNKFN[IV]YPD[QS][IL]P[SD]W[LV][FMV]D[FW][MI]P[SE][RE]GGY[FL]IGN[LV][QS]PA[HR]MDFR	4.4e-393
	Motif 4	41	[ED]IV[KR]NF[LI]L[HK]TL[QR]LQSWEKT[VMI]DC[FY][SK][PL]G[QE]G[LV]MPASFKV[RL][TH][VD]PL	1.4e-316
	Motif 5	41	[ED][VT][LI]D[PA]DFGE[SA]AIGRVAPVDSG[LF]WWIILLRAY[GT]K[CSI][TS]GD[LY]S[LV][QA]	6.9e-316
	Motif 6	50	R[PI][EQ][IL]A[AQR][KR]A[IVA][EA][LI]AE[KS]R[LI][SR][KMS]D[NGR]WPEYYD[TG][KR][LRT][GA][RK][FY][IMV]GKQ[AS][RQ][LK][YF]QTW[ST][IV]AG[YF]L[VT]	3.3e-288
	Motif 7	50	F[YF][SM]AL[RL]C[AS][RL]E[ML]L[AK][PV][DEN]DG[GS][KA][ED][LF][VI][RE][AR][IL]N[NK]RL[HSV]ALS[FY]H[IM]R[ES]Y[YF]W[LV]D[MFL][KR][KQ]LN[ED]I	5.3e-265
	Motif 8	21	NYDQVF[IV]RDFVPS[AG][LI]AFL[LM][KN]G	5.2e-158
	Motif 9	29	[IM][EV][EK]EAW[ERT][LT]L[RK][RDN]S[VM]VY[YF][CR][GN][NQ]PVGT[IL]AAND[PH]	1.2e-162
	Motif 10	29	GN[LC][WV][SA]I[VL]SSL[AG]TP[EK]Q[SAN]HAI[LM][DN]LIE[SA][KR]W[DE][ED]	1.1e-154
	Motif 11	41	K[LM]LL[AE][ND]P[SE][KA]A[KA][LN]L[FV][WN][ED]ED[SY]EL[LV][EN][AI][CF][SV]C[AM][IL]S[AK]S[PG]R[KR]K[CR][GS][RW][KG][AN]	4.9e-072
	Motif 12	35	[ML]SCKC[QE]QAES[LFI][RS]G[AS]T[EAT][EK]D[QE][HN]G[AET][VIW]FVD[KES][ST][DK]K[AF][VIN][SPT][VFI][PN]	4.8e-040
	Motif 13	41	[TL][SA][ES][RVA][VF]R[QN][VL][MS][ST][GS][AI][LE][PT]R[LVF][GN][CDE][FN][ND][IF][CE][LKR][ISR][NY][MVI][NQ]G[VG][VI][SN]VK[PS][GL]V[NE][NRT][AI][RD]K	5.5e-043
	Motif 14	41	[TP][ST]PD[IN][DG]E[LF]K[VA]DQQL[NK][QH]E[VD]GGFGSN[TS]K[PA]TAARK[KS]KGST[RQ]KSK	3.0e-025
	Motif 15	50	MKPTCRILNR[CR]RNSAFFGFPRPA[KT]WLHGLTKTGNSSSFCVNFEQN[CS]QYHA	2.0e-024

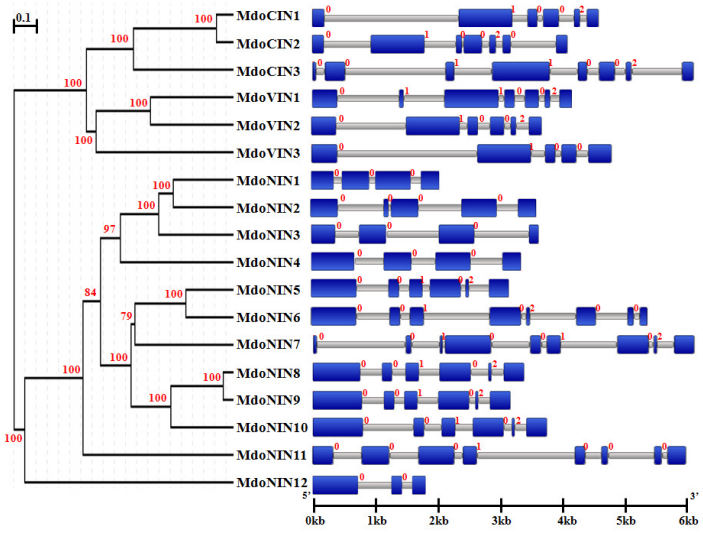


Fig 1. Phylogenetic tree and gene structures of acid invertases and neutral/alkaline invertases in the domestic apple (*Malus* × *domestica* Borkh.). Default values were used except for 100 bootstraps. Only bootstrap scores >70 are shown. Exons are drawn as boxes and connecting thin boxes indicate the positions of introns. Numbers above introns indicate the phase of the intron.

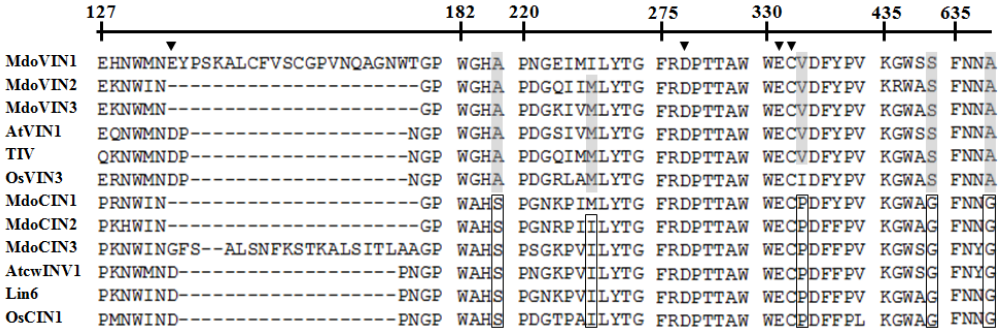


Fig 2. Alignment of the conserved regions from known acid invertases of selected higher plants. The five boxed amino acids are consistently different between cell-wall and vacuolar invertases. Arrows show the four residues (Asp133, Asp277, Glu331, and Cys332) that correspond to the enzyme active site residues proposed by Alberto et al. (2004). AtVIN1, *Arabidopsis* vacuolar invertase (At1g12240); TIV, tomato vacuolar invertase (AF465612); OsVIN3, rice vacuolar invertase (Os02g0106100); AtceINV1, *Arabidopsis* cell-wall bound invertase (At3g13790); Lin6, tomato cell-wall bound invertase (AAM28823); OsCIN1, rice cell-wall bound invertase (Os02g0534400).

Materials and methods

Identification of invertase genes

To identify members of apple invertase family, multiple database searches were performed using the Basic Local Alignment Search tool (BLAST) algorithms BLASTp and tBLASTn available on PLAZA 2.0 (http://bioinformatics.psb.ugent.be/plaza) and the Genome Database for Rosaceae (http://www.rosaceae.org/). We used the amino acid sequences of the invertase family in *Arabidopsis* as seed to BLAST all databases. The domestic apple nucleotides and protein sequences, as well as information regarding the gene structure were obtained from Apple GBrowse (http://www.rosaceae.org/gb/gbrowse/malus_x_domestica/).

Computational analysis of invertase proteins

The amino acid sequences of all invertase proteins were analyzed for domain search (SMART, http://smart.embl-heidelberg.de/) and predicted subcellular localization (SignalP, http://www.cbs.dtu.dk/services/SignalP/; MitoProt, http://ihg.gsf.de/ihg/mitoprot.html; PSORT, http://psort.hgc.jp/; TargetP, http://www.cbs.dtu.dk/services/TargetP/). The program MEME was used for the recognition of motifs in invertases. MEME was run from the web server (http://meme.sdsc.edu/meme4_6_1/cgi-bin/meme.cgi) with the following parameters:

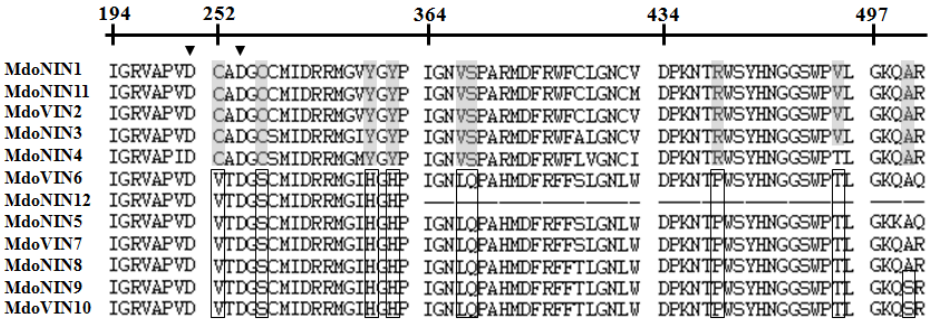


Fig 3. Alignment of the conserved regions from neutral/alkaline invertases. The nine boxed amino acids are consistently different between subgroup α and subgroup β invertases. Arrows show the two residues that correspond to the enzyme active site residues proposed by Ji et al. (2005).

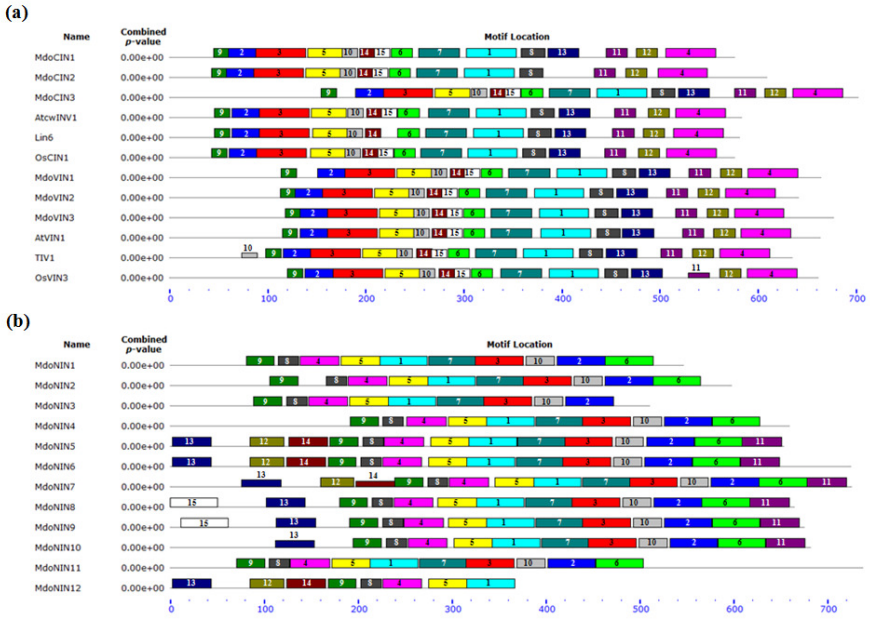


Fig 4. Motif distribution in apple invertases. Motifs of acid invertases (a) and neutral/alkaline invertases (b) were investigated using the MEME web server. The different motifs are represented by numbers.

minimum motif width, 6aa; maximum width, 50 aa; maximum motif number, 15.

Multiple-sequence alignment and phylogenetic analysis of invertase sequences

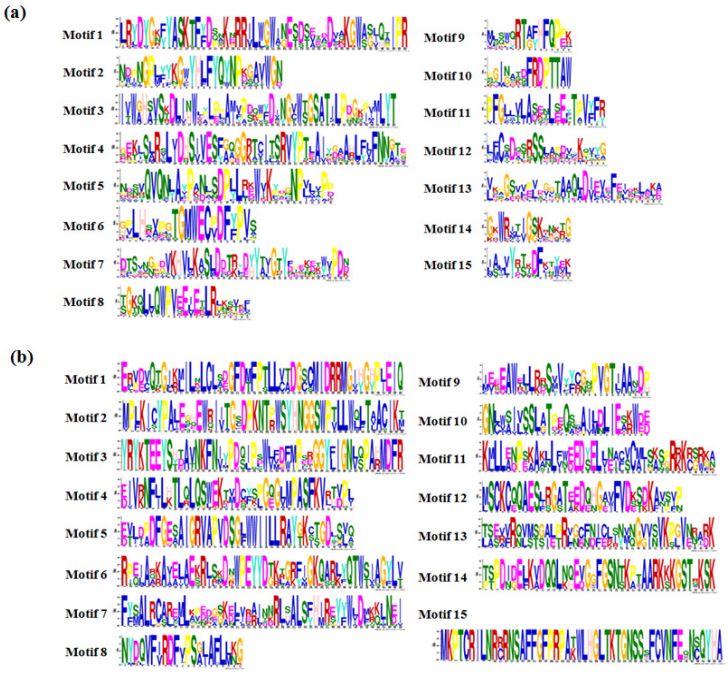
Multiple-sequence alignments of invertase amino acid sequences were performed using ClustalW (http://bioinformatics.ubc.ca/resources/tools/clustalx) and were manually corrected. Phylogenetic analysis was carried out using the neighbor-joining method, and the phylogenetic tree was displayed using TreeTop (http://www.genebee.msu.su/services/phtree_reduced.html).

Gene structure analysis of the invertase gene family

The intron distribution and gene organization studies in the invertase family were performed using the construction of an intron-exon organization map. The positional conservation of different classes of intronic phases was analyzed to infer the evolutionary relatedness among the members of the invertase family. The intron phases of different introns were determined using Wise 2.0 (http://www.ebi.ac.uk/Tools/Wise2). For this, amino acid and corresponding total gene nucleotide sequences were aligned to determine the position of introns.

Conclusion

This study compiles a full list of the invertase families in *Malus* × *domestica* Borkh. *In silico* analysis of public genomic databases using BLASTp and tBLASTn resulted in the identification of six and 12 potential non-redundant members belonging to the acid invertases and neutral/alkaline invertases, respectively. A further and motivating challenge would be to check the protein activities and functions of these genes using transgenic and antisense approaches. In addition, we have found conserved motifs which may be useful for classification of neutral/alkaline invertases into subgroups α and β. Our genomic and bioinformatics analysis supports a solid



Supplementary data 1. Sequence logos of apple invertase motifs analyzed by the MEME program. Over-represented motifs in acid invertases (a) and neutral/alkaline invertases (b) were identified by MEME analysis. The overall height of the stack indicates the level of sequence conservation. The height of residues within the stack indicates the relative frequency of each residue at that position.

foundation for further functional characterization of invertases in *Malus* × *domestica* Borkh., and will provide the basis for future research on the regulation of source/sink-relations via the invertase family.

References

Alberto F, Bignon C, Sulzenbacher G, Henrissat B, Czjzek M (2004) The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. J Biol Chem. 279:18903-18910

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 34: W369-W373

Barratt DHP, Barber L, Kruger N, Smith AM, Wang TL, Martin C (2001) Multiple, distinct isoforms of sucrose synthase in pea. Plant Physiol. 127: 655-664

Baud S, Vaultier M-N, Rochat C (2004) Structure and expression profile of the sucrose synthase multigene family in *Arabidopsis*. J Exp Bot. 55: 397-409

Bocok PN, Morse AM, Dervinis C, Davis JM (2008) Evolution and diversity of invertase genes in *Populus trichocarpa*. Planta. 227: 565-576

Bonfig, K. B., Schreiber, U., Gabler, A., Roitsch, T., and Berger, S. (2006). Infection with virulent and avirulent *P. Syringae* strains differentially affects photosynthesis and sink metabolism in *Arabidopsis* leaves. Planta. 225: 1-12

Bournay A-S, Hedley PE, Maddison A, Waugh R, Machray GC (1996) Exon skipping induced by cold stress in a potato invertase gene transcript. Nucleic Acids Res. 24: 2347-2351

Carlson M, Botstein D (1982) Two differentially regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase. Cell. 28: 145-154

Cevik V, Ryder CD, Popovich A, Manning K, King GJ, Seymour GB (2010) A FRUITFULL-like gene is associated with genetic variation for fruit flesh firmness in apple (*Malus domestica* Borkh.). Tree Genet Genomes. 6:271-279

Charkazi F, Ramezanpour S, Soltanloo H (2010) Expression pattern of two sugar transporter genes (SuT4 and SuT5) under salt stress in wheat. Plant Omics J. 3: 194-198

Chauve C, Doyon J, El-Mabrouk N (2008) Gene family evolution by duplication, speciation, and loss. J Comput Biol. 15: 1043-1062

Di Maro A, Dosi R, Ferrara L, Rocco M, Sepe J, Ferrari G, Parente A (2011) Free amino acid profile of *Malus domestica* Borkh cv. Annurca from the Campania Region and other Italian vegetables. Aust J Crop Sci. 5(2):154-161.

Ehness R, Ecker M, Godt D, Roitsch T (1997) Glucose and stress independently regulate source/sink relations and defense mechanisms via signal transduction pathways involving protein phosphorylation. Plant Cell. 9: 1825-1841

Ehrmann J Jr, Kolek A, Kod'ousek R, Zapletalova J, Lisova S, Murray PG, Drabek J, Kolar Z (2003) Identification of the gene for beta-fructofuranosidase of *Bifidobacterium lactis* DSM10140(T) and characterization of the enzyme expressed in *Escherichia coli*. Curr Microbiol. 46:391-397

Farrokhi J, Darvishzadeh R, Naseril L, Azar MM, Maleki HH (2011) Evaluation of genetic diversity among Iranian apple (*Malus domestica* Borkh.) cultivars and landraces using simple sequence repeat markers. Aust J Crop Sci .5: 815-821

Fedorov A, Suboch G, Bujakov M, Fedorova L (1992) Analysis of nonuniformity in intron phase distribution. Nucleic Acids Res. 20: 2553-2557

Feng S, Wang Y, Yang S, Xu Y, Chen X (2011) Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. Planta. 232: 245-255

Gallagher JA, Pollock CJ (1998) Isolation and characterization of a cDNA clone from *Lolium temulentum* L. encoding for a sucrose hydrolytic enzyme which shows alkaline/neutral invertase activity. J Exp Bot. 49: 789-795

Godt DE, Roitsch T (1997) Regulation and tissue-specific distribution of mRNA for three extracellular invertase isoenzymes of tomato suggests an important function in establishing and maintaining sink metabolism. Plant Physiol. 115: 273-282

Goetz M, Roitsch T (1999) The different pH optima and substrate specificities of extracellular and vacuolar invertases from plants are determined by a single amino-acid substitution. Plant J. 20:707-711

Haouazine-Takvorian N, Tymowska-Lalanne Z, Takvorian A, Tregear J, Lejeune B, Lecharny A, Kreis M (1997) Characterization of two members of the *Arabidopsis thaliana* gene family, *At_fruct3* and *At_fruct4*, coding for vacuolar invertases. Gene. 197:239-251

Hyun TK, Hoffmann A, Sinha AK, Roitsch T (2009) Tomato mitogen activated protein kinases regulate the expression of extracellular invertase *Lin6* in response to stress related stimuli. Funct Plant Biol. 36:1088-1097

Ji X, Van den Ende W, Laere AV, Cheng S, Bennett J (2005) Structure, evolution and expression of two invertase gene families of rice. J Mol Evol. 60:615-634

Kron P, Husband BC (2009) Hybridization and the reproductive pathways mediating gene flow between native Malus coronaria and domestic apple, *M. domestica*. Botany. 87: 864-874

Lammens W, Roy KL, Schroeven L, Laere AV, Rabijins A, Van den Ende W (2009) Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. J Exp Bot. 60: 727-740.

Lee HS, Sturm A (1996) Purification and characterization of neutral and alkaline invertase from carrot. Plant Physiol. 112:1513-1522

Murayama S, Handa H (2007) Genes for alkaline/neutral invertase in rice: alkaline/neutral invertases are located in plant mitochondria and also in plastids. Planta. 225: 1193-1203

Nonis A, Ruperti B, Pierasco A, Canaguier A, Adam-Blondon A-F, Gaspero GD, Vizzotto G (2008) Neutral invertases in grapevine and comparative analysis with *Arabidopsis*, poplar and rice. Planta. 229: 129-142

Pan QH, Yu XC, Zhang N, Zou X, Peng CC, Wang XL, Zou KQ, Zhang DP (2006) Activity, but not expression, of soluble and cell wall-bound acid invertases is induced by abscisic acid in developing apple fruit. J Integrat Plant Biol. 48: 536-549

Pan Q-H, Zou K-Q, Peng C-C, Wang X-L, Zhang D-P (2005) Purification, biochemical and immunological characterization of acid invertases from apple fruit. J Integr Plant Biol. 47: 50-59

Qi X, Wu Z, Li J, Mo X, Wu S, Chu J, Wu P (2007) AtCYT-INV1, a neutral invertase, is involved in osmotic stress-induced inhibition on lateral root growth in Arabidopsis. Plant Mol Biol. 64: 575-587

Rae AL, Casu RE, Perroux JM, Jackson MA, Grof CPL (2011) A soluble acid invertase is directed to the vacuole by signal anchor mechanism. J Plant Physiol. 168:983-989

Roitsch T, Balibrea ME, Hofmann M, Proels R, Shina AK (2003) Extracellular invertase: Key metabolic enzyme and PR protein. J Exp Bot. 54: 513-524

Roitsch T, Gonzalez MC (2004) Function and regulation of plant invertases: sweet sensations. Trends Plant Sci. 9: 606-613

Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV (2003) Remarkable interkingdom conservation of intron positions and massive lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol. 13: 1512-1517

Schroeven L, Lammens W, Van Laere A, Van den Ende W (2008). Transforming wheat vacuolar invertase into a high affinity sucrose:sucrose 1-fructosyltransferase. New Phytol. 180: 822-831

Sinha AK, Hofmann MG, Römer U, Köckenberger W, Elling L, Roitsch T (2002) Metabolizable and non-metabolizable sugars activate different signal transduction pathways in tomato. Plant Physiol. 128: 1480-1489

Sherson SM, Alford HL, Forbes SM, Wallace G, Smith SM (2003) Roles of cell-wall invertases and monosaccharide transporters in the growth and development of Arabidopsis. J Exp Bot. 54: 525-531

Sturm A, Chrispeels MJ (1990) cDNA cloning of carrot extracellular beta-fructosidase and its expression in response to wounding and bacterial-infection. Plant Cell. 2:1107-1119

Sturm A, Hess D, Lee HS, Lienhard S (1999) Neutral invertase is a novel type of sucrose-cleaving enzyme. Physiol Plantarum. 107: 159-165

Tetlow IJ, Morell MK, Emes MJ (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. J Exp Bot 55: 2131-2145

Vargas W, Cumino A, Salerno GL (2003) Cyanobacterial alkaline/neutral invertases. Origin of sucrose hydrolysis in the plantcytosol? Planta. 216:951-960

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nature Genet. 42: 833-841

Welham T, Pike J, Horst I, Flemetakis E, Katinakis P, Kaneko T, Sato S, Tabata S, Perry J, Parniske M, Wang TL (2009) A cytosolic invertase is required for normal growth and cell development in the model legume, *Lotus japonicas*. J Exp Bot. 60: 3353-3365

Winter H, Huber SC (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. Crit Rev Biochem Mol Biol. 35: 253-289

Xiang L, Le Roy K, Bolouri-Moghaddam M-R, Vanhaecke M, Lammens W, Rolland F, Van den Endel W (2011) Exploring the neutral invertase–oxidative stress defence connection in *Arabidopsis thaliana*. J Exp Bot. 62: 3849-3862

Zhuang J, Yao QH, Xiong A-S, Zhang J (2011) Isolation, phylogeny and expression pattern of AP2-like genes in apple (*Malus* × *domestica* Borkh). Plant Mol Biol Rep. 29: 209-216