Mol Genet Genomics (2012) 287:663–678
DOI 10.1007/s00438-012-0707-7

ORIGINAL PAPER

A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the *Ma* locus is associated with low fruit acidity in apple

Yang Bai · Laura Dougherty · Mingjun Li · Gennaro Fazio · Lailiang Cheng · Kenong Xu

Received: 12 May 2012 / Accepted: 3 July 2012 / Published online: 18 July 2012

Abstract Acidity levels greatly affect the taste and flavor of fruit, and consequently its market value. In mature apple fruit, malic acid is the predominant organic acid. Several studies have confirmed that the major quantitative trait locus *Ma* largely controls the variation of fruit acidity levels. The *Ma* locus has recently been defined in a region of 150 kb that contains 44 predicted genes on chromosome 16 in the Golden Delicious genome. In this study, we identified two aluminum-activated malate transporter-like genes, designated *Ma1* and *Ma2*, as strong candidates of *Ma* by narrowing down the *Ma* locus to 65–82 kb containing 12–19 predicted genes depending on the haplotypes. The *Ma* haplotypes were determined by sequencing two bacterial artificial chromosome clones from G.41 (an apple rootstock of genotype *Mama*) that cover the two distinct haplotypes at the *Ma* locus. Gene expression profiling in 18 apple germplasm accessions suggested that *Ma1* is the major determinant at the *Ma* locus controlling fruit acidity as *Ma1* is expressed at a much higher level than *Ma2* and the *Ma1* expression is significantly correlated with fruit titratable acidity ($R^2 = 0.4543$, $P = 0.0021$). In the coding sequences of low acidity alleles of *Ma1* and *Ma2*, sequence variations at the amino acid level between Golden Delicious and G.41 were not detected. But the alleles for high acidity vary considerably between the two genotypes. The low acidity allele of *Ma1*, *Ma1-1455A*, is mainly characterized by a mutation at base 1455 in the open reading frame. The mutation leads to a premature stop codon that truncates the carboxyl terminus of Ma1-1455A by 84 amino acids compared with Ma1-1455G. A survey of 29 apple germplasm accessions using marker $CAPS_{1455}$ that targets the SNP_{1455} in *Ma1* showed that the $CAPS_{1455A}$ allele was associated completely with high pH and highly with low titratable acidity, suggesting that the natural mutation-led truncation is most likely responsible for the abolished function of *Ma* for low pH or high acidity in apple.

Keywords *Malus* (apple) · Fruit acidity · *Ma* · Haplotype · Aluminum-activated malate transporter (ALMT)

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-012-0707-7) contains supplementary material, which is available to authorized users.

Y. Bai · L. Dougherty · K. Xu (✉)
Department of Horticulture, New York State Agricultural Experiment Station, Cornell University, Geneva, NY 14456, USA
e-mail: kx27@cornell.edu

M. Li · L. Cheng (✉)
Department of Horticulture, Cornell University, Ithaca, NY 14853, USA
e-mail: lc89@cornell.edu

G. Fazio
Plant Genetic Resources Unit USDA ARS, New York State Agricultural Experiment Station, Cornell University, Geneva, NY 14456, USA

Introduction

Organic acids, many of which are intermediates in metabolic processes, play significant roles in fruit growth, maturation, ripening and softening. The level of organic acids greatly affects the taste and flavor of fruit, and consequently its market value. The major determinants of fruit acidity include malic acid, citric acid and tartaric acid. In

mature apple fruit, malic acid is the predominant organic acid although other organic acids such as citric acid, fumaric acid and quinic acid are detectable (Zhang et al. 2010). Apple fruit varies widely in pH and titratable acidity (TA) levels. However, the acceptable range for dessert apple fruit is often measured within a range of 3.1–3.8 in pH or 3.0–10.0 mg/ml in TA, beyond either end of which, fruit acidity is either too high or too low for fresh consumption (Brown and Harvey 1971; Nybom 1959; Visser and Verhaegh 1978).

Inheritance of high pH or low TA in apple fruit was attributed to a recessive gene in early studies (Brown and Harvey 1971; Nybom 1959; Visser and Verhaegh 1978). The acidity locus was mapped to linkage group (LG) 16 and designated as *Ma* (*malic acid*), where *Ma* is noted for the dominant low pH or high acidity allele and *ma* for high pH or low acidity allele (Maliepaard et al. 1998). In other species, major genes or quantitative trait locus (QTL) similar to *Ma* in controlling fruit acidity include *acitric* in citrus (Fang et al. 1997), *SS* in pomegranate (Jalikop 2007) and *pH* in sweet melon (Lerceteau-Köhler et al. 2012), where low acidity is also inherited recessively. The major gene *D* in peach, however, acts differently with low acidity being dominant over high acidity (Boudehri et al. 2009) although both peach and apple are members of the Rosaceae family. In tomato, complex and multiple QTLs are reported in conditioning fruit acidity levels (Fulton et al. 2002).

The primary role of the *Ma* locus in determining fruit pH and TA in apple was also demonstrated in QTL studies as a major QTL was consistently detected on LG 16 (Kenis et al. 2008; Liebhard et al. 2003; Xu et al. 2011). In addition to the *Ma* QTL, multiple minor QTLs of significant effect on acidity were identified in these studies. Although the minor QTLs are less consistent, the notion that the *Ma* locus and minor QTLs collectively determine fruit acidity levels is widely accepted. Consistent with this notion, a recent report finds that a mixed model of a major gene and polygenes fits best in explaining the apple acidity variation in a complex breeding population among four models (mixed, Mendelian, polygenic and environmental) tested (Iwanami et al. 2012).

Malate metabolism in fruit cells may involve several pathways according to recent reviews (Beruter 2004; Sweetman et al. 2009). Malate synthesis is considered to occur locally in fruit. The primary path is glycolysis of hexoses derived from sucrose and/or sorbitol, which are imported from leaves, in the cytosol of parenchyma cells of fruit. Depending upon developmental stages, pathways of photosynthesis in the chloroplast, the tricarboxylic acid (TCA) cycle in mitochondrion, and glyoxylate cycle in glyoxysome in fruit cells also appear to be important for malate synthesis. For degradation of malate, gluconeogenesis and the TCA cycle are likely the main pathways. It is possible that the various enzymes involved in malate synthesis and degradation, such as phosphoenolpyruvate carboxylase (PEPC), NADP-dependent malic enzyme (NADP-ME), and NAD-dependent malate dehydrogenase (NAD-MDH) and many others, may play a role in regulating malate metabolism in fruit cells, thus acidity of fruit. In addition, the vacuolar transporters, such as the vacuolar pumps, e.g. V-ATPase (Schumacher and Krebs 2010), tonoplast dicarboxylate transporter, e.g. AttDT (Emmerlich et al. 2003), and members of the aluminum-activated malate transporter1 (ALMT1) family proteins (Barbier-Brygoo et al. 2011), e.g. AtALMT9 (Kovermann et al. 2007) and AtALMT6 (Meyer et al. 2011), may also play critical roles in determining fruit acidity as they can regulate the malate accumulation in and release from the vacuole in plant cells.

In apple, the pattern of malate accumulation and degradation is similar in developing fruits of several high/medium acid varieties studied, i.e. malic acid level significantly increases in young fruit (around 4 weeks after full bloom) and then progressively decreases through maturity although the total content per fruit increases along with fruit development (Beruter 2004; Hulme and Wooltorton 1957; Ulrich 1970; Zhang et al. 2010). Several recent studies have attempted to identify candidate genes and/or enzymes that may be associated with the acidity variations in apple fruit. Using a cDNA-AFLP-based approach, a gene designated *Mal-DDNA* (DQ417661) of unknown function previously appeared to be associated with low acid in a population segregating for fruit acidity (Yao et al. 2007). Direct profiling of expression patterns and enzyme activities of genes putatively involved in malate metabolism, including *MdPEPC* (EU315246, for PEPC), *MdcyME* (DQ280492, for NADP-ME) and *MdVHA-A* (EF128033, for subunit A of vacuolar H^+-ATPase), found that there were differences between low and high acid genotypes (Yao et al. 2009). Involvement of genes encoding NADP-ME (GD254910, degradation of malate) and NAD-MDH (GD254856, synthesis of malate) in malate accumulation and degradation was also reported in a cDNA microarray analysis of 1,536 genes (Soglio et al. 2009). Moreover, a gene encoding NAD-MDH (DQ221207) has been functionally demonstrated to be involved in malate synthesis in apple (Yao et al. 2011). Overall, these data suggest that the genes and/or enzymes studied above may contribute to the variation of fruit acidity.

However, a detailed analysis of a low acid variety Usterapfel and its high acid mutant (Beruter 1998, 2004) indicated that key enzymes in malic acid metabolism, PEPC, NAD-MDH and NADP-ME, may not play a key role in determining the difference in fruit acidity because there was no difference in the catalytic activity of these

enzymes between the two contrasting genotypes. Examining the localities of these genes in the apple genome (Velasco et al. 2010) appeared to support that these enzymes and genes involved in malate metabolism may not be *Ma* because none of those studied above, including *Mal-DDNA*, is on chromosome 16 where the *Ma* gene resides.

To uncover the genes underlying *Ma*, we had defined the *Ma* locus to a region of 150 kb encompassing 44 predicted genes on chromosome 16 in the Golden Delicious genome in a previous study (Xu et al. 2011). In this study, we report the identification of two aluminum-activated malate transporter (ALMT)-like genes, *Ma1* and *Ma2*, as strong candidates of *Ma*. We show that the *Ma* region is reduced to a genomic segment of 65 kb containing 19 predicted genes in Golden Delicious by developing three new markers and analyzing two more populations. In two bacterial artificial chromosome (BAC) clones that are distinguishable with haplotype *ma* and *Ma* from apple rootstock G.41, the *Ma* region harbors 12 predicted genes, including *Ma1* and *Ma2*, although it spans over 71 kb in haplotype *ma* and 82 kb in haplotype *Ma*. We further show that the expression of *Ma1* is significantly correlated with fruit acidity levels, whereas *Ma2* is expressed constantly at low levels across high and low acidity fruit. Finally, we show that a single nucleotide mutation in the open reading frame of *Ma1* that leads to truncation of Ma1 by 84 amino acids is perfectly associated with high pH and highly with low TA in 29 apple germplasm accessions studied.

Materials and methods

Plant materials and fruit pH and TA evaluation

Four half-sib F_1 populations of interspecific crosses were used to further narrow down the *Ma* locus, namely GMAL 4590, GMAL 4592, GMAL 4595 and GMAL 4596 (Table 1). The seed parent of the four populations is Royal Gala (*Mama*), a widely grown apple cultivar (*Malus* × *domestica* Borkh.). The pollen parents are elite clones of *M. sieversii* (i.e. of fruit size close to cultivated apple) collected from Kazakhstan (Forsline et al. 2003), including PI 613971 (*Mama*), PI 613978 (*mama*), PI 613988 (*Mama*) and PI 613979 (*Mama*), respectively. *M. sieversii* has been proven to be the major progenitor species of *M.* × *domestica* (Velasco et al. 2010). The four F_1 populations were derived from controlled crosses made in 2002 and planted on their own seedling roots in 2004 in the USDA-ARS Apple Germplasm Repository, Geneva, NY, USA. Populations GMAL 4590 of 216 individuals and GMAL 4595 of 222 genotypes were used in a previous study (Xu et al. 2011), but 36 and 23 individuals that did not bear fruit in 2010 from the two crosses, respectively, were not included previously. These individuals bore fruit in 2011 and were added in this study (Table 1). Populations GMAL 4592 (155 genotypes) and GMAL 4596 (215 genotypes) were used for the first time. Overall, there are 724 fruiting individuals in a total of 808 genotypes in the four populations (Table 1).

Evaluation of fruit maturity and fruit acidity (pH paper estimates and instrumental measurements of pH and TA) was conducted similarly as described previously (Xu et al. 2011). Briefly, fruit maturity was determined via starch test that corresponds to Cornell Starch Index 4.0–6.0 (Blanpied and Silsby 1992). For pH estimates, pH paper (Hydrion Papers, pH 3.0–5.5, Micro Essential Laboratory Inc., Brooklyn, NY, USA) was applied onto the fruit cuts at maturity in the orchard. For pH and TA instrumental measurements, fruit juice samples were prepared by pooling 5–10 fruits per genotype at maturity. The pooled juices were then measured with a pH meter (Accumet AB15, Fisher Scientific, Pittsburgh, PA, USA), and subsequently, an autotitrator (Metrohm 848 Titrino Plus and Metrohm 869 Compact Sample Changer, Herisau, Switzerland). Evaluation for most genotypes was conducted either in 2010 or 2011. But for the informative recombinants between markers CH05c06 and CH02a03 or CH05a09 (Figs. 1a–d, 2), pH meter-based measurements were obtained in both years if fruits were available.

Table 1 List of populations segregating for fruit pH

Population	Seed parent	Pollen parent	All genotypes	Fruiting genotypes	pH ≤ 3.8 (*Ma_*)	pH ≥ 3.9 (*mama*)	Ratio (*Ma_*:*mama*)	*P* (χ^2)
GMAL 4590[a]	Royal Gala (*Mama*)	PI 613971 (*Mama*)	216	190	143	47	3:1	0.9300 (0.007)
GMAL 4592	Royal Gala (*Mama*)	PI 613978 (*mama*)	155	133	82	51	1:1	0.0070 (7.226)
GMAL 4595[b]	Royal Gala (*Mama*)	PI 613988 (*Mama*)	222	213	157	56	3:1	0.3400 (0.189)
GMAL 4596	Royal Gala (*Mama*)	PI 613979 (*Mama*)	215	198	156	42	3:1	0.2200 (1.515)
Total			808	734	538	196		

Estimated by pH paper. But pH meter reads for the informative recombinants (Figs. 1, 2) and the 190 genotypes of GMAL 4595 (measured in 2010) were used

[a] The fruiting genotypes include one set of 36 bearing fruit in 2011 and the other set of 154 in 2010

[b] The fruiting genotypes include one set of 23 bearing fruit in 2011 and the other of 190 in 2010

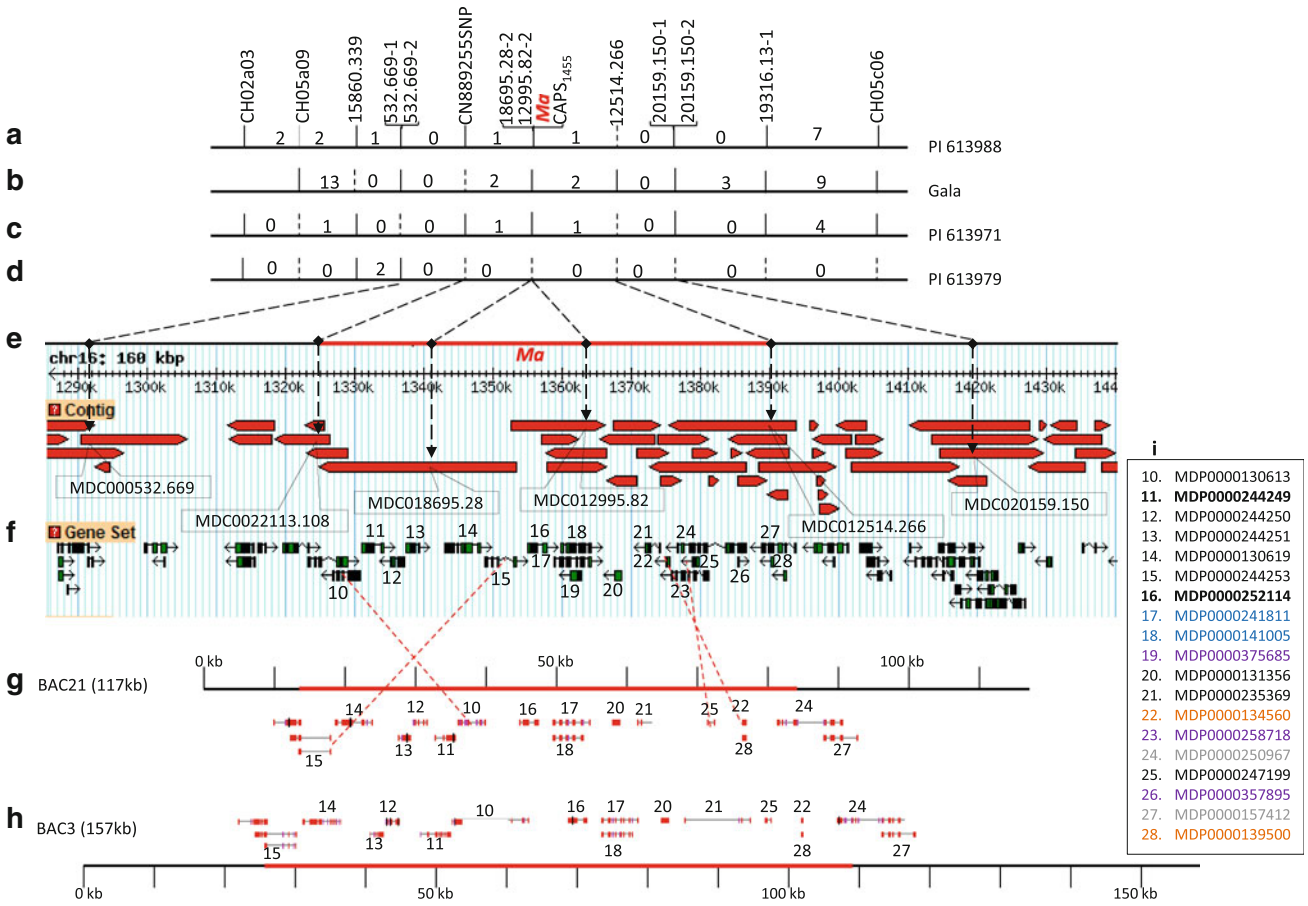

Fig. 1 Fine genetic and haploid specific physical maps of the *Ma* locus on chromosome 16. Fine genetic maps of *Ma* in PI 613988 (**a**), Royal Gala (**b**), PI 613971 (**c**) and PI 613979 (**d**). The number between the markers stands for the number of informative recombinants found in the interval. The *solid vertical lines* indicate the position of mapped markers, and the *broken vertical lines* are for positions of the presumed markers. **e** Physical map of the *Ma* region (a Genome Brower snapshot from the GDR website) in Golden Delicious (GD). The *Ma* region of 65 kb between markers CN889255SNP and 12514.266 is shown with a red solid bar. The labeled contigs indicate the source sequences, from which the markers were developed. **f** Predicted genes in the *Ma* region of GD. There are 19 predicted genes, which are conveniently labeled with #10–28, respectively. **g** A sequenced clone BAC21 of G.41 covering the *Ma* region. The *numbers* show the physical locations of the corresponding genes predicted in GD. **h** A sequenced clone BAC3 of G.41. **i** A list of the 19 genes predicted. Genes not present in G.41 (in *purple*): MDP0000375685 (#19), MDP0000258718 (#23) and MDP0000357895 (#26). Genes outside of the *Ma* region in G.41 (in *grey*): MDP0000250967 (#24) and MDP0000157412 (#27). Genes spliced alternatively (in *blue*): MDP0000241811 (#17) and MDP0000141005 (#18). Genes with duplicated IDs (in *orange*): MDP0000134560 (#22) and MDP0000139500 (#28). Candidate genes of *Ma* (in *black*): 12 genes, including MDP0000252114 (*Ma1*) and MDP0000244249 (*Ma2*) (colour figure online)

To examine the association between fruit acidity and the mutation at base 1455 in gene *Ma1*, pH and TA of mature fruit were evaluated for 29 representative apple cultivars and accessions (Table S1, including three progeny from GMAL 4595) grown in the USDA-ARS Apple Germplasm Repository, Geneva, NY, USA.

Marker development and genetic mapping

New simple sequence repeat (SSR) markers linked to *Ma* were developed using the same strategy as described in Xu et al. (2011). Briefly, DNA sequences of contigs between the two existing markers 532.669-2 and 20159.150-1 (Xu et al. 2011) were downloaded from the Genome Database for Rosaceae (GDR, http://www.rosaceae.org/), and analyzed for the presence of potential SSRs markers using the web-based program BatchPrimer3 (http://probes.pw.usda.gov/batchprimer3/index.html) (You et al. 2008). Genomic DNA isolation, PCR and SSR analyses were conducted as described previously (Xu et al. 2011).

For single nucleotide polymorphism (SNP) marker development, we targeted expressed sequence tags (ESTs) that are present in the region between the two markers 532.669-2 and 20159.150-1. The presence and segregation of SNP were determined by direct sequencing of the PCR products amplified from the five parents and the

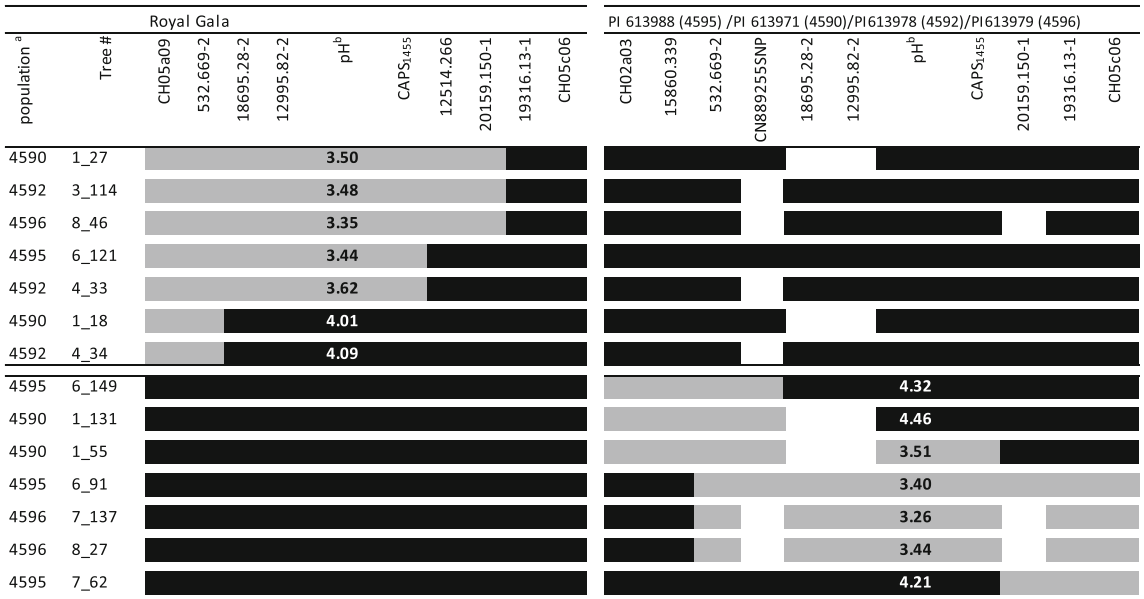

Fig. 2 Key informative recombinants identified from populations GMAL 4590, 4592, 4595 and 4596 and their marker genotypes. [a] 4590 = GMAL 4590; 4592 = GMAL 4592; 4595 = GMAL 4595; 4596 = GMAL 4596. Marker genotype linked to the low pH (high acid) allele *Ma* in coupling phase is shown in *grey*, and those linked to the high pH (low acid) allele *ma* in *black*; [b] pH data for these recombinants were measured by a pH meter in 2010 (for GMAL 4595) and 2011 (for the rest)

informative recombinants between markers CH05c06 and CH02a03 or CH05a09.

$CAPS_{1455}$ is a cleaved amplified polymorphic sequence (CAPS) marker targeting base 1455 in the open reading frame of gene *Ma1*. The PCR program includes 2 min at 98 °C, 35 cycles of 10 s at 98 °C, 15 s at 55 °C and 90 s at 72 °C, and a final 5 min at 72 °C. PCR were conducted in a volume of 20 µl, which includes 1× PrimeSTAR® MAX DNA Polymerase (R045A, Takara/Clontech, Mountain View, CA, USA), 0.5 mM of each primer and 30 ng of genomic DNA. Restriction digestion was performed at 37 °C for overnight in a volume of 20 µl that contains 10 µl PCR products, 2 U of BspHI (New England Biolabs, Ipswich, MA, USA), 1× NEBuffer 4. Gel analysis of $CAPS_{1455}$ was conducted with agarose gels of 1.5 % (w/v).

Mapping of markers in relation to the *Ma* locus was conducted with the informative recombinants between SSR markers Hi22f06 and CH02a03 or CH05a09, which were identified from the four populations described above. The informative recombinants, as explained previously (Xu et al. 2011), refer to individual trees developed from zygotes that combined a parental-type gamete of an allele of *ma* (non-recombinant) with a recombined gamete near the *Ma* locus. Recombinants derived from zygotes that include a parental-type gamete of an allele of *Ma* (non-recombinant) are considered non-informative in this study. This is because the strong dominance effect of allele *Ma* from the parental-type gamete would make the effect of allele *Ma* or *ma* from a recombined gamete difficult, if not impossible, to detect by pH or TA values.

Identification of BAC clones and sequencing

The BAC library was constructed from G.41, an apple rootstock developed from an interspecific cross Malling 27× Robusta 5 (Cummins et al. 2006). The mature fruits of G.41 are small (2–3 cm in diameter) and have astringent taste (not edible) and high acidity (pH 3.1, TA = 13 mg/ml), suggesting that G.41 has a genotype of *MaMa* or *Mama*. The BAC library was constructed by Amplicon Express (Pullman, WA) using a restriction enzyme/vector combination of MboI/pECBAC1. It has a total of 41,472 clones with an average insert size of 120 kb, which provides approx. 6.6× coverage of the apple genome. The library was pooled at two levels with a total of five dimensions. The first level is the nine super pools, each of which comprises 12 plates containing a total of 4,608 (12 × 384) clones. The second level is the nine sets of matrix pools, and each set was pooled from the 12 plates associated with one of the nine super pools. One set of matrix pool includes eight matrix plate pools (P:1.2.3, P:4.5.6, P:7.8.9, P:10.11.12, P:1.5.9, P:2.6.10, P:3.7.11 and P:4.8.12) pooled from three of the 12 individual plate pools, eight matrix row pools (R:A.B.C.D, R:E.F.F.G, R:I.J.K.L, R:M.N.O.P, R:A.E.I.M, R:B.F.J.N, R:C.G.K.O and R:D.H.L.P) pooled from four of the 16 individual row pools, ten matrix column pools (C:1.2.3.4.5.6, C:7.8.9.10.11.12, C:13.14.15.16.17.18, C:19.20.21.22.23.24, C:1.7.13.19, C:2.8.14.20, C:3.9.15.21, C:4.10.16.22, C:5.11.17.23 and C:6.12.18.24) pooled from six or four of the 24 individual column pools, and ten matrix diagonal pools

(D:1.2.3.4.5.6, D.7.8.9.10.11.12, D.13.14.15.16.17.18, D.19.20.21.22.23.24, D:1.7.13.19, D:2.8.14.20, D:3.9.15.21, D:4.10.16.22, D:5.11.17.23 and D:6.12.18.24) pooled from six or four of the 24 individual diagonal column pools. Different from a common individual column pool, which is pooled from the same column across a stack of 12 plates, an individual diagonal column pool comprises 12 varying columns on the diagonal line from the stack of 12 plates. For example, diagonal column pool D1 is pooled from column (C) 1 in plate (P) 1, C2 in P2… and C12 in P12, and pool D2 is from C2 in P1, C3 in P2… and C1 in P12.

Screening of BAC clones was conducted on the library super pools and their associated matrix pools using the *Ma*-linked PCR-based markers we developed. BAC clones originated from the *Ma* region were restricted with endonuclease BamHI and NotI (New England Biolabs, Ipswich, MA, USA) and then analyzed by pulse field gel electrophoresis (PFGE) using CHEF-DR II System (Bio-Rad, Hercules, CA, USA) for preliminary fingerprinting and size estimation. BAC sequencing was conducted using a 454 GS FLX system at Cornell Biotechnology Center and assembled with the Newbler Assembly (454 Life Sciences, Branford, CT, USA).

Gene prediction and annotation at the *Ma* locus

Genes predicted in the *Ma* region of Golden Delicious (Velasco et al. 2010) were adopted and their coding sequences (CDS) and deduced protein sequences were downloaded from GDR. Confirmation of gene annotation was carried out by searching the GenBank non-redundant protein database using the BLASTP program with a cutoff expected value of 10^{-9}. Putative functions of the predicted genes were annotated with the GenBank accession numbers of the highest similarities and associated functions if known.

Quantitative (q) RT-PCR assay of *Ma* candidate genes

Total RNA from mature fruits of 18 of the 29 apple accessions (Table S1) was isolated using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) with three biological replicates. Reverse transcription reactions were carried out with 1.8 μg of total RNA using the Superscript III RT (Invitrogen, Carlsbad, CA, USA). The resulting first strand cDNA was diluted by fivefold, and then used as templates for qRT-PCR analysis, in which a *Malus* (Gala) actin gene/EST (EB136338) served as a reference with primers Actin F (5′-GGCTGGATTTGCTG GTGATG-3′) and Actin R (5′-TGCTCACTATGCCGTGC TCA-3′).

Two rounds of qRT-PCR were performed. In the initial round, all 12 genes predicted at the *Ma* locus were screened with their gene-specific primers (Table S2). Four low acid (Britegold, KAZ 96 08-17, Novosibirski Sweet and Sweet Delicious) and four high acid (Cox's Orange Pippin, Golden Delicious, Marshall McIntosh and Winter Majetin) apple accessions were used. cDNA of each genotype was bulked evenly from the three replicates and then used for qRT-PCR. In the second round, three selected genes (*Ma1*, *Ma2* and *MDP0000141005*) were analyzed in detail with all 18 apple accessions. The gene-specific primers (Table S2) for *Ma1* are Ma1F (5′-CGTCATGGTGTCTGGAA CAT-3′) and Ma1R (5′-CTCCATGGCAAAAACCTGTC-3′), and those for *Ma2* are Ma2F (5′-TCGGAAGACGGCC TAATGGA-3′) and Ma2R (5′-TTGAAGCCGGGCAACAA ACT-3′). These gene-specific primers were designed to cover the known alleles of *Ma1* and *Ma2* (Figs. S2, S3).

qRT-PCR was conducted using Roche (Indianapolis, IN) LightCycler 480 Real-Time PCR System. For each qRT-PCR, a final volume of 20 μl was used, which contained 5 μl of the cDNA dilutions, 0.5 μM of the forward and reverse primers, and 1× SYBR Green Master Mix (Roche Cat. # 04707516001). The qRT-PCR program included an initial denaturation step of 10 min at 94 °C, 45 cycles of amplification using 10 s at 94 °C, 30 s at 58 °C, and 25 s at 72 °C, and a dissociation stage of 5 s at 95 °C, 60 s at 60 °C, and 15 s at 97 °C. Expression quantification and data analysis were performed by LightCycler 480 Software (Version 1.5) using the comparative cycle threshold method (Pfaffl 2001). Regression analysis between the gene expression and fruit acidity variation was performed using MS Excel 2007.

Phylogenetic analysis

Phylogenetic analysis of the deduced protein sequences of the *Ma* candidate genes, *Ma1* and *Ma2*, was conducted along with the members of the ALMT1 family in *Arabidopsis*, which sequences were downloaded from TAIR 10 (http://www.arabidopsis.org/index.jsp), using MEGA4 (Tamura et al. 2007).

Results

Delimiting the *Ma* locus to a 65 kb genomic segment on chromosome 16

Segregation of fruit pH in populations GMAL 4590 and GMAL 4595 had been studied previously, and the three parents Royal Gala, PI 613971 and PI 613988 had been determined of heterozygous genotype *Mama* (Xu et al. 2011). With additional fruiting individuals included, i.e. 36 in GMAL 4590 and 23 in GMAL 4595, the low pH ($\leq$3.8) and high pH ($\geq$3.9) segregation remained unchanged with

the expected ratio of 3:1 (Table 1). In population GMAL 4596, fruit pH segregated similarly with the ratio 3:1 (156:42, $P = 0.22$), suggesting PI 613979, the pollen parent of GMAL 4596, is of a heterozygous genotype *Mama* as well. However, 82 low and 51 high pH genotypes were scored in population GMAL 4592, indicating a significant deviation from the 3:1 ratio ($P = 0.0004$). Examining the markers linked to *Ma* (Fig. 1) revealed that none of them segregated for the pollen parent PI 613978 while all segregated normally for Royal Gala. Moreover, the markers that segregate for Royal Gala alone predicted the segregation of pH (data not shown), suggesting that PI 613978 has a genotype of *mama*. Given the known genotype *Mama* of Royal Gala, pH is expected to segregate 1:1 in population GMAL 4592. But the observed ratio of 82:51 distorted significantly from 1:1 ($P = 0.007$) (Table 1).

Three new markers, including two SSRs 12514.266 and 12995.82-2, and one SNP CN889255SNP, were developed (Fig. 1a–d; Table 2) between the existing two markers 532.669-2 and 20159.150-1 that defined the *Ma* region previously (Xu et al. 2011). For map integration, the three new markers were assessed with a total of 52 informative recombinants between markers CH05c06 and CH02a03 or CH05a09, including 17 mapped in GMAL 4590, 7 in GMAL 4592, 19 in GMAL 4595, and 9 in GMAL 4596 (Fig. 1a–d). Out of the 52 informative recombinants, 14 were the most informative in ordering the markers (Fig. 2). SSR marker 12995.82-2 along with the existing marker 18695-28-2 cosegregated with *Ma*, and markers CN889255SNP and 12514.266 flanked *Ma* immediately to narrow the *Ma* locus down to a smaller genetic interval on chromosome 16 (Figs. 1a–d, 2). This genetic interval of *Ma* was supported with four most informative recombinants, including GMAL 4595-6-149 and GMAL 4590-1-131 between marker CN889255SNP and *Ma*, and GMAL 4592-4-33 and GMAL 4595-6-121 between *Ma* and marker 12514.266 (Fig. 2). In physical terms, the *Ma* interval corresponds to a genomic segment of 65 kb on chromosome 16 in Golden Delicious (Fig. 1e), which was reduced from a 150 kb region defined previously (Xu et al. 2011).

Haplotypes of the *Ma* locus

The draft sequence of the Golden Delicious genome does not provide clear haplotype information although *M.* × *domestica* is a highly heterozygous species. To understand the possible sequence variation and local genomic structure and organization that may discriminate allele *Ma* from *ma*, we identified two BAC clones BAC3 and BAC21 from the BAC library of apple rootstock G.41 using three markers 18695.28-2, 12995.82-2 and 12514.266 simultaneously. The two BAC clones were confirmed to contain not only the three makers used to screen the BAC library, but also the PCR amplicon source for marker CN889255SNP (Fig. S1a), suggesting both BAC clones cover the *Ma* locus completely. Based on the band patterns associated with markers 18695.28-2, 12995.82-2 and 12514.266 (Fig. S1b) and the restricted bands generated by BamHI and NotI digestions (Fig. S1c), the two BACs are clearly of different haploid origin although the genotype of G.41 could be either *Mama* or *MaMa*. The estimated sizes for BAC3 and BAC21 were 150–160 kb and 110–120 kb, respectively (Fig. S1c).

Sequencing of the two BAC clones revealed that the *Ma* region extends over a larger segment of 71 kb in BAC21 and 82 kb in BAC3 in G.41 (Fig. 1g, h). Sequence alignment using BLAST demonstrated that BAC21 had higher overall sequence identity with the Golden Delicious contigs than BAC3 (data not shown), suggesting that BAC21 represents a haplotype likely closer to the two haplotypes in Golden Delicious than BAC3.

Identification of *Ma1* and *Ma2*

The 65 kb genomic region of *Ma* contains 19 predicted genes in Golden Delicious (#10–28 in the 44 genes listed in Xu et al. (2011), Fig. 1f, i; Table S3). Aligning the 19 predicted genes with the two BACs indicated that three genes (*MDP0000375685* (#19), *MDP0000258718* (#23) and *MDP0000357895* (#26)) were not found in the two BAC sequences, and two (*MDP0000250967* (#24) and *MDP0000157412* (#27) reside outside of the *Ma* region defined by the two markers CN889255SNP and 12514.266 in G.41. Moreover, *MDP0000134560* (#22) and *MDP0000139500* (#28) are duplicated gene IDs for a single gene, and *MDP0000241811* (#17) and *MDP0000141005* (#18) are alternatively spliced variants from another single gene. Therefore, the two BAC clones harbor 12 predicted genes at the *Ma* locus (Fig. 1g–i; Tables S2, 3), which include *MDP0000252114* (#16), designated *Ma1*, and *MDP0000244249* (#11), designated *Ma2*. Proteins Ma1 and Ma2 are putative members of the ALMT1 family and, respectively, share 57 % (338/595) and 55 % (302/553) of identity in amino acid sequence with AtALMT9, an *Arabidopsis* protein known to be a vacuolar malate channel involved in maintaining the cytosolic malate homeostasis (Kovermann et al. 2007). A search of *Malus* EST databases in GenBank found that there are 20 EST accessions of the origin of *Ma1* (Table S4) and one EST (CN929391) matching with *Ma2*, suggesting both *Ma1* and *Ma2* are expressed genes, and therefore strong candidate genes of *Ma*.

There are two inversions in gene orders between Golden Delicious and G.41 (Fig. 1f, g): one between genes *MDP0000130613* (#10) and *MDP0000244253* (#15), and the other between genes *MDP0000134560* (#22) and

Table 2 Primer sequences and other relevant information of markers developed in the *Ma* region

Marker name	Marker type	Primer-F/R (from 5′ to 3′)	IDs of source sequences	Targeted SSRs or bases	Expected size (bp)	Allele size (bp) or base for *Ma*[a]	Allele size (bp) or base for *md*[b]	Detection	Genome mapped
CN889255SNP	SNP	GGAGGGTCTCCATCCAATTTA/ TCCCCACACATCTCATATTCC	CN889255/ MDC022113.108	Base 330[c]	496(cDNA)/ 792(gDNA)	C	T	Sequencing	P2, P4
12995.82-2	SSR	AAAGCTTCTTCACACCAAGCA/ TGGTGATGATGGTGGTAGTCA	MDC012995.82	$(TC)_9$	173	164 (P1), 162 (P4, P5)	168 (P1–P5)	PAGE	P1,P4, P5
12514.266	SSR	AGGTATTGCCTAAATGTGTGTG/ TCACATCATAATGTTTTTCCGAAT	MDC012514.266	$(TC)_{13}$	192	282 (P1)	290 (P1)	PAGE	P1
$CAPS_{1455}$	CAPS	GCCGCTTCTGGACTATCACTA/ TTCTTCAACCGCAAACTCCT	*Ma1*	TGG_{1455}/ TGA_{1455}	2,013	2,013 (P1, P2, P4, P5)	1,764 + 249 (P1–P5)	BspHI digestion + agarose gel electrophoresis	P1, P2, P4, P5

[a] Allele linked to *Ma* in coupling phase

[b] Allele linked to *ma* in coupling phase

[c] The 330th base from the first base of the fwd primer. The preceding sequence is AGGAATGGATTTTGGCTTCTAGGCTTGCAGCTTCTGATCAATTGGGTCT(T/C)$_{330}$. Genotype of Royal Gala is "T_{330}/T_{330}", and that of both PI 613971 and PI 613988 is "T_{330}/C_{330}"

P1 Royal Gala, *P2* PI 613971, *P3* PI 613978, *P4* PI 613988, *P5* PI 613979, *PAGE* polyacrylamide gel electrophoresis

MDP0000247199 (#25) (Fig. 1f, g). Genes *Ma1* and *Ma2* are physically separated by one gene *MDP0000130613* (#10) in both haplotypes of G.41 (Fig. 1g, h), but by four genes *MDP0000244250* (#12), *MDP0000244251* (#13), *MDP0000130619* (#14) and *MDP0000244253* (#15) in Golden Delicious (Fig. 1f).

qRT-PCR analysis of genes predicted at the *Ma* locus

To investigate the expression patterns in mature fruit, genes *Ma1* and *Ma2* as well as the other ten genes in the *Ma* region were screened alongside four low acid (Britegold, KAZ 96 08-17, Novosibirski Sweet and Sweet Delicious) and four high acid (Cox's Orange Pippin, Golden Delicious, Marshall McIntosh and Winter Majetin) apple germplasm accessions (Table S1) using qRT-PCR. Gene *Ma1* was expressed at much higher levels in high acid fruit than in low acid fruit while *Ma2* was expressed consistently at low levels across both low and high acid fruit (data not shown). The correlation between gene expression and TA among the eight apple accessions was highly significant for *Ma1* ($R^2 = 0.9430$, $P = 0.0001$) but non-significant for *Ma2* ($R^2 = 0.0559$, $P = 0.5729$). Among the other ten genes, *MDP0000141005*, which encodes a putative serine/threonine protein phosphatase 2A (PP2A) regulatory subunit A, was expressed at high levels and showed a significant correlation with fruit acidity ($R^2 = 0.7428$, $P = 0.0059$). The remaining nine genes were expressed at low levels and did not show correlations with fruit acidity (data not shown), and therefore we did not analyze them further.

A more comprehensive qRT-PCR analysis of *Ma1*, *Ma2* and *MDP0000141005* indicated that the relative expression levels of *Ma1* remained high and were significantly correlated with TA ($R^2 = 0.4543$, $P = 0.0021$) and pH ($R^2 = 0.4630$, $P = 0.0019$) in fruit of 18 apple germplasm accessions (Fig. 3a, b). In contrast, the expression of *Ma2* was low and showed no correlation with TA ($R^2 = 0.0086$, $P = 0.7148$) and pH ($R^2 = 0.0356$, $P = 0.4531$) (Fig. 3c, d). These data suggest that *Ma1* may be the major factor in determining fruit acidity and the role of *Ma2* would be limited if any. The correlation of *MDP0000141005* expression with acidity was reduced to a non-significant level ($R^2 = 0.1497$, $P = 0.1126$ for TA; $R^2 = 0.0916$, $P = 0.2222$ for pH), allowing *MDP0000141005* to be excluded from subsequent analyses.

Allelic variations of *Ma1* and *Ma2*

The *Ma1* allele in BAC3, designated *Ma1-G41*, differed by eight bases from that in BAC21, designated *ma1-G41* (Fig. S2, Table 3). Examining the coding sequence of

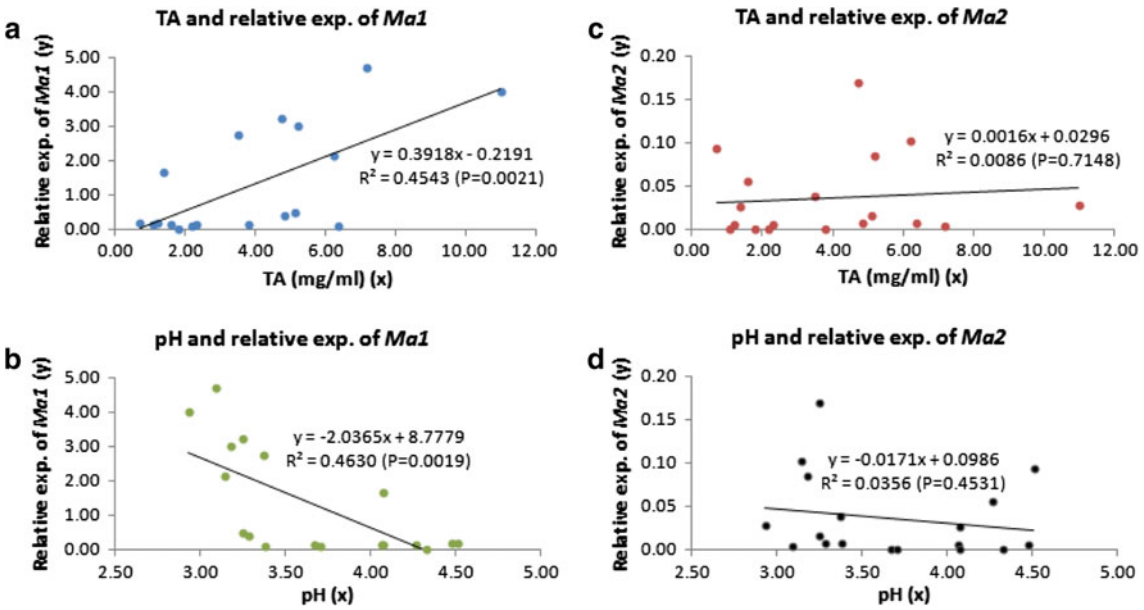

Fig. 3 Regression between fruit acidity (TA and pH) and relative gene expression (*Ma1* and *Ma2*) in 18 apple germplasm accessions. **a–d** Self explainable

Table 3 DNA and amino acid sequence variations in the *Ma1* alleles of G.41 and Golden Delicious (GD)

Base position[a]	DNA					AA				
	Ma1-G41[b]	*ma1-G41*[b]	*Ma1-GD*[c]	*ma1-GD*[c]	*MDP0000252114*[c]	*Ma1-G41*	*ma1-G41*	*Ma1-GD*	*ma1-GD*	*MDP0000252114*
108	A	C	C	C	C	A	A	A	A	A
118	C	C	A	C	M = A, C	H	H	N	H	H, N
162	C	C	A	C	M = A, C	N	N	K	N	N, K
814	G	A	A	A	A	V	I	I	I	I
834	G	G	T	G	K = G, T	T	T	T	T	T
1011	A	T	A	T	W = A, T	A	A	A	A	A
1032	T	C	C	C	C	H	H	H	H	H
1286	A	G	A	G	R = A, G	K	R	K	R	K, R
1304	T	T	G	T	K = G, T	V	V	G	V	V, G
1394	T	C	C	C	C	V	A	A	A	A
1455	G	A	G	A	R = A, G	W	STOP	W	STOP	W/STOP
1645	G	G	C	C	C	A	A	P	P	P
1688	C	G	G	G	G	T	S	S	S	S

[a] Counted from the 1st base in the coding sequences

[b] There are eight base variations in the coding sequence between *Ma1-G41* and *ma1-G41*. Out of the eight base variations, three are silent mutations and five are pronounced, including the one at the 1455th base that led to a stop codon in ma1-G41 for premature termination. As a result, another pronounced mutation at base 1688 was beyond the coding sequence of *ma1-G41*

[c] There are seven ambiguous nucleotides in the coding sequence of MDP0000252114. The 1455th base is R, an ambiguous base for A or G, suggesting a similar premature stop codon in *ma1-GD*

MDP0000252114 showed that nucleotides at seven positions are ambiguous, i.e. M = A/C (bases 118 and 162); K = G/T (bases 834 and 1304); W = A/T (base 1011) and R = A/G (bases 1286 and 1455), presumably caused by the two different haplotypes in Golden Delicious (Fig. S2, Table 3). To distinguish the two alleles of *Ma1* in Golden Delicious, we compared the sequence of MDP0000252114 with both *Ma1-G41* and *ma1-G41*. Excluding the seven ambiguous positions, *MDP0000252114* differed by one base from *ma1-G41*, but by six bases from *Ma1-G41*,

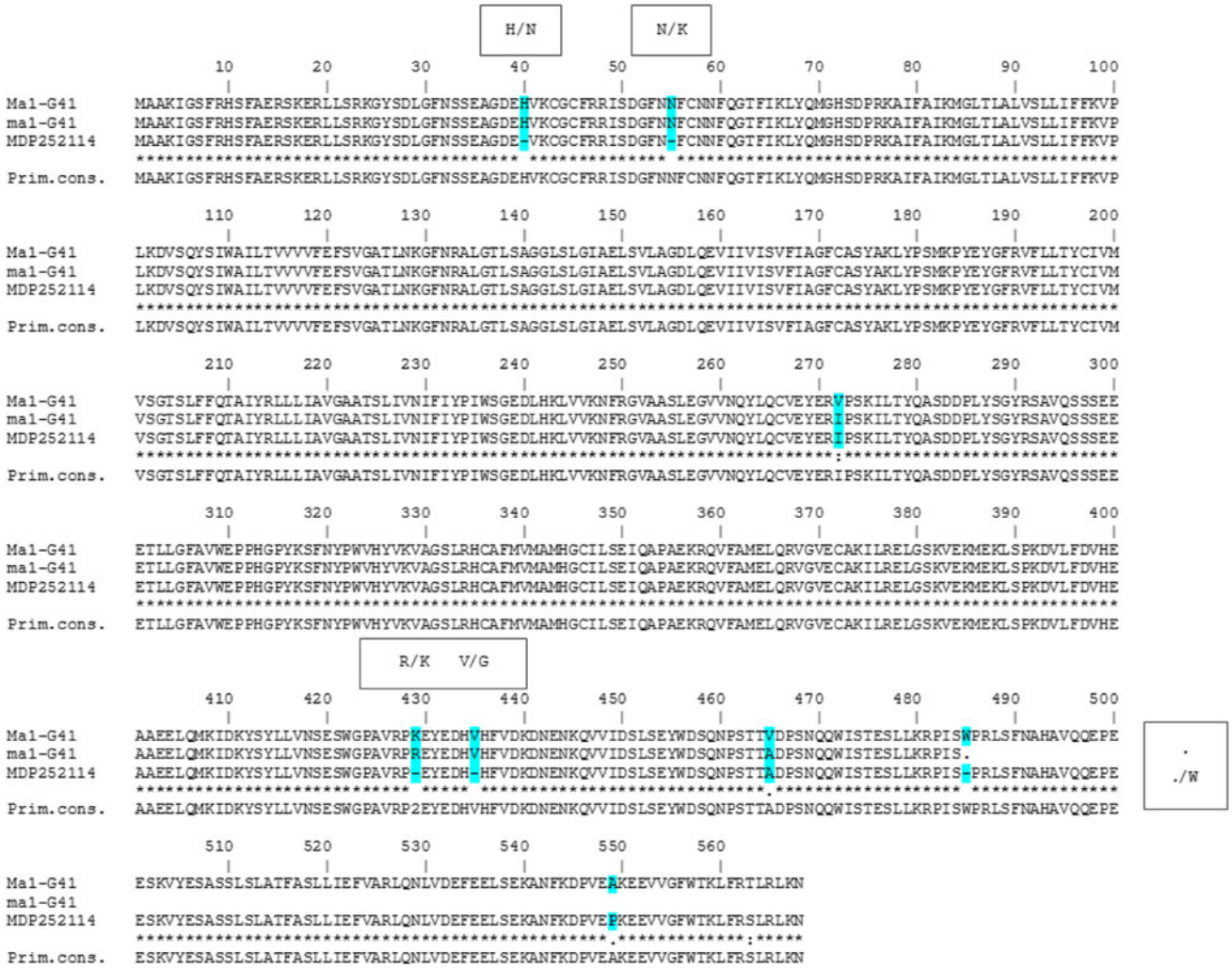

Fig. 4 Alignment of the Ma1 deduced protein sequences. MDP252114 stands for the Golden Delicious protein MDP0000252114, which combines proteins Ma1-GD and ma1-GD. Each sign "–" in MDP252114 is for two possible amino acid residues, and annotated accordingly as shown. Amino acid residues that vary are highlighted in blue. The stop codon TGA_{1455} caused by SNP_{1455A} leads to a truncation of 84 amino acids at the carboxyl terminus in proteins ma1-G41 and ma1-GD compared with proteins Ma1-G41 and Ma1-GD

suggesting *MDP0000252114* is much closer to *ma1-G41* than to *Ma1-G41*. When the seven ambiguous positions were considered, one of the two possible bases at each of the seven positions matches with the base at their corresponding positions in *ma1-G41*. This set of seven bases was therefore inferred to be co-present in one allele of Golden Delicious, designated *ma1-GD*. The other set of seven bases was concluded to be co-present in the other allele of Golden Delicious, designated *Ma1-GD* (Fig. S2; Table 3).

In the deduced amino acid sequences, Ma1-G41 and Ma1-GD diverge by seven residues while there is no difference between ma1-G41 and ma1-GD (Fig. 4; Table 3). However, both ma1-G41 and ma1-GD are truncated by 84 amino acids at the carboxyl terminus compared with either Ma1-G41 or Ma1-GD (Fig. 4). This truncation is due to a nucleotide mutation from G to A at the 1455th base (SNP_{1455}) in the open reading frame, leading to a pronounced change from a tryptophan (W) codon TGG_{1455} to a stop codon TGA_{1455} (Fig. 4; S2; Table 3).

The allelic variations of gene *Ma2* were investigated similarly (Figs. S3–4; Table S5). Briefly, the *Ma2* allele in BAC3 and that in BAC21 were designated *Ma2-G41* and *ma2-G41*, respectively, whereas *Ma2-GD* and *ma2-GD* were assigned as two alleles for Golden Delicious based on the *MDP0000244249* sequence of four ambiguous positions, i.e. R = A/G (base 26), W = A/T (bases 165 and 951) and M = A/C (base 1245). There are 24 different bases (17 aa) between *Ma2-G41* and *ma2-G41*, 4 bases (2 aa) between *Ma2-G41* and *Ma2-GD*, and 22 bases (17 aa) between *Ma2-G41* and *ma2-GD*. The coding sequences in alleles *ma2-G41* and *ma2-GD* are identical (Figs. S3, 4; Table S5).

Allelic association of the *Ma1* and *Ma2* alleles with *Ma* and *ma*

To uncover which *Ma1* and *Ma2* allele is associated with *Ma* or *ma*, a CAPS marker, named $CAPS_{1455}$, was developed to target SNP_{1455} using endonuclease BspHI, which cleaves site $TCATGA_{1455}$ in the truncated alleles *ma1-G41* or *ma1-GD*, but not $TCATGG_{1455}$ in the intact alleles *Ma1-G41* or *Ma1-GD* (Table 2). Agarose gel assay of maker $CAPS_{1455}$ in population GMAL 4595 and the informative recombinants indicated that homozygous genotype $CAPS_{1455G}CAPS_{1455G}$ cosegregated with *MaMa*, $CAPS_{1455G}CAPS_{1455A}$ with *Mama*, and $CAPS_{1455A}CAPS_{1455A}$ with *mama* (Figs. 1, 2, 5a), suggesting the intact allele of *Ma1* (*Ma1-1455G*), such as *Ma1-G41* or *Ma1-GD*, is associated with the high acid allele *Ma* while the truncated allele of *Ma1* (*ma1-1455A*), such as *ma1-G41* or *ma1-GD*, with the low acid allele *ma*. Consequently, alleles *Ma2-G41* and *Ma2-G41* are associated with *Ma* while *ma2-G41* and *ma2-GD* with *ma*.

Together with the analyses in the haplotypes at the *Ma* locus and allelic variations in the two genes *Ma1* and *Ma2* in G.41 and Golden Delicious, the allelic associations identified here conclude that BAC3 stands for a haplotype of *Ma* for high acidity while BAC21 represents a haplotype of *ma* for low acidity, and that the allele diversity is higher for the high acidity alleles, but none or low for the low acidity alleles.

Association of the mutation-led truncation in Ma1 with low fruit acidity in apple germplasm

To see how SNP_{1455} may explain the acidity levels in the other apple germplasm, a set of 29 (Table S1) representative apple germplasm accessions were analyzed with marker $CAPS_{1455}$ (Figs. 5b, 6). Genotype $CAPS_{1455A}$-$CAPS_{1455A}$ is associated either exclusively with high pH (7/7, Fig. 6a) or tightly with low TA (7/9, Fig. 6b). Genotypes $CAPS_{1455G}CAPS_{1455G}$ and $CAPS_{1455G}CAPS_{1455A}$, however, are associated either completely with low pH (22/22, Fig. 6a) or highly with high TA (20/22, Fig. 6b). These data indicate a complete or highly tight association between the mutation-led truncation in *Ma1* (*ma1-1455A*) with low acidity in these apple accessions.

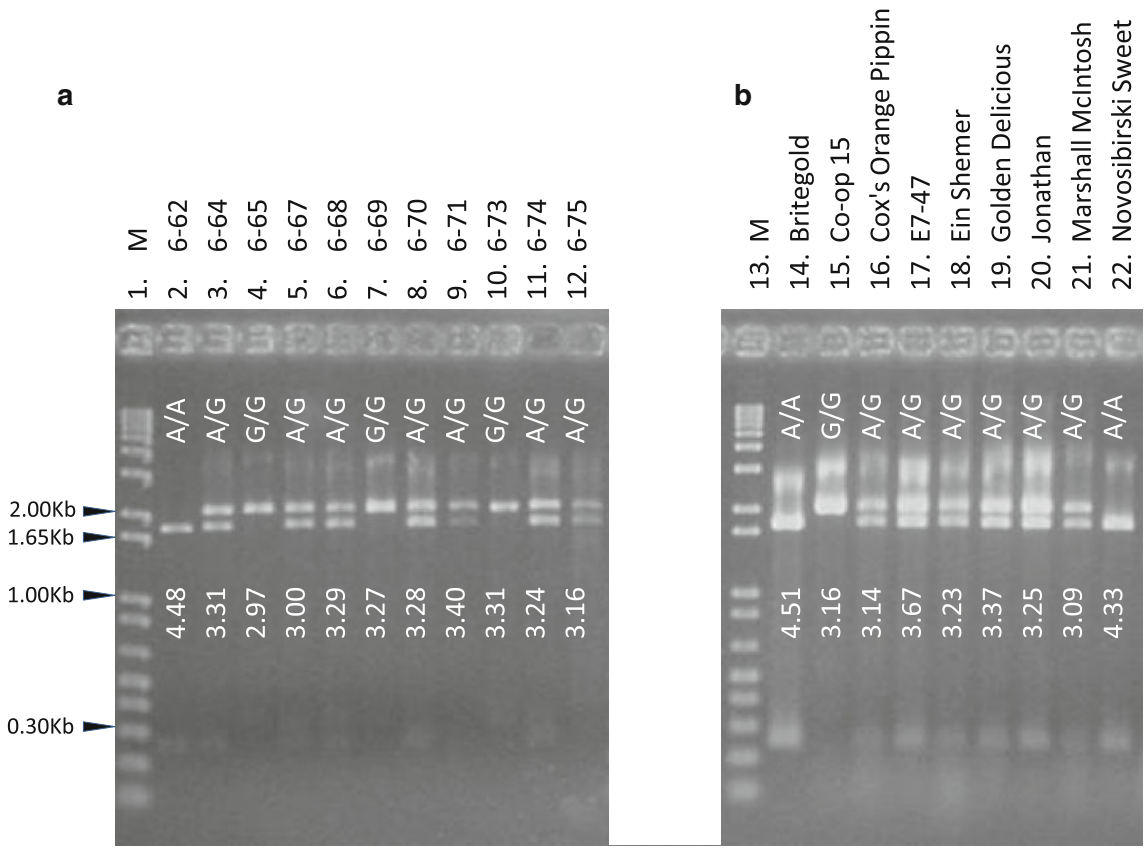

Fig. 5 Agarose gel analysis of marker $CAPS_{1455}$. Bands of 2,013 bp correspond to allele $CAPS_{1455G}$, i.e. the *Ma1-1455G* allele for high acidity. The combined bands of 1,764 and 249 bp are expected for allele $CAPS_{1455A}$, i.e. the *ma1-1455A* allele for low acidity. A/A = genotype $CAPS_{1455A}$ $CAPS_{1455A}$, A/G = genotype $CAPS_{1455A}$ $CAPS_{1455G}$, and G/G = genotype $CAPS_{1455G}$ $CAPS_{1455G}$. The *numbers* indicate fruit pH. **a** *Lane 1*: 1 kb Plus DNA Ladder (Invitrogen, CA). *Lanes 2–12*: 11 progeny of GMAL 4595. **b** *Lane 13*: 1 kb Plus DNA Ladder. *Lanes 14–22*: nine apple germplasm accessions as shown

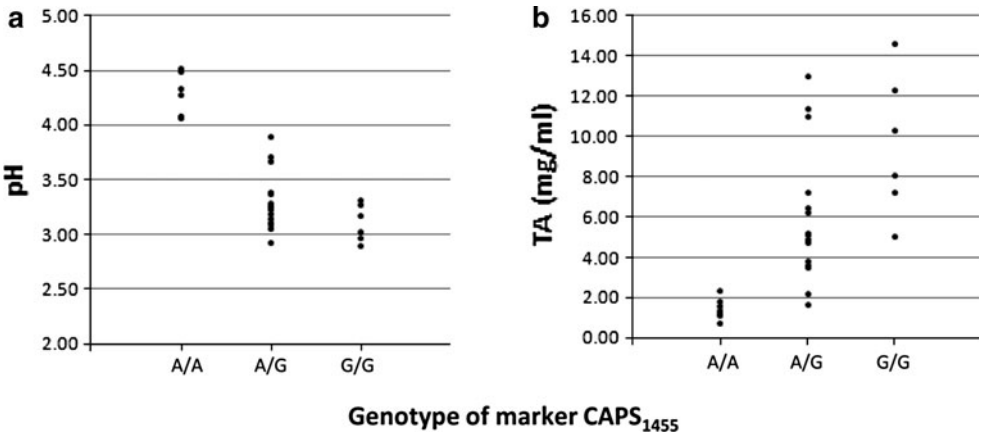

Fig. 6 Survey of marker $CAPS_{1455}$ genotypes and their association with fruit pH (**a**) and TA (**b**) in 29 apple germplasm accessions (Table S1). A/A = genotype $CAPS_{1455A}$ $CAPS_{1455A}$, A/G = genotype $CAPS_{1455A}$ $CAPS_{1455G}$, and G/G = genotype $CAPS_{1455G}$ $CAPS_{1455G}$

Discussion

Delimiting the *Ma* locus to a 65 kb genomic segment and identification of two ALMT-like genes *Ma1* and *Ma2*

By developing three new markers and analyzing two additional populations, we delimited the *Ma* locus between markers CN889255SNP and 12514.266. The genetic interval was supported by four recombinants with GMAL 4595-6-149 and GMAL 4590-1-131 between marker CN889255SNP and *Ma*, and GMAL 4592-4-33 and GMAL 4595-6-121 between *Ma* and marker 12514.266 (Figs. 1a, 2) among the 52 informative recombinants identified. The *Ma* locus between markers CN889255SNP and 12514.266 corresponds to a homologous genomic segment of 65 kb in Golden Delicious, enabling us to reduce the number of candidate genes of *Ma* from 44 identified previously (Xu et al. 2011) to 19 in the present study.

Since the draft sequence of the apple genome does not provide clear haplotype specific information (Velasco et al. 2010), we identified two BAC clones of different haploid origin from apple rootstock G.41, BAC3 and BAC21, which completely cover the *Ma* locus. Sequencing the two BAC clones revealed that the *Ma* locus spanned 71 kb in BAC3 and 82 kb in BAC21. A more detailed analysis showed that out of 19 predicted genes in Golden Delicious, three were not present in the two BACs and two were beyond the *Ma* interval. In the remaining 14 predicted genes, two were duplicated, leading to 12 predicted genes for *Ma* in both BACs, including *Ma1* and *Ma2* (Fig. 1g–i; Table S3). Although the draft sequence of the apple genome is of high quality (Velasco et al. 2010), the local general structure of the *Ma* locus revealed by the two sequenced BACs from G.41 may be more representative. Given the limited number of genes in the *Ma* locus and the putative functions of ALMT genes in maintaining the malate homeostasis in plant cells, e.g. *AtALMT9* (Kovermann et al. 2007) and *AtALMT6* (Meyer et al. 2011), *Ma1* and *Ma2* are considered to be strong candidate genes of *Ma*.

Putative function of Ma1 and Ma2 as vacuolar malate channels/transporters in apple fruit

The first member of the ALMT1 family unique to plants is TaALMT1 that confers wheat tolerance to soil aluminum toxicity (Sasaki et al. 2004). TaALMT1 protein facilitates malate efflux from root apices and is localized on the plasma membrane (Yamaguchi et al. 2005). The counterpart of *TaALMT1* that shows similar aluminum tolerance function includes *AtALMT1* in *Arabidopsis* (Hoekenga et al. 2006), *ScALMT1-M39.1* and *ScALMT1-1135.1* (a hybrid gene) in rye (Collins et al. 2008), and *BnALMT1* and *BnALMT2* in rape (Ligaba et al. 2006). The *Arabidopsis* genome encodes 14 *ALMT1* genes, which are distributed in four of the five clades in the ALMT1 family (Barbier-Brygoo et al. 2011). Phylogenetic analysis of the deduced protein sequences of Ma1 and Ma2 together with the 14 Arabidopsis ALMT1 proteins showed that the two apple proteins belong to clade 2 that includes five members AtALMT3–6, 9 (Fig. 7).

AtALMT9 is a vacuolar membrane protein functioning as a vacuolar malate channel for maintaining cell malate homeostasis (Kovermann et al. 2007), differing from AtALMT1 (Yamaguchi et al. 2005) and AtALMT12 (Meyer et al. 2010), which are plasma membrane proteins. AtALMT9 is expressed in all organs, but its expression in leaves is specifically in mesophyll cells. AtALMT6, another member in clade 2 that has been characterized recently, is expressed in guard cells of leaves as well as in flower organs and stems, but not in roots (Meyer et al. 2011). The AtALMT6 protein is also targeted to the vacuolar membrane, and it functions as a malate influx or efflux channel that is highly regulated by vacuolar pH and cytosolic malate (Meyer et al. 2011). It has been shown that low malate content in low acid fruit is the result of a

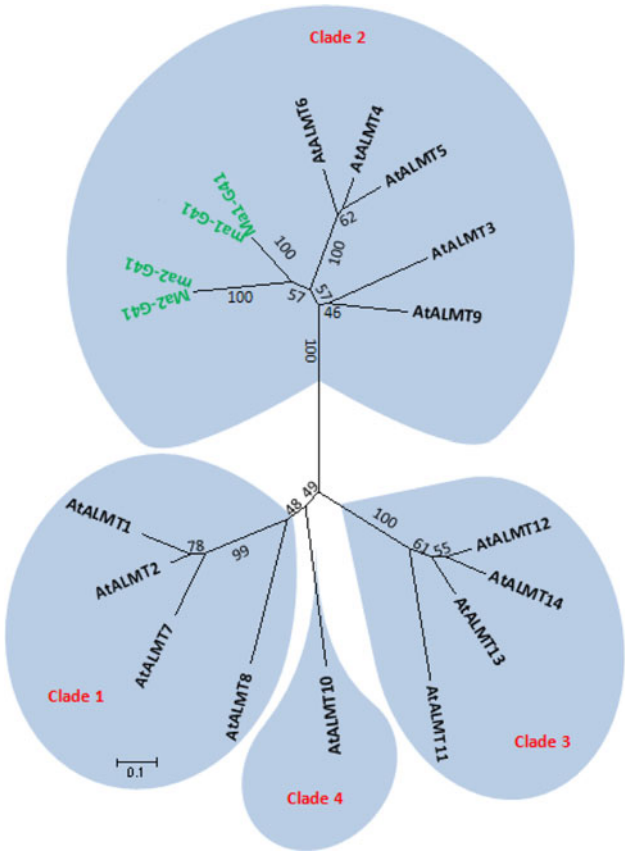

Fig. 7 Phylogenetic analysis of Ma1 and Ma2 proteins. The 14 members AtALMT1-14 of the AtALMT1 family were retrieved from TAIR 10 (http://www.arabidopsis.org/). The protein sequences were aligned with ClustalW and the trees were constructed with the MEGA4 program (Tamura et al. 2007) using the neighbor joining method. To test the phylogeny, "bootstrap samples of 1000" was set during the analysis. The tree is drawn to scale and the evolutionary distances are in the units of the number of amino acid substitutions per site. Naming system of the clades as described previously (Barbier-Brygoo et al. 2011) is adapted here

restricted ability to accumulate malate in apple parenchyma cells (Beruter 2004). As members of clade 2, Ma1 and Ma2, especially Ma1, are likely vacuolar malate channels/transporters with primary function in maintaining malate homeostasis by regulating the malate levels in vacuole and cytosol in the parenchyma cells of apple fruit, thereby controlling fruit acidity levels.

Haplotypes of *Ma* and allelic association of the *Ma1* and *Ma2* alleles with *Ma* and *ma*

Sequencing of the two BAC clones from apple rootstock G.41 provided the first view of the *Ma* locus at the DNA sequence level with distinction between haplotypes *Ma* and *ma*. The difference between the *Ma* (BAC3) and *ma* (BAC21) haplotypes is significant in both size (82 vs. 71 kb) and the coding sequences of predicted genes (Fig. 1g–h). In the *Ma1* and *Ma2* sequences, the alleles (*Ma1-G41* and *ma1-G41*, and *Ma2-G41* and *ma2-G41*) are clearly distinguishable. This made it possible to infer their allelic counterparts (*Ma1-GD* and *ma1-GD*, and *Ma2-GD* and *ma2-GD*) in genes *MDP0000252114* and *MDP0000244249* of Golden Delicious, respectively. Comparison of the allelic sequences of *Ma1* and *Ma2* revealed that there are no variations in the deduced amino acid sequences in alleles (*ma1-G41* and *ma1-GD*, and *ma2-G41* and *ma2-GD*) associated with *ma* for low acidity, whereas the variations are considerable for alleles (*Ma1-G41* and *Ma1-GD*, and *Ma2-G41* and *Ma2-GD*) associated with *Ma* for high acidity. A similar trend exists in the entire *Ma* region when the sequences at the *Ma* locus between Golden Delicious and G.41 were compared as BAC21 is much closer to Golden Delicious than BAC3. One possible explanation is that the natural or human selection of fruit acidity has mostly acted upon the high acid allele *Ma* rather than the low acidity allele *ma* due to its recessive nature, leading to a greater diversity in high acidity allele *Ma*. Whether or not the high diversity among the *Ma* alleles plays a role in large fruit acidity variations in different apple cultivars would be of great interest for future investigation.

One of the most important findings of this work is the discovery of the mutation at the 1455th base of *Ma1*, which turns the tryptophan (W) codon TGG_{1455} in *Ma1-G41* into a stop codon TGA_{1455} in *ma1-G41*, leading to a premature termination and truncation of ma1-G41 by 84 deduced amino acids at the C terminus. The presence of the mutation in Golden Delicious is confirmed with the ambiguous base R_{1455}, which stands for G_{1455}/A_{1455} in *Ma1* (Fig. S2). In view of the dramatic implication of this mutation and the critical role of the C-terminus in regulating the function and activity of TaALMT1 in wheat (Furuichi et al. 2010; Ligaba et al. 2009), marker $CAPS_{1455}$ was developed to target SNP_{1455}. Analysis using marker $CAPS_{1455}$ showed that it segregates in a codominant fashion and accurately predicts genotypes *MaMa*, *Mama* and *mama* in population GMAL 4595 and the informative recombinants (Figs. 1, 2, 5a). Moreover, the marker shows a perfect association with pH and a highly tight association with TA in 29 apple accessions studied (Figs. 5b, 6). Overall, these data strongly suggested that SNP_{1455} is critical in determining the function of the *Ma1* alleles.

It should be pointed out that the plant materials used in this study are restricted in *M. sieversii*, *M. domestica* and some of its hybrids. Since there are at least 23 species in *Malus* (Robinson et al. 2001), understanding the role of the *Ma* locus and SNP_{1455} in the remaining species would be an interesting extension of this work.

Expression of *Ma1* and *Ma2*

Compared with *Ma2* expression in mature fruit, *Ma1* expression is much higher (Fig. 3). This trend appeared to

be consistent with the number of ESTs identified for the two genes in the *Malus* EST database of 336,017 accessions in GenBank. There are 20 EST accessions for *Ma1* (Table S4) and one for *Ma2*, i.e. CN929391 derived from pre-opened floral bud of Royal Gala. The tissue source for the 20 *Ma1* ESTs includes fruit (9 accessions), flower (3), leaf (3), root (3), stem xylem (1) and bud (1) from nine apple varieties, such as Royal Gala (6), GoldRush (5), Granny Smith (2), M.9 (1, rootstock) and others. Therefore, in addition to higher expression levels, *Ma1* is also evidenced to be expressed in a wider range of organs than *Ma2*, suggesting a broader role of *Ma1* in apple.

Significant correlations between gene expression and fruit acidity were observed for gene *Ma1* but not for *Ma2* (Fig. 3). This suggests that *Ma1* is the major factor in determining fruit acidity levels. Since alleles *Ma1-1455G* and *ma1-1455A* are associated with *Ma* and *ma*, respectively, the strong positive correlation between *Ma1* expression and fruit acidity would suggest that transcripts of *Ma1-1455G* be more readily detected than those of *ma1-1455A*. Examining the presence of SNP_{1455} in the 20 ESTs of *Ma1* supported this reasoning. SNP_{1455G} appeared in all seven ESTs (CO723101, CX024250, CN494439, GO547092, GO509271, GO562003 and HM641023) that span over base 1455 in *Ma1* while SNP_{1455A} was not detected (Table S4). It appears, therefore, that both SNP_{1455} and expression levels of *Ma1* are important in apple fruit acidity. To elucidate the role of *Ma2*, more dedicated studies are needed.

Gene *MDP0000141005* encodes a putative serine/threonine protein phosphatase 2A (PP2A) regulatory subunit A and its expression was initially found to be correlated with fruit acidity. *MDP0000141005* was excluded in allelic variation analysis since the correlation became non-significant when 18 apple accessions were analyzed. We examined the coding sequences of *MDP0000141005* in BAC3 and BAC21 of G.41, which did not confer variations in the amino acid sequences. Although PP2A is involved in many plant processes (Ahn et al. 2011; Leivar et al. 2011; Skottke et al. 2011), its subunit genes, including regulatory subunit A, have been used as reference genes for qRT-PCR analysis in plants (Czechowski et al. 2005; Navascues et al. 2012; Obrero et al. 2011). The constitutive expression of the PP2A regulatory subunit A gene and the inconsistent correlation between the *MDP0000141005* expression and fruit acidity make it unlikely the gene responsible for fruit acidity variation.

In conclusion, we discovered two ALMT-like genes, *Ma1* and *Ma2*, at the *Ma* locus of 65–82 kb containing 12–19 predicted genes that controls fruit acidity levels in apple. Expressions of *Ma1* and *Ma2* contrast sharply in the 18 apple germplasm accessions studied. *Ma1* was expressed at much higher levels than *Ma2* in mature fruit, especially in those of high acidity. Moreover, the *Ma1* expression is significantly correlated with fruit acidity, whereas the *Ma2* expression remains at low levels regardless of fruit acidity variations. These data suggest that *Ma1* is the major determinant at the *Ma* locus controlling fruit acidity. Sequencing of clones BAC3 and BAC21 that cover the two distinct haplotypes at the *Ma* locus allowed us to determine specific alleles of both *Ma1* and *Ma2* for high or low acid phenotype. A single nucleotide mutation at base 1455 in the open reading frame of *Ma1* led to a premature stop codon TGA_{1455}, which truncates the carboxyl terminus of Ma1 by 84 amino acids. A survey of 29 apple germplasm accessions using marker $CAPS_{1455}$ targeting SNP_{1455} found that the $CAPS_{1455A}$ allele is associated completely with high pH and tightly with low TA, suggesting that the natural mutation-led truncation is most likely responsible for the abolished function of *Ma* for low pH or high TA in apple.

Acknowledgments The authors are grateful for Mr. Phil Forseline and Dr. Herb Aldwinkle who developed the four mapping populations and made them available for this study.

References

Ahn CS, Han JA, Lee HS, Lee S, Pai HS (2011) The PP2A regulatory subunit Tap46, a component of the TOR signaling pathway, modulates growth and metabolism in plants. Plant Cell 23: 185–209

Barbier-Brygoo H, De Angeli A, Filleur S, Frachisse J-M, Gambale F, Thomine S, Wege S (2011) Anion channels/transporters in plants: from molecular bases to regulatory networks. Annu Rev Plant Biol 62:25–51

Beruter J (1998) Carbon partitioning in an apple mutant deficient in malic acid. Acta Hort 446:23–28

Beruter J (2004) Carbohydrate metabolism in two apple genotypes that differ in malate accumulation. J Plant Physiol 161:1011–1029

Blanpied GD, Silsby KJ (1992) Predicting harvest date windows for apples. Information Bulletin 221. Cornell Cooperative Extension, Cornell University, Ithaca

Boudehri K, Bendahmane A, Cardinet G, Troadec C, Moing A, Dirlewanger E (2009) Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. BMC Plant Biol 9:14

Brown AG, Harvey DM (1971) Nature and inheritance of sweetness and acidity in cultivated apple. Euphytica 20:68–80

Collins NC, Shirley NJ, Saeed M, Pallotta M, Gustafson JP (2008) An ALMT1 gene cluster controlling aluminum tolerance at the *Alt4* locus of rye (*Secale cereale* L.). Genetics 179:669–682

Cummins J, Aldwinckle H, Robinson T, Fazio G (2006) Apple tree rootstock named 'G.41'. In: Office TUSPaT (ed) The United States Patent and Trademark Office, Cornell Research Foundation Inc, USA. USPP17139

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol 139:5–17

Emmerlich V, Linka N, Reinhold T, Hurth MA, Traub M, Martinoia E, Neuhaus HE (2003) The plant homolog to the human sodium/dicarboxylic cotransporter is the vacuolar malate carrier. P Natl Acad Sci USA 100:11122–11126

Fang DQ, Federici CT, Roose ML (1997) Development of molecular markers linked to a gene controlling fruit acidity in citrus. Genome 40:841–849

Forsline PL, Aldwinckle HS, Dickson EE, Luby JJ, Hokanson SC (2003) Collection, maintenance, characterization and utilization of wild apples of Central Asia. Horticultural Rev 29:1–61

Fulton TM, Bucheli P, Voirol E, Lopez J, Petiard V, Tanksley SD (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica 127:163–177

Furuichi T, Sasaki T, Tsuchiya Y, Ryan PR, Delhaize E, Yamamoto Y (2010) An extracellular hydrophilic carboxy-terminal domain regulates the activity of TaALMT1, the aluminum-activated malate transport protein of wheat. Plant J 64:47–55

Hoekenga OA, Maron LG, Pineros MA, Cancado GMA, Shaff J, Kobayashi Y, Ryan PR, Dong B, Delhaize E, Sasaki T, Matsumoto H, Yamamoto Y, Koyama H, Kochian LV (2006) AtALMT1, which encodes a malate transporter, is identified as one of several genes critical for aluminum tolerance in *Arabidopsis*. P Natl Acad Sci USA 103:9738–9743

Hulme AC, Wooltorton LSC (1957) The organic acid metabolism of apple fruits: changes in individual acids during growth on the tree. J Sci Food Agr 8:117–122

Iwanami H, Moriya S, Kotoda N, Mimida N, Takahashi-Sumiyoshi S, Abe K (2012) Mode of inheritance in fruit acidity in apple analysed with a mixed model of a major gene and polygenes using large complex pedigree. Plant Breed 131:322–328

Jalikop SH (2007) Linked dominant alleles or inter-locus interaction results in a major shift in pomegranate fruit acidity of 'Ganesh' × 'Kabul Yellow'. Euphytica 158:201–207

Kenis K, Keulemans J, Davey M (2008) Identification and stability of QTLs for fruit quality traits in apple. Tree Genet Genomes 4:647–661

Kovermann P, Meyer S, Hortensteiner S, Picco C, Scholz-Starke J, Ravera S, Lee Y, Martinoia E (2007) The *Arabidopsis* vacuolar malate channel is a member of the ALMT family. Plant J 52:1169–1180

Leivar P, Antolin-Llovera M, Ferrero S, Closa M, Arro M, Ferrer A, Boronat A, Camposa N (2011) Multilevel control of *Arabidopsis* 3-hydroxy-3-methylglutaryl Coenzyme A reductase by protein phosphatase 2A. Plant Cell 23:1494–1511

Lerceteau-Köhler E, Moing A, Guérin G, Renaud C, Petit A, Rothan C, Denoyes B (2012) Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. Theor Appl Genet 124:1059–1077

Liebhard R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003) Mapping quantitative physiological traits in apple (*Malus* × *domestica* Borkh). Plant Mol Biol 52:511–526

Ligaba A, Katsuhara M, Ryan PR, Shibasaka M, Matsumoto H (2006) The BnALMT1 and BnALMT2 genes from rape encode aluminum-activated malate transporters that enhance the aluminum resistance of plant cells. Plant Physiol 142:1294–1303

Ligaba A, Kochian L, Pineros M (2009) Phosphorylation at S384 regulates the activity of the TaALMT1 malate transporter that underlies aluminum resistance in wheat. Plant J 60:411–423

Maliepaard C, Alston FH, van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73

Meyer S, Mumm P, Imes D, Endler A, Weder B, Al-Rasheid KAS, Geiger D, Marten I, Martinoia E, Hedrich R (2010) AtALMT12 represents an R-type anion channel required for stomatal movement in *Arabidopsis* guard cells. Plant J 63:1054–1062

Meyer S, Scholz-Starke J, De Angeli A, Kovermann P, Burla B, Gambale F, Martinoia E (2011) Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. Plant J 67:247–257

Navascues J, Perez-Rontome C, Sanchez DH, Staudinger C, Wienkoop S, Rellan-Alvarez R, Becana M (2012) Oxidative stress is a consequence, not a cause, of aluminum toxicity in the forage *legume Lotus corniculatus*. New Phytol 193:625–636

Nybom N (1959) On the inheritance of acidity in cultivated apples. Hereditas 45:332–350

Obrero A, Die JV, Roman B, Gomez P, Nadal S, Gonzalez-Verdejo CI (2011) Selection of reference genes for gene expression studies in Zucchini (*Cucurbita pepo*) using qPCR. J Agr Food Chem 59:5402–5411

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45

Robinson JP, Harris SA, Juniper BE (2001) Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant Syst and Evol 226:35–58

Sasaki T, Yamamoto Y, Ezaki B, Katsuhara M, Ahn SJ, Ryan PR, Delhaize E, Matsumoto H (2004) A wheat gene encoding an aluminum-activated malate transporter. Plant J 37:645–653

Schumacher K, Krebs M (2010) The V-ATPase: small cargo, large effects. Curr Opin Plant Biol 13:724–730

Skottke KR, Yoon GM, Kieber JJ, DeLong A (2011) Protein phosphatase 2A controls ethylene biosynthesis by differentially regulating the turnover of ACC synthase isoforms. PLoS Genet 7:e1001370

Soglio V, Costa F, Molthoff J, Weemen-Hendriks W, Schouten H, Gianfranceschi L (2009) Transcription analysis of apple fruit development using cDNA microarrays. Tree Genet Genomes 5:685–698

Sweetman C, Deluc LG, Cramer GR, Ford CM, Soole KL (2009) Regulation of malate metabolism in grape berry and other developing fruits. Phytochemistry 70:1329–1344

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Ulrich R (1970) Organic acids. In: Hulme A (ed) The biochemistry of fruit and their products. Academic Press, London and New York, pp 89–118

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, Chagne D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouze P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel C-E, 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). Nat Genet 42:833–839

Visser T, Verhaegh JJ (1978) Inheritance and selection of some fruit characters of apple.1. Inheritance of low and high acidity. Euphytica 27:753–760

Xu K, Wang A, Brown S (2011) Genetic characterization of the Ma locus with pH and titratable acidity in apple. Mol Breeding. doi: 10.1007/s11032-011-9674-7

Yamaguchi M, Sasaki T, Sivaguru M, Yamamoto Y, Osawa H, Ahn SJ, Matsumoto H (2005) Evidence for the plasma membrane localization of Al-activated malate transporter (ALMT1). Plant Cell Physiol 46:812–816

Yao YX, Li M, Liu Z, Hao YJ, Zhai H (2007) A novel gene, screened by cDNA-AFLP approach, contributes to lowering the acidity of fruit in apple. Plant Physiol Bioch 45:139–145

Yao YX, Li M, Liu Z, You CX, Wang DM, Zhai H, Hao YJ (2009) Molecular cloning of three malic acid related genes MdPEPC, MdVHA-A, MdcyME and their expression analysis in apple fruits. Sci Hortic 122:404–408

Yao Y-X, Li M, Zhai H, You C-X, Hao Y-J (2011) Isolation and characterization of an apple cytosolic malate dehydrogenase gene reveal its function in malate synthesis. J Plant Physiol 168:474–480

You F, Huo N, Gu Y, Luo M-c, Ma Y, Hane D, Lazo G, Dvorak J, Anderson O (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253

Zhang YZ, Li PM, Cheng LL (2010) Developmental changes of carbohydrates, organic acids, amino acids, and phenolic compounds in 'Honeycrisp' apple flesh. Food Chem 123:1013–1018