Genomic analyses of the CAM plant pineapple

Jisen Zhang1,2, Juan Liu1 and Ray Ming1,3,*

1 FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, Fujian, 350002, China
2 College of Life Sciences, Fujian Normal University, Fuzhou, 350108, China
3 Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

* To whom correspondence should be addressed. E-mail: rming@life.uiuc.edu

Received 12 December 2013; Revised 28 January 2014; Accepted 10 February 2014

Abstract

The innovation of crassulacean acid metabolism (CAM) photosynthesis in arid and/or low CO2 conditions is a remarkable case of adaptation in flowering plants. As the most important crop that utilizes CAM photosynthesis, the genetic and genomic resources of pineapple have been developed over many years. Genetic diversity studies using various types of DNA markers led to the reclassification of the two genera Ananas and Pseudananas and nine species into one genus Ananas and two species, A. comosus and A. macrodentes with five botanical varieties in A. comosus. Five genetic maps have been constructed using F1 or F2 populations, and high-density genetic maps generated by genotype sequencing are essential resources for sequencing and assembling the pineapple genome and for marker-assisted selection. There are abundant expression sequence tag resources but limited genomic sequences in pineapple. Genes involved in the CAM pathway has been analysed in several CAM plants but only a few of them are from pineapple. A reference genome of pineapple is being generated and will accelerate genetic and genomic research in this major CAM crop. This reference genome of pineapple provides the foundation for studying the origin and regulatory mechanism of CAM photosynthesis, and the opportunity to evaluate the classification of Ananas species and botanical cultivars.

Key words: Crassulacean acid metabolism (CAM) photosynthesis, epigenetics, genetic diversity, linkage mapping, reference genome.

Genetic and genomic analysis of pineapple

The pineapple (Ananas comosus L.) is a perennial monocot belonging to the family Bromeliaceae, subfamily Bromelioidae, in the order Bromeliales. It is a tropical plant native to South America and the third most important tropical fruit crop after banana and mango. The world pineapple production was 21.9 million tonnes in 2012, which is approximately six times the production in 1961 (3.8 million tonnes) (http://faostat.fao.org). In addition to fresh fruit consumption, pineapple is used for canned slices, juice and juice concentrates, extraction of bromelain (a meat-tenderizing enzyme), high-quality fiber, and animal feed.

Based on traditional morphology, the pineapple includes nine species distributed in two genera, Ananas and Pseudananas (Smith and Downs, 1979). Later, of these nine species, A. monstrosus was excluded from the two genera on the basis of leaf morphology (Leal, 1990). During the past
decade, the classification was further refined, and the two genera and eight species were downgraded into two species, *A. macrodontes* and *A. comosus*, and five botanical varieties of *A. comosus*, var. *comosus*, *ananassoides*, *parguazensis*, *erectifolius*, and *bracteatus* (Coppens d’Eeckenbrugge and Leal, 2003). Evidence based on molecular markers confirmed this revised classification (Coppens d’Eeckenbrugge and Leal, 2003). *A. macrodontes* is a self-fertile tetraploid with $2n=4x=100$ chromosomes, whereas *A. comosus* is a mostly self-incompatible diploid with $2n=2x=50$ chromosomes (Marchant, 1967; Brown and Gilmartin, 1986; Brown et al., 1997). Triploid and tetraploid clones were observed in var. *comosus* and tetraploid clones in var. *ananassoides* (Sharma and Ghosh, 1971; Lin, 1987; Dujardin, 1991; Cotias-de-Oliveira et al., 2000).

Pineapple has a special photosynthetic pathway, crassulacean acid metabolism (CAM), the biochemistry of which is well documented. A CAM plant such as pineapple has the unique ability to store carbon dioxide as malic acid within the plant, allowing it to fix carbon dioxide at night as malate acids, which are then released during the day. This mechanism increases the efficiency of photosynthesis while preventing excessive water loss due to transpiration from open stomata. Pineapple is considered an ‘obligate’ CAM plant, using an exclusively CAM pathway during photosynthesis. Some CAM plants are able to switch between C$_3$ and CAM subject to changes in the environment. Pineapple is a non-climacteric fruit without the autocatalytic ethylene burst and increase in respiration.

The available genetic and genomic resources and a large collection of pineapple germplasm make pineapple an excellent model for studying obligate CAM photosynthesis and molecular mechanisms of non-climacteric ripening. Thanks to high-throughput sequencing technologies and the development of efficient molecular markers, significant progress has been made in developing genomic resources, leading to the sequencing of the pineapple genome. Here we review the genetic and genomic advances in pineapple and CAM photosynthesis and assess the current state of knowledge in these fields for annotating and analysing the pineapple genome.

**Genetic diversity**

The genetic diversity among *Ananas* germplasm was initially investigated using isozyme markers (DeWald et al., 1988; Garcia, 1988; Duval and Coppens d’Eeckenbrugge, 1992; Aradhya et al., 1994). In the study of DeWald et al. (1988), 15 of 27 *A. comosus* cultivars were identified by five enzymatic systems, two peroxidases and three phosphoglucomutases. In the study of Aradhya et al. (1994), 161 pineapple accessions from the Hawaiian collection, including four different species of *Ananas* and one species of *Pseudananas*, were identified by six isozyme systems involving seven putative loci (Smith and Downs, 1979). Multivariate analyses of the isozyme variation revealed an inconsistency with the *A. comosus* group classifications based only on phenotypic traits. Isozyme evidence also showed that *A. erectifolius* is a conspecific variant of *A. comosus*, and that among other wild species, *A. ananassoides* is more closely related to *A. comosus* than *A. bracteatus*. *Pseudananas* was shown to be genetically distinct from all species of *Ananas*. However, enzymatic systems have limited use for analysing the genetic diversity of *Ananas* accessions. In a test for 37 enzymatic systems, only eight were confirmed to be useful among *Ananas* accessions (Duval and Coppens d’Eeckenbrugge, 1992).

More recently, DNA-based markers have been used to study the phylogenetic relationships between *Ananas* and related genera. Restriction fragment length polymorphism (RFLP) markers were analysed among 301 accessions of *Ananas* and related genera including 168 *A. comosus* accessions, suggesting that *A. comosus* has lower levels of polymorphism than wild *Ananas* species (Duval et al., 2001). Similarly, based on amplified fragment length polymorphism (AFLP) markers pattern of Mexican germplasm collections, *A. comosus* accessions were reported to have a low level of diversity (Paz et al., 2005). In contrast, in a study of 148 *A. comosus* accessions and 14 related genera analysed by AFLP, *A. comosus* showed a higher degree of genetic variations (Kato et al., 2005).

According to isozyme, RFLP, and AFLP results (Duval and Coppens d’Eeckenbrugge, 1992; Duval et al., 2001), *A. comosus* and *A. macrodontes* can be separated clearly, but *A. comosus* var. *bracteatus* appears relatively uniform and better differentiated from the other varieties. In the study of Kato et al. (2005), *A. comosus* var. *comosus* is most closely related to *A. comosus* var. *ananassoides*; however, early AFLP results demonstrated that *A. comosus* var. *ananassoides*, some of *A. comosus* var. *parguazensis* accessions, and *A. comosus* var. *comosus* were clustered in same group (Duval et al., 2001). Chloroplast DNA (cpDNA) diversity of *Ananas* and related genera were evaluated by PCR-RFLP (Duval et al., 2003), suggesting that the genetic diversity of *Ananas* was relative to the geographical origin of the accessions but not the species. These results supported the pineapple classification by Coppens d’Eeckenbrugge and Leal (2003) and enable us to generate a dendrogram for pineapple classification (Fig. 1). The level of genetic diversity among commercial cultivars is still unclear. Three commercial cultivar groups, ‘Cayenne’, ‘Queen’, and ‘Spanish’, were investigated by random amplified polymorphic DNA (RAPD). ‘Cayenne’ and ‘Queen’ cultivars were grouped into two separate clusters, whereas ‘Spanish’ failed to form a monophyletic group (Sripaoraya et al., 2001). However, major cultivar groups of the 148 *A. comosus* accessions of pineapple, such as ‘Cayenne’, ‘Spanish’, and ‘Queen’, could not be distinctively separated by AFLP (Kato et al., 2005). Simple sequence repeat (SSR) markers further confirmed the incongruence between taxonomic and molecular markers (Shoda et al., 2012; Feng et al., 2013). Thirty one pineapple accessions could be differentiated by the 18 SSR markers, but could not be clustered into distinct groups with morphological classification (Shoda et al., 2012). Similar results were reported in 48 pineapple varieties that were clustered into four subgroups using SSR markers (Feng et al., 2013). In all these studies, genetic diversity was not in good accordance with morphological classification, indicating that the phenotypic traits used to characterize different cultivar groups could have evolved from a small number of mutations in different genetic backgrounds. Re-sequencing
of selected cultivars from each group could reveal the nature of genomic variation within and among cultivar groups.

Genetic mapping

Five genetic maps of pineapple were constructed using F₁ and F₂ populations with molecular markers developed over the years. The first genetic map of pineapple was generated using an F₁ population with 46 progenies derived from a cross between *A. comosus* var. *comosus* (cv. ‘Rondon’, clone BR50) and var. *bracteatus* (cv. ‘Brancodo mato’, clone BR20) (Carlier et al., 2004). The map of the female parent BR50 consisted of 156 molecular markers (33 RAPD, 115 AFLP, and 8 SSR) in 30 linkage groups, which spans over 1311 cM covering 31.6% of 3693 cM genome length estimated using maximum likelihood method (Chakravarti et al., 1991). The map of the male parent BR20 consisted of 335 molecular markers in 50 linkage groups, which spans over 2111 cM and covered 57.2% of the genome (Carlier et al., 2004).

The second release of the map is an integrated map that combined these two F₁ maps and an F₂-based genetic map (Carlier et al., 2005). The integrated map contains a total of 46 linkage groups with 574 markers, which spans more than 2421 cM and covers 62% of the genome. Using 142 F₂ individuals, a genetic map was constructed with 412 markers in 50 linkage groups. The total span of the map is 2458 cM, which corresponds to 62.7% of the genome (Botella and Smith, 2008).

Recently, an F₂-based genetic map was constructed with 492 DNA markers in 40 linkage groups, covering approximately 80% of the genome (Carlier et al., 2012). The latest integrated genetic map combines F₁ and F₂-based maps, with 741 markers in 28 linkage groups, and spans 2113 cM (de Sousa et al., 2013) (Table 1).

A reference genetic map of pineapple is a valuable tool for positional cloning and for marker-assisted selection. A high-density genetic map is also required for physical mapping and *de novo* genome assembly. The current genetic maps have contributed to building the genomic resources for pineapple, but are still mostly partial maps. With the advance of sequencing technologies, a sequence-tagged high-density genetic map covering the vast majority of the pineapple genome is necessary for genome assembly and comparative genomic research in monocots.

**EST resources**

Expression sequence tags (ESTs) are an essential resource for gene discovery and genome annotation. Pineapple ESTs were generated from cDNA libraries of roots, fruit, and aerial tissues (Neuteboom et al., 2002), green mature fruits (Moyle et al., 2005b), and nematode infected gall (Moyle et al., 2005a). PineappleDB includes more than 5500 ESTs with 3383 consensus sequences that were analysed with splice variants and further compared with putative *Arabidopsis* homologues for functional annotation (http://genet.imb.uq.edu.au/Pineapple/index.html). The comprehensive sequence, bioinformatic and functional classification of EST resources are available for text or sequence-based searches.

Recently, a pineapple microarray with 9277 elements was used for gene expression profiling between mature green and yellow stages of pineapple fruit ripening (Koia et al., 2012). A total of 271 unique cDNAs expressed at least 1.5-fold difference between the two stages. Among the 237 pineapple sequences with identifiable homologues, 160 were upregulated and 77 were downregulated during pineapple fruit ripening. The ESTs showing differential expression were predicted to

---

**Fig. 1.** The dendrogram of pineapple species and cultivars based on published diversity studies (Smith and Downs, 1979; Duval et al., 2001; Coppens d’Eeckenbrugge and Leal, 2003; Kato et al., 2005).

**Table 1. Summary of five pineapple genetic maps**

<table>
<thead>
<tr>
<th>No.</th>
<th>Population</th>
<th>Markers</th>
<th>Linkage groups</th>
<th>Coverage of genome</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F₁ male</td>
<td>335</td>
<td>50</td>
<td>57.2%</td>
<td>Carlier et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>F₁ female</td>
<td>156</td>
<td>30</td>
<td>31.6%</td>
<td>Carlier et al. (2004)</td>
</tr>
<tr>
<td>2</td>
<td>F₁ and F₂</td>
<td>574</td>
<td>46</td>
<td>62%</td>
<td>Carlier et al. (2006)</td>
</tr>
<tr>
<td>3</td>
<td>F₂</td>
<td>412</td>
<td>50</td>
<td>62.7%</td>
<td>Botella and Smith (2008)</td>
</tr>
<tr>
<td>4</td>
<td>F₂</td>
<td>492</td>
<td>40</td>
<td>80%</td>
<td>Carlier et al. (2012)</td>
</tr>
<tr>
<td>5</td>
<td>F₁ and F₂</td>
<td>741</td>
<td>28</td>
<td>86%</td>
<td>de Sousa et al. (2013)</td>
</tr>
</tbody>
</table>
be genes enriched for redox activity, organic acid metabolism, metalloenzyme activity, glycolysis, vitamin C biosynthesis, antioxidant activity, and cysteine peptidase activity (Koia et al., 2012).

Using the Illumina sequencing platform, about 4.7 million paired-end reads were generated and assembled for transcriptsomes of ripe yellow pineapple fruit flesh (Ong et al., 2012). The paired-end reads were assembled into a total of 28,728 unique transcripts with average size of approximately 200 bp. Of these, 16,932 unique transcripts (58.93%) BLAST hit to homologue sequences in the NCBI database. A total of 13,598 unique transcripts (47.33%) were mapped to 126 pathways in the genomes pathway database (http://www.genome.jp/kegg/). This study provided an EST resource to be further utilized in gene expression, gene predictions, and other functional genomic studies in pineapple.

In these reported experiments, the molecular basis of ripening was the primary objective for both the Australian and Malaysian groups. However, the ESTs from Sanger methods covered only 5500 transcripts, whereas the RNAseq experiments revealed 28,728 unique transcripts, reflecting the rapid advance of sequencing technologies and increased resolution of genomic research. For the purposes of gene prediction and genome annotation, it would be necessary to sample widely across many tissues and development stages to obtain maximum coverage of pineapple transcripts.

Gene discovery and identification

Gene discovery and identification will facilitate pineapple breeding, through either genetic transformation or marker-assisted selection. In pineapple, genes relative to the ripening of pineapple fruit and plant development have been cloned and characterized. Ethylene biosynthetic genes 1-aminocyclopropane-1-carboxylate (ACC) synthase and ACC oxidase have been cloned and their expression levels analysed at different developmental stages of fruit (Cazzonelli et al., 1998). The ACC synthase gene (AcACSS2) was shown to be involved in the regulation of time of flowering of pineapple by gene silencing (Trusov and Botella, 2006). A somatic embryogenesis receptor-like kinase (SERK) gene was isolated from pineapple, and expression analysis indicated that it may play a role in the transition from somatic cell to embryogenic cell and development of competent cell to global embryo (Ma et al., 2011). Moreover, a full-length cDNA of a putative aspartic acid protease (AcAPI) has been isolated. The gene was thought to be involved in resistance to post-harvest chilling stress based on the expression level between two pineapple varieties differing in their resistance to blackheart (Raimbault et al., 2013).

Several genes related to modern medicinal use and conferring microbial resistance have been studied, including bromelain inhibitors (Sawano et al., 2002), the cysteine protease inhibitor cistatin (Shyu et al., 2004), and aspartic acid protease (AcAPI) (Antony et al., 2008). A pineapple fruit bromelain gene (BAA) has been cloned and transferred into Chinese cabbage (Brassica rapa) resulting in enhanced resistance to bacterial soft rot (Firoozabady et al., 2005). The transcript expression of bromelain and AcCYSI were shown to directly correlate with the resistance to blackheart development in pineapple fruit (Carter et al., 2000; Raimbault et al., 2013).

Genes involved in sugar transport pathways play essential roles for sugar accumulation, affecting fruit quality. Three putative vacuolar sugar transporters—hexose transporter (AcMST1), inositol transporter (AcINT1), and sucrose transporter (AcSUT1)—have been cloned from the leaves of pineapple (Raimbault et al., 2013). The subcellular localization of the three transporters has been investigated, and both AcMST1 and AcINT1 observed to localize in the tonoplast, whereas AcSUT1 localized to prevacuolar compartments. The expression profile of AcINT1 showed day–night changes (Raimbault et al., 2013).

The discovery of key genes will play an important role in developing strategies for genetic improvement and understanding the mechanism of complex traits in pineapple. According to the physiological features of pineapple, genes related to the CAM pathway and non-climacteric ripening could be of primary interest. For better understanding of the functional diversity of these target genes, it would be necessary to investigate gene family members and allelic variations in Ananas.

Genome sequencing in pineapple

The pineapple genome was sequenced by our group using the whole genome shotgun (WGS) approach with both Illumina and 454 sequencing technologies and pooled BACs with Illumina only to high depth on pineapple variety F153. The highly heterozygous nature and self-incompatibility of pineapple poses substantial challenge for genome assembly. We used a combination of bioinformatic techniques to overcome the difficulty of assembling the pineapple genome, which will be described in the genome manuscript.

The resulting draft genome has a contig N50 of 9.5 kb and a scaffold N50 of 408 kb, and covers about 375 Mb (62%) of the estimated 526 Mb genome. Most of the missing sequences are repeat sequences, since between 88 and 92% of pineapple protein-coding genes are covered in the current draft assembly based on evaluation of RNAseq and CEGMA coverage, respectively. The pineapple draft genome contains 43% repetitive sequences, and is predicted to have 25862 genes using MAKER assisted by RNAseq transcripts. The availability of a high-quality reference genome will serve as an indispensable resource for pineapple researchers.

Genomic analysis of the CAM pathway

Origin of CAM

CAM plants comprise at least 343 genera in 35 families of angiosperms, including both monocots and eudicots, but are mainly characterized in Crassulaceae, Bromeliaceae, Orchidaceae, and Agavaceae (Silvera et al., 2010; Matiz et al., 2013). Phylogenetic reconstruction of CAM and C₃ species indicates that C₃ photosynthesis is the ancestral state
of CAM. Almost all identified genes involved in the CAM pathway are already present in C₅ plants, and CAM photosynthesis likely evolved by regulation and reorganization of gene expression in the ancestral C₃ pathway (West-Eberhard et al., 2011). CAM plants in 35 families evolved independently, even within the same family. Phylogenetic analysis in Orchidaceae demonstrated that CAM has evolved at least 10 times independently with several reversals, along with parallel evolution in subfamilies (Crayn et al., 2004; Silvera et al., 2009).

Genes and gene network involved in the CAM pathway

CAM photosynthesis is characterized by nocturnal CO₂ fixation by phosphoenolpyruvate carboxylase (PEPC) to generate oxaloacetate (OAA), which is rapidly converted to malate and transported into the vacuole. During the day, when stomata are closed, the C₃ organic acids are remobilized from the vacuoles and decarboxylated, generating CO₂ as a substrate for Rubisco in the Calvin cycle (C₃).

Few genes involved in the CAM pathway have been studied in pineapple. The related genes from several CAM plants provide a good reference for CAM pathway analysis in pineapple. Large-scale mRNA expression profiles in CAM-inducing and C₃-inducing Mesembryanthemum crystallinum revealed a range of metabolic genes related to CAM, including genes for C₄ metabolism, glycolysis/glucogenesis, starch synthesis/degradation, and transporters (Cushman et al., 2008). Transcriptome sequencing of Agave deserti and Agave tequilana, two drought-tolerant CAM plants, and comparative analysis with the proteomes of four monocotyledonous grass species (Brachypodium distachyon, Oryza sativa, Sorghum bicolor, Zea mays) revealed 4325 orthologous groups common in both Agave species but absent in the four grass species (Gross et al., 2013), with some of them likely attributed to CAM photosynthesis.

Transporters are critical in the CAM pathway. Vacuolar sugar transporters characterized in pineapple appear to be functioning to maintain the carbohydrate partition (Antony et al., 2008). The vacuolar H⁺-ATPase provides the proton-motive force for malic acid accumulation in vacuoles. Recently, dynamic changes in the proteome of M. crystallinum leaves showed that the majority of subunits of V-ATPase (vacuolar H⁺-ATPase) holoenzyme were significantly affected during the transition to CAM (Costantino et al., 2013). While switching to CAM, all known subunits of V-ATPase showed a sharp increase, but subunits E and c have the highest rise. The data show that the V-ATPase holoenzyme performs different functions in CAM and C₃ photosynthesis via changes in subunit structure (Costantino et al., 2013).

Na⁺/H⁺ antiporters play a role in Na⁺ compartmentation during plant adaptation to high salinity and expression levels show a temporal correlation with salt accumulation in leaves but not in roots. These results indicate that Na⁺/H⁺ antiporters are probably involved in CAM metabolism (Costantino et al., 2010).

It has been demonstrated that aquaporins (AQPs) play a role in maintaining water balance during CAM. The protein abundance of AQPs showed a persistent rhythm with CAM. However, the transcript profiles were not associated with the protein abundance, which suggests post-transcriptional regulation of these AQPs in the CAM pathway (Vera-Estrella et al., 2012).

Proteomic profiling of CAM-inducing pineapple demonstrated five major spots that were different from C₃-inducing plants. Two of them, OEE1 (oxygen involving enhancer 1) and OEE2, were reported for the first time in Ananas comosus, whereas the other three are subunits of Rubisco. OEE1 and OEE2 are subunits of the PSII oxygen-evolving system. The same study also demonstrated that transcripts of the antioxidant enzyme catalase and OEE1 are related to the CAM pathway in pineapple (Aragon et al., 2013). The relationship between CAM and oxidative metabolism has been observed in many CAM plants, but molecular mechanisms are still not clear.

It has been demonstrated that the common ice plant has a similar multigene loop oscillator with Arabidopsis (Boxall et al., 2005). Transcripts of circadian central gene TOC1 and LATE ELONGATED HYPOCOTYL/CIRCADIAN CLOCK ASSOCIATED 1 (LHY/CCA1) oscillated with a 24-h periodicity, the same as the expression profiles in C₃ plants. However, a recent report suggested that CCA1 and PRR9 showed a 12-h oscillation in another CAM plant, the cactus Opuntia ficus-indica (Mallona et al., 2011). Different expression rhythms in M. crystallinum and O. ficus-indica may be the result of independent evolution under distinct environmental cues.

Based on the available data, ZEITLUPE (ZTL) was proposed as a candidate protein in regulating the expression of CAM. This is a core component of ZEITLUPE/FLAVIN-BINDING, KELCH REPEAT, F-BOX 1/LOV KELCH PROTEIN 2 (ZTL/FKF1/LKP2), one of the three families of LOV blue light receptors in green plants. AtZTL has been reported to be expressed independent of light–dark cycles and the circadian clock in Arabidopsis (Somers et al., 2000), while oscillation of McZTL transcripts was observed not only in C₃-induced but also in CAM-induced M. crystallinum (Boxall et al., 2005). ZTL specifically interacts with TOC1 and PRR5 and mediates their proteolytic degradation (Suetsumu and Wada, 2013). TOC1 is a component of an autoregulatory negative feedback loop, together with two Myb transcription factors LHY and CCA1, whereas PRR5 is essential for robust oscillation (Nagel and Kay, 2012). It has been hypothesized that these rhythm changes of CCA1 and PRR9 in O. ficus-indica may be related to the gene ZTL or blue light.

TOC1, considered a core circadian gene in multigene loop oscillator for CAM regulation, has been cloned and overexpressed in Kalanchee fedtschenkoi (Hartwell, 2005; Borland et al., 2009). Studies in O. ficus-indica indicated that TOC1, as in Arabidopsis, oscillates with a 12-h rhythm, peaking during the night (Mallona et al., 2011). Therefore, TOC1 probably plays a minor role in the CAM pathway or is regulated post-transcriptionally. In addition, diel expression of the other blue light photoreceptors (PHOT1 and PHOT2) under circadian clock control have been found in Clusia rosea (Barrera Zambrano, 2012) (Table 2).
### Table 2. Genes involved in the CAM pathway

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession number</th>
<th>Function</th>
<th>Species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key enzymes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphoenolpyruvate</td>
<td>AJ312631.1</td>
<td>Initial fixation of CO₂ into OAA in C₄ and CAM</td>
<td>Ananas comosus; Phalaenopsis</td>
<td>Aragon et al. (2012, 2013); Chen et al. (2008);</td>
</tr>
<tr>
<td>carboxylase (PEPC1)</td>
<td>(EC 4.1.1.31)</td>
<td></td>
<td>aphrodite; Mesembryanthemum</td>
<td>Cushman et al. (2008); Davies and Griffiths</td>
</tr>
<tr>
<td></td>
<td>X63774.1</td>
<td></td>
<td>crystallinum; Agave deserti;</td>
<td>(2012); Gross et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>AF158091</td>
<td></td>
<td>Agave tequilana</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAPP1_MESCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>K01595 (KEGG)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphoenolpyruvate</td>
<td>NM_119948</td>
<td>Phosphorylation of PEPC</td>
<td>A. comosus; M. crystallinum;</td>
<td>Aragon et al. (2012, 2013)</td>
</tr>
<tr>
<td>carboxykinase (PEPCK)</td>
<td></td>
<td></td>
<td>Agave deserti; Agave tequilana</td>
<td></td>
</tr>
<tr>
<td>Rubisco</td>
<td></td>
<td>Carboxylation</td>
<td>M. crystallinum</td>
<td>Cushman et al. (2008); Davies and Griffiths</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2012)</td>
</tr>
<tr>
<td>NAD⁺-malic enzyme (NAD⁺-ME)</td>
<td></td>
<td>Catalyses decarboxylation of malate</td>
<td>P. aphrodite; Rosularia eymaitica; Kalanchoe daigremontiana</td>
<td>Chen et al. (2008); Habibi and Hajiboland (2011); Peckmann et al. (2012)</td>
</tr>
<tr>
<td>NADP-malic enzyme (NADP-ME)</td>
<td></td>
<td>Catalyses decarboxylation of malate</td>
<td>Opuntia ficus-indica</td>
<td>Mallona et al. (2011)</td>
</tr>
<tr>
<td>Large subunit of Rubisco (LSU)</td>
<td></td>
<td>Photosynthetic carbon fixation</td>
<td>A. comosus</td>
<td>Aragon et al. (2013)</td>
</tr>
<tr>
<td>Small subunit of Rubisco (SSU)</td>
<td></td>
<td>Photosynthetic carbon fixation</td>
<td>A. comosus</td>
<td>Aragon et al. (2013)</td>
</tr>
<tr>
<td>Pyruvate phosphate dikinase</td>
<td></td>
<td>Controls the phosphorylation state of PEPC</td>
<td></td>
<td>Mallona et al. (2011)</td>
</tr>
<tr>
<td>(PPDK)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmitters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V-ATPase</td>
<td></td>
<td>Provides the proton-motive force for malic acid</td>
<td>M. crystallinum</td>
<td>Cosentino et al. (2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>accumulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugar transporters</td>
<td></td>
<td>Regulate carbon flow</td>
<td>A. comosus</td>
<td>Antony et al. (2008)</td>
</tr>
<tr>
<td>Na⁺/H⁺ antiporter</td>
<td></td>
<td>High salinity Na⁺ compartmentation</td>
<td>M. crystallinum</td>
<td>Cosentino et al. (2010)</td>
</tr>
<tr>
<td>Oxidative related</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malate dehydrogenase (MDH)</td>
<td>DT336869</td>
<td>Reversibly catalyses the oxidation of malate</td>
<td>A. comosus</td>
<td>Freschi et al. (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to oxaloacetate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxygen evolving enhancer 1</td>
<td></td>
<td>A subunit of PSII oxygen evolving system</td>
<td>A. comosus</td>
<td>Aragon et al. (2013)</td>
</tr>
<tr>
<td>(OEE1)</td>
<td>ABQ52657.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxygen evolving enhancer 2</td>
<td></td>
<td>A subunit of PSII oxygen evolving system</td>
<td>A. comosus</td>
<td>Aragon et al. (2013)</td>
</tr>
<tr>
<td>(OEE2)</td>
<td>P12302.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catalase (CAT)</td>
<td>GU266543.1</td>
<td>Key enzyme of the antioxidative system</td>
<td>A. comosus; Sedum album; R. elymatica</td>
<td>Aragon et al. (2013); Habibi and Hajiboland (2011); Hajiboland (2010)</td>
</tr>
<tr>
<td></td>
<td>EC 1.11.1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascorbate peroxidase (APX)</td>
<td>GU266541.1</td>
<td>Key enzyme of the antioxidative system</td>
<td>A. comosus; S. album; R. elymatica</td>
<td>Aragon et al. (2013); Habibi and Hajiboland (2011); Hajiboland (2010)</td>
</tr>
<tr>
<td></td>
<td>EC 1.11.1.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superoxide dismutase (SOD)</td>
<td>AJ250667.1</td>
<td>Key enzyme of the antioxidative system</td>
<td>A. comosus; S. album; R. elymatica</td>
<td>Aragon et al. (2013); Habibi and Hajiboland (2011); Hajiboland (2010)</td>
</tr>
<tr>
<td></td>
<td>EC 1.15.1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: a = annotated genes; K01595 (KEGG) = KEGG number; EC = Enzyme Commission number.
Table 2. Continued

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession number</th>
<th>Function</th>
<th>Species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrate reductase (NR)</td>
<td>EC 1.6.6.1</td>
<td>NO production and stimulates the expression of CAM</td>
<td>Pineapple</td>
<td>Freschi et al. (2009, 2010)</td>
</tr>
<tr>
<td>Photosynthesis and circadian clock related</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phototropin 1/2 (phot1/2)</td>
<td>CC_MiF_011E05b</td>
<td>Blue light receptor</td>
<td>Clusia rosea</td>
<td>Barrera Zambrano (2012)</td>
</tr>
<tr>
<td></td>
<td>CC_MiF_011G04b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timing of Cab expression 1 (TOC1)</td>
<td>BI543444</td>
<td>Core circadian gene</td>
<td>M. crystallinum</td>
<td>Boxall et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>BI543434</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryptochrome 2</td>
<td>CC_MuF_004G08b</td>
<td>Harvesting light</td>
<td>C. rosea</td>
<td>Barrera Zambrano (2012)</td>
</tr>
<tr>
<td>Zeitupe (ZTL)</td>
<td>Ay371290</td>
<td>Blue light acceptor</td>
<td>M. crystallinum</td>
<td>Boxall et al. (2005)</td>
</tr>
<tr>
<td>Circadian clock associated 1 (CCA1)</td>
<td>Contig2293s</td>
<td>Myb transcription factors, controlling the circadian clock</td>
<td>O. ficus-indica; M. crystallinum</td>
<td>Malona et al. (2011)</td>
</tr>
<tr>
<td>Pseudo response regulator 9 (PRR9)</td>
<td>Contig18061b</td>
<td>Controlling the circadian clock</td>
<td>O. ficus-indica</td>
<td>Malona et al. (2011)</td>
</tr>
</tbody>
</table>

* Communicate with Malona et al. for sequence information.
  
* Was a Clusia EST database, http://xyala.cap.ed.ac.uk/Gene_Pool/Katherine_Shorrock

Epigenetic control of CAM

DNA methylation plays a role in adaptation to salt stress in the halophyte *M. crystallinum*. Under salt stress, *M. crystallinum* switches from C_{3} to CAM, coupled with increase in the expression of *Ppc1*, a CAM-specific isoform of the PEPC gene family. Nine cytosines in the promoter region and 5′-untranslated region changed their methylation pattern during the transition to CAM (Huang et al., 2010). Further research demonstrated that the methylation level of a CTG site, which is 9 bp downstream of TATA-box in the promoter of *Ppc1*, coincided with the expression of *Ppc1* (Huang et al., 2010). The methylation of this specific site activated the interaction of AT-rich DNA-binding factor with the promoter of *Ppc1* (Cushman and Bohnett, 1992).

Hypermethylation of CCWGG sequences in the nuclear genome have an effect on adaptation to salt stress in *M. crystallinum* (Dyachenko et al., 2006). The CpNpG-hypermethylation occurs in the satellite DNA when switching to the CAM pathway, probably promoting a specialized chromatin structure, which could change the expression pattern of a large number of genes and result in shifting from a C_{3} to CAM pathway. Transcriptome sequencing of *A. deserti* and *A. tequilana* revealed that the class II KNOX transcription factors have a higher expression in the distal leaf, where elevated CAM photosynthesis was detected compared with leaf base (Gross et al., 2013). The expression of *KNOX1*, associated with organogenesis during bulbil formation, was epigenetically regulated by the histone modification marker H3K9me3 in *Agave* (De-la-Peña et al., 2012). In addition, a genome-wide study in *Arabidopsis* showed that H3K4Me3 abundance is proportional to the expression of genes in response to drought (van Dijk et al., 2010). Histone modification might be more important in CAM plants, which mostly live in arid environments.

Most epigenetic variation is reset between generations, although in some cases it can be transgenerationally inheritable, which is beneficial for the offspring on adaptation to environmental change (Robertson and Wolf, 2012). It has been shown that *A. tequilana* offsets from rhizomes and bulbils on the inflorescence showed an overall increase in methylation (Díaz-Martínez et al., 2012). Although the methylation patterns between mother plant and offspring differed in three forms of asexual reproduction, *in vitro* cultured plants showed patterns specific to each generation.

Epigenetic allele (epiallele) stability and inheritability were also tested in *Arabidopsis* under salt stress. The results showed that epigenetic marker H3K9ac is less abundant in the progeny, with expression analysis identifying downregulated genes in salt-stressed plants (Farahani, 2013). This suggests that histone modifications play a part in establishment of trans-generational stress memory. Research aimed at discovering epigenetic effects on adaptation and evolution showed that epigenetic variation in three allopolyploid orchids had been restructured during genome doubling, forming species-specific patterns adapting to specific eco-environmental cues, especially water availability and temperature (Paun et al., 2010).

Perspective

Pineapple is the most important crop with CAM biosynthesis, parthenocarpic fruit development, and non-climacteric fruit ripening. Its economic importance and these biological features are justification for sequencing the pineapple genome. The reference genome covered about 90% of genes, and the remaining unassembled genome is mostly repetitive sequences. The difficulty in assembling a near-complete genome is due to heterozygosity caused by self-incompatibility and vegetative propagation and short reads of current sequencing technology. Nevertheless, the reference genome will accelerate genetic and genomic research of pineapple and CAM photosynthesis, as well as helping in understanding the
molecular basis of self-incompatibility in monocots, parthenocarpy, and non-climacteric fruit ripening.

The available genetic and genomic resources and a large collection of pineapple germplasm make pineapple an excellent model for studying the genomics of CAM photosynthesis. Based on the potential reference genome of pineapple, analysis of the genes involved in CAM photosynthesis by comparing gene expression between pineapple seedlings (utilize C₃) and mature plants (utilize CAM), it might be possible to predict the origin and divergence of CAM plants. Generating mutant populations through irradiation and/or EMS treatment to identify mutants that do not express CAM biosynthesis would be a better approach for identifying genomic regions containing regulatory elements controlling CAM photosynthesis. Such information could also be generated from comparative analysis of ice plants M. crystallinum under normal and stressed conditions that switch between C₃ and CAM biosynthesis and then generating mutant populations in this model species for studying CAM biosynthesis. The information from multiple species with independent origins of CAM photosynthesis will contribute to the identification of genes and regulatory elements controlling this photosynthesis pathway, and expedite the conversion between C₃ and CAM pathways on targeted crops.

Inconsistencies between morphological classification and molecular diversity could be explained by similar mutations in different genetic backgrounds that resulted in different morphological traits (Duval et al., 2001; Kato et al., 2005). Utilizing the range of germplasm collections from the centre of diversity in Brazil and production regions worldwide, it may be possible to discover domestication genes via genome-wide association studies (GWAS). The ongoing re-sequencing of Ananas accessions representing all species and cultivar groups from the previous and current taxonomic classification will enable us to verify the current classification. Re-sequencing of multiple cultivars from ‘Smooth Cayenne’, ‘Queen’, and ‘Spanish’ will provide the opportunity to test the hypothesis that pineapple cultivars are selected from somatic mutations.

Acknowledgements

This work was supported by a grant from EBI and startup funds from Fujian Agriculture and Forestry University to RM. We thank Drs Michael Schatz and Haibao Tang for assembling the pineapple genome that was briefly described.

References


Robertson AL, Wolf DE. 2012. The role of epigenetics in plant adaptation. Trends in Evolutionary Biology 4, 64.


