


# SCIENTIFIC REPORTS

OPEN

## Metabolomics and genomics combine to unravel the pathway for the presence of fragrance in rice

Venea Dara Daygon<sup>1</sup>, Mariafe Calingacion<sup>1</sup>, Louise C. Forster<sup>2</sup>, James J. De Voss<sup>2</sup>, Brett D. Schwartz<sup>2</sup>, Ben Ovenden<sup>3</sup>, David E. Alonso<sup>4</sup>, Susan R. McCouch<sup>5</sup>, Mary J. Garson<sup>2</sup> & Melissa A. Fitzgerald<sup>1</sup> 

Received: 6 March 2017

Accepted: 27 June 2017

Published online: 18 August 2017

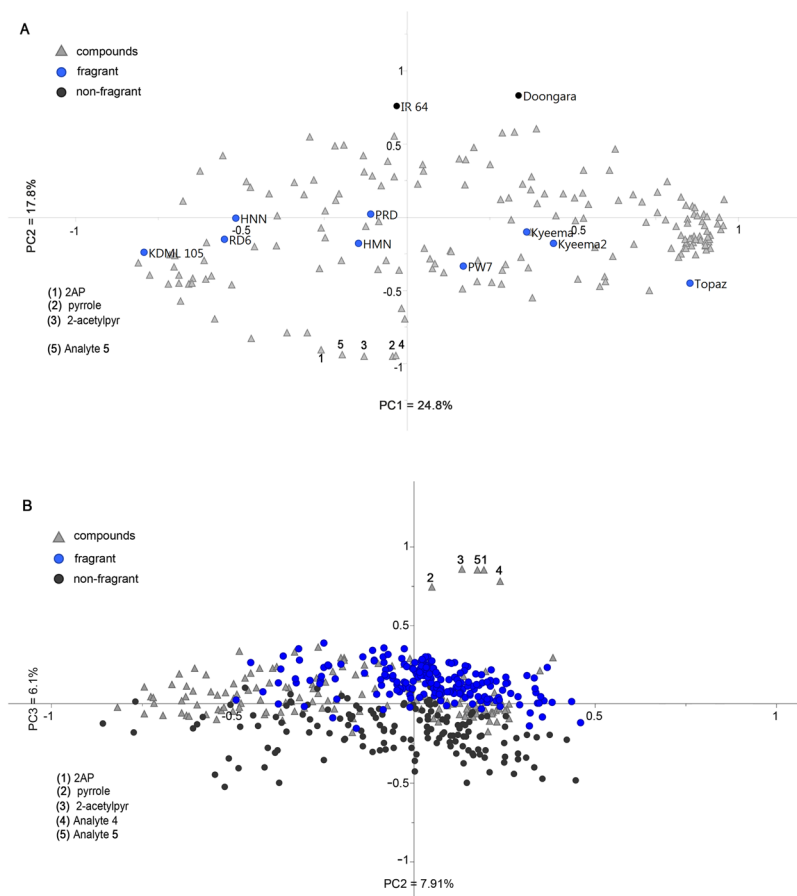
Since it was first characterised in 1983, 2-acetyl-1-pyrroline (2AP) has been considered to be the most important aroma compound in rice. In this study, we show four other amine heterocycles: 6-methyl, 5-oxo-2,3,4,5-tetrahydropyridine (6M5OTP), 2-acetylpyrrole, pyrrole and 1-pyrroline, that correlate strongly with the production of 2AP, and are present in consistent proportions in a set of elite aromatic rice varieties from South East Asia and Australia as well as in a collection of recombinant inbred lines (RILs) derived from *indica* Jasmine-type varieties, Australian long grain varieties (*temperate japonica*) and Basmati-type rice (*Grp V*). These compounds were detected through untargeted metabolite profiling by two-dimensional gas chromatography-time-of-flight mass spectrometry (GC × GC-TOF-MS), and their identity were confirmed by comparison with authentic standards analysed using gas chromatography mass spectrometry (GC-MS) and High Resolution GC × GC-TOF-MS (GC × GC HRT-4D). Genome-wide association analysis indicates that all compounds co-localised with a single quantitative trait locus (QTL) that harbours the *FGR* gene responsible for the production of GABA. Together, these data provide new insights into the production of 2AP, and evidence for understanding the pathway leading to the accumulation of aroma in fragrant rice.

Fragrant (or aromatic) rice commands the highest prices in the global rice market<sup>1</sup>. The trait of fragrance has been selected multiple times by ancient farmers throughout South and Southeast Asia, and multiple, independent alleles have been found in the same gene in different genetic backgrounds<sup>2</sup>. The strength of fragrance differs between varieties<sup>3,4</sup>, and the most highly fragrant varieties tend to be the most popular among consumers<sup>1</sup>. For example, in Lao PDR, there is a fragrant rice variety called ‘Four Houses’, meaning that when this rice is cooking, its fragrance is enjoyed by people four houses away<sup>5</sup>. Emphasising strong fragrance through variety naming indicates the value that consumers place on highly fragrant rice.

Fragrance in rice is currently defined by a single compound, 1-(3,4-dihydro-2H-pyrrol-5-yl) ethanone, usually known by its synonym 2-acetyl-1-pyrroline (2AP)<sup>4,6</sup>. Contrary to previous speculations that 2AP is produced in rice as a result of Maillard reaction<sup>7</sup>, it has been demonstrated that the presence of 2AP is due to loss-of-function mutations in the coding region of the *FGR* gene (Os08g0424500)<sup>2,8,9</sup>. Suppression of the *FGR* allele in non-aromatic rice varieties results in production of 2AP and leads to fragrance<sup>9</sup>, while transgenic expression of the *FGR* allele in *fgr*-carrying genotypes converts aromatic lines into non-aromatic lines<sup>10</sup>. These two studies conclusively confirm that alleles of the *FGR* gene are essential for the presence of 2AP in rice. Similarly, the inactivation of *FGR* orthologues in other plant species such as *Bassia latifolia*<sup>11</sup>, soybean<sup>12</sup>, and cucumber<sup>13</sup> also leads to the production of 2AP.

The *FGR* locus encodes an amino aldehyde dehydrogenase (AADH) which catalyses the oxidation of 4-aminobutanal, the product of oxidative deamination<sup>14</sup>, to  $\gamma$ -aminobutyric acid (GABA)<sup>9,15</sup> as a stress response of plants<sup>16</sup>. Varieties carrying *fgr*, the recessive allele, produce a truncated AADH protein which cannot catalyse the oxidation of 4-aminobutanal to GABA<sup>15</sup>. It has been reported that 4-aminobutanal

<sup>1</sup>School of Agriculture and Food Sciences, The University of Queensland, St Lucia, QLD, 4072, Australia. <sup>2</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD, 4072, Australia. <sup>3</sup>NSW Department of Primary Industries, Yanco Agricultural Institute, Yanco, NSW, 2703, Australia. <sup>4</sup>LECO Corporation Life Science and Chemical Analysis Centre, 1850 Hilltop Rd, Saint Joseph, MI, 49085, USA. <sup>5</sup>Department of Plant Breeding & Genetics, Cornell University, Ithaca, NY, 14853, USA. Correspondence and requests for materials should be addressed to M.A.F. (email: [m.fitzgerald2@uq.edu.au](mailto:m.fitzgerald2@uq.edu.au))



**Figure 1.** PCA biplot of the volatile metabolites detected in rice by GC × GC-TOF-MS. Metabolites are the triangles, overlaid with the rice varieties (circles). (A) Set 1 (n = 11), PC1 = 24.8% and PC2 = 17.8%. (B) Set 2 (n = 341) PC2 = 7.91% and PC3 = 6.1%.

accumulates, and cyclisation and dehydration result in the formation of 1-pyrroline, which becomes acetylated to form 2AP<sup>10,11,15</sup>. It is also widely proposed that 1-pyrroline is directly acetylated to 2AP by methylglyoxal<sup>17,18</sup> or acetyl CoA<sup>15,19</sup>. Moreover, reports have demonstrated that the addition of 1-pyrroline to *Bacillus aureus* culture medium<sup>20</sup> and to unheated rice callus<sup>21</sup> leads to a significant increase in 2AP, although in these studies, the recovery of 2AP is low. Together, these studies suggest firstly that 1-pyrroline (or a biological equivalent) could participate in the pathway leading to 2AP; and secondly that the temperature sensitivity of 2AP formation implies an enzyme-mediated synthesis.

Previous analyses of the volatile compounds in rice have consistently detected an unknown compound in fragrant rice varieties<sup>22–24</sup>. The similarity of the mass spectral electron ionisation (EI) fragmentation pattern of the unknown compound to that of 2AP has led to suggestions that the compound is an isomer of 2AP<sup>23,24</sup>. However, both the identity of the compound, and any role it plays in 2AP synthesis, are yet to be determined. There remains, therefore, much to understand about the processes regulating the accumulation of 2AP in fragrant rice grains.

The objectives of the present paper, using a collection of advanced breeding lines developed from diverse rice varieties, were to (i) perform metabolite profiling using several analytical platforms to identify metabolites in the pathway of 2AP synthesis; (ii) conduct genome-wide association studies (GWAS) to identify Quantitative Trait Loci (QTL) associated with the accumulation of relevant metabolites; and (iii) further understand the pathway leading to aroma in rice.

## Results

**Untargeted metabolite profiling of volatile compounds.** To identify metabolites that relate to the pathway of 2AP, we performed metabolite profiling of volatile compounds in a test set of 10 commercial rice varieties using GC × GC-TOF-MS. A total of 175 putative compounds were detected and were used for multivariate analysis (Fig. 1A). The principal components analysis (PCA) biplot shows that 24.8% of total variation is explained by PC1. The samples separated according to germplasm class (*indica* vs *japonica* varieties) along PC1, with the *indica* samples clustering in the left of the axis and the *japonica* varieties on the right of the axis, with one exception. KDML 105, a Thai *indica* variety, grouped with the *japonica* samples, as was also observed in our previous study<sup>22</sup> using a similar set of varieties. A clear separation between the fragrant and the non-fragrant rices was observed along PC2, which explains 17.8% of total variation. To determine which of the compounds are the most discriminating between aromatic and nonaromatic rice, OPLS-DA was conducted (Figure S1). The five

Set 1	(1)	(2)	(3)	(4)	(5)
2-acetyl-1-pyrroline (1)	1				
pyrrole (2)	0.68 (0.04)	1			
2-acetylpyrrole (3)	0.99 (<0.0001)	0.64 0.06	1		
Analyte 4	0.1 0.79	0.78 0.01	0.08 0.84	1	
Analyte 5	1 (<0.0001)	0.66 0.05	0.99 (<0.0001)	0.09 0.83	1
<b>Set 2</b>					
2-acetyl-1-pyrroline (1)	1				
pyrrole (2)	0.62 (<0.0001)	1			
2-acetylpyrrole (3)	0.61 (<0.0001)	0.85 (<0.0001)	1		
Analyte 4	0.47 (<0.0001)	0.65 (<0.0001)	0.56 (<0.0001)	1	
Analyte 5	0.98 (<0.0001)	0.62 (<0.0001)	0.62 (<0.0001)	0.44 (<0.0001)	1

**Table 1.** Correlation coefficients of 2AP linked compounds in Set 1 (n = 11) and Set 2 (n = 341). *P* values are given in italics.

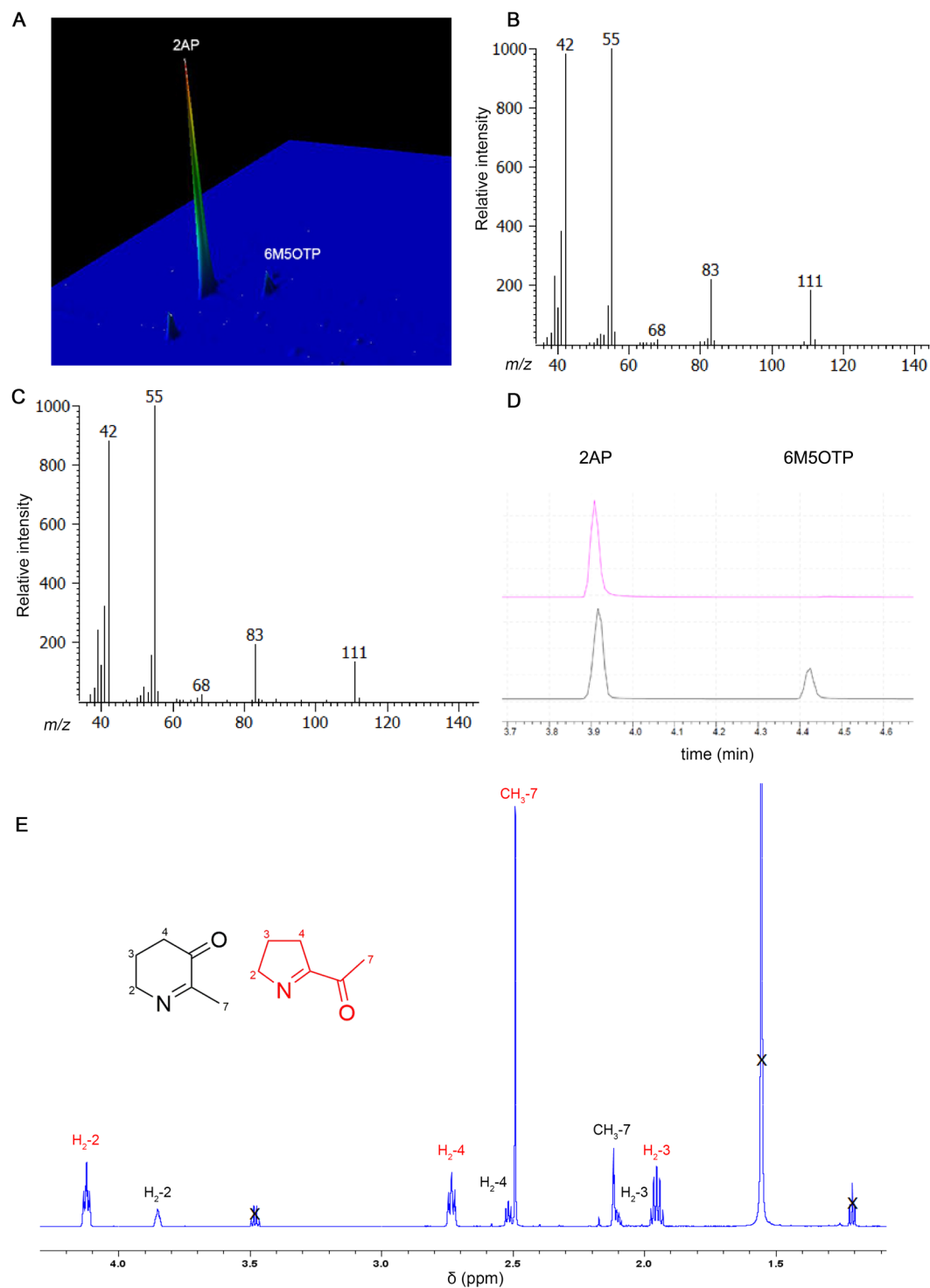
compounds with the highest variable importance in projection (VIP) scores in the aromatic cluster were: 2-acetyl-1-pyrroline (Analyte 1), pyrrole (Analyte 2), 2-acetylpyrrole (Analyte 3), Analyte 4 and Analyte 5. Analytes 1, 2, and 3 had convincingly high similarity matches to the library (NIST 11 v2.0) mass spectra (94%, 90.9% and 94.5%, respectively) and the annotations were subsequently validated against analytical standards (2AP synthesised in this study; pyrrole, 98% and 2-acetylpyrrole, 99% both Sigma-Aldrich, MO, USA) (Figure S2). Analytes 4 and 5 (identified as 1-ethenylaziridine and 1-piperidine by the library search) were treated as unannotated compounds due to the relatively low similarity matches of 81.1% and 77.7% respectively. The mass spectral data of 4: *m/z* (%) 69 (48), 68 (24), 54 (2), 42 (27), 41 (100), 40 (17), 39 (22), 38 (11); and 5: 111 (10), 84 (2), 83 (17), 66 (2), 55 (100), 54 (14), 52 (6), 42 (84), 41 (30), 40 (14), 39 (33) matched the library spectra poorly. The spectral data and retention time of Analyte 5 matches that of an unknown compound (Analyte 9) in our previous study<sup>22</sup>. A correlation analysis between the five compounds detected in the aromatic rice shows a strong correlation between 2AP and 2-acetylpyrrole ( $R^2 = 0.99$ , Table 1), 2AP and Analyte 5 ( $R^2 = 1$ ), and 2-acetylpyrrole and Analyte 5 ( $R^2 = 0.99$ ). A weak correlation was found between Analyte 4 with all the compounds except for pyrrole ( $R^2 = 0.78$ ).

A collection of 341 rice samples (Set 2), planted in a different season and location (see Methods section), was used to confirm the metabolites detected on the elite commercial samples. PCA showed that 2AP and the linked compounds discriminate between fragrant and non-fragrant rice samples in the third PC, which explains 6.1% of total variation observed (Fig. 1B). Furthermore, the strong correlations between the five compounds were once again confirmed (Table 1). There was a strong correlation between 2AP and Analyte 5, while the observed positive correlation between 2AP and 2-acetylpyrrole was seen to decrease in the bigger set ( $R^2 = 0.61$  from  $R^2 = 0.99$ ). The strength of the correlation between Analyte 4 and the other compounds generally increased in Set 2, although the highest association was only 0.65, again with pyrrole. Next, the relative amounts of the five compounds were compared between both sets of samples. A full model, accounting for sample group and compound, was fit to the data. In the usual fashion, a second, nested model, which omitted sample group as predictor variable, was fit, and the two models compared. There was no significant loss of fit to the data by removal of the sample group as predictor, indicating that variation between the two groups was not significant ( $\chi^2 = 7.3193$ ,  $df = 5$ ,  $p = 0.198$ ). The proportion of 2AP to Analyte 5 was calculated in the fragrant accessions of Set 2, and an apparent 4:1 ratio was consistently noted (0.79:0.21 with a standard deviation of 0.05, Table S1). These analyses confirm the linkage between all five compounds, and emphasised the need to further characterise and identify Analytes 4 and 5.

**Identification of unknown compounds.** *Analyte 4.* Using high resolution MS (EI) analysis by HRT 4D, Analyte 4 had ions at *m/z* C<sub>2</sub>H<sub>3</sub>N 41.0260, C<sub>4</sub>H<sub>6</sub>N 68.0497 and C<sub>4</sub>H<sub>7</sub>N 69.0573 (Figure S3A) with mass accuracies of 0.01, 0.59 and  $-0.01$  ppm, respectively. The molecular ion corresponds to a molecular formula of C<sub>4</sub>H<sub>7</sub>N (calc: 69.0578), which is expected for 1-pyrroline, the putative precursor in the biosynthesis of 2AP<sup>15</sup>. The reported EI mass spectral fragmentation pattern of 1-pyrroline *m/z* (%) (69 (68), 68 (31), 42 (41), 41 (100))<sup>25,26</sup> is similar to that observed, and therefore 1-pyrroline became a logical candidate for the identity of Analyte 4. We therefore synthesised an authentic sample of 1-pyrroline using literature methods<sup>26</sup> and found that it possessed identical mass spectrometric properties (Figure S4A,B) and spectroscopic (Proton nuclear magnetic resonance,<sup>1</sup>H NMR, Figure S4C) to those reported in the literature<sup>25,26</sup>. When analysed by GC × GC TOF-MS, the retention time of 1-pyrroline was identical to that of Analyte 4, at 370 s and 0.7 s in the primary and the secondary columns, respectively. The MS EI fragmentation of Analyte 4 (*m/z* (%) 69 (48), 68 (30), 41(100)) detected in rice, matched that of the synthetic sample (Figure S4), confirming the identification of 1-pyrroline.

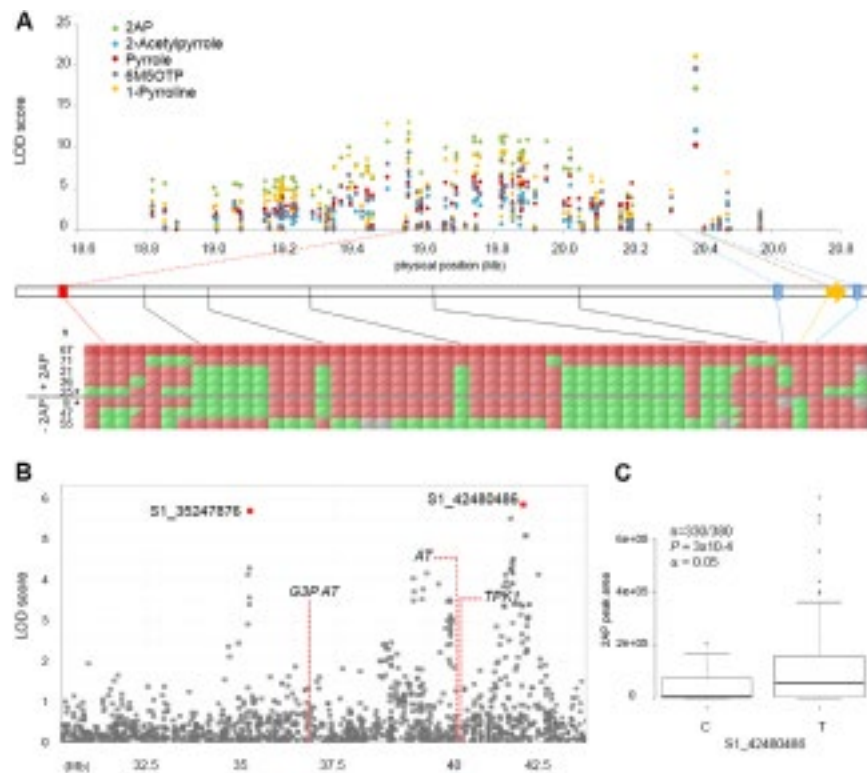
*Analyte 5.* HRT 4D analysis of Analyte 5 shows ions at *m/z* C<sub>2</sub>H<sub>4</sub>N 42.0463, C<sub>3</sub>H<sub>5</sub>N 55.0416, C<sub>5</sub>H<sub>9</sub>N 83.0729 and a molecular ion of C<sub>6</sub>H<sub>9</sub>NO 111.0680, with mass accuracies of  $-0.01$ , 1.21,  $-0.61$  and  $-0.92$  ppm, respectively (Figure S3B). High resolution MS analysis (CI) showed an accurate mass value of 112.07572 [MH]<sup>+</sup> which correlated to a molecular formula of C<sub>6</sub>H<sub>10</sub>NO, with mass accuracy of 0.29 ppm. Given that it was unconvincingly annotated by the mass spectral library search, and since C<sub>6</sub>H<sub>9</sub>NO is the same formula as that of 2AP, we considered an isomeric structure of 2AP as the candidate for Analyte 5.

We synthesised authentic 2AP standards (Supplementary Information S1), which produced a mixture of 2AP and a second compound (Fig. 2A). When analysed by GC × GC TOF-MS, the retention time (707.5 s, 1.2 s) and



**Figure 2.** Identification of Analyte 5 as (6M5OTP). (A) GC x GC-TOF-MS 3D chromatogram of synthesised 2AP standard showing 2AP (RT: 627.5 s, 1.025 s) and 6M5OTP (RT: 707.5 s, 1.27 s). MS Electron ionisation fragmentation pattern of Analyte 5 detected by GC x GC-TOF-MS (B) synthesised 2AP standard, and (C) rice. (D) GC-MS trace comparison of ratio between 2AP and 6M5OTP before (above) and after acid-base extraction (below). (E) Expansion of the  $^1\text{H}$  NMR of 2AP (red) and 6M5OTP (black). Solvent peaks are denoted in X.

mass spectral fragmentation pattern of the second compound (Fig. 2B) is identical to that of Analyte 5 detected in rice (Fig. 2C). Furthermore, 2AP and the second compound was found to be in a 99:1 ratio by GC - MS. Subjecting the reaction product to a single round of acid-base partitioning increased the concentration of the second compound to a ratio of 76:24 (GC-MS) (Fig. 2D). A second acid-base extraction further increased the



**Figure 3.** GWAS of 2AP and linked compounds. (A) Physical distribution of the most significant SNPs in chromosome 8. The functional markers for *FGR* (diamonds) outperformed the MS SNPs (S8\_19555833) for each of the five compounds. Lines connecting (A,B) denote the position of the S8\_19555833 (previously MS SNP, red line) and the SNPs closest to *FGR* (SNPs S8\_20311425 and S8\_20408314, blue lines) relative to *FGR* (orange arrow). (B) Representative haplotypes of the population. Red cells represent SNPs identical to the aromatic line KDML 105, heterozygotes are represented by half-filled cells, while grey are missing data. The number of lines carrying each haplotype is indicated in the table. Asterisk denotes the haplotypes where linkage breaks between the MS SNP S8\_19555833 are not predictive of the presence or absence of aroma, explaining the lower LOD score of SNP S8\_19555833. (C) Zoom in of the QTL region on chromosome 1. The most significant SNPs (MS SNPs) are denoted in red. Location of the candidate genes are shown by dotted lines. (D) Box plot for the 2AP levels at different alleles of the MS SNP S1\_42480486. The middle line indicates the median, the box indicates the range of the 25th to 75th percentiles of the total data, the whiskers indicate the interquartile range and the outer dots are outliers.  $n$  shows the number of samples used to make the plot over the total number of samples,  $p$  and  $\alpha$  denote the  $p$ -values and alpha, respectively.

product ratio to approximately 73:27 (GC-MS). The  $^1\text{H}$  NMR spectrum of the second compound ( $^1\text{H}$  NMR:  $\delta$  3.78–3.85 (m, 2H), 2.48 (t,  $J=6.9$  Hz, 2H), 2.08 (s, 3H), 2.03–2.10 (m, 2H), Fig. 2E) matched that reported by De Kimpe<sup>27</sup>, and by Harrison and Dake<sup>28</sup> for 6-methyl-5-oxo-2,3,4,5-tetrahydropyridine, as did the  $^{13}\text{C}$  NMR data obtained from 2D (HSQC, HMBC) NMR spectra (Figure S5). Taken together, these data all indicate that Analyte 5 is 6-methyl-5-oxo-2,3,4,5-tetrahydropyridine (6M5OTP), an isomer of 2AP.

**GWA analysis for 2AP and linked compounds.** The milled rice grains from a collection of 341 rice lines was phenotyped to determine the relative concentrations of the five volatile compounds in each line (as described above) and the same plant was genotyped to obtain a genetic profile for each line. Using Genotyping-by-Sequencing (GBS), a total of 233,214,413 short sequencing reads were produced from the 384 barcoded samples, with an average number of reads per individual of 572,545. Only one sample (Y20140298) generated <10% of the mean barcode reads value (MBRV) and was discarded, while blank samples generated <1% of the total MBRV. 1,888,334 tags were aligned with the Nipponbare reference genome (MSU version 7.0) and 87.6% aligned to unique positions. A total of 240,250 Single Nucleotide Polymorphisms (SNPs) were discovered in the initial allele calling. SNPs were filtered on call rates ( $\geq 75\%$ ) and minimum allele frequency ( $\geq 0.05$ ), resulting in a dataset consisting of 33,725 SNPs that were well-distributed across the genome. This SNP matrix was used for further data analysis and association mapping.

For GWAS on all 340 lines, a large QTL for 2AP was identified on chromosome 8 (logarithm of odds, LOD score = 14) near the *FGR* locus<sup>8</sup>. Co-located QTLs were found in the same region for the presence of 2-acetylpyrrole, pyrrole, 1-pyrroline and 6M5OTP (LOD scores of = 10.7, 9.3, 16.7 and 16.6, respectively) (Fig. 3A, Figure S6). The Manhattan and QQ plots for this GWAS are given in Figure S6. The most significant SNP (MS SNP) S8\_19555833 (LOD score = 14) mapped 750 kb ( $\sim 3$  cM) away from the *FGR* locus (Fig. 3). No

SNP markers from the GBS dataset mapped within the coding region of *FGR*. Two SNPs – S8\_20311425 and S8\_20408314 – were found to be closer to *FGR* (68 kb and 29 kb from the translational start site, respectively), but both were present at low frequency and did not segregate with the phenotype. To test the hypothesis that polymorphism in the *FGR* gene might be responsible for the variation observed in the levels of the five compounds, 262 of the lines were further screened using allele-specific “functional” markers for *FGR* targeting 8 bp indel and 3 SNPs in seventh exon of the *FGR* gene<sup>29</sup>. Ninety seven progenies were found to carry the functional form of the gene (*Fgr* allele conferring no fragrance) while 165 carried the recessive (mutated) version (*fgr* allele conferring fragrance), and the marker perfectly predicted the phenotype (Table S1, Fig. 3B). Upon addition of the *FGR* functional marker as a covariate to the GWAS model, the peak shifted toward the *FGR* gene for all five compounds, and the significance of the functional marker surpassed that of SNP S8\_19555833 (Fig. 3A).

A second, minor-effect QTL for levels of 2AP was observed on chromosome 1 (Figure S6A). Further examination of this region revealed two QTL peaks, with MS SNPs S1\_35247876 (LOD score = 6.06) and S1\_42480486 (LOD score = 5.96), approximately 7 Mb apart (Fig. 3C). SNP S1\_35247876 has a slightly higher LOD score than S1\_42480486, however, the former also has more ‘N’ genotypes or missing values (72 versus 10 missing genotypes, respectively). Therefore we used S1\_42480486 in looking at the contribution of the different alleles (C/T) to the levels of 2AP (Fig. 3D). Samples carrying a ‘T’ SNP show significantly more 2AP production ( $p < 0.0003$ ) compared to those carrying a ‘C’ SNP. Scanning the region between the two MS SNPs, we found several genes which are directly or indirectly involved in the transfer of either an acyl group or a two-carbon subunit, such as *thiamine pyrophosphate kinase 1* (*TPK1*, OS01G0931400, LOC\_Os01g70580 start site 1:40871884), *acyltransferase* (*AT*, OS01g0931300, LOC\_Os01g70570 start site 1:40868356), and *glycerol-3-phosphate acyltransferase* (*G3PAT*, OS01g0855000, LOC\_Os01g63580 start site 1:36866375) (Fig. 3C). While none of these genes are in strong linkage disequilibrium (LD) with the MS SNPs examined here, it would be worth investigating whether allelic variation at one or more of these candidate genes may impact levels of 2AP.

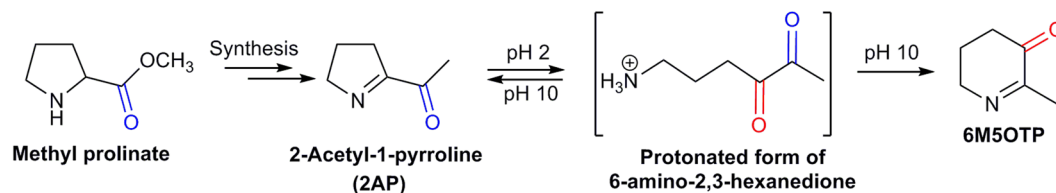
## Discussion

Previous studies<sup>2, 8, 23, 24</sup> on rice aroma have focused on 2AP as the major aromatic compound. Here, we present four other compounds: pyrrole, 2-acetylpyrrole 1-pyrroline, and 6-methyl-5-oxo-2,3,4,5-tetrahydropyridine (6M5OTP) that are highly discriminating between aromatic and non-aromatic rices, and which associate strongly with 2AP, both chemically and genetically (Fig. 1). The genetic and metabolomic data presented here indicate strongly that these five compounds all participate in one pathway. Two of the compounds, 6M5OTP and 1-pyrroline, are identified here for the first time in rice. The compound 6M5OTP appears to have been previously detected in other aromatic rices<sup>22–24</sup> but never correctly identified. The compound 1-pyrroline has long been considered the precursor of 2AP, but has not been previously detected in rice. In the past, the floral, popcorn and toast smell of rice has been attributed to 2AP alone. However, its structural isomer 6M5OTP has been reported to have a similar scent perceivable through gas chromatography olfactory analysis<sup>24</sup>. It is now evident from our chemical study that 2AP co-occurs with 6M5OTP. This means that 6M5OTP, presumably produced concomitantly with 2AP in rice, contributes to the distinguishing aroma of fragrant rice.

Strong correlations were found between 2AP, 2-acetylpyrrole and 6M5OTP in the Asian and Australian elite aromatic varieties; these were also observed in the larger collection grown in a different location and experimental season (Table 1), suggesting that this occurrence could be universal to all aromatic rices. 2AP can easily oxidise to 2-acetylpyrrole at room temperature<sup>30</sup>, which could explain the correlation between these two compounds. The same oxidation process is likely to explain the correlation between 1-pyrroline and pyrrole<sup>31</sup>. Both 2AP and 2-acetylpyrrole have sweet aroma reminiscent of cooked popcorn scent<sup>6</sup>. However, the odour threshold of 2-acetyl pyrrole is 170000 ppb as compared to 2AP (0.06 ppb)<sup>6</sup>, suggesting that the oxidation of 2AP leads to loss of favourable aroma in rice.

The major QTL peak for all five heterocycles localised in the region of the *FGR* gene. This could be explained by several hypotheses: (i) there is allelic variation for a master switch that pleiotropically governs the accumulation of multiple compounds; (ii) there is allelic variation for a gene that governs the expression of a critical node in one pathway, and variation in that pathway indirectly affects other pathways that determine the accumulation of the other compounds; (iii) there is an array of linked genes within the QTL peak. To test for the possibility of a master switch, the population was screened for the major *FGR* allele<sup>2</sup> using previously developed markers<sup>29</sup>. Upon rerun of the GWAS, the functional *FGR* marker was found to be more significantly associated with both fragrance and levels of all 5 compounds than the MS SNP (S8\_19555833). To understand whether this was due to recombination separating S8\_19555833 from the putative functional gene, *FGR*, we examined the genotypic profiles and identified 33 recombinants between S8\_19555833 and the functional *FGR* marker (Fig. 3A). This explained the lower level of significance for the genotype-phenotype association, and supported the conclusion that the loss of function of *FGR* is indeed the master switch that determines the increase in levels of all five compounds.

This study shows, for the first time in rice, that 1-pyrroline is a volatile compound that is detected only in fragrant rice (Fig. 1A). If 1-pyrroline is the immediate precursor of 2AP, it would be expected to correlate strongly with 2AP. However, this study shows that the correlations between 1-pyrroline and either 2AP or 6M5OTP are the weakest amongst the compounds (Table 1). Likewise, previous attempts to synthesise 2AP *in vivo* and *in vitro* from 1-pyrroline gave very poor yields. *In vitro* synthesis of 2AP from 1-pyrroline-5-carboxylic acid and methylglyoxal yielded only 6.3%<sup>32</sup>, and a similar synthesis of a 2AP analogue from phenylglyoxal gave a 3.7% recovery<sup>26</sup>. *In vivo*, addition of 1-pyrroline to non-fragrant rice at physiological pH has the highest yield of 0.05% conversion to 2AP in unheated callus at pH 8.5<sup>21</sup>. These studies indicate that acceptable amounts of 2AP do not form *in vitro* or in the callus at physiological pH from 1-pyrroline. These previous reports, together with the weak correlation between 2AP and 1-pyrroline reported here, suggest that direct acetylation of 1-pyrroline may not be the primary pathway for the formation of 2AP.



**Figure 4.** Chemical formation of 6M5OTP through *in vitro* acid/base workup. 2AP ring opens in excess acid to form an unstable diketone, which cyclises back to 2AP or to 6M5OTP under basic conditions.

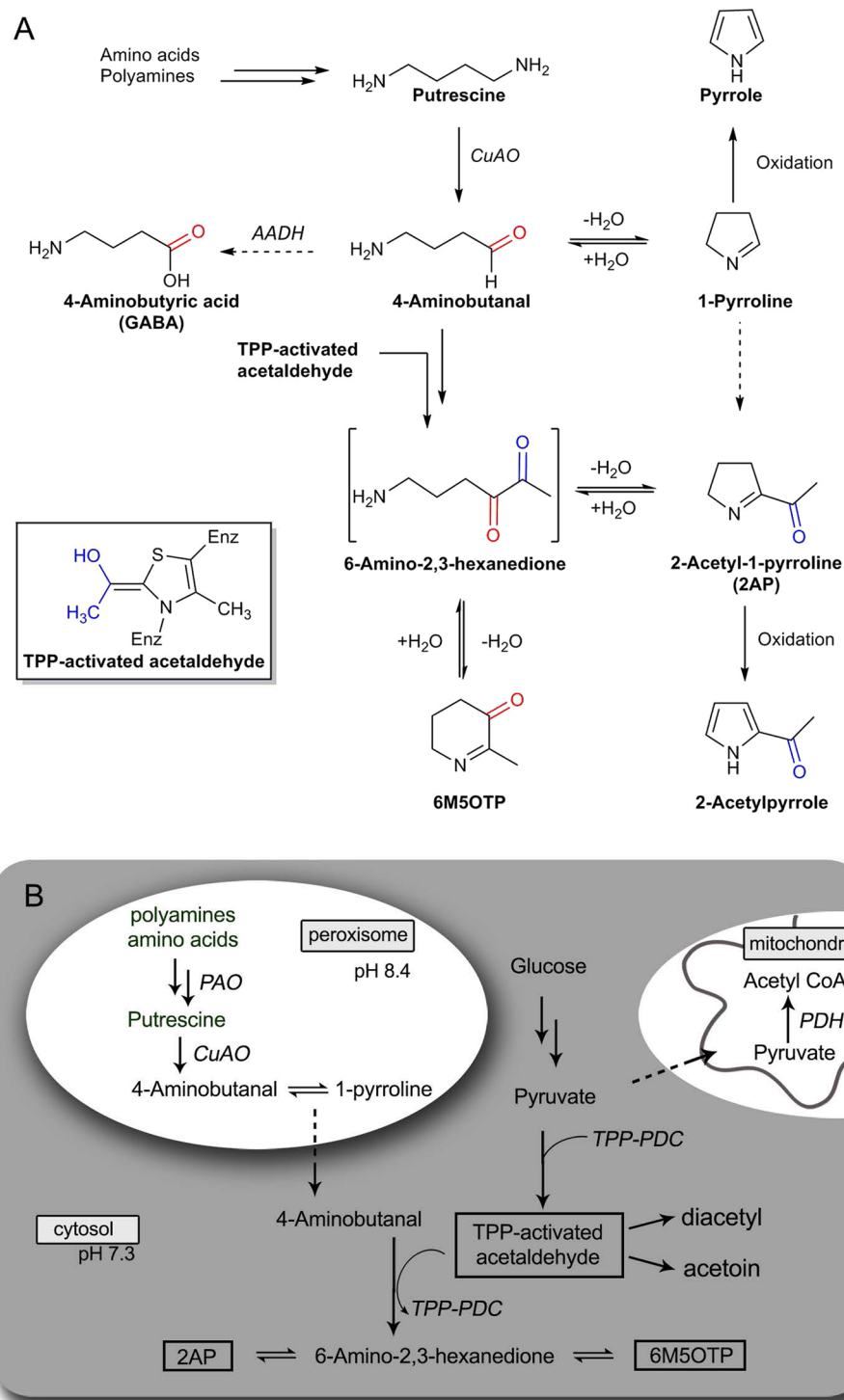
The present paper, as well as several previous studies<sup>27, 28</sup> supports the possibility of ring opening *in vitro* when 2AP is subjected to acidic conditions. De Kimpe *et al.*<sup>27</sup> isolated 6M5OTP as a minor product during acid-catalysed deprotection of imidoyl- or diethoxyethyl-pyrrolines. Likewise, during deprotection of *N*-Boc-protected pyrroline, Harrison and Dake<sup>28</sup> reported the isolation of 6M5OTP. In the present study, an acid-base workup of the synthesised 2AP led to an increase in the proportion of 6M5OTP, indicating ring opening of 2AP followed by ring closure. The ring opening of 2AP leads to an unstable, linear diketone intermediate – 6-amino-2,3-hexanedione (Fig. 4), which differentially cyclises at carbon 2 (C2) or at C3 resulting in 2AP or 6M5OTP, respectively, when the reaction solution is made basic. These experiments demonstrated the importance of the formation of the diketone intermediate in the production of 2AP and 6M5OTP.

The clear chemical and genetic linkage between 6M5OTP and 2AP raises questions about the biosynthesis of 2AP. It is well known that 4-aminobutanal is derived from putrescine in the pathway of polyamine oxidation to GABA, and in rice, this process occurs only in the peroxisome<sup>33, 34</sup>. The transfer of two carbon units, such as is required to form 2AP from 4-aminobutanal or 1-pyrroline, has been observed in several biochemical processes, such as the fatty acid  $\beta$ -oxidation producing acetyl CoA, the non-oxidative stages of the pentose phosphate pathway in the cytoplasm, and the synthesis of diacetyl and acetoin in microbes. Previous reports have speculated that acetyl CoA could be a source of the acetyl group in 2AP<sup>15, 20</sup>. However, the biochemical role of acetyl CoA is usually associated with  $\beta$ -keto thioester formation (eg. fatty acid biosynthesis) rather than  $\alpha$ -imino ketone or 1,2-diketone formation, as is required in 2AP biosynthesis, and therefore acetyl CoA is not necessarily the source of the two carbon units of 2AP. Previous precursor studies implicated acetaldehyde (from ethanol<sup>19</sup>) or pyruvate<sup>35</sup> as the most likely donor for the acetyl group of 2AP.

Mechanistic considerations suggest that the pathway for the transfer of two carbon units to 4-aminobutanal could be similar to the way acetoin and diacetyl are biosynthesised<sup>36, 37</sup>. The production of these from pyruvate is initiated by the pyruvate dehydrogenase complex or pyruvate decarboxylase<sup>38, 39</sup> utilising a cofactor – thiamine diphosphate, formerly called thiamine pyrophosphate (TPP). In these reactions, the thiazolium ring of TPP is involved in the transfer of two carbon units via an hydroxyethyl-TPP enzyme complex<sup>40, 41</sup> that corresponds to an “activated acetaldehyde” unit. This process has been shown to occur in developing pea seeds<sup>36</sup> and wheat grains<sup>37</sup>. The enzymes required to synthesise acetoin are expressed in developing rice grains<sup>42, 43</sup>, and acetoin is one of the compounds found in grains in the present study (Table S1). Therefore, following the mechanistic action of the TPP-catalysed formation of diacetyl and acetoin, 2AP synthesis in rice could be occurring as follows (Fig. 5A). The “active acetaldehyde” could react with the carbonyl of 4-aminobutanal. Reaction of the aldehyde produces a hydroxyethyl intermediate, which would rapidly oxidise to the diketone 6-amino-2,3-hexanedione. The primary amine of the dione would then attack either of the carbonyl functions of the diketone moiety to produce either 2AP, when the ring closes at C3, or if the ring closes at C2, 6M5OTP would be formed. Further studies would be useful to replicate previous attempts to synthesise 2AP from 1-pyrroline<sup>21, 26</sup> in the presence of TPP.

Pyruvate is produced by glycolysis in the cytoplasm. It is then transported into mitochondria and oxidised by the pyruvate dehydrogenase complex, forming acetyl CoA<sup>44</sup>. It can also be decarboxylated by pyruvate decarboxylase (PDC) in the cytoplasm<sup>45</sup>. Both enzymes require TPP, which is ubiquitous<sup>46</sup>. If pyruvate is decarboxylated to form active acetaldehyde, it must be proximal to 4-aminobutanal or 1-pyrroline in order to react and form 2AP. In non-fragrant rice, the GABA produced from 4-aminobutanal exits the peroxisome<sup>34</sup> through a channel-forming protein that allows small molecules, below 400 Da, to cross the peroxisomal membrane passively<sup>33, 47</sup>. In fragrant rice, the channel-forming protein may allow 4-aminobutanal or 1-pyrroline to exit the peroxisome into the cytosol in the same way as other small molecules, where a reaction with activated acetaldehyde from PDC could lead to the production of 2AP and 6M5OTP (Fig. 5B). At physiological pH, 4-aminobutanal is a minor component (<5%) of a complex equilibrium, which contains several other chemical species, including a hydrate, a hemiaminal and the cyclic 1-pyrroline<sup>48</sup>. While acetylation of 1-pyrroline could, in theory, lead directly to 2AP, the formation of 6M5OTP under physiological conditions requires the presence of the linear dicarbonyl, suggesting formation from the aldehyde. Future studies to confirm the movement of GABA, 4-aminobutanal or 1-pyrroline from the peroxisome to the cytoplasm are underway.

Minor effect QTLs were detected for the presence of 2AP on chromosome 1 (Fig. 3C). Candidate genes within this region are *TPK1*, *G3PAT* and *AT*. The two latter genes, however, are involved in the transfer of long chain acyl groups (greater than C10<sup>49</sup>), therefore would be less likely to be involved in the production of 2AP. *TPK1* encodes thiamine pyrophosphate kinase, responsible for catalysing the conversion of thiamine to TPP, and thus is a more probable candidate gene in terms of function, and is consistent with the likely pathway of acetylation (Fig. 5B). The presence of regulatory motifs in the upstream or downstream regions of this gene may also contribute to different levels of fragrance in rice, a hypothesis that could be explored by further sequencing and molecular assays.



**Figure 5.** Putative biosynthesis of 2AP. Enzymes are in italics: CuAO – copper amine oxidase; PAO – polyamine oxidase; AADH – amino aldehyde dehydrogenase; PDC – Pyruvate decarboxylase; PDH – pyruvate dehydrogenase complex. **(A)** TPP-catalysed transfer of a C2 unit to 4-aminobutanal. The reaction with 4-aminobutanal forms an unstable diketone 6-amino-2,3-hexanedione, which cyclises to either 2AP or 6M5OTP. **(B)** Cellular localisation of the reactions leading to the formation of 2AP and 6M5OTP.

Nonetheless, new insights into the genetic and biochemical route of 2AP production would enable breeders to select more accurately for highly fragrant rice.



Variety	Germplasm Class	Origin	Phenotype	FGR genotype
Khao Khor 6 (RD 6)	Indica	Thailand	Fragrant	<i>fgr</i>
Hom Mali Niaw (HMN)*	Indica	Thailand	Fragrant	<i>fgr</i>
Khao Dawk Mali 105 (KDML105)	Indica	Thailand	Fragrant	<i>fgr</i>
Hom Nang Neuang (HNN)	Indica	Laos	Fragrant	<i>fgr</i>
Phka Rum Duol (PRD)	Indica	Cambodia	Fragrant	<i>fgr</i>
IR64	Indica	Philippines	Non-fragrant	<i>Fgr</i>
Doongara	Temperate Japonica	Australia	Non-fragrant	<i>Fgr</i>
Kyeema*	Temperate Japonica	Australia	Fragrant	<i>fgr</i>
Pandan Wangi 7	Tropical Japonica	Indonesia	Fragrant	<i>fgr</i>

**Table 2.** Elite rice varieties from Southeast Asia and Australia used in this study (Set 1). \*Two samples from different plots have been used in Set 1.

## Methodology

**Rice materials.** The first set of samples (Set 1) consisted of seven aromatic and two non-aromatic varieties of rice (Table 2). The samples were sown in Mackay Queensland, in the summer season from December 2013, and harvested in April 2014. A second set of samples (Set 2) consisted of an association mapping population comprised of multiple families derived from crosses between the varieties in Set 1 and elite Australian long grain varieties, giving a diverse collection of 380 genotypes (Table S1). These were grown as F7 RILs at the Leeton Field Station (Yanco, NSW), the Australian Rice Breeding Program between November 2013 and April 2014. Field management, irrigation, pesticide and fertiliser treatments all followed typical commercial practices. Single plants were tagged, a single leaf from each was preserved for DNA extraction and genotyping, and the grain from two panicles was collected for metabolomics profiling. Out of the 380 samples, 39 did not produce enough grains for metabolomic profiling, or had unfilled panicles due to photoperiod and/or cold sensitivity, giving a final number of 341 samples. All panicles were dried to 14% moisture and then the grain was shipped to The University of Queensland as paddy to minimise any possible release of volatile compounds or potential oxidation of lipids by damaging the bran. The paddy rices were dehulled using an Otake dehuller (Model FC2K, Aichi, Japan), and polished with a Kett Mill (Pearlest Grain Polisher, Tokyo, Japan). The white rices were cryoground to fine powder using a Geno Grinder (2010; Stanmore, UK). Rice flour was stored at  $-80^{\circ}\text{C}$  until analysis.

**Metabolite profiling.** *Comprehensive GC  $\times$  GC-TOF-MS.* Volatile compounds in all of the rice samples (Sets 1 and 2) were measured on one gram of rice flour by static headspace extraction using GC  $\times$  GC-TOF-MS (Pegasus 4D, LECO Corporation; St. Joseph, MI, USA) following previously described protocols<sup>22</sup>. The full details of the metabolite profiling including GC  $\times$  GC program settings, MS conditions and chromatogram alignment parameters are given in Table S2 and Supplementary Information S2. Quality control standards (commercial rice flour) were injected every 10 samples to monitor the performance of the instrument and to check for batch effects. Data preprocessing, baseline correction and identification of putative compounds were done using the software LECO ChromaTOF 4.50 with library match searching to Nist 11 v2.0. Commercial standards of several heterocycles, alcohols, ketones and aldehydes were either purchased or synthesized and used for retention time, identification of compounds, and confirmation of library annotation. All samples were analysed in two technical replicates.

*High Resolution GC  $\times$  GC-TOF-MS (GC  $\times$  GC-HRT 4D).* Five rice samples exhibiting high levels of 2AP were analysed with a Pegasus HRT 4D at LECO Life Science and Chemical Analysis Centre Michigan, USA. The acquisition parameters were identical to those reported above. The mass spectrometer was set to electron ionisation (EI) mode to enable investigation of the fragmentation pattern of the compounds and allow comparison against the NIST library. Samples were further analysed using chemical ionisation (CI) detection to preserve the molecular ion and to identify the accurate mass of the compound. The acquisition mode was at high resolution  $R = 25,000$  full width at half maximum (FWHM). The mass spectra were recorded at a mass range of 35–400  $m/z$  and an acquisition rate of 200 spectra/s.

**Synthesis of authentic standards.** In order to understand further the relationship between 2AP and analytes that associated with it in rice samples, 2AP and 1-pyrroline were both synthesised from proline<sup>26, 50, 51</sup>. For 2AP, the carbomethoxy substituent of methyl proline was converted into an acetyl group by (i) *N*-chlorination followed by elimination of HCl using  $\text{Et}_3\text{N}$ ; and (ii) addition of  $\text{MeMgBr}$  in anhydrous  $\text{Et}_2\text{O}$  at  $0^{\circ}\text{C}$  followed by reaction workup using aqueous  $\text{NH}_4\text{Cl}$ ; (iii) purification by flash chromatography on base-washed silica using diethyl ether: pentane (1:3) as eluent. For 1-pyrroline, proline was oxidised with sodium metaperiodate and the resulting product purified by alumina chromatography using *n*-pentane:diethyl ether (8:2) as eluent. The full details of each synthesis are given in Supplementary Information S1.

The behaviour of the synthesised 2AP upon exposure to acid-base conditions was examined by both GC-MS (Shimadzu GC-MS-QP2010) and  $^1\text{H}$  NMR (Bruker Avance DRX 700 spectrometer). A synthetic sample of 2AP was subjected to two rounds of acid-base partitioning by extraction into 1M HCl, followed by basification to pH 10, and extraction into chloroform. The accompanying GC-MS and  $^1\text{H}$  NMR data are given in Supplementary Information S2.

**Genotyping.** *Genotyping-By-Sequencing (GBS).* Total genomic DNA was extracted from freeze dried leaf materials using Qiagen 96-plex DNEasy Plant DNA Extraction kit (Qiagen Hilden, Germany) following the manufacturer's protocol. The DNA was quantified to 50–100 ng/ul using a Qubit 2.0 fluorometer (Invitrogen, USA). DNA quality was tested by digesting 500 ng DNA with 5 units of methylation insensitive restriction enzyme *HindIII*. The digested and uncut DNA samples were run on 0.8–1% agarose gel using  $\lambda$  *HindIII* as size standard. High-quality DNA samples that showed complete digestion on the gel were submitted for GBS.

The 384-plex GBS libraries were prepared at the BioResource Center at Cornell University (Ithaca, NY) exactly as previously described<sup>52</sup>. The samples were digested with the restriction enzyme *ApeKI*. DNA fragments were barcoded and sequenced using Illumina HiSeq. 2500 (Illumina, San Diego, CA). Contamination was checked using a blank sample randomly placed in the sequencing plate (one per plate); blanks should have less than 10% of the mean barcode reads value (MBRV). Sample Y20140298 has less than 10% of the MBRV and was removed from the data set.

The raw GBS reads were analysed using the default parameter of the TASSEL pipeline version 5.0<sup>52–54</sup> (<http://www.maizegenetics.net>). Tags were aligned to the reference genome (Nipponbarre, MSU version 7.0) using the Burrows-Wheeler Aligner version 0.7.8-r455 (BWA)<sup>55</sup>. SNPs were identified from the Tags using the plugin TagsToSNPByAlignment and identical SNPs were merged using the MergeDuplicateSNPs plugin<sup>54</sup>. Imputation of missing data was performed with the FastImputationBitFixedWindow plugin using the default settings as of Begum<sup>56</sup>. SNPs missing in more than 25% of the data set were filtered along with SNPs with minor allele frequency (MAF) of 5%, with a resulting data matrix of 33,725 SNPs.

*Genome-wide Association Study.* Association mapping was performed in TASSEL 5.0<sup>54</sup> using the mixed-linear model (MLM) with the  $\log_{10}$ -transformed peak area of the target metabolites as the 'trait'. Corrections for confounding effects due to subpopulation structure and relatedness between individuals were done by performing PCA<sup>57</sup> and calculating the kinship matrix (k)<sup>58</sup>, using the (PCA + K) workflow in TASSEL 5.0<sup>54</sup>. The first five components of the PCA were used as covariates in the MLM and the kinship matrix was generated using the Centered\_IBS method<sup>59</sup>. Manhattan plots and qq plots were also generated in TASSEL 5.0. Significance thresholds were set at SNPs with LOD scores  $\geq 5$ , where there were also at least five other SNPs with LOD scores  $\geq 4$  within 200kb from the MS SNP. LD in the QTL on chromosome 1 was estimated by standardised disequilibrium coefficients ( $D'$ )<sup>60</sup>. Plots were generated using TASSEL.

*Genotyping of FGR locus.* Several mutations that lead to a non-functional AADH have been identified in rice on different genetic backgrounds<sup>2, 8, 61, 62</sup>. In this study, we screen for the major allele the recessive *fgr* allele, conferring fragrance, which is characterised by a 7 bp deletion and three SNPs in the seventh exon of the *FGR* gene<sup>8</sup>. We utilised previously developed markers<sup>29</sup> to genotype Set 2 collection for this allele. This data was added to the 33,275 SNP matrix and GWAS was rerun for target compounds.

**Statistical analyses.** After pre-processing the metabolomics data, the mean values of the two technical replicates were used for further statistical analysis and GWA. For multivariate analysis, the data were log transformed to reduce the skewness of the variables<sup>63</sup>. The data was then pareto scaled to moderate the influence of compounds produced at high concentrations and recognise the effects of compounds produced at relatively lower concentrations in the overall characterisation of rice volatile compounds<sup>64</sup>. To obtain an overview of the data sets, the data were represented by unsupervised modelling through principal components analysis (PCA) using Simca-P. Next, to determine the discriminating compounds in the aromatic rice varieties used in this study (Set 1), orthogonal partial least squares-discriminant analysis (oPLS-DA) was conducted, again in Simca-P. OPLS-DA maximises the covariance of the matrix of predictor data, X, and the response block, Y, thus providing a more powerful prediction of among-group variation compared to PCA<sup>65</sup>. Compounds with the top five VIP scores were examined in greater detail. Finally, correlation analysis was performed for the top five discriminating compounds of both sample sets using the "cor" function in R<sup>66</sup>. To compare the relative amounts of the five amine heterocycles between the two sample groups (Set 1 n = 11; and Set 2 n = 341), a full model, accounting for sample group and compound, was fit to the data using the linear mixed effects 'lme4' R package<sup>67</sup>. In the usual fashion, a second, nested model, which omitted sample group as predictor variable, was fit, and the two models were compared.

**Data availability.** Data files are accessible through the UQ espace digital repository <http://dx.doi.org/10.14264/uql.2017.174>.

## References

- Calingacion, M. *et al.* Diversity of Global Rice Markets and the Science Required for Consumer-Targeted Rice Breeding. *PLoS ONE* **9**, e85106, doi:10.1371/journal.pone.0085106 (2014).
- Kovach, M. J., Calingacion, M. N., Fitzgerald, M. A. & McCouch, S. R. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences USA* **106**, 14444–14449, doi:10.1073/pnas.0904077106 (2009).
- Fitzgerald, M. A., Sackville-Hamilton, N. R., Calingacion, M. N., Verhoeven, H. A. & Butardo, V. Jr. Is there a second gene for fragrance in rice? *Plant Biotechnology Journal* **6**, 416–423, doi:10.1111/j.1467-7652.2008.00327.x (2008).
- Wakte, K. *et al.* Thirty-three years of 2-acetyl-1-pyrroline, a principal basmati aroma compound in scented rice (*Oryza sativa* L.): a status review. *Journal of the Science of Food and Agriculture*. doi:10.1002/jsfa.7875 (2016).
- Schiller, J. M. *Rice in Laos* (International Rice Research Institute, 2006).
- Buttery, R. G., Turnbaugh, J. G. & Ling, L. C. Contribution of volatiles to rice aroma. *Journal of Agricultural and Food Chemistry* **36**, 1006–1009, doi:10.1021/jf00083a025 (1988).
- Blank, I., Devaud, S., Matthey-Doret, W. & Robert, F. Formation of Odorants in Maillard Model Systems Based on l-Proline as Affected by pH. *Journal of Agricultural and Food Chemistry* **51**, 3643–3650, doi:10.1021/jf034077t (2003).
- Bradbury, L. M. T., Fitzgerald, T. L., Henry, R. J., Jin, Q. & Waters, D. L. E. The gene for fragrance in rice. *Plant Biotechnology Journal* **3**, 363–370, doi:10.1111/j.1467-7652.2005.00131.x (2005).

9. Vanavichit, A., Tragoonrun, S., Toojinda, T., Wanchana, S. & Kamolsukyunyong, W. Transgenic rice plants with reduced expression of Os2AP and elevated levels of 2-acetyl-1-pyrroline. *United States of America patent US7319181 B2* (2008).
10. Chen, S. *et al.* *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. *The Plant Cell* **20**, 1850–1861, doi:10.1105/tpc.108.058917 (2008).
11. Wakte, K. V., Kad, T. D., Zanan, R. L. & Nadaf, A. B. Mechanism of 2-acetyl-1-pyrroline biosynthesis in *Bassia latifolia* Roxb. flowers. *Physiology and molecular biology of plants* **17**, 231–237, doi:10.1007/s12298-011-0075-5 (2011).
12. Arikrit, S. *et al.* Deficiency in the amino aldehyde dehydrogenase encoded by GmAMADH2, the homologue of rice Os2AP, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (*Glycine max* L.). *Plant Biotechnology Journal* **9**, 75–87, doi:10.1111/j.1467-7652.2010.00533.x (2011).
13. Pramnoi, P., Somta, P., Chankaew, S., Juwattanasomran, R. & Srinives, P. A single recessive gene controls fragrance in cucumber (*Cucumis sativus* L.). *Journal of Genetics* **92**, 147–149 (2013).
14. Fogel, W. A., Biegański, T. & Maśliński, C. Effects of inhibitors of aldehyde metabolizing enzymes on putrescine metabolism in guinea pig liver homogenates. *Agents and Actions* **9**, 42–44, doi:10.1007/bf02024104 (1979).
15. Bradbury, L., Gillies, S., Brushett, D., Waters, D. & Henry, R. Inactivation of an aminoaldehyde dehydrogenase is responsible for fragrance in rice. *Plant Molecular Biology* **68**, 439–449, doi:10.1007/s11103-008-9381-x (2008).
16. Kinnersley, A. M. & Turano, F. J. G. A. Acid (GABA) and Plant Responses to Stress. *Critical Reviews in Plant Sciences* **19**, 479–509, doi:10.1080/07352680091139277 (2000).
17. Huang, T.-C. *et al.* Biosynthetic Mechanism of 2-Acetyl-1-pyrroline and Its Relationship with  $\Delta^1$ -Pyrroline-5-carboxylic Acid and Methylglyoxal in Aromatic Rice (*Oryza sativa* L.) Callus. *Journal of Agricultural and Food Chemistry* **56**, 7399–7404, doi:10.1021/jf8011739 (2008).
18. Hinge, V. R., Patil, H. B. & Nadaf, A. B. Aroma volatile analyses and 2AP characterization at various developmental stages in Basmati and Non-Basmati scented rice (*Oryza sativa* L.) cultivars. *Rice* **9**, 1–22, doi:10.1186/s12284-016-0113-6 (2016).
19. Costello, P. J. & Henschke, P. A. Mousy Off-Flavor of Wine: Precursors and Biosynthesis of the Causative N-Heterocycles 2-Ethyltetrahydropyridine, 2-Acetyltetrahydropyridine, and 2-Acetyl-1-pyrroline by *Lactobacillus hilgardii* DSM 20176. *Journal of Agricultural and Food Chemistry* **50**, 7079–7087, doi:10.1021/jf020341r (2002).
20. Adams, A. & De Kimpe, N. Chemistry of 2-Acetyl-1-pyrroline, 6-Acetyl-1,2,3,4-tetrahydropyridine, 2-Acetyl-2-thiazoline, and 5-Acetyl-2,3-dihydro-4H-thiazine: Extraordinary Maillard Flavor Compounds. *Chemical Reviews* **106**, 2299–2319, doi:10.1021/cr040097y (2006).
21. Poonlaphdecha, J. *et al.* Biosynthesis of 2-acetyl-1-pyrroline in rice calli cultures: Demonstration of 1-pyrroline as a limiting substrate. *Food Chemistry* **197**, Part A, 965–971, doi:10.1016/j.foodchem.2015.11.060 (2016).
22. Daygon, V. D. *et al.* Understanding the Jasmine phenotype of rice through metabolite profiling and sensory evaluation. *Metabolomics* **12**, 1–15, doi:10.1007/s11306-016-0989-6 (2016).
23. Grimm, C. C., Champagne, E. T., Lloyd, S. W., Easson, M. & Condon, B. Analysis of 2-Acetyl-1-Pyrroline in Rice by HSSE/GC/MS. *Cereal Chemistry* **88**, 271–277, doi:10.1094/cchem-09-10-0136 (2011).
24. Yoshihashi, T. Quantitative analysis on 2-acetyl-1-pyrroline of an aromatic rice by stable isotope dilution method and model studies on its formation during cooking. *Journal of Food Science* **67**, 619–622, doi:10.1111/j.1365-2621.2002.tb10648.x (2002).
25. Ottinger, H. & Hofmann, T. Quantitative model studies on the efficiency of precursors in the formation of cooling-active 1-pyrrolidinyl-2-cyclopenten-1-ones and bitter-tasting cyclopenta-[b]azepin-8(1H)-ones. *Journal of Agricultural and Food Chemistry* **50**, 5156–5161 (2002).
26. Schieberle, P. Quantitation of Important Roast-Smelling Odorants in Popcorn by Stable Isotope Dilution Assays and Model Studies on Flavor Formation during Popping. *Journal of Agricultural and Food Chemistry* **43**, 2442–2448, doi:10.1021/jf00057a024 (1995).
27. De Kimpe, N. & Keppens, M. Novel Syntheses of the Major Flavor Components of Bread and Cooked Rice. *Journal of Agricultural and Food Chemistry* **44**, 1515–1519, doi:10.1021/jf950728c (1996).
28. Harrison, T. J. & Dake, G. R. A. Expeditious, High-Yielding Construction of the Food Aroma Compounds 6-Acetyl-1,2,3,4-tetrahydropyridine and 2-Acetyl-1-pyrroline. *The Journal of Organic Chemistry* **70**, 10872–10874, doi:10.1021/jo051940a (2005).
29. Bradbury, L. M. T., Henry, R. J., Jin, Q., Reinke, R. F. & Waters, D. L. E. A perfect marker for fragrance genotyping in rice. *Molecular Breeding* **16**, 279–283 (2005).
30. Belitz, H. D., Grosch, W. & Schieberle, P. *Food Chemistry* Vol. 4th rev. and expand; H.-D. Belitz, W. Grosch, P. Schieberle: [translat from the German by M.M. Burghagen].; 4. rev. and extend; 4 (Springer, 2009).
31. Fuhlhage, D. W. & VanderWerf, C. A. Studies on the Formation and Reactions of 1-Pyrroline. *Journal of the American Chemical Society* **80**, 6249–6254, doi:10.1021/ja01556a022 (1958).
32. Huang, T.-C., Huang, Y.-W., Hung, H.-J., Ho, C.-T. & Wu, M.-L.  $\delta^1$ -Pyrroline-5-carboxylic Acid Formed by Proline Dehydrogenase from the *Bacillus subtilis* ssp. natto Expressed in *Escherichia coli* as a Precursor for 2-Acetyl-1-pyrroline. *Journal of Agricultural and Food Chemistry* **55**, 5097–5102, doi:10.1021/jf0700576 (2007).
33. Baker, A. *et al.* Peroxisomal ABC transporters: functions and mechanism. *Biochemical Society Transactions* **43**, 959–965, doi:10.1042/BST20150127 (2015).
34. Zarei, A., Trobacher, C. P. & Shelp, B. J. Arabidopsis aldehyde dehydrogenase 10 family members confer salt tolerance through putrescine-derived 4-aminobutyrate (GABA) production. *Scientific Reports* **6**, 35115, doi:10.1038/srep35115 (2016).
35. Wongpia, A. *et al.* Proteomic Analysis of Isogenic Rice Reveals Proteins Correlated with Aroma Compound Biosynthesis at Different Developmental Stages. *Molecular Biotechnology* **58**, 117–129, doi:10.1007/s12033-015-9906-x (2016).
36. Davies, M. E. Acetolactate and Acetoin Synthesis in Ripening Peas. *Plant Physiology* **39**, 53–59 (1964).
37. Singer, T. P. & Pensky, J. Mechanism of acetoin synthesis by  $\alpha$ -carboxylase. *Biochimica et Biophysica Acta* **9**, 316–327, doi:10.1016/0006-3002(52)90167-4 (1952).
38. Speckman, R. A. & Collins, E. B. Diacetyl biosynthesis in *Streptococcus diacetylactis* and *Leuconostoc citrovorum*. *Journal of bacteriology* **95**, 174–180 (1968).
39. Baggetto, L. G. & Lehninger, A. L. Formation and utilization of acetoin, an unusual product of pyruvate metabolism by Ehrlich and AS30-D tumor mitochondria. *Journal of Biological Chemistry* **262**, 9535–9541 (1987).
40. Jordan, F. Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions. *Natural Product Reports* **20**, 184–201, doi:10.1039/B111348H (2003).
41. Frey, P. A. & Hegeman, A. D. *Enzymatic reaction mechanisms* (Oxford University Press, 2007).
42. Yamakawa, H. & Hakata, M. Atlas of Rice Grain Filling-Related Metabolism under High Temperature: Joint Analysis of Metabolome and Transcriptome Demonstrated Inhibition of Starch Accumulation and Induction of Amino Acid Accumulation. *Plant and Cell Physiology* **51**, 795–809, doi:10.1093/pcp/pcq034 (2010).
43. Xu, S. B. *et al.* Dynamic Proteomic Analysis Reveals a Switch between Central Carbon Metabolism and Alcoholic Fermentation in Rice Filling Grains. *Plant Physiology* **148**, 908–925, doi:10.1104/pp.108.125633 (2008).
44. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry* 5th edn (W. H. Freeman and CO., 2002).
45. Karl, T., Curtis, A. J., Rosenstiel, T. N., Monson, R. K. & Fall, R. Transient releases of acetaldehyde from tree leaves – products of a pyruvate overflow mechanism? *Plant, Cell & Environment* **25**, 1121–1131, doi:10.1046/j.1365-3040.2002.00889.x (2002).
46. Fraccascia, P., Sniekers, M., Casteels, M. & Van Veldhoven, P. P. Presence of thiamine pyrophosphate in mammalian peroxisomes. *BMC Biochemistry* **8**, 10, doi:10.1186/1471-2091-8-10 (2007).

47. Antonenkov, V. D. & Hiltunen, J. K. Peroxisomal membrane permeability and solute transfer. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1763**, 1697–1706, doi:10.1016/j.bbamcr.2006.08.044 (2006).
48. Struve, C. & Christophersen, C. Structural equilibrium and ring-chain tautomerism of aqueous solutions of 4-aminobutyraldehyde. *Heterocycles* **60**, 1907–1914 (2003).
49. Zhu, S. Q., Zhao, H., Zhou, R., Ji, B. H. & Dan, X. Y. Substrate selectivity of glycerol-3-phosphate acyl transferase in rice. *Journal of Integrative Plant Biology* **51**, 1040–1049, doi:10.1111/j.1744-7909.2009.00876.x (2009).
50. Srinivas, P. & Gurudutt, K. N. Process for the preparation of 2-acetyl-1-pyrroline, the basmati rice flavorant. *USA patent* (2004).
51. De Kimpe, N. G., Stevens, C. V. & Keppens, M. A. Synthesis of 2-acetyl-1-pyrroline, the principal rice flavor component. *Journal of Agricultural and Food Chemistry* **41**, 1458–1461, doi:10.1021/jf00033a020 (1993).
52. Elshire, R. J. *et al.* A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **6**, e19379, doi:10.1371/journal.pone.0019379 (2011).
53. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635, doi:10.1093/bioinformatics/btm308 (2007).
54. Glaubitz, J. C. *et al.* TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **9**, e90346, doi:10.1371/journal.pone.0090346 (2014).
55. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595, doi:10.1093/bioinformatics/btp698 (2010).
56. Begum, H. *et al.* Genome-Wide Association Mapping for Yield and Other Agronomic Traits in an Elite Breeding Population of Tropical Rice (*Oryza sativa*). *PLoS ONE* **10**, e0119873, doi:10.1371/journal.pone.0119873 (2015).
57. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006). [http://www.nature.com/ng/journal/v38/n8/supinfo/ng1847\\_S1.html](http://www.nature.com/ng/journal/v38/n8/supinfo/ng1847_S1.html)
58. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821–824, doi:10.1038/ng.2310 (2012).
59. Müller, D., Technow, F. & Melchinger, A. E. Shrinkage estimation of the genomic relationship matrix can improve genomic estimated breeding values in the training set. *Theoretical and Applied Genetics* **128**, 693–703, doi:10.1007/s00122-015-2464-6 (2015).
60. Lewontin, R. C. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* **49**, 49 (1964).
61. Shao, G. N. *et al.* A new deletion mutation of fragrant gene and the development of three molecular markers for fragrance in rice. *Plant Breeding* **130**, 172–176 (2011).
62. Shi, W., Yang, Y., Chen, S. & Xu, M. Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. *Molecular Breeding*. doi:10.1007/s11032-008-9165-7 (2008).
63. Van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142–142, doi:10.1186/1471-2164-7-142 (2006).
64. Allwood, J. W. *et al.* In *Methods in Enzymology* Vol. 500 (eds Daniel Jameson, Malkhey Verma, & V. Westerhoff Hans) 299–336 (Academic Press, 2011).
65. Worley, B., Halouska, S. & Powers, R. Utilities for quantifying separation in PCA/PLS-DA scores plots. *Analytical Biochemistry* **433**, 102–104, doi:10.1016/j.ab.2012.10.011 (2013).
66. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2015).
67. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48, doi:10.18637/jss.v067.i01 (2015).

## Acknowledgements

The authors would like to acknowledge Andrew Barfield for growing the samples in Mackay, Queensland; Arthur Riedel for the inputs in statistical analysis; Alan Jones for the technical support; Brett McVittie for genotyping the *FGR* locus; and Peter Gorst-Allman for the assistance in metabolite profiling by GC × GC HRT 4D. VD gratefully acknowledges financial support from the Rural Industries Research and Development Corporation. This study was funded by the Rural Industries Research and Development Corporation Project PRJ008568 and the Industry Transformation Training Centre Australian Research Council Project IC130100011.

## Author Contributions

M.A.F. designed the study. V.D.D. performed the untargeted metabolite profiling and data analysis. D.E.A. performed the GC × GC HRT 4D profiling. B.O. developed and grew the mapping population. V.D.D. and S.R.M. did the genotyping and association mapping. L.C.F., J.J.D.V., B.D.S. and M.J.G. synthesised the authentic chemical standards. L.C.F., B.S. and M.J.G. analysed the chemical standards by GC-MS and NMR. M.J.G., J.J.D.V. and M.A.F. formulated the pathway of 2AP biosynthesis. V.D.D., M.J.G. and M.A.F. wrote the paper. The paper was edited by all.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-07693-9

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017