

Research article

Open Access

Correlation exploration of metabolic and genomic diversity in rice

Keiichi Mochida*¹, Taku Furuta², Kaworu Ebana³, Kazuo Shinozaki¹ and Jun Kikuchi*^{1,2,4}

Address: ¹RIKEN Plant Science Center, 1-7-22, Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, ²Graduate School of BionanoSciences, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, ³National Institute of Agrobiological Sciences, 2-1-2, Kannondai, Tsukuba, 305-8602 Japan and ⁴Graduate School of Bioagriculture Sciences, Nagoya University, 1-1 Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Email: Keiichi Mochida* - mochida@psc.riken.jp; Taku Furuta - lyrebird@tsurumi.yokohama-cu.ac.jp; Kaworu Ebana - ebana@aafc.go.jp; Kazuo Shinozaki - sinozaki@psc.riken.jp; Jun Kikuchi* - kikuchi@psc.riken.jp

* Corresponding authors

Published: 1 December 2009

Received: 25 May 2009

BMC Genomics 2009, 10:568 doi:10.1186/1471-2164-10-568

Accepted: 1 December 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/568>

© 2009 Mochida et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: It is essential to elucidate the relationship between metabolic and genomic diversity to understand the genetic regulatory networks associated with the changing metabolo-phenotype among natural variation and/or populations. Recent innovations in metabolomics technologies allow us to grasp the comprehensive features of the metabolome. Metabolite quantitative trait analysis is a key approach for the identification of genetic loci involved in metabolite variation using segregated populations. Although several attempts have been made to find correlative relationships between genetic and metabolic diversity among natural populations in various organisms, it is still unclear whether it is possible to discover such correlations between each metabolite and the polymorphisms found at each chromosomal location. To assess the correlative relationship between the metabolic and genomic diversity found in rice accessions, we compared the distance matrices for these two "omics" patterns in the rice accessions.

Results: We selected 18 accessions from the world rice collection based on their population structure. To determine the genomic diversity of the rice genome, we genotyped 128 restriction fragment length polymorphism (RFLP) markers to calculate the genetic distance among the accessions. To identify the variations in the metabolic fingerprint, a soluble extract from the seed grain of each accession was analyzed with one dimensional ¹H-nuclear magnetic resonance (NMR). We found no correlation between global metabolic diversity and the phylogenetic relationships among the rice accessions ($r_s = 0.14$) by analyzing the distance matrices (calculated from the pattern of the metabolic fingerprint in the 4.29- to 0.71-ppm ¹H chemical shift) and the genetic distance on the basis of the RFLP markers. However, local correlation analysis between the distance matrices (derived from each 0.04-ppm integral region of the ¹H chemical shift) against genetic distance matrices (derived from sets of 3 adjacent markers along each chromosome), generated clear correlations ($r_s > 0.4$, $p < 0.001$) at 34 RFLP markers.

Conclusion: This combinatorial approach will be valuable for exploring the correlative relationships between metabolic and genomic diversity. It will facilitate the elucidation of complex regulatory networks and those of evolutionary significance in plant metabolic systems.

Background

It is essential to elucidate the relationships between metabolic and genomic diversity to understand the genetic causes of phenotypic variation among natural populations. Visible and chemical variations associated with genomic diversity provide the information required to identify key genes associated with such phenotypic changes [1]. Patterns of nucleotide polymorphisms and a population structure figured in a natural population have allowed us to understand the evolutionary significance of genetic variation under the influence of geographic factors or different environments [2,3].

The technical development of metabolite profiling has provided quantitative phenotypic information on the metabolome among strains and identified genes involved in metabolic networks [4]. Therefore, a comprehensive exploration of the correlations between metabolic and genomic diversity, achieved by superimposing data from these two "omics" approaches, could provide information regarding both the broad and specific relationships between metabolo-phenotypes and genotypes. This information would aid the identification of genetic associations between the metabolic and/or visible phenotypes [5,6].

Metabolite quantitative trait loci (mQTLs) analysis using segregated populations has been applied to various plant species as a popular forward genetics approach [4,5,7-9]. Although mQTLs analysis has been used to dissect the genetic loci involved in metabolo-phenotypic changes in mapping populations of various organisms, no methods have been developed to identify the correlations between genetic polymorphisms and metabolo-phenotypic differences in divergent accessions or individuals from natural populations. To date, several attempts have been performed to find correlative relationships between genetic and metabolic diversity in various organisms [10-13]. It is still unclear whether it is possible to discover such correlations between each metabolite and the polymorphic pattern found at each chromosomal location to explain natural variation. Combinatorial approaches of population genomics coupled with metabolo-phenotyping should play a significant role in the exploration of the association of genetic variation with metabolic changes that have a significant impact on evolution and adaptation [3].

In this study, to assess the correlative relationship between metabolic and genomic diversity found in 18 rice accessions selected from the world rice collection, we compared the distance relationships of accessions for these two omics patterns, globally, with marker polymorphisms vs. the metabolomic fingerprints of NMR spectra, as well as locally, for each chromosomal region vs. each

spectra. To perform these analyses, we developed a robust procedure to explore the relationship between metabolic and genomic diversity. As a result, we present the first correlation map of metabolic and genomic diversity among rice accessions.

Results and Discussion

To select the appropriate rice accessions for the comparison of metabolic profiles with patterns of genetic polymorphisms, we initially investigated the population structure of the world rice collection. The accessions of the world rice collection have been classified into three major groups, japonica, indica I, and indica II [14]. To determine in greater detail the genetic structure of these populations, we performed population structure analysis using Structure software on data generated by genotyping 128 restriction-fragment-length polymorphism (RFLP) markers. This analysis showed that the japonica type accessions were composed of 4 sub-groups; temperate, and tropical I, II, and III. For the two indica groups, the members of indica I were not divided, while those of indica II were divided into two sub-groups; indica IIa and indica IIb (Fig. 1). We chose 18 rice accessions from 6 representative sub-groups to cover a diverse range of cultivated rice populations, on the basis of population structure, to avoid regional bias (Fig. 1 and Additional file 1).

The RFLP genotyping data were also used to generate global and local genetic distance matrices along each chromosome in combination with physical allocation data on the rice genome (Additional file 2). Soluble extracts of seeds from each accession were analyzed by 1D ¹H-nuclear magnetic resonance (NMR) analysis to elucidate the metabolic diversity among the accessions (Fig. 2 and Additional file 3).

To assess the global correlation between metabolic and genomic diversity, we calculated Spearman's coefficient of correlation (r_s) between the distance matrices calculated from the 4.29- to 0.71-ppm ¹H chemical shifts and genetic distance among accessions with polymorphisms at the 128 RFLP markers. We found no correlation between global metabolic diversity and the phylogenetic relationships among the rice accessions ($r_s = 0.14$). Although the phylogenetic tree of the genetic distance matrix showed the expected 3 typical clusters (corresponding to japonica, indica I, and indica II; Fig. 3a), the metabolic diversity tree showed no identifiable clusters corresponding to the phylogenetic clusters (Fig. 3b). This result indicates that there are no relationships between the genomic and metabolic diversity of solution metabolites of rice grain in the comparison of holistic data. In higher plants, the diversity of the metabolome might tend to be unrelated with genomic diversity. In sesame, which has been an example for the comparison of metabolic and

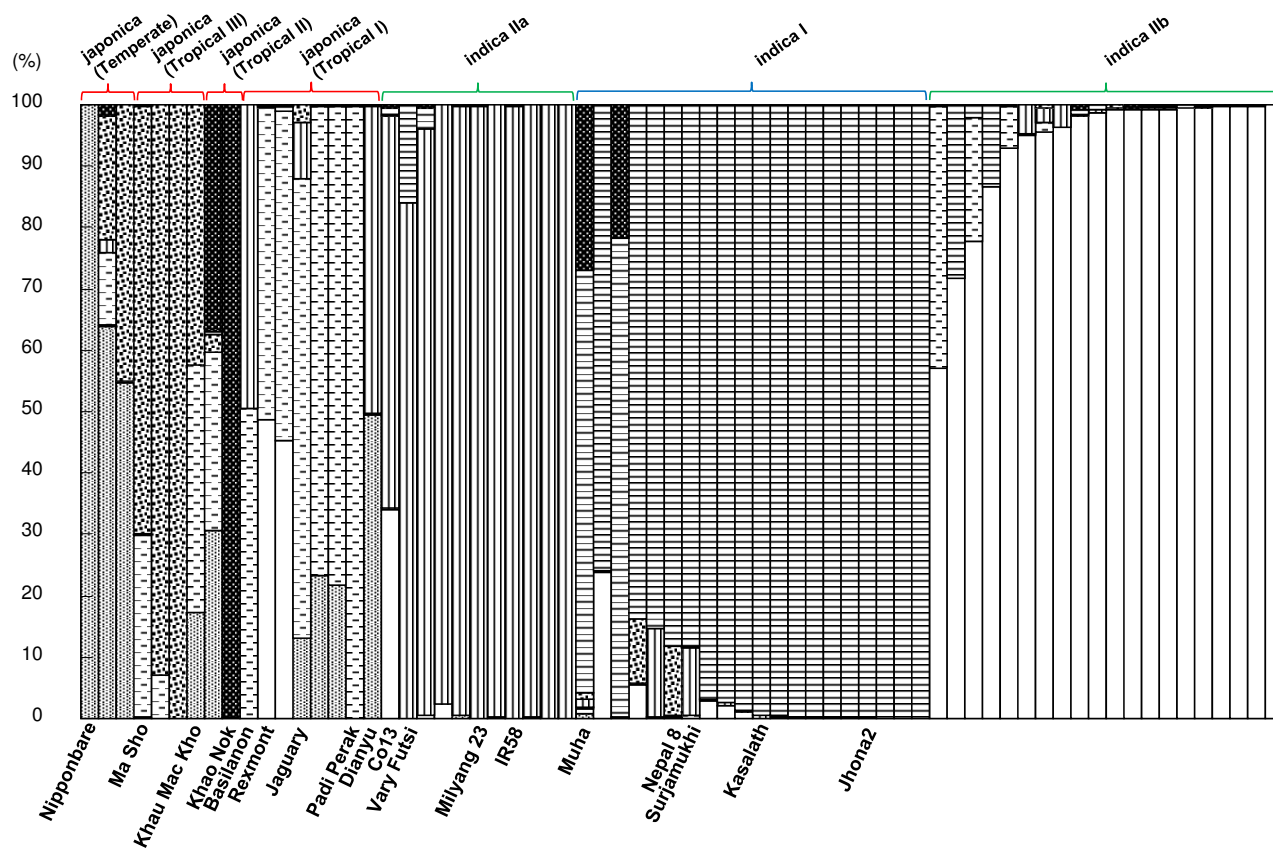


Figure 1
Population structure of 68 rice accessions from around the world (including the 18 used in this study) based on the genotype data of 128 RFLP markers determined with Structure software.

genomic diversity in higher plants, the relationship patterns generated for the AFLPs and the seed metabolic profiles also differed [10]. Seed metabolic profiles are derived from the transcriptome that is specifically activated in the embryo and endosperm during seed development. Genes involved in the abundance of solution metabolites of rice seed might consist of small members expressed in seed specific transcriptome. Although there have been few examples of the global assessment of correlations between metabolic and genomic diversity in higher plants, the recent updating of the infrastructure for population genomics and metabolomics should allow us to determine correlation patterns at different levels i.e., from genomic segments to small chromosomal regions in each organism along with their evolutionary patterns. In Arabidopsis, a large number of metabolic QTLs were identified in a recombinant inbred line (RIL) of two parental ecotypes, which had been selected from 14 ecotypes based on their metabolic profiles [15]. Most of the metabolic variation observed among the ecotypes was successfully mapped using this RIL population, indicating that meta-

bolic variation can be associated with simplified polymorphic patterns at specific chromosomal regions.

We then calculated the r_s of Euclidian distance matrices derived from each 0.04-ppm integral region of the ^1H chemical shift against the genetic distance matrices derived from sets of three adjacent markers along each chromosome (Additional file 4). The results of these local analyses are presented as a 2D correlation map (Fig. 4). To assess the statistical significance of each correlation coefficient value in this dataset, we calculated the p-value for each r_s value and determined the relationship between the r_s value and the p-value for all combinations. As a result of this analysis, an r_s value of 0.4 was supported with a p-value of $2.99\text{E-}07$; the value of r_s at the 0.1% level of the p-value was between ≥ 0.3 and ≤ 0.4 . We also calculated the correlation coefficient values between each genetic distance matrix of three randomly sampled RFLP markers and the metabolic profile of each of the 104 NMR spectra in 10,000 trials to estimate the probability of a high correlation by random chance (Additional file 5). The results

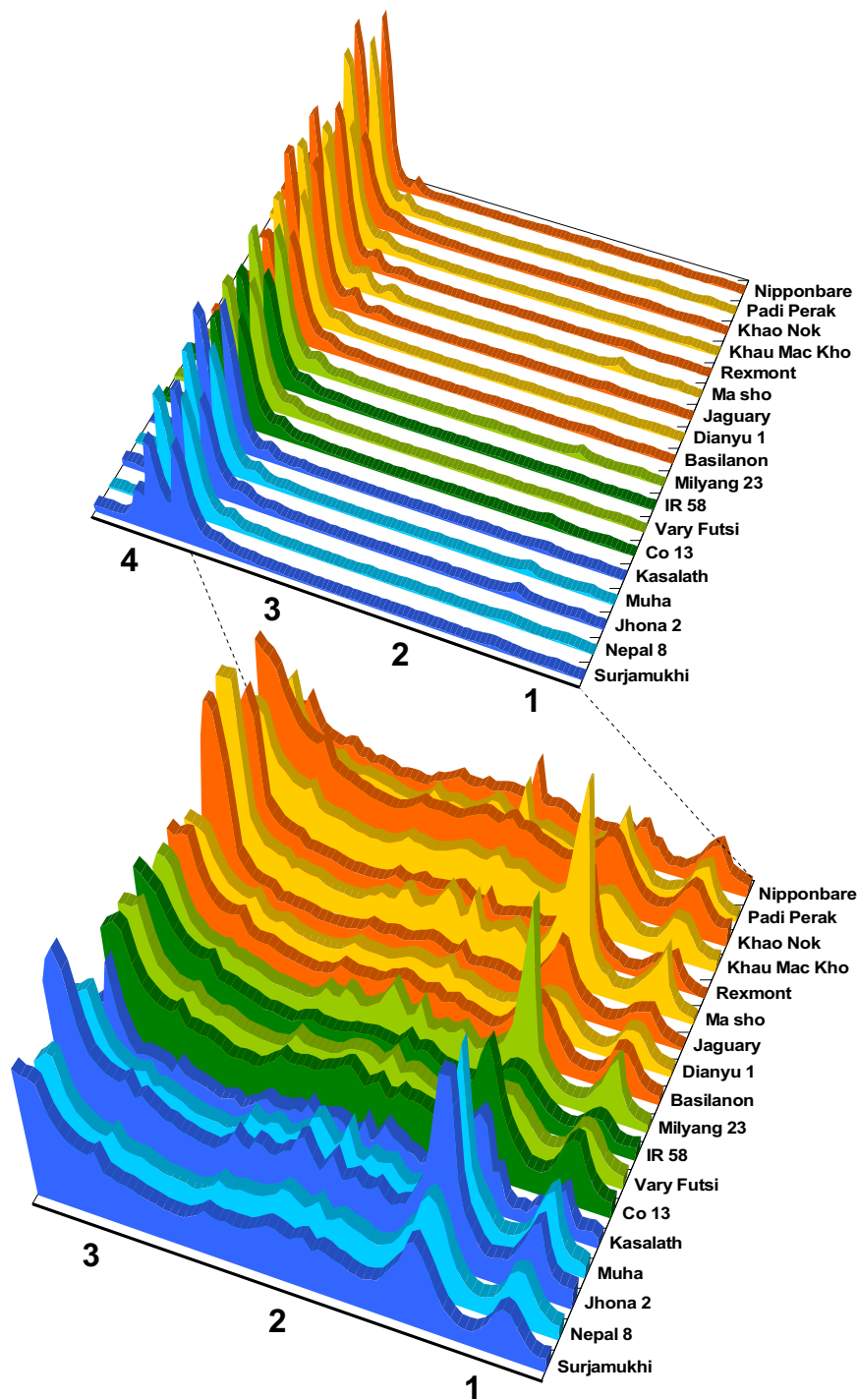


Figure 2

Diversity of soluble metabolites among the 18 rice accessions. The global metabolic profiles of soluble metabolites were determined from 0.04-ppm integral values based on ^1H NMR spectra. The 1D ^1H NMR spectra in the range of chemical shifts from 4.29 to 0.71 ppm are shown, along with a close-up of the spectra from 3.3 ppm. Orange, japonica; blue, indica I; green, indica II.

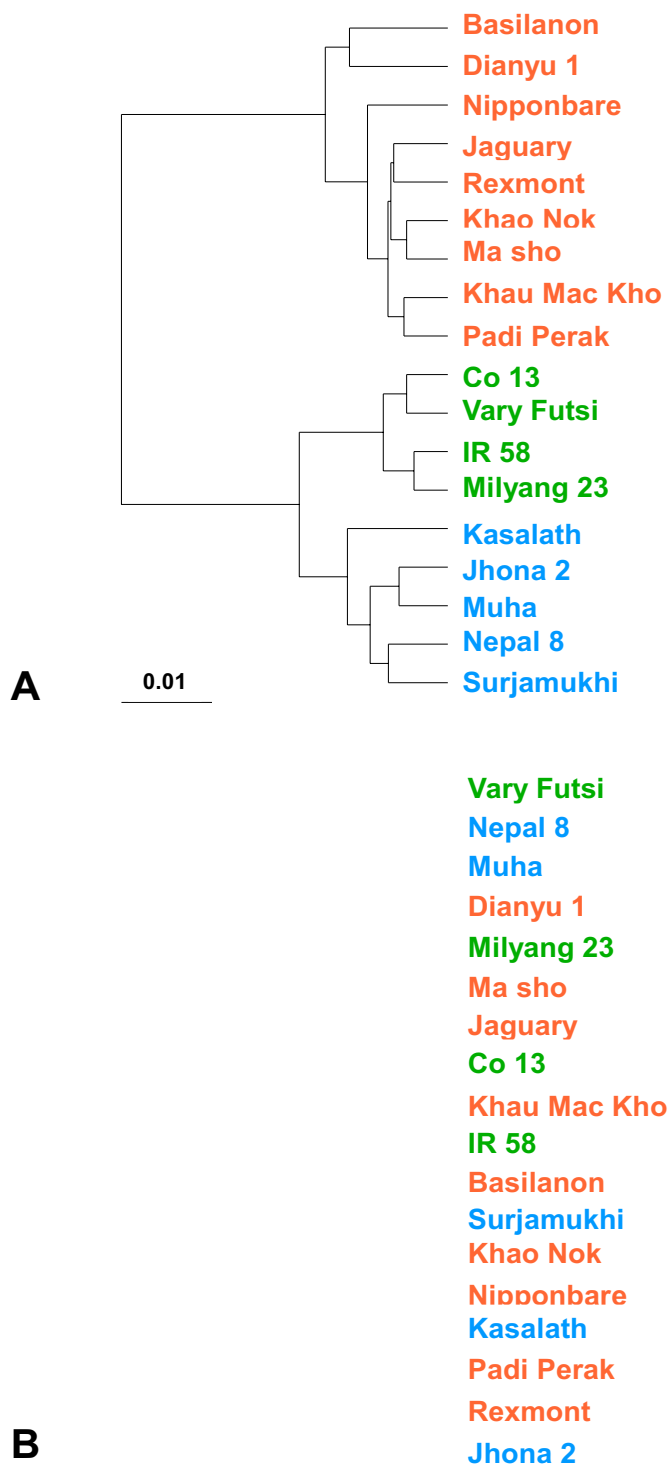


Figure 3

Global genomic and metabolo-phenotypic diversity among the 18 rice accessions. Spearman's coefficient of correlation between the genetic and Euclidian distance matrices of the 1D NMR data was calculated. (A) Dendrogram of the phylogenetic tree obtained by the UPGMA method. (B) Metabolomic distance based on normalized data from the integral regions of the ¹H NMR spectra is shown in Additional file 3 Table S2.

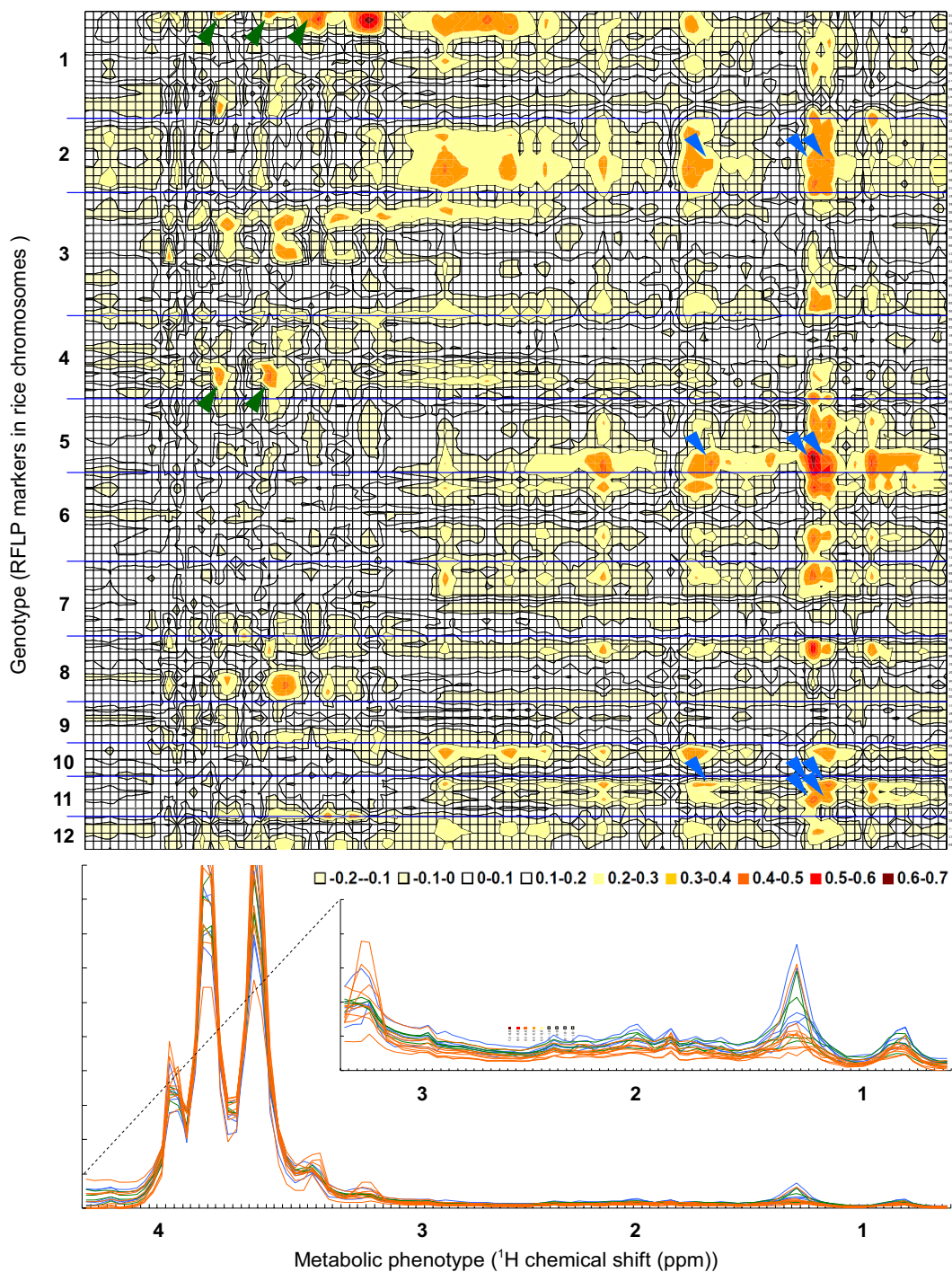


Figure 4
Correlation display between genomic and metabolic diversity in rice. Top: Heatmap displaying coefficients of correlation between metabolite abundance based on the ¹H chemical shift (horizontal) and polymorphisms along each of the 12 rice chromosomes (vertical) in the 18 rice accessions. Bottom: Overlay of ¹H-NMR spectra of metabolite extracts from the 18 rice accessions. Orange, japonica; blue, indica I; green, indica II. Intra-molecular correlations against more than two NMR signals can be determined; these are allocated on the chromosomal regions in this correlation display (green and blue arrows for sucrose and lipids, respectively).

showed that the probability of a correlation coefficient $r_s \geq 0.4$ in the randomized data was less than 1% (0.71%). Therefore, in this analysis, we applied $r_s = 0.4$ as the threshold to detect correlative combinations between metabolic and genomic diversity.

The presentation of the correlations as a 2D heat map showed large non-correlative areas studded with correlative areas, which can be explored in omics-wide comparisons. The heat map showed clear correlations between the profiles of soluble metabolites and local genomic diversity at each chromosomal location. To identify metabolites, we performed 2D 1H-13C NMR analysis (Additional file 6, 7 and 8). Diversity in the choline profile (3.2 ppm) showed a high correlation with genetic diversity around the R1613 RFLP marker, located at 176.5 cM on chromosome (Chr.) 1 ($r_s = 0.67$, $P < 0.01$). The lipid methylene signal (1.22 ppm) showed correlations with polymorphic sites at multiple chromosomal locations, with the highest located around the C43 RFLP marker at 92.8 cM on Chr. 5 ($r_s = 0.66$, $P < 0.01$), and the second highest around R1943 at 123.7 cM on Chr. 8 ($r_s = 0.58$, $P < 0.01$). The correlation peaks corresponding to each of the lipids (1.00, 1.16, 1.22, and 1.60 ppm) with Chr. 2 (153.6 and 95.7 cM), Chr. 5 (41.4, 92.8, 103.5, 112.3, and 9.8 cM), and Chr. 11 (123.3 and 85.8 cM) were in good accordance with the covalent bond networks within molecules of each compound. Furthermore, sucrose (3.40, 3.59, and 3.78 ppm) with Chr. 1 (186.8 cM) and Chr. 4 (31.4 cM) also showed good accordance. Similar intramolecular covalent bond networks were reported in NMR experiments of biological metabolite mixtures, including covariance total correlation spectroscopy (TOCSY) analysis of insect venom [16] and TOCSY analysis of intact human gut biopsies [17].

Local comparison between the dendrogram based on the correlative RFLP polymorphisms and NMR spectra showed a similar tree structure between both genetic and metabolic variation among the accessions. For example, the dendrogram of genetic polymorphisms around the C43 RFLP marker (92.8 cM on Chr. 5) and the diversity of the lipid methylene signal indicated the genetic and metabolic divergence between indica I and two other groups, japonica and indica II (Additional file 9).

To display the chromosomal locations of the correlative regions at 34 RFLP markers, the results of the correlation analysis were superimposed onto a genetic map of rice (Additional file 10). This genetic map should allow us to identify metabolic QTLs associated with each of the metabolic abundances allocated on a genetic map. It would be beneficial if future genetic analyses narrowed down candidate genes and/or isolated such metabolic QTLs in combination with other genomic resources for rice such as

genetic marker systems and genome annotation databases.

Throughout this study, we developed a robust approach to explore the correlation between genomic and metabolic diversity observed in natural variation, which is shown as a schematic flow in additional file 11. The procedure is summarized as follows: (1) selection of appropriate accessions based on population structures; (2) acquisition of the metabolic profile of selected accessions using various types of metabolomics platforms and the calculation of the distance among accessions on the basis of metabolite abundance or spectrum intensities; (3) genotyping of selected accessions using DNA markers covering the whole genome and the calculation of local genetic distance among accessions from genotype data of adjacent marker sets; (4) to discover the correlative relationships between each metabolite and chromosomal region and the coefficients of correlation for metabolic and genetic distances in all-against-all combinations and to display their distribution on a genetic map. In this analysis of rice, we did not find a correlation between metabolic and genomic diversity as observed in previously reported studies. On the other hand, local correlation analysis allowed us to identify small correlative areas mostly in non-correlative spaces. This suggests that local correlation analysis could be an effective method to discover such correlative areas allocated to chromosomal regions together with linkage maps or variation data of genomic sequences [18,19].

This study illustrated a new integrative approach to explore the links between metabolomics and genomics. The procedure could be extended to a higher resolution by using denser polymorphic marker maps in various species.

Conclusion

This study assessed the correlative relationships between genetic and metabolic diversity among rice accessions. The data presents the first correlative relationship for the chromosomal distribution of each metabolic profile in rice. The combinatorial use of genomic and metabolomic data demonstrated here should facilitate the elucidation of complex regulatory networks and those of evolutionary significance in plant metabolic systems.

Methods

Rice seed materials

The rice accessions were chosen from the RFLP-based Rice Discovery Research Set of rice germplasms, which contains 68 rice accessions and is developed and maintained by Genebank http://www.gene.affrc.go.jp/about_en.php at the National Institute of Agrobiological Sciences (NIAS). The accessions were classified into three major

groups by principal components analysis (PCA) of the RFLP data: japonica, indica I, and indica II [14]. Population structure analysis using Structure software (v. 2.1, http://pritch.bsd.uchicago.edu/software/structure2_1.html) was also conducted on the world rice collection to determine the population structure in more detail, especially for the japonica type [20,21]. Based on the results of the PCA and Structure analyses we selected nine accessions each of japonica (temperate, tropical I, II, and III) and indica (indica I and indica II) to be included in the correlation analysis between genetic and metabolic diversity.

RFLP genotyping and phylogenetic analysis of rice accessions

We used 128 RFLP markers to evaluate the genetic diversity of the 18 accessions by the Southern hybridization method of Kojima et al. (2005). Each of the marker alleles was scored as Nipponbare-type (1), Kasalath-type (2), or other types (3-7) (Additional file 3), and the presence or absence of each allele was recorded. The genetic distances between the 18 rice accessions were calculated with the "restdist" program of the PHYLIP (v. 3.67) package <http://evolution.genetics.washington.edu/phylip.html>; we used the distance matrix to calculate the coefficients of correlation with the metabolic diversity matrix and to calculate a phylogenetic tree by the UPGMA method using the "neighbor" program of PHYLIP. The physical position of each RFLP marker was retrieved from the Rice Annotation Project Database (RAP-DB) <http://rapdb.dna.affrc.go.jp/> [19].

NMR spectroscopy and quantitative analysis

We prepared NMR samples from rice seeds, as previously described [22,23] with slight modifications. All 1D Watergate [24] spectra were acquired at 298 K on a Bruker Avance DRX 500 NMR spectrometer operating at 500.13 MHz and equipped with a ¹H inverse triple-resonance probe with triple-axis gradients.

The 1D NMR spectra were integrated between 0.0 and 10.0 ppm over a series of 0.04-ppm regions by our custom integration software [25]. After exclusion of the water and DSS (2,2-dimethyl-2-silapentane-5-sulfonate) resonances, each integral region was normalized to the total integral region (Additional file 2). The data were analyzed by a partial least-squares projection, based on the spectral bins obtained from 1D spectral analysis and by using a hierarchical clustering analysis package running on R software.

Correlation display of genetic and metabolic variation

A Euclidian distance matrix based on the 0.04-ppm integral regions of the chemical shift from 0.708 to 4.292 ppm (104 bins) in the 1D NMR spectra was calculated by

using the "dist" command of the R package. Genetic diversity among the 18 rice accessions was calculated from the genotype data of all alleles of the 128 scored RFLP markers. To assess the correlation between genome- and metabolome-wide diversity among the 18 accessions, we calculated Spearman's coefficient of correlation between the genetic and Euclidian distance matrices. Local metabolic diversity among the 18 accessions was calculated as a Euclidian distance matrix in each of the 104 bins. Local genomic diversity was also calculated as a genetic distance matrix using sliding bins that included three adjacent RFLP markers along each rice chromosome. To explore the local correlations between genomic and metabolic diversity, we calculated Spearman's r_s and p-values for all combinations of the local metabolic and genomic distance matrices using the "cor.test" program of the R package (Additional file 4). An RFLP map with correlative chromosomal positions was generated with MapChart ver. 2.2 [26].

Annotation of candidate metabolites by 2D NMR of ¹³C-labelled rice seeds

¹³C-labelled rice extracts were prepared as previously described methods with minor modifications [27,28]. In brief, seeds were powdered, 20 mg of which was suspended in 600 μ L standard buffer (100 mM potassium phosphate, pH 7.0, and 1.0 mM 2,2-dimethyl-2-silapentane-5-sulfonate in D₂O), heated at 65°C for 15 min, and centrifuged at 12,000 \times g for 5 min. The supernatant (500 μ L) was decanted into a 5 mm diameter NMR tube.

Two-dimensional (2D) heteronuclear single quantum coherence (HSQC) [29] spectra were recorded on a Bruker Avance 500 spectrometer equipped with an inverse triple resonance CryoProbe with a Z-axis gradient, operating at 500.13 MHz for ¹H frequency (176.061 MHz for ¹³C frequency), and the temperature of the NMR samples was maintained at 298 K. A total of 128 complex f1 (¹³C) and 1024 complex f2 (¹H) points were recorded with 96 scans per f1 increment, resulting in a total recording time of about 6 h. The spectral window and offset frequency in the f1 dimension were 7042.593 Hz (40 ppm) and 11971.59 Hz (68 ppm), respectively. The spectral window in the f2 dimension was 11160.7 Hz (16 ppm). The offset frequency in the f2 dimension was 3330 Hz (4.75 ppm).

The 2D HSQC spectra of the ¹³C-labelled rice samples were processed using the NMRPipe software package [30]. To quantify the signal intensities, a Lorentzian-to-Gaussian window with a Lorentzian line width of 10 Hz and a Gaussian line width of 15 Hz was applied in both dimensions before Fourier transformation. An automatic polynomial baseline correction was subsequently applied in the f1 dimension. The indirect dimension was zero-filled to 4096 points in the final data matrix. Cross peaks of

each metabolite were matched to our standard $^1\text{H}/^{13}\text{C}$ chemical shift database [31]. The database is implemented with an in-house Java program, which allows for the systematic batch identification of large numbers of metabolites by simply matching the observed ^{13}C -HSQC peaks with peaks in the database. Our chemical shift data are continuously updated and are available on the PRIME website <http://prime.psc.riken.jp/> [32]. The queried peaks were classified as annotated on the basis of whether the chemical shift difference in each dimension between the observed peak and the peak in the database was less than a specified tolerance value. Typically, tolerances of 0.03 ppm for ^1H and 0.53 ppm for ^{13}C were used in this study (Additional file 8). From these candidate metabolites, an identification or assignment was defined as unique if there was only one candidate in the database within the specified tolerances for an observed peak (Additional file 6).

Authors' contributions

KM and JK designed the project, performed data analysis, and drafted the manuscript. TF performed NMR data processing and analyses. EK provided RFLP polymorphism data and population structure of the rice core collection. KS served as the principal investigator of the project. All the authors have contributed to writing this manuscript and have read and approved the contents of the final submitted version.

Additional material

Additional file 1

Figure S1. Rice seed material used in this study and their origins.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S1.PDF>]

Additional file 2

Table S1. The scored genotype data from the 128 RFLP markers.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S2.XLS>]

Additional file 3

Table S2. The numeric metabolite profiling data from rice seed extract based on the ^1H NMR spectra.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S3.XLS>]

Additional file 4

Table S3. The coefficients of correlation values between genetic and metabolic diversity in rice.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S4.XLS>]

Additional file 5

Figure S2. Histogram of the correlation coefficient for the randomized data from 10,000 trials.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S5.PDF>]

Additional file 6

Table S4. The data of the annotated metabolites.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S6.XLS>]

Additional file 7

Figure S3. The ^{13}C -HSQC spectra of seed extracts from the japonica and indica rice varieties.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S7.PDF>]

Additional file 8

Figure S4. The ^{13}C -HSQC spectrum from a seed extract of ^{13}C -labelled Nipponbare.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S8.PDF>]

Additional file 9

Figure S5. The correlative local genetic and metabolo-phenotypic diversity in rice.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S9.PDF>]

Additional file 10

Figure S6. The correlative chromosomal regions for each NMR spectrum allocated on an RFLP map of rice.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S10.PDF>]

Additional file 11

Figure S7. A schematic representation of the work flow used to explore the correlations between metabolic and genomic diversity with genetic resources.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-568-S11.PDF>]

Acknowledgements

We thank E. Chikayama (RIKEN) for technical assistance and data analysis, and T. Hirayama (RIKEN) and D. Saisho (Okayama University) for their valuable discussion and suggestions in this study. We also thank Y. Nishizuka (RIKEN) for providing technical assistance in rice RFLP marker analysis. This work was supported in part by the Research and Development Program for New Bio-industry Initiatives of the Bio-oriented Technology Research Advancement Institution (BRAIN).

References

1. Dumas ME, Wilder SP, Bihoreau MT, Barton RH, Fearnside JF, Argoud K, D'Amato L, Wallis RH, Blancher C, Keun HC, et al.: **Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models.** *Nat Genet* 2007, **39**:666-672.
2. Mitchell-Olds T, Schmitt J: **Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis.** *Nature* 2006, **441**:947-952.
3. Keurentjes JJ: **Genetical metabolomics: closing in on phenotypes.** *Curr Opin Plant Biol* 2009, **12**:223-230.
4. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, et al.: **Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement.** *Nat Biotechnol* 2006, **24**:447-454.
5. Schauer N, Semel Y, Balbo I, Steinfath M, Reipsilber D, Selbig J, Pleban T, Zamir D, Fernie AR: **Mode of inheritance of primary metabolic traits in tomato.** *Plant Cell* 2008, **20**:509-523.
6. Fu J, Keurentjes JJ, Bouwmeester H, America T, Verstappen FW, Ward JL, Beale MH, de Vos RC, Dijkstra M, Scheltema RA, et al.: **System-wide molecular evidence for phenotypic buffering in Arabidopsis.** *Nat Genet* 2009, **41**:166-167.
7. Liseic J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Torjek O, Selbig J, Altmann T, et al.: **Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations.** *Plant J* 2008, **53**:960-972.
8. Morreel K, Goeminne G, Storme V, Sterck L, Ralph J, Coppieters W, Breynne P, Steenackers M, Georges M, Messens E, et al.: **Genetical metabolomics of flavonoid biosynthesis in Populus: a case study.** *Plant J* 2006, **47**:224-237.
9. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ: **Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome.** *Plant Cell* 2008, **20**:1199-1216.
10. Laurentin H, Ratzinger A, Karlovsky P: **Relationship between metabolic and genomic diversity in sesame (Sesamum indicum L.).** *BMC Genomics* 2008, **9**:250.
11. Wolde-Meskel E, Terefework Z, Frostegard A, Lindstrom K: **Genetic diversity and phylogeny of rhizobia isolated from agroforestry legume species in southern Ethiopia.** *Int J Syst Evol Microbiol* 2005, **55**:1439-1452.
12. Wolde-Meskel E, Terefework Z, Lindstrom K, Frostegard A: **Metabolic and genomic diversity of rhizobia isolated from field standing native and exotic woody legumes in southern Ethiopia.** *Syst Appl Microbiol* 2004, **27**:603-611.
13. Seymour FA, Cresswell JE, Fisher PJ, Lappin-Scott HM, Haag H, Talbot NJ: **The influence of genotypic variation on metabolite diversity in populations of two endophytic fungal species.** *Fungal Genet Biol* 2004, **41**:721-734.
14. Kojima Y, Ebana K, Fukuoka S, Nagamine T, Kawase M: **Development of an RFLP-based Rice Diversity Research Set of Germplasm.** *Breeding Science* 2005, **55**:431-440.
15. Keurentjes JJ, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, Plas LH van der, Jansen RC, Vreugdenhil D, Koornneef M: **The genetics of plant metabolism.** *Nat Genet* 2006, **38**:842-849.
16. Zhang F, Dossey AT, Zachariah C, Edison AS, Bruschweiler R: **Strategy for automated analysis of dynamic metabolic mixtures by NMR. Application to an insect venom.** *Anal Chem* 2007, **79**:7748-7752.
17. Wang Y, Cloarec O, Tang H, Lindon JC, Holmes E, Kochhar S, Nicholson JK: **Magic angle spinning NMR and IH-31P heteronuclear statistical total correlation spectroscopy of intact human gut biopsies.** *Anal Chem* 2008, **80**:1058-1066.
18. Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, et al.: **Gramene: a resource for comparative grass genomics.** *Nucleic Acids Res* 2002, **30**:103-105.
19. Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, Numa H, Sakai H, Wu J, Itoh T, Sasaki T, et al.: **The Rice Annotation Project Database (RAP-DB): 2008 update.** *Nucleic Acids Res* 2008:DI028-1033.
20. Ebana K, Kojima S, Fukuoka S, Uga Y, Kawase K, Okuno K: **DNA polymorphism detected in the rice diversity research set of germplasm.** *5th International Rice Genetics Symposium* 2005.
21. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945-959.
22. Kikuchi J, Shinozaki K, Hirayama T: **Stable isotope labeling of Arabidopsis thaliana for an NMR-based metabolomics approach.** *Plant Cell Physiol* 2004, **45**:1099-1104.
23. Kikuchi J, Hirayama T: **Practical aspects of uniform stable isotope labeling of higher plants for heteronuclear NMR-based metabolomics.** *Methods Mol Biol* 2007, **358**:273-286.
24. Piotto M, Saudek V, Sklenar V: **Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions.** *J Biomol NMR* 1992, **2**:661-665.
25. Liu X, Deng Z, Gao S, Sun X, Tang K: **Molecular cloning and characterization of a glutathione S-transferase gene from Ginkgo biloba.** *DNA Seq* 2007, **18**:371-379.
26. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs.** *J Hered* 2002, **93**:77-78.
27. Sekiyama Y, Kikuchi J: **Towards dynamic metabolic network measurements by multi-dimensional NMR-based fluxomics.** *Phytochemistry* 2007, **68**:2320-2329.
28. Tian CJ, Chikayama E, Tsuboi Y, Shinozaki K, Kikuchi J, Hirayama T: **Top-down phenomics of Arabidopsis thaliana-One and two-dimensional NMR metabolic profiling and transcriptome analysis of albino mutants.** *J Biol Chem* 2007, **282**:18532-18541.
29. Bodenhausen G, Ruben DJ: **Natural Abundance N-15 Nmr by Enhanced Heteronuclear Spectroscopy.** *Chemical Physics Letters* 1980, **69**:185-189.
30. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A: **Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes.** *J Biomol NMR* 1995, **6**:277-293.
31. Chikayama E, Suto M, Nishihara T, Shinozaki K, Hirayama T, Kikuchi J: **Systematic NMR analysis of stable isotope labeled metabolite mixtures in plant and animal systems: coarse grained views of metabolic pathways.** *PLoS ONE* 2008, **3**:e3805.
32. Akiyama K, Chikayama E, Yuasa H, Shimada Y, Tohge T, Shinozaki K, Hirai MY, Sakurai T, Kikuchi J, Saito K: **PRIME: a Web site that assembles tools for metabolomics and transcriptomics.** *In Silico Biol* 2008, **8**:339-345.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

