

# SCIENTIFIC REPORTS



OPEN

## Population transcriptomics uncovers the regulation of gene expression variation in adaptation to changing environment

Received: 10 February 2016

Accepted: 19 April 2016

Published: 06 May 2016

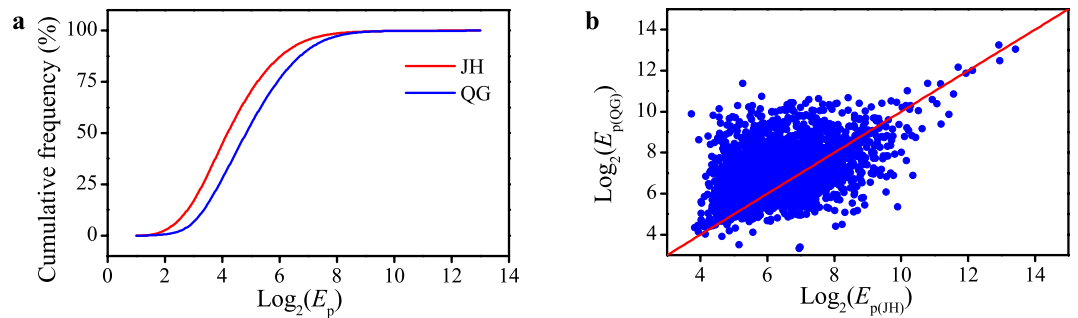
Qin Xu<sup>1,\*</sup>, Caiyun Zhu<sup>2,3,\*</sup>, Yangyang Fan<sup>1,3</sup>, Zhihong Song<sup>1,3</sup>, Shilai Xing<sup>2,3</sup>, Wei Liu<sup>2</sup>, Juan Yan<sup>4</sup> & Tao Sang<sup>1,2</sup>

Expression variation plays an important role in plant adaptation, but little is known about the factors impacting the expression variation when population adapts to changing environment. We used RNA-seq data from 80 individuals in 14 *Miscanthus lutarioriparius* populations, which were transplanted into a harsh environment from native habitat, to investigate the expression level, expression diversity and genetic diversity for genes expressed in both environments. The expression level of genes with lower expression level or without SNP tended to be more changeable in new environment, which suggested highly expressed genes experienced stronger purifying selection than those at lower level. Low proportion of genes with population effect confirmed the weak population structure and frequent gene flow in these populations. Meanwhile, the number of genes with environment effect was the most frequent compared with that with population effect. Our results showed that environment and genetic diversity were the main factors determining gene expression variation in population. This study could facilitate understanding the mechanisms of global gene expression variation when plant population adapts to changing environment.

The molecular and physiological phenotypes of plant, such as protein levels or trait phenotypes can be coordinated by genetic and environmental factors via *cis*- or *trans*- expression regulation and environmentally induced epigenetic regulation<sup>1–6</sup>. Therefore, gene expression level is considered an intermediate phenotype from which conclusions about the emergence of high level traits can be drawn<sup>7,8</sup>. Generally, abrupt environmental changes can induce the change of gene expression programs, which can help adjust the regulation of plant growth and metabolism to the new environment in return and increase population-level phenotypic variation<sup>9–13</sup>. The increase of phenotypic variation is capable of producing alternative phenotypes and improving the potential of genetic evolution due to artificial and natural selection, and thus enables the evolution of heritable adaptive traits<sup>14,15</sup>. Hence, modulation of gene expression level has a central role in plant architecture and adaptation to changing environment. Exploring the characterization of population gene expression variation in changing environment is very important for understanding the roles and mechanisms of gene expression in plant adaptation to new environment.

Expression variation among populations is found to be pervasive in many plants such as maize, tomato and grapevine<sup>9,16–21</sup>. Those studies address the variation in gene expression within and between natural populations for studying genome evolution. Substantial expression variations, even among isogenic individuals or cells in seriously controlled common environment, can still be observed in response to different environmental conditions<sup>9,22</sup>. In general, gene expression level can be considered as a quantitative trait since it differs between individuals with respect to genetic factors, based on which many studies had mapped the genetic variation determining expression<sup>1,23,24</sup>. However, environment has been shown to have a stronger effect on gene expression than genetic

<sup>1</sup>Key Laboratory of Plant Resources and Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>2</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. <sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>4</sup>Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.S. (email: sang@ibcas.ac.cn)



**Figure 1.** The comparison of gene expression abundance in population between JH and QG. (a) The cumulative frequency of gene expression abundance in JH and QG.  $E_p$ s were transformed to the logarithm to the base 2. (b) The correlation of  $E_p$ s for genes between JH and QG.

variation<sup>17,25</sup>. The powerful functionality of environment on gene expression may be due to cryptic genetic variation, which has no discernable effect under common ecological environment and can be amplified in new stressful environment, facilitating evolutionary adaptation if this variation happens to be pre-adapted to a new environment and revealing advantageous<sup>26–31</sup>. Epigenetic modifications are another source of expression variation<sup>32</sup>. The establishment and maintenance of epigenetic modifications can contribute to the variability of phenotypes by regulating gene expression in response to the stress stimulation<sup>33</sup>.

For the past 40 years, it has gained wide acceptance that the genetic variation rate is determined mainly by the importance of protein function<sup>34,35</sup>. It has also been reported that expression abundance could negatively induce about 20–40% of variation in protein evolutionary rates<sup>36</sup>. In addition, genetic variation rates seem to correlate strongly with measures related to expression level and expression variation<sup>37,38</sup>. Many studies find that highly expressed genes have been under strong selective constraint for translation robustness<sup>39</sup>, and broadly expressed genes have been constrained by the need of function in several biochemical environments<sup>40</sup>.

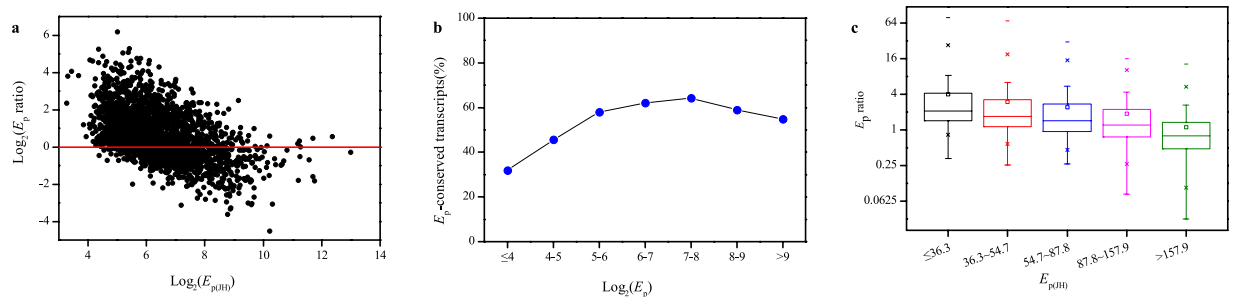
Although these hypotheses are proposed, the genome-wide gene expression variation pattern has not been well characterized when population adapts to new environment. With the fast development of next generation sequencing platforms, high throughput transcriptome sequencing makes exploring global gene expression variation possible. In our previous experiment, we collected 14 populations of *Miscanthus lutarioriparius*, the candidate wild progenitor of second-generation energy crops, across its natural distribution and planted them in two experimental fields, one near its native habitat in Jiangxia of Hubei Province (JH) and the other at the domestication site Qingyang of Gansu Province (QG) with much colder and drier climates and poorer soil conditions. As a C4 perennial grass capable of producing high biomass on marginal land, *Miscanthus* has emerged to be a promising candidate of dedicated energy crops<sup>41–46</sup>. The following question is how to achieve fast adaptation to marginal and harsh land for *Miscanthus*. Field experiment showed that *M. lutarioriparius* was able to establish and to produce higher biomass in stressful environment than in native habitat<sup>45</sup>. To characterize the transcriptomics of *M. lutarioriparius*, a total of 40 individuals originated from the same natural populations and at the same growth stage were sampled from each field site for RNA-seq respectively. The comparison of population transcriptomes had demonstrated that a number of gene expression levels changed across environments<sup>47,48</sup>. Genetic diversity significantly decreased and the expression diversity significantly increased when population was transplanted from the native habitat to the unfavorable and harsh environment.

However, these above studies did not focus on characterizing the patterns and factors impacting gene expression variation. Moreover, previous studies of gene expression variation usually adopted the samples under strictly controlled environment, which suppressed the variability in population level<sup>49</sup>. Here we used population transcriptome of two natural field conditions imbuing various levels of variation to uncover the global gene expression pattern and expression variation in the process of adaptation. We aimed to carry out a thorough and comprehensive analysis to elucidate the expression variation pattern. In detail, we explored the relationship among the gene expression level ( $E_p$ , see method), expression variation ( $E_p$  ratio), expression diversity ( $E_d$ ), and genetic diversity ( $\pi$ ) based on population transcriptomes. We tried to answer the question that what factors determine or correlate with the gene expression variation when plant population adapts to a changing environment.

## Results

### Expression of lower expressed genes tended to be more changeable in new environment.

Expression level ( $E_p$ , see method) and expression diversity ( $E_d$ , see method) in the population were analyzed based on 40 *M. lutarioriparius* individuals from each environment. There were 15367 genes expressed in at least half of the individuals in each environment, and 11400 of them were expressed in all the individuals at both sites. The distribution of  $E_p$  for all genes within each environment was represented in Fig. 1. It was shown that  $E_p$ s in transplanted or target domestic site QG were higher than those in near native habitat JH (Fig. 1a), and they were significantly related with each other between the two distinct environments ( $r = 0.98$ ,  $P < 0.01$ ) (Fig. 1b). To study the expression variation between two sites, we calculated  $E_p$  ratio for each gene ( $E_{p(QG)}/E_{p(JH)}$ ). There were 58.3% (8961 genes) of genes having conserved  $E_p$  ( $E_p$  ratio ranged from 0.5 to 2) in new environment. Enriched Pfam group of these genes mainly included Ubiquitin-conjugating enzyme, FKBP-type peptidyl-prolyl cis-trans isomerase, Zinc finger, C3HC4 type (RING finger), RNA recognition motif, (Aka RRM, RBD, or RNP domain), Ubiquitin family, and Thioredoxin, all of which had more than 75% members having conserved  $E_p$  value (Supplementary Table S1).



**Figure 2. The comparison of gene expression abundance in population and expression variation between two distinct field sites.** (a) The  $\log_2(E_p \text{ ratio})$  was plotted  $\log_2 E_p$  value for gene in JH, and  $E_p$  ratio for each genes was calculated using  $E_{p(QG)}/E_{p(JH)}$ . (b) The proportion of genes with conserved  $E_p$  between the two distinct field sites. The  $E_p$  ratio ranging from 0.5 to 2 was considered as conserved  $E_p$ . (c) The relationship between  $E_p$ s in JH and  $E_p$  ratio. The data was divided equally into five groups according to  $E_p$ s in JH.

Using GO functional classification, we found these genes were mainly ATP hydrolysis coupled proton transport, acid-amino acid ligase activity, unfolded protein binding, transcription initiation, small GTPase mediated signal transduction, transcription initiation, protein folding, protein modification process, metalloendopeptidase activity, cell redox homeostasis and cellular protein metabolic process.

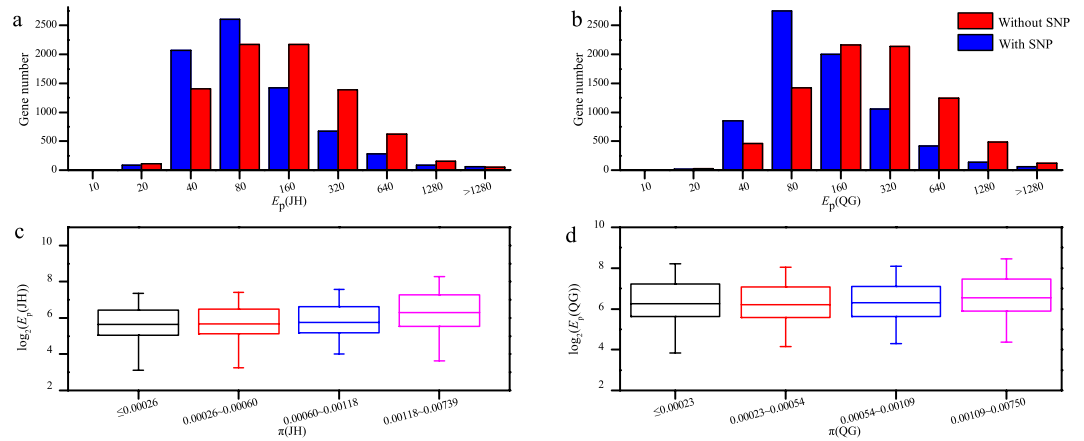
To analyze the relationship between expression variation and expression level,  $E_p$  ratio was plotted against  $E_p$ s in near native habitat site JH (Fig. 2a). It was found that  $E_p$  ratio tended to decrease with the increase of  $E_p$ s in JH. However, the percentage of genes with conserved  $E_p$  did not persistently increased along with the increase of  $E_p$ , and reached the highest value (65%) when  $\log_2(E_p)$  was at 8. In order to obtain a significance test for the relationship between  $E_p$ s and  $E_p$  ratio, all genes were divided into 5 groups, each of which had the same number of genes, according to their  $E_p$ s value in JH (Fig. 2c). It was shown that  $E_p$  ratio significantly decreased with the increase of  $E_p$  between any two groups (Wilcoxon-test,  $P < 0.001$ ).  $E_p$  ratios for the group with lower  $E_p$  value was significantly larger than that with higher  $E_p$  value, which suggested that highly expressed genes were relatively insensitive to environmental change.

**Expression level of genes with SNP was lower than that without SNP.** We analyzed the relationship between expression level ( $E_p$ ) and genetic diversity ( $\pi$ ). Of the 15367 genes, only 7226 and 7248 genes had genetic variation (SNP) in near native habitat JH and transplanted site QG respectively. The genetic diversity ( $\pi$ ) ranged from  $5.96E-06$  to  $7.39E-03$  in JH and from  $5.56E-06$  to  $7.50E-03$  in QG, with an average value of  $8.57E-04$  and  $7.98E-04$ , respectively. Based on the presence or absence of SNP, genes were grouped into two categories, one representing genes with SNP and the other representing genes without SNP. First, we tested whether the  $E_p$ s were significantly different between gene groups with and without SNP. The distribution of  $E_p$ s showed that the group without SNP had a greater number of genes than that with SNP when  $E_p$ s were at a given level of 160 or greater both in JH and in QG, suggesting that  $E_p$ s for genes without SNP tended to be larger than those with SNP (Fig. 3a,b). Further test of significance showed a significant difference of  $E_p$ s between the two gene groups (Wilcoxon test,  $P < 0.001$ ). These indicated that the genes with SNP usually expressed at a lower level.

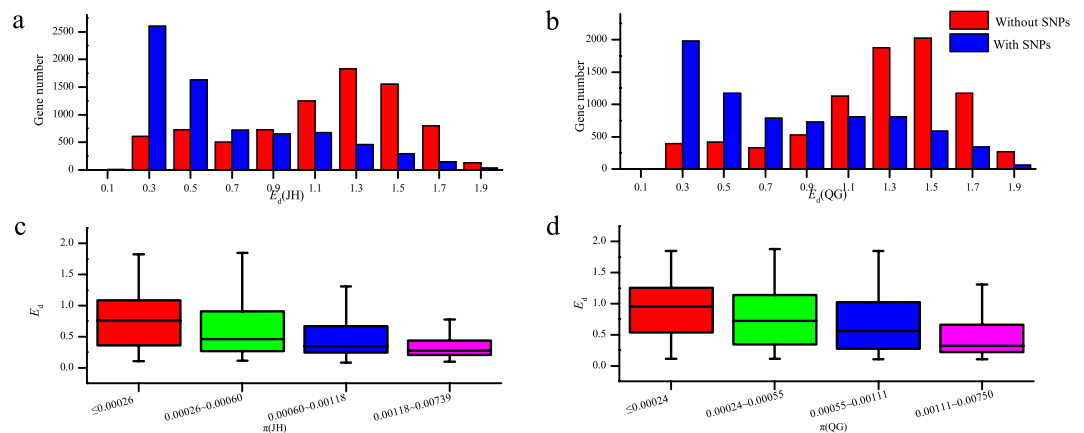
Then we further tested the relationship between  $E_p$ s and  $\pi$ . The  $E_p$ s was plotted against  $\pi$  using the gene sets with SNP. Although it seemed low correlation between them, we further statistically tested the relationship by dividing the gene sets into four groups, each of which included the same number of genes. There were no significant difference in  $E_p$ s between the first 3 groups, but the  $E_p$ s of last group were significantly higher with the increase of  $\pi$  both in JH and in QG (Fig. 3c,d, Wilcoxon test,  $P < 0.01$ ). In addition, since a longer sequence may have higher possibility of presence of SNP, we tested whether SNP was correlated with sequence length. It was shown that the distributions of gene length for the genes with or without SNP were almost consistent, which suggested that the presence or absence of SNP had no correlation with gene length. Thus, the contribution of genetic variation to expression level in population was mainly due to the presence or absence of SNP rather than the level of genetic diversity.

**Genes with lower genetic diversity tended to have higher expression diversity.** To address the relationship between expression diversity ( $E_d$ ) and genetic diversity ( $\pi$ ), first we tested whether expression diversity was correlated with presence or absence of SNP. The group with SNP had a greater number of genes at a low level of  $E_d$  and a fewer number of genes at a high level of  $E_d$  both in JH and in QG when compared with the group without SNP (Fig. 4a,b). This suggested that the expression diversity of genes with SNP was always smaller than those without SNP. Using significance test, we found that it was significantly different (Wilcoxon test,  $P < 0.001$ ). This result suggested that the genes without SNP always had a broader range of expression level in population than those with SNP.

To further test the relationship between expression diversity and genetic diversity, we divided genes with SNPs into 4 groups according to the range of  $\pi$  value, each of which had the same number of genes. It was shown that  $E_d$ s decreased along with the increase of  $\pi$  both in JH and QG (Fig. 4c,d). Using significance test method, we found it was significantly different between any two groups (Wilcoxon test,  $P < 0.001$ ). Together, our results showed that expression diversity was negatively correlated with genetic diversity.



**Figure 3. The relationship between gene genetic diversity and expression abundance in population.** In each environment site, the genes were divided into two groups based on presence or absence of SNP. The distribution of  $E_p$  for each group was shown in (a) JH and (b) QG. The genes harboring SNP in each environment were divided into four groups according to their genetic diversity  $\pi$ . The distribution of  $E_p$  for genes in each divided group was shown in (c) JH and (d) QG.

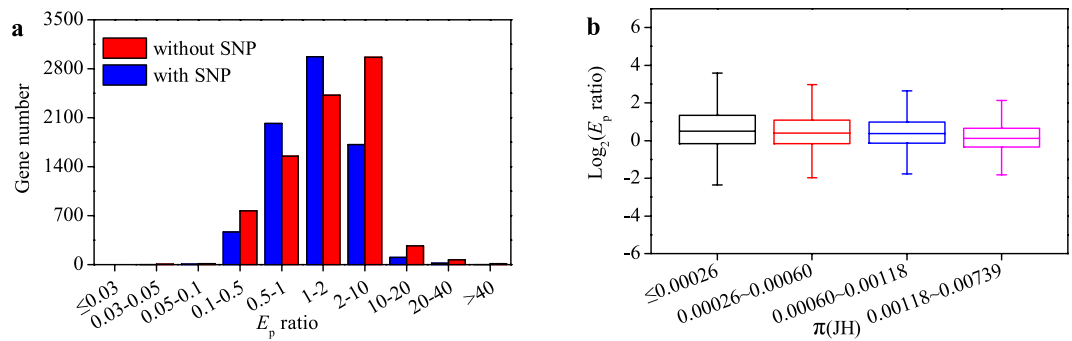


**Figure 4. The relationship between gene genetic diversity and expression diversity in population.** The genes were divided into two groups based on presence or absence of SNP. The distribution of gene expression diversity for each group was shown in (a) JH and (b) QG. The genes harboring SNPs in each environment were divided into four groups according to the  $\pi$  value, and the distribution of the expression diversity for genes in each group was shown in (c) JH and (d) QG.

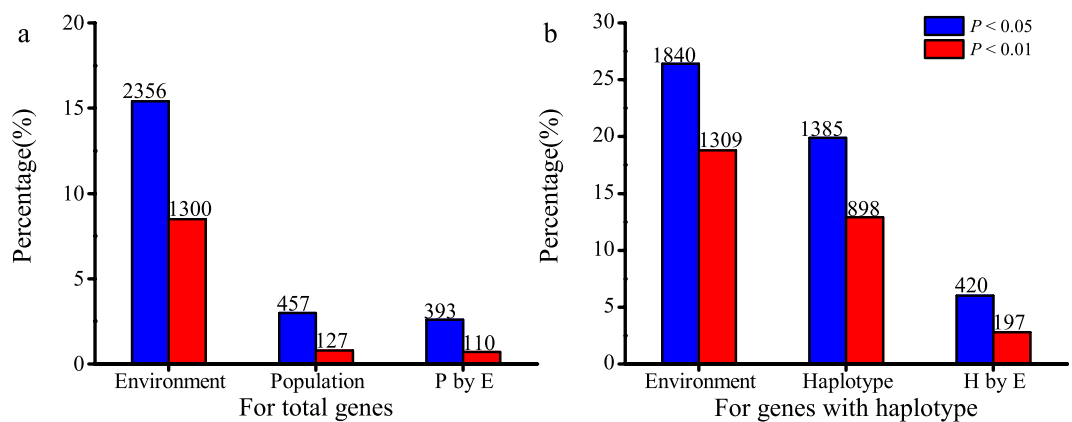
**Expression level of gene with lower genetic diversity tended to be changed more easily in new environment.** In order to examine the relationship between expression variation and genetic diversity, first we compared the  $E_p$  ratio between genes with and without SNP. The distribution of  $E_p$  ratio showed that the group without SNP had a greater number of nonconserved- $E_p$  genes ( $E_p$  ratio  $< 0.5$  or  $E_p$  ratio  $> 2$ ) and a fewer number of conserved- $E_p$  genes ( $0.5 < E_p$  ratio  $< 2$ ) when compared with the group with SNP (Fig. 5a), and this difference was significant (Wilcoxon test,  $P < 0.01$ ). This suggested that the expression level of the genes without SNP were more easily to be changed in new environment compared with those with SNP.

Secondly we analyzed the relationship between expression variation and genetic variation by plotting  $E_p$  ratio against  $\pi$  for the genes set with SNP (Fig. 5b). All the genes harboring SNP were divided into 4 groups, each of which had the same number of genes. It was found that the  $E_p$  ratio decreased along with the increase of  $\pi$ . Using significance test, we found that it was significantly different between any two groups except for the comparison of the middle two (Wilcoxon-test,  $P < 0.001$ ). Together, these results showed that the expression variation tended to decrease with the increase of genetic diversity, which suggested that genes with lower genetic diversity were more sensitive to environmental change at expression level.

**Environment was the leading contributor to gene expression variation compared with population and haplotype.** We performed analysis of variation (ANOVA) to evaluate the factors impacting gene expression variation using the groups of two environments and 14 populations. Out of 15367 genes, we found that 2365 (15.4%), 457 (3.0%) and 393 (2.6%) genes had environmental effect, habitat population effect, and



**Figure 5. The relationship between genetic diversity and  $E_p$  ratio.** (a) The genes were divided into two groups based on presence or absence of SNP. The number of gene in a specific range of  $E_p$  ratio was shown. (b) The genes harboring SNPs in each environment were divided into four groups according to  $\pi$  value in JH. The distribution of  $E_p$  ratio value was shown.



**Figure 6. Comparison of genes with environment effect, population effect, haplotype effect and their interaction effect.** (a) All transcripts were used for analysis of the genes with environment effect, population effect, and their interaction effect. P by E indicates the genes with population by environment effect; (b) The phased gene set were used for analysis of the genes with environment effect, haplotype effect, and their interaction effect. H by E indicates the genes with haplotype by environment effect. Significance test was set at the level of  $P < 0.05$  (blue) and  $P < 0.01$  (red).

population by environment effect on the expression variation, respectively ( $P < 0.05$ ) (Fig. 6a). Enriched Pfam group of these genes was also shown in Supplementary Table S2. This suggested that environment played a much broader role in gene expression variation than population effect.

In addition, we phased the haplotypes based on the population transcriptome data to analyze the haplotype effect. Totally 6964 genes were phased for ANOVA of environment and haplotype effects on gene expression variation. It was found that 1840 (26.4%), 1385 (19.9%) and 420 (6.0%) genes had environment, haplotype and haplotype by environment effect, respectively (Fig. 5b, Supplementary Table S3,  $P < 0.05$ ), which suggested that environment played a more important role in gene expression variation than haplotype. Together, environment played the leading role in gene expression variation compared with population and haplotype.

To compare the relative effect of population and haplotype on expression variation, we compared the percentage of genes with population effect and haplotype effect. 3.0% (457 genes) and 19.8% (1385 genes) out of their gene sets were identified to have population effect and haplotype effect, respectively (Fig. 5, Supplementary Table S3,  $P < 0.05$ ), which suggested that genotype effect (haplotype) was more than population effect on gene expression variation.

Since environment and haplotype were the two leading factors impacting gene expression, we classified genes into 8 groups according to whether genes had one or more of the environment, haplotype or their interaction effects (Supplementary Fig. S1, Supplementary Table S3). We found that genes with different kinds of effect exhibited different expression level, genetic diversity, expression variation, and expression diversity. For example, genes with both environment and haplotype effects had the highest genetic diversity, while genes with both haplotype and haplotype by environment effects had the lowest genetic diversity.

**The implication of population expression level and expression diversity under different environments.** Since  $E_p$  and  $E_d$  were two indexes potentially weighting the roles in adaptation, we divided genes into 9 groups based on the changes of  $E_p$  and  $E_d$  (Supplementary Fig. S2). About 43% of transcripts were



conserved in both  $E_p$  and  $E_d$ . This group contained the genes that were insensitive to environmental change and were probably essential to plant growth and development. About 9% of transcripts had conserved  $E_d$  and upregulated  $E_p$ , implying these genes had a common response to environment but without individual differentiation. About 24% of transcripts had upregulated  $E_p$  and enlarged  $E_d$ . This group of genes probably had key functions for plant development and was easy to be artificially captured for adaptation (Supplementary Fig. S3).

## Discussion

**Expression level, genetic diversity and evolutionary rate.** It is widely accepted that highly expressed genes evolve slowly<sup>50</sup>. Our result found that the genes without SNP tended to have higher expression level (Fig. 3a,b), which was consistent with previous conclusion. This could be explained by the stronger purifying selection for the highly expressed genes than those expressed at a lower level<sup>51</sup>. However, when we extended to study the relationship between expression level and genetic diversity, no correlation was observed when genetic diversity was at a low or middle level. Although we found a significantly difference of  $E_p$ s for the last group having a high  $\pi$  value compared with the remained groups, this phenomenon may be caused by detection sensitivity for the genes with extreme high  $E_p$ s in the last group. Thus the observed result that the genes with high genetic diversity had high  $E_p$ s may be an illusion. Overall, our data suggested that the genetic contribution to differential expression level was mainly due to the presence or absence of SNP rather than the quantity of SNP.

Although there was very weak relationship between expression level and the quantity of SNPs in our data, we dissected the genetic variation of genes into nonsynonymous and synonymous variation. It was found a trend that higher proportion of nonsynonymous substitution was occurred in highly expressed genes (Supplementary Fig. S4). This phenomenon may be explained by that synonymous substitutions affect the stability of mRNA secondary structures and the genes with high mRNA level tend to need more stable secondary structures to prevent aggregation of mRNA molecules and improve the efficiency of translation elongation<sup>52,53</sup>. Thus, synonymous substitution in highly expressed genes may experience purifying selection. These suggested that strength of selection was different across expression level and the selection on the highly expressed gene was always stronger than the genes with low expression level.

**Negative relationship between expression diversity and genetic diversity.** In our data, we found evidence for the negative relationship between expression diversity and genetic diversity, and even our newly sequencing data also support this conclusion. This seems inconsistent with some previous studies<sup>54</sup>. There was one explanation for the negative relationship between expression diversity and genetic diversity. *M. lutarioriparius* experienced chromosomal duplication in its evolution history<sup>55,56</sup>, therefore many genes had multiply copy number in genome. Duplicated genes often significantly increase gene expression diversity within and between species compared with single-copy genes<sup>57,58</sup>. However, duplicated genes have relatively weak genetic diversity compared with single-copy genes, since they evolve slower than single-copy genes<sup>59</sup>. Thus, the negative relationship between genetic diversity and expression diversity was observed in our results.

We speculated expression diversity may be an important contributor to adaptation to changing environment through the mechanism underlying phenotypic robustness, in which the process of development is balanced in an unpredictable world. Especially some potential adaptation-related genes with both low genetic diversity and high expression diversity had been detected in our previous studies<sup>47,48</sup>. Previous study found that reduced expression diversity within regulatory networks could promote the accumulation of genetic variation<sup>60</sup>. Thus we speculated that expression diversity had played the potential role in canalization to the environment. Under the balance of genetic variation and expression variation, *M. lutarioriparius* had experienced long-term natural selection and stayed at the state of canalization in its native region. In such case, this relationship between genetic and expression diversity could explain the genetic canalization mechanism that species had kept relatively invariant when individuals of the same single or multilocus genotype differ in their genetic background. Genes with higher expression diversity may have buffered the effects of lower genetic variation, and thus maintained the stability of phenotype in population consequently.

**Relationship between expression variation and genetic diversity.** In our data, large expression variation ( $E_p$  ratio  $< 0.5$  or  $E_p$  ratio  $> 2$ ) was observed mainly in the genes without SNP, and  $E_p$  ratio tended to decrease with the increase of genetic diversity  $\pi$  (Fig. 5). This implied that the genes without SNP or having low genetic diversity were more susceptible to environmental changes in expression and were subjected to high environment effect<sup>4</sup>, which also supported that stress or environment responsive genes were often genetically conserved among populations and species<sup>4,61</sup>. Since the genes with low genetic diversity had high environment effect on expression, it can be inferred that gene expression variation was regulated by environment via trans-regulation.

In addition, out of the genes with and without SNP, 1.6% and 4.3% had  $E_p$  ratio of over 10, respectively, the extent to which may be considered as stress or environment responsive genes. Thus, these genes with large  $E_p$  ratio may represent adaptive responses to environmental change, and they were expected to experience purifying or strong positive selection and to harbor lower genetic diversity in population.

**Environment was the leading contributor to expression variation.** Previous studies showed that environment was one of the most decisive factors on expression variation<sup>4</sup>. It was also argued that the environmental effect was stronger than the genetic effect<sup>25,62</sup>. To analyze the relative contribution of these factors, we utilized all transcript data and phased haplotype data respectively. We found the genes with environment effect were more frequent compared with those with population effect or haplotype effect, which supported that environment was the dominant contributor to gene expression variation. In our data, genetic diversity limitedly contributed to expression variation, and we found population stratification even had a much weaker role by comparing the

relative proportion of genes with haplotype effect and population effect. This was consistent with the results of STRUCTURE analysis<sup>48</sup>, and suggested that although these plants had experienced long-period local adaptation in their native habitats, there was weak population structure and there had been frequent gene flow in these *M. lularioriparius* populations<sup>63</sup>.

In our data, we found that environment was the leading contributor to expression variation compared with genetic diversity and population effects. It seems there was an amplification of environment effect and shrinkage of genetic effect and population effect compared with previous studies<sup>64–68</sup>. The main reason may be that the natural field conditions we adopted included much more variety of stresses than strictly controlled environment previous studies usually used. Since the haplotype had limited contribution to gene expression variation, we inferred that the adaptability to new environment was mainly attributed to some other factors rather than genetically induced adaptive genes, which was consistent with our result that environmentally induced expression diversity played a potential role in adaptation<sup>48</sup>.

**The implication of gene expression regulation in the process of adaptation to a new environment.** Our previous study reported that gene expression diversity played a potentially positive role in adaptation to new environment by comparing the genetic diversity and expression diversity. We ranked Pfam families by the proportion of genes with enlarged  $E_d$  in QG, and found that the top five Pfam families were Legume lectin domain, WRKY DNA domain, 2OG-Fe(II) oxygenase superfamily, non-haem dioxygenase in morphine synthesis N-terminal, and Sugar (and other) transporter (Supplementary Fig. S3). The increased expression diversity of WRKY gene family, 2OG-Fe(II) oxygenase superfamily, and Sugar transporters, which could play important roles correlating with the environment, could help species to adapt to the changing environment<sup>48</sup>. Here we found high proportion of genes with environment and environment by haplotype (or genetic) effects on gene expression, which were stress responsive genes and local adaptation genes, respectively. Our data implied that gene expression mainly regulated by environment and haplotype contributed to adaptation to new environment. Our result may facilitate understanding of global gene expression variation in the process of plant adaptation to new environment and supply evidence for the contribution of gene expression variation to adaptation.

## Materials and Methods

**Data resource.** The data were generated from our previous publication<sup>48</sup>. In brief, we collected *M. lularioriparius* across its natural distribution and planted in two experimental fields, one near its native habitat JH and the other at the domestication site QG. The location of these materials had been described in previous study<sup>48</sup>. The same 14 populations of *M. lularioriparius*, 3 individuals for each population were both randomly sampled in JH and QG. The fourth mature leaf of each individual at the same growth stage between two sites was chosen for RNA-seq. The 100-bp paired-end library was constructed and sequencing was performed on Illumina HiSeq 2000. Since in both sites 2 individuals were discarded due to the quality of their reads, a total of 40 individuals in each site were used for transcriptomic analysis ultimately. The raw data had been released at NCBI's Short Read Archive under three BioProjects, PRJNA227191, PRJNA227195, and PRJNA226258. We trimmed and filtered out reads showing more than 3% of bases with quality less than Q20. Then a high quality reference transcriptome was obtained using the population-based assembling method<sup>48</sup>. RNA-seq of *M. lularioriparius* from each of the two experimental fields generated a total of ~2.76 billion 80 bp paired-end reads after the quality control. Sequence coverage for individual samples of the 80 individuals of *M. lularioriparius* related to reference transcriptome ranged from 41.2% to 74.7%, with an average of 60.4% for the 80 individuals of *M. lularioriparius*. Furthermore, the sequencing depth was saturated when the number of 80-bp reads of an individual used for assembly reached about 40 million<sup>48</sup>.

**Population expression level analysis.** Clean reads of each individual were mapped to the Bowtie-build indexed reference transcriptome of *M. lularioriparius* using TopHat and Cufflinks<sup>69,70</sup>. After excluding the genes with half individual with expression level equal to zero, 15367 genes remained. Expression level and population expression diversity were estimated as the formula  $E_p = \frac{\sum_{i=1}^n E_i}{n}$  and  $E_d = \frac{\sum_{i=1}^n |E_i - E_p|}{(n-1)E_p}$ , where  $n$  represents the number of individuals,  $E_i$  represents the FPKM of a given gene of the  $i$ th individual in the population and  $E_p$  represents the expression level of a given gene. The expression level in population for genes in near native site JH and transplanted or target domestic site QG were abbreviated as  $E_{p(JH)}$  and  $E_{p(QG)}$  respectively, and the expression diversity for those genes were abbreviated as  $E_{d(JH)}$  and  $E_{d(QG)}$  respectively. The Wilcoxon test implemented in R 3.2.1 was applied to compare the gene expression level  $E_p$ , gene expression diversity  $E_d$ , and the ratio of  $E_{p(QG/JH)}$  among divided groups.

SNPs were identified using SAMtools with default settings<sup>71</sup>. After excluded SNPs with quality score  $\leq 10$ , missing data  $\geq 10\%$  or minor allele frequency  $\leq 0.05$ , the remained SNPs were used for further study. Genetic diversity ( $\pi$ ) for each gene of *M. lularioriparius* was calculated based on remained SNPs. The genetic diversity for genes in JH and QG were abbreviated as  $\pi_{(JH)}$  and  $\pi_{(QG)}$  respectively. As the native site, JH was used as the denominator for calculating related ratios between two environments.

To distinguish the extent of conservation, we considered genes with  $E_p$  ratio ranged from 0.5 to 2 as conserved  $E_p$  genes, and genes with  $E_d$  change ( $E_{d(QG)} - E_{d(JH)}$ ) of  $-0.25$  to  $0.25$  as conserved  $E_d$  genes.

**Haplotype inference and ANOVA analysis of haplotypes, population and environment effects.** Population transcriptome SNP data sets were used to infer gene haplotype for individual from population data. We used PHASE v2.1.1 to analyze the transcriptome data of the 80 individuals and to infer haplotypes<sup>72</sup>. PHASE was based on the approximate coalescent prior, which considered that the gene sequence of a mutant offspring differs only slightly from the progenitor sequence, and takes both homozygous and heterozygous positions

into account when considering whether close-matching haplotypes are sought. Considering the high heterozygosity of *M. lutheri* genome, we inferred the haplotype only for genes with less than 10 SNPs to ensure the accuracy of inference. For each gene, SNPs of all 80 individuals were pooled as input data for PHASE, and SNPs were ordered according to their relative position on the genes. The new model that makes explicit allowance for recombination was chosen for haplotype reconstruction with the default value for the initial estimate of the background recombination parameter of 0.0004. The program was implemented with default parameters (the default number of iterations is 100, with 100 burn-in, and a thinning interval of 1) to obtain the haplotype group with the highest probability. The haplotype group of best reconstruction with the highest probability in each gene was considered as the “best guess” estimate of the true underlying patterns of haplotype structure and was used in the analyses of haplotype expression.

Two-way ANOVA was carried out to study the environment effect and population effect on expression variation. In this analysis, the two-factor referred to the 14 populations and two environments. Additionally, since the specific haplotype of individuals for genes was inferred from population, a two-way ANOVA was also carried out to study the environment effect and haplotype effect on expression variation. Thus, in each gene, we fit a fixed effect general linear model including a term for population, haplotype, environment and their interaction. In all analyses, we controlled for multiple testing using a positive false discovery rate of 0.05.

## References

1. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).
2. Des Marais, D. L., Hernandez, K. M. & Juenger, T. E. Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. *Ann. Rev. Ecol. Syst.* **44**, 5–29 (2013).
3. Fay, J. C., McCullough, H. L., Sniegowski, P. D. & Eisen, M. B. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* **5**, R26 (2004).
4. Lasky, J. R. *et al.* Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **31**, 2283–2296 (2014).
5. Angers, B., Castonguay, E. & Massicotte, R. Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. *Mol. Ecol.* **19**, 1283–1295 (2010).
6. Skinner, M. K., Mohan, M., Haque, M. M., Zhang, B. & Savenkova, M. I. Epigenetic transgenerational inheritance of somatic transcriptomes and epigenetic control regions. *Genome Biol.* **13**, R91 (2012).
7. Ackermann, M., Sikora-Wohlfeld, W. & Beyer, A. Impact of natural genetic variation on gene expression dynamics. *PLoS Genet.* **9**, e1003514 (2013).
8. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250 (2009).
9. López-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* **9**, 583–593 (2008).
10. Berger, D., Bauerfeind, S. S., Blanckenhorn, W. U. & Schaefer, M. A. High temperatures reveal cryptic genetic variation in a polymorphic female sperm storage organ. *Evolution* **65**, 2830–2842 (2011).
11. Kim, S. Y., Noguera, J. C., Tato, A. & Velando, A. Vitamins, stress and growth: the availability of antioxidants in early life influences the expression of cryptic genetic variation. *J. Evol. Biol.* **26**, 1341–1352 (2013).
12. Ledon-Rettig, C. C., Pfennig, D. W. & Crespi, E. J. Diet and hormonal manipulation reveal cryptic genetic variation: implications for the evolution of novel feeding strategies. *Proc. Royal Soc. B Biol. Sci.* **277**, 3569–3578 (2010).
13. McGuigan, K., Nishimura, N., Currey, M., Hurwit, D. & Cresko, W. A. Cryptic genetic variation and body size evolution in threespine stickleback. *Evolution* **65**, 1203–1211 (2011).
14. Anderson, P. S., Renaud, S. & Rayfield, E. J. Adaptive plasticity in the mouse mandible. *BMC Evol. Biol.* **14**, 85 (2014).
15. Chevin, L. M., Collins, S. & Lefèvre, F. Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field. *Funct. Ecol.* **27**, 967–979 (2013).
16. Dal Santo, S. *et al.* The plasticity of the grapevine berry transcriptome. *Genome Biol.* **14**, r54 (2013).
17. Fraser, H. B. Gene expression drives local adaptation in humans. *Genome Res.* **23**, 1089–1096 (2013).
18. Koenig, D. *et al.* Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. USA* **110**, E2655–E2662 (2013).
19. Pfennig, D. W. *et al.* Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol. Evol.* **25**, 459–467 (2010).
20. Swanson-Wagner, R. *et al.* Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. USA* **109**, 11878–11883 (2012).
21. Whitehead, A. & Crawford, D. L. Variation within and among species in gene expression: raw material for evolution. *Mol. Ecol.* **15**, 1197–1211 (2006).
22. Newman, J. R. S. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
23. Li, Y. *et al.* Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* **2**, 2155–2161 (2006).
24. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
25. Gibson, G. The environmental contribution to gene expression profiles. *Nat. Rev. Genet.* **9**, 575–581 (2008).
26. Hayden, E. J., Ferrada, E. & Wagner, A. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* **474**, 92–U120 (2011).
27. Ledon-Rettig, C. C., Pfennig, D. W., Chunco, A. J. & Dworkin, I. Cryptic genetic variation in natural populations: A predictive framework. *Integr. Comp. Biol.* **54**, 783–793 (2014).
28. Masel, J. Cryptic genetic variation is enriched for potential adaptations. *Genetics* **172**, 1985–1991 (2006).
29. Paaby, A. B. & Rockman, M. V. Cryptic genetic variation: evolution's hidden substrate. *Nat. Rev. Genet.* **15**, 247–258 (2014).
30. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
31. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
32. Grativol, C., Hemerly, A. S. & Gomes Ferreira, P. C. Genetic and epigenetic regulation of stress responses in natural plant populations. *Biochim. Biophys. Acta.* **1819**, 176–185 (2012).
33. Sides, C. B. *et al.* Revisiting Darwin's hypothesis: does greater intraspecific variability increase species' ecological breadth? *Am. J. Bot.* **101**, 56–62 (2014).
34. McInerney, J. O. The causes of protein evolutionary rate variation. *Trends Ecol. Evol.* **21**, 230–232 (2006).
35. Zuckerkandl, E. Evolutionary processes and evolutionary noise at the molecular level. *J. Mol. Evol.* **7**, 269–311 (1976).



36. Park, S. G. & Choi, S. S. Expression breadth and expression abundance behave differently in correlations with evolutionary rates. *BMC Evol. Biol.* **10**, 241 (2010).
37. Drummond, D. A., Raval, A. & Wilke, C. O. A single determinant dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* **23**, 327–337 (2006).
38. Zhang, L. & Li, W.-H. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol. Biol. Evol.* **21**, 236–239 (2004).
39. Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008).
40. Choi, S. S. & Hannenhalli, S. Three independent determinants of protein evolutionary rate. *J. Mol. Evol.* **76**, 98–111 (2013).
41. Liu, W. *et al.* Long-term water balance and sustainable production of Miscanthus energy crops in the Loess Plateau of China. *Biomass and Bioenergy* **62**, 47–57 (2014).
42. Liu, W. & Sang, T. Potential productivity of the Miscanthus energy crop in the Loess Plateau of China under climate change. *Environ. Res. Lett.* **8**, 044003 (2013).
43. Liu, W., Yan, J., Li, J. & Sang, T. Yield potential of Miscanthus energy crops in the Loess Plateau of China. *GCB Bioenergy* **4**, 545–554 (2012).
44. Mi, J. *et al.* Carbon sequestration by Miscanthus energy crops plantations in a broad range semi-arid marginal land in China. *Sci. Total Environ.* **496**, 373–380 (2014).
45. Yan, J. *et al.* Variability and adaptability of Miscanthus species evaluated for energy crop domestication. *GCB Bioenergy* **4**, 49–60 (2012).
46. Yan, J. *et al.* High photosynthetic rate and water use efficiency of Miscanthus lutarioriparius characterize an energy crop in the semiarid temperate region. *GCB Bioenergy* **7**, 207–218 (2015).
47. Fan, Y. *et al.* Transcriptome-wide characterization of candidate genes for improving the water use efficiency of energy crops grown on semiarid land. *J. Exp. Bot.* **66**, 6415–6429 (2015).
48. Xu, Q. *et al.* Population transcriptomics reveals a potentially positive role of expression diversity in adaptation. *J. Integr. Plant Biol.* **57**, 284–299 (2015).
49. Bhosale, R. *et al.* Predicting gene function from uncontrolled expression variation among individual wild-type Arabidopsis plants. *Plant Cell* **25**, 2865–2877 (2013).
50. Liao, B.-Y. & Zhang, J. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol. Biol. Evol.* **23**, 1119–1128 (2006).
51. Atkin, O., Botman, B. & Lambers, H. The relationship between the relative growth rate and nitrogen economy of alpine and lowland Poa species. *Plant Cell Environ.* **19**, 1324–1330 (1996).
52. Zur, H. & Tuller, T. Strong association between mRNA folding strength and protein abundance in *S. cerevisiae*. *EMBO Rep.* **13**, 272–277 (2012).
53. Chursov, A., Frishman, D. & Shneider, A. Conservation of mRNA secondary structures may filter out mutations in *Escherichia coli* evolution. *Nucleic Acids Res.* gkt507 (2013).
54. Jiang, S.-Y., Ma, Z., Vanitha, J. & Ramachandran, S. Genetic variation and expression diversity between grain and sweet sorghum lines. *BMC Genomics* **14**, 18 (2013).
55. Hodkinson, T., Klaas, M., Jones, M., Prickett, R. & Barth, S. Miscanthus: a case study for the utilization of natural genetic variation. *Plant Genet. Resour.* **13**, 219–237 (2015).
56. Ma, X.-F. *et al.* High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid Miscanthus sinensis. *PLoS One* **7**, e33821 (2012).
57. Ha, M., Kim, E.-D. & Chen, Z. J. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc. Natl. Acad. Sci. USA* **106**, 2295–2300 (2009).
58. Gu, Z., Rifkin, S. A., White, K. P. & Li, W.-H. Duplicate genes increase gene expression diversity within and between species. *Nat. Genet.* **36**, 577–579 (2004).
59. Jordan, I. K., Wolf, Y. I. & Koonin, E. V. Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol. Biol.* **4**, 1 (2004).
60. Garfield, D. A. *et al.* The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. *PLoS Biol.* **11**, e1001696 (2013).
61. Des Marais, D. L. *et al.* Physiological genomics of response to soil drying in diverse Arabidopsis accessions. *Plant Cell* **24**, 893–914 (2012).
62. Dermitzakis, E. T. & Stranger, B. E. Genetic variation in human gene expression. *Mamm. Genome* **17**, 503–508 (2006).
63. Yan, J. *et al.* Genetic variation and bidirectional gene flow in the riparian plant Miscanthus lutarioriparius, across its endemic range: Implications for adaptive potential. *GCB Bioenergy* doi: 10.1111/gcbb.12278 (2015).
64. Granados-Cifuentes, C., Bellantuono, A. J., Ridgway, T., Hoegh-Guldberg, O. & Rodriguez-Lanetty, M. High natural gene expression variation in the reef-building coral *Acropora millepora*: potential for acclimative and adaptive plasticity. *BMC Genomics* **14**, 228 (2013).
65. Lai, Z., Kane, N. C., Zou, Y. & Rieseberg, L. H. Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. *Genetics* **179**, 1881–1890 (2008).
66. Müller, L. *et al.* Population transcriptomics of *Drosophila melanogaster* females. *BMC Genomics* **12**, 81 (2011).
67. Storey, J. D. *et al.* Gene-expression variation within and among human populations. *Am. J. Hum. Genet.* **80**, 502–509 (2007).
68. Swanson-Wagner, R. *et al.* Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. USA* **109**, 11878–11883 (2012).
69. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
70. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
71. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
72. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162–1169 (2003).

## Acknowledgements

We thank the Beijing Center for Physical and Chemical Analysis for generating the sequencing data, the Beijing Computing Center for assisting with computational infrastructure for data analysis. The work was supported by grants from the National Natural Science Foundation of China (No. 31500186, No. 31400284) and Key Program of the National Natural Science Foundation of China (No. 91131902).

## Author Contributions

T.S. conceived and designed the experiments. Q.X., C.Z. and W.L. carried out the experiments and analyzed data. Y.F. and Z.S. prepared figure. Q.X., J.Y. and T.S. wrote the manuscript. All authors reviewed the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Xu, Q. *et al.* Population transcriptomics uncovers the regulation of gene expression variation in adaptation to changing environment. *Sci. Rep.* **6**, 25536; doi: 10.1038/srep25536 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>