

RESEARCH ARTICLE

Open Access



Systematic comparison of lncRNAs with protein coding mRNAs in population expression and their response to environmental change

Qin Xu^{1†}, Zhihong Song^{1,3†}, Caiyun Zhu^{2,3}, Chengcheng Tao^{1,3}, Lifang Kang¹, Wei Liu², Fei He⁴, Juan Yan⁵ and Tao Sang^{1,2*} 

Abstract

Background: Long non-coding RNA (lncRNA) is a class of non-coding RNA with important regulatory roles in biological process of organisms. The systematic comparison of lncRNAs with protein coding mRNAs in population expression and their response to environmental change are still poorly understood. Here we identified 17,610 lncRNAs and calculated their expression levels based on RNA-seq of 80 individuals of *Miscanthus lutarioriparius* from two environments, the nearly native habitats and transplanted field, respectively.

Results: lncRNAs had significantly higher expression diversity and lower expression frequency in population than protein coding mRNAs in both environments, which suggested that lncRNAs may experience more relaxed selection or divergent evolution in population compared with protein coding RNAs. In addition, the increase of expression diversity for lncRNAs was always significantly higher and the magnitude of fold change of expression in new stress environment was significantly larger than protein-coding mRNAs. These results suggested that lncRNAs may be more sensitive to environmental change than protein-coding mRNAs. Analysis of environment-robust and environment-specific lncRNA-mRNA co-expression network between two environments revealed the characterization of lncRNAs in response to environmental change. Furthermore, candidate lncRNAs contributing to water use efficiency (WUE) identified based on the WUE-lncRNA-mRNA co-expression network suggested the roles of lncRNAs in response to environmental change.

Conclusion: Our study provided a comprehensive understanding of expression characterization of lncRNAs in population for *M. lutarioriparius* under field condition, which would be useful to explore the roles of lncRNAs and could accelerate the process of adaptation in new environment for many plants.

Keywords: Population transcriptome, lncRNAs, *Miscanthus lutarioriparius*, Co-expression, Environmental response, Expression diversity

* Correspondence: sang@ibcas.ac.cn

†Equal contributors

¹Key Laboratory of Plant Resources and Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

²State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

Full list of author information is available at the end of the article



Background

The genome complexity of higher eukaryotes is revealed by a high proportion of non-coding for proteins regulatory elements, of which a class of processed long non-coding RNA (lncRNA) is attracting increasing attention. The lncRNAs are highly heterogeneous transcripts in length, varying from 200 bp to tens of thousands of nucleotides. They pervasively distribute in genomes, not only in intron and intergenic region, but also in some antisense transcripts, pseudogenes and retrotransposons [1]. lncRNAs have lower level of sequence conservation than protein-coding mRNAs [2]. They were originally considered as transcriptional by-products or “expression noise” of protein coding genes and were often dismissed in analyses of transcriptome [3], however recent studies uncovered the increasing body of evidences that lncRNAs strictly regulated and played important roles in biological process of organisms [4].

Recently, a great number of lncRNAs had been identified in animals and humans [5–11]. They participated in numerous biological processes, such as X-chromosome inactivation [12], and human diseases [6–8, 13, 14], although only a few of them have been functionally annotated. Compared with researches about lncRNAs in animals and humans, studies on plants are relatively infrequent, and only restrict to some model plants, such as *Arabidopsis*, rice, maize, and *Populus* [15–19]. Most of the annotated lncRNAs are related with regulation of the development of plants.

lncRNAs are found to be differentially expressed between different organs, development stage and environment. The fact that lncRNAs are expressed in a strong state-specific manner was revealed by the recent analyses of lncRNAs expression under abiotic stress conditions [20–22]. Thus it is reasonable to identify the lncRNAs associated with stressful environment by comparing the lncRNA expression pattern across environments. Indeed, the identified differentially expressed lncRNAs during abiotic stress had suggested the important role of lncRNAs in plant stress response [17]. To date, some powdery mildew infection and heat stress-responsive lncRNAs were identified in wheat [23], and some lncRNAs induced by drought, high salt, cold, and abscisic acid were identified in *Arabidopsis* [24]. These studies provide us a good starting point of understanding the role of lncRNAs in the process of abiotic stress tolerance in plants [21].

Although several studies have identified plant lncRNAs that are activated under various environmental conditions, only a few of them have been characterized in population under field condition. Due to the low level of expression frequency and strong state-specific expression manner of lncRNAs, only a limited number of lncRNAs can be identified using RNA-seq for a few samples, and the

components that are important in specific condition may been overlooked. Genome and population wide approaches could provide a larger scale of lncRNAs identification and a broader picture of the role of lncRNAs as regulators of the massive protein coding genes responding to stress under field environment.

With markedly accelerated consumption of fossil energy and its worsening environmental impact, the development of energy crops to provide renewable feedstock for clean energy and safe material offers an appealing solution to the sustainability problems facing the society [25]. The extent to which this can solve the problem depends heavily on whether the crops can be produced in large scales without threat to the food security and at little cost of natural ecosystem function [26]. A promising candidate of dedicated energy crops meeting these requirements is *Miscanthus* [25, 27], a group of C4 perennial grasses capable of producing high biomass on marginal land [28]. The challenge now becomes whether and how the new crop can adapt to meet environmental requirements. The integration of population genetics and new genomic technologies holds a great potential to meet the challenge.

Of more than a dozen wild *Miscanthus* species, *M. lutarioriparius* that produces the highest biomass stood out as a desirable wild progenitor for crop adaptation. Fourteen populations of *M. lutarioriparius* were collected across the distributional range of this endemic species in central China and were planted in two experimental fields in 2009, one located aside its native habitat in Jiangxia of the Hubei Province (JH) and the other located in Qingyang of the Gansu Province (QG) in the Loess Plateau [29]. Loess Plateau, one of the most seriously eroded regions of the world, possesses remarkably large area of marginal land for producing second-generation energy crops such as *Miscanthus* [28]. It has been shown that *M. lutarioriparius* was not only able to establish in QG but also to produce higher biomass than in JH, while the annual precipitation and temperature is nearly two-third and 10-degree lower in QG than in JH [29–36].

In this study, we used population RNA-seq data to identify lncRNAs in *Miscanthus* leaves and to investigate their roles in adaptation to environmental change. Here, we obtained a comprehensive list of 17,610 lncRNAs. Systematic expression analysis revealed that the expression of lncRNAs displayed more variation than protein coding mRNAs. Integrating with our published datasets of water use efficiency (WUE) [32], we highlighted potential contributions of lncRNAs to improve WUE and adaptation to environmental change.

Results

De novo assembly and identification of putative lncRNAs

We performed RNA-seq to investigate 80 individuals of *Miscanthus* transcriptomes. Non-directional paired-end

RNA-seq was carried out and ~2.76 billion 80 bp paired-end reads were generated after quality control. Population transcripts were assembled using the Pipeline PopART which combined multiple *kmers* and multiple individuals [33, 37]. Totally we obtained 818,491 transcripts, out of which 18,503 unique mRNAs from the 80 individuals were identified by blasting against the latest annotation of transcriptome references of close related species including *Sorghum bicolor*, *Zea mays*, *Oryza sativa*, and *Brachypodium distachyon*. To identify lncRNAs, we used Coding Potential Calculator software to evaluate the protein-coding potential for transcripts longer than 200 bp [38], and transcripts with the evidence of protein-coding potential were discarded. All the remaining transcripts were aligned against Rfam database to discriminate lncRNAs from previously annotated miRNA, rRNA, or other small noncoding RNA transcripts [39]. To obtain a more reliable data of lncRNAs and a better comparison between two environments, we discarded the lncRNA candidates that expressed in less than 20 individuals out of the 80 individuals. Under this criterion, a total of 17,610 putative lncRNAs were remained. This amount was similar to that of lncRNAs in *Arabidopsis*, rice and maize [16, 24].

The lengths of lncRNAs ranged from 200 to 5196 bp, the majority (92%) of which were around 200–600 bp (Fig. 1a and Additional file 1). The length of lncRNA was 683 bp in average and 352 bp in median, respectively, which was longer than those in *Arabidopsis* [24]. The length of protein coding mRNAs was 1601bp in average and 1425bp in average respectively. lncRNAs were significantly shorter than mRNAs in length (Wilcoxon test or *t*-test, $P < 2.2e-16$; Fig. 1a). In addition, GC content for lncRNAs and mRNAs were 46% and 51% in median, respectively. The lncRNAs showed significant lower GC content than mRNAs (Wilcoxon test or *t*-test, $P < 2.2e-16$; Fig. 1a and Additional file 1). In order to examine whether these candidate lncRNAs sequence

assembly were reliable, 8 lncRNAs were randomly selected and validated by full-length PCR experiments. PCR results showed that the sequence identity ranged from 94% to 100% compared with the reference sequences assembled from population data (Table 1 and Additional file 2). As the reference sequences of these lncRNAs were assembled from the population transcriptome data, there could be some single nucleotide polymorphisms or indels (insertions and deletions). For example, the sequence of lncRNA_MluLR14524 by PCR method has an 84bp deletion compared with the reference sequence derived from with RNA-seq, which may be due to an alternative splicing of this individual or a chimera between two similar lncRNAs in the assembling data. In general, the results suggested a high quality assembly of the lncRNAs and they could be used for further analysis.

Higher expression diversity of lncRNAs compared with that of mRNAs within and across environments

We first set out to characterize the global expression pattern of lncRNAs in JH and QG. The expression level of lncRNAs in population (E_p) was calculated. Then, the expression levels of lncRNAs and those of protein coding mRNAs were compared. It was found that E_{pS} of lncRNAs in QG shifted significantly toward to lower levels, suggesting an overall depression of lncRNAs expression from JH to QG. The change of lncRNAs expression between two environments was examined by E_p ratios of QG to JH for each lncRNA, and it was found that the resulting distribution of $\log_2(E_p \text{ ratios})$ deviated significantly from 0 and skewed toward bottom (Wilcoxon test or *t*-test, $P < 2.2e-16$; Fig. 2 and Additional file 3), indicating that lncRNAs decreased expression level in QG. Compared with protein coding mRNAs, lncRNAs had significantly lower E_p ratios (Wilcoxon test or *t*-test, $P < 2.2e-16$; Fig. 2, Additional files 3 and 4). Meanwhile, lncRNAs had a

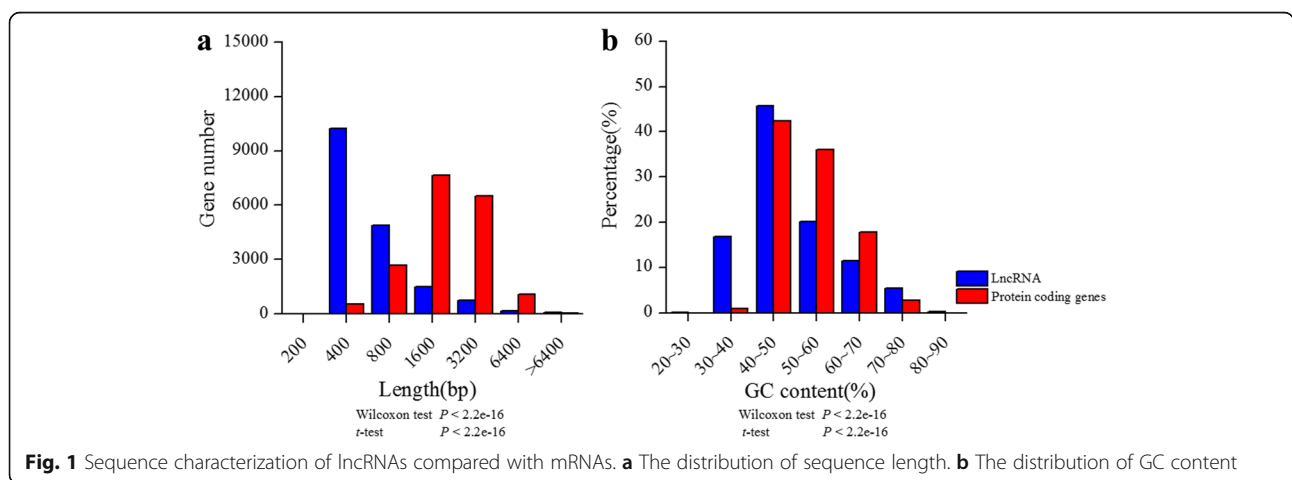


Table 1 Summary of the PCR validation for assembled lncRNAs

lncRNA	Assemble length	PCR length	Coverage	Identity
lncRNA_MluLR14385	960	907	94%	98%
lncRNA_MluLR14524	748	625	93%	94%
lncRNA_MluLR6328	687	593	86%	99%
lncRNA_MluLR15001	538	393	73%	99%
lncRNA_MluLR14468	586	516	88%	99%
lncRNA_MluLR14141	699	360	51%	100%
lncRNA_MluLR2250	781	281	35%	100%
lncRNA_MluLR6864	499	69	13%	100%

wider range of E_p ratios than that of protein coding mRNAs (Fig. 2, Additional files 3 and 4).

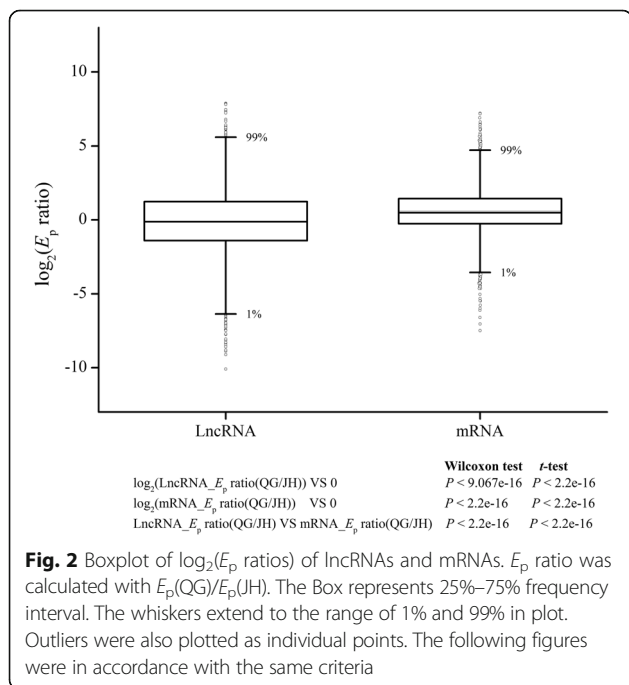
To evaluate and compare population variation of lncRNAs expression within and between environments, we calculated and compared expression diversity (E_d) between JH and QG. Compared with protein coding mRNAs, lncRNAs had significantly higher E_d in both environments (Wilcoxon test or t -test, $P < 2.2e-16$; Fig. 3a, Additional files 3 and 4). Moreover, the distributions of E_d of lncRNAs were compared between two environments, and the result showed that E_d were significantly higher in QG than in JH with 60.3% of E_d s being elevated in QG (Wilcoxon test or t -test, $P < 2.2e-16$; Fig. 3, Additional files 3 and 4). The significant differentiation was also detected by E_d ratios between JH and QG for each lncRNA, and $\log_2(E_d \text{ ratios})$ of lncRNAs were compared with 0, and the distribution of $\log_2(E_d \text{ ratios})$ significantly deviated from 0 and skewed toward top

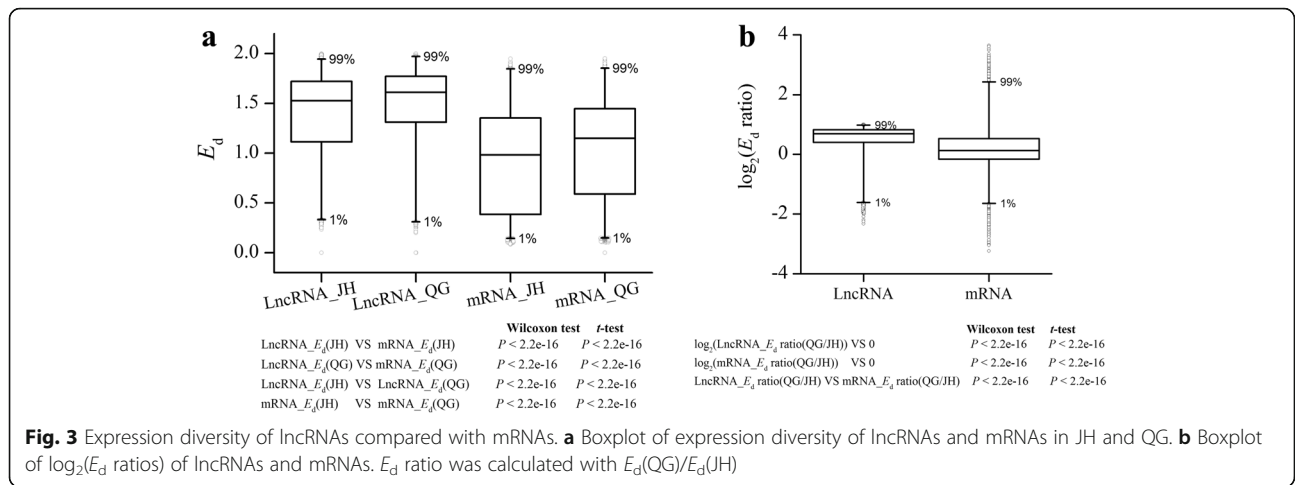
(Wilcoxon test or t -test, $P < 2.2e-16$; Fig. 3b, Additional files 3 and 4). In addition, the E_d ratios of lncRNAs were significantly higher than protein coding mRNAs (Wilcoxon test or t -test, $P < 2.2e-16$; Fig. 3b, Additional files 3 and 4). The higher expression diversity of lncRNAs may suggest a relaxed expression regulation compared with protein coding mRNAs.

To evaluate the expression variation of lncRNAs in the population, we measured the expression frequency for each lncRNA in two environments. It was found that lncRNAs expressed at significantly lower frequency than protein coding mRNAs in both environments (Wilcoxon test or t -test, $P < 2.2e-16$; Fig. 4a and Additional file 5). Expression frequency in QG showed a lower level than JH for both lncRNAs and mRNAs (Wilcoxon test, $P < 2.2e-16$; t -test $P < 0.01794$; Fig. 4a and Additional file 5), suggesting an overall decrease of both lncRNAs and mRNA expression frequency from JH to QG. The change of lncRNAs and mRNAs expression frequency between two environments was examined by difference of QG to JH for each lncRNA and mRNA, respectively. It was found that lncRNAs had a broader range of the difference than mRNAs (Wilcoxon test or t -test, $P < 2.2e-16$; Fig. 4b and Additional file 5). The lower level and larger difference of expression frequency for lncRNAs might suggest the loosen expression regulation compared with protein coding mRNAs.

Differential expression of lncRNAs in environment responsive regulation

Emerging evidences showed that lncRNAs participate in stress responsive regulation when plants are in the face of stressful environment [15, 23], thus we analyzed the differential expression and variation of lncRNAs between the two environments. First, we plotted the number of individuals in which lncRNA expressed between two environments (Fig. 4c). Only 1376 (7.8%) of the 17610 lncRNAs expressed in all individuals (Fig. 4d and Additional file 5), and only 45 out of the 80 individuals could be detected the lncRNA expression in median (Additional file 5), suggesting the specific and differential expression manner of lncRNAs in our population. Specially, it was found that 2 and 59 lncRNAs were detected only in QG and JH (Additional file 5), respectively. These newly arisen lncRNAs in those regions or environment-specific lncRNAs may be related with environmental adaptation. It was also possible that they just lost from the other region and had no function. In addition, except the low frequency of lncRNAs which expressed in less than 10 individuals in JH, the expression frequency of lncRNAs between 2 environments was positively correlated (Fig. 4c), suggesting that most of these lncRNAs expressed at robust or similar

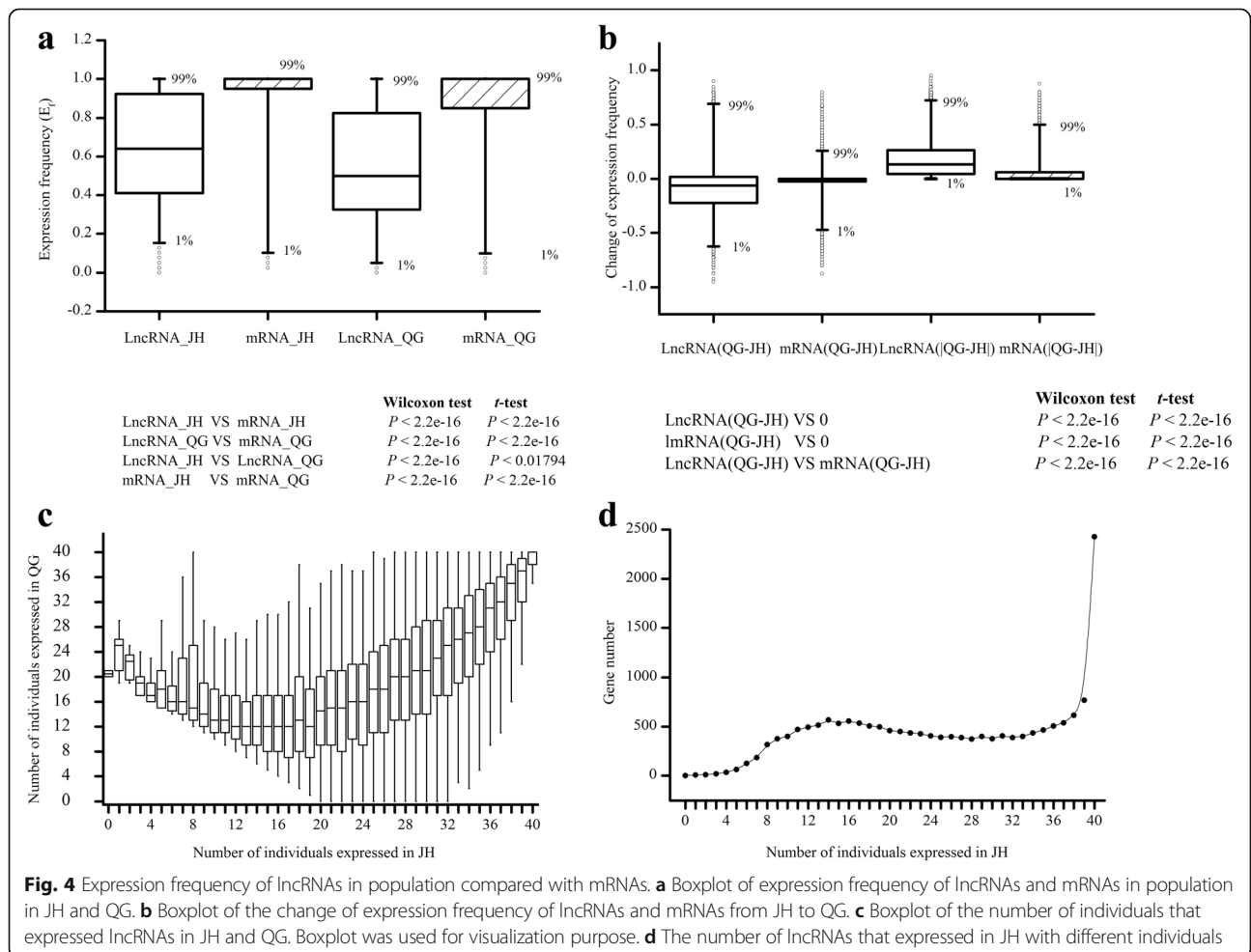




expression frequency between environment, and also suggesting the accuracy of the estimation of lncRNAs expression.

On the other hand, the differentially expressed genes between two sites were tested to identify candidate genes responding to environmental stress. The expression level

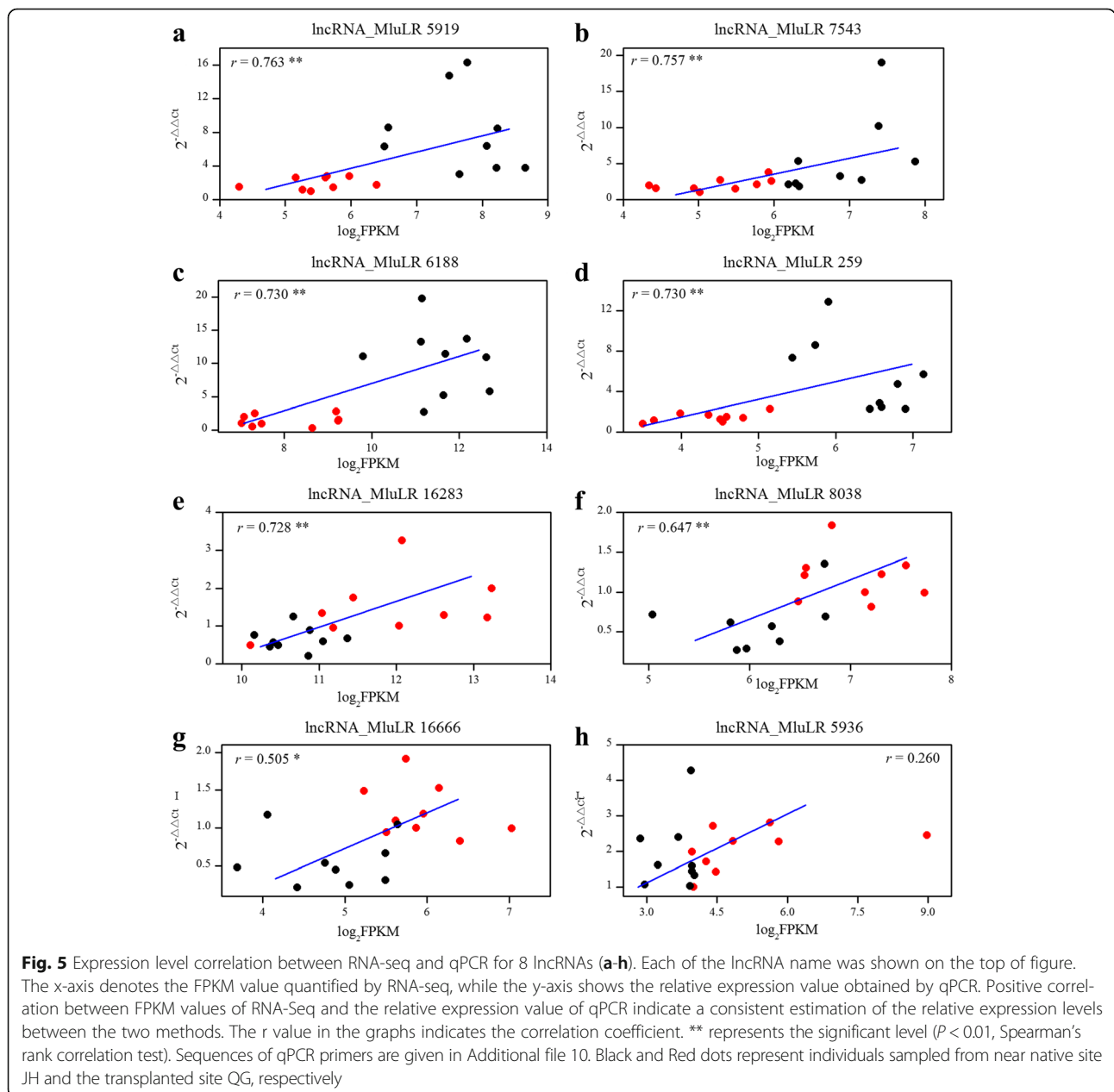
changes (E_p ratios) of 17,610 lncRNAs between two environments were calculated. The magnitude of fold change (\log_2) in the expression level of lncRNAs under new environment was observed between -9 to 10.8 (Additional file 3). We obtained a total of 2,063 lncRNAs that were over 2- fold up- or down- regulated with FDRs



less than 0.01, including 821 up-regulated and 1242 down-regulated lncRNAs (Additional file 3). Because genes with substantially increased expression diversity in new stress environment QG could have increased phenotypic variation upon which natural and artificial selection could act to improve adaptability of the *Miscanthus* species, we also calculated the expression diversity changes (E_d ratios) of lncRNAs between two environments for lncRNAs. The magnitude of fold change in the expression diversity of lncRNAs under stress condition ranged from 0.12 to 6.65 (Additional file 3). A total of 931 stress responsive lncRNAs with E_d ratio of more than 2-fold change were obtained. These lncRNAs may have

higher potential in regulation of genes expression for contribution to adaptation.

To validate the reliability of the expression level, quantitative real-time PCR (qPCR) was performed to assay the accuracy of the RNA-seq by using 9 individuals from each field site for 8 lncRNAs which were differentially expressed between two environments. The relative quantitation of expression levels which was calculated using $2^{-\Delta\Delta Ct}$ method was compared with the FPKM values in RNA-seq data in each sample [32, 40–42] (Fig. 5). It was found that the relative expression levels determined by the two methods were significantly correlated except for lncRNA_MluLR5936 which had outlier value in 2



samples (Fig. 5). The results confirmed that the relative gene expression levels were reliable and could be used for further analysis.

Functions of differentially expressed lncRNAs based on lncRNA-mRNA co-expression network

In order to explore the correlation of lncRNAs and mRNAs, we used the transcripts with E_d more than 0.6 to perform pairwise correlation. In total, 3,086 lncRNA-mRNA pairs in both environments had the correlation coefficient of more than 0.9 ($P < 0.001$; Table 2 and Additional file 6), which included 1,431 mRNAs and 1,601 lncRNAs, respectively. These lncRNA-mRNA expression pairs which would represent the consensus or robust relationship between environments were constructed, and the max number of node was 8 for mRNAs and 33 for lncRNAs. These results suggested that lncRNAs in the robust relationship may play more pivotal roles in the process of regulation network. In addition, there were 215,251 lncRNA-mRNA pairs having the correlation coefficient of more than 0.95 in QG but smaller than 0.1 in JH. Similarly, there were 241,459 lncRNA-mRNA pairs having the correlation coefficient of more than 0.95 in JH but smaller than 0.1 in QG. These differentially co-expressed lncRNA-mRNA pairs may involve in environment-specific regulation ($P < 0.001$; Table 2).

The environmental-specific lncRNA-mRNA co-expression network between two environments can be used to identify key lncRNAs responding to environmental change. As genes with increased expression diversity in new environment may be relevant with adaptation, we filtered out 2003 lncRNAs and 4108 mRNAs with E_d ratio larger than 1.5. Finally 2 272 lncRNA-mRNA pairs with Spearman correlation coefficients larger than 0.7, including 1 052 mRNAs and 1 023 lncRNAs (Additional file 7), were identified for network construction. The similar amount of lncRNAs and mRNAs potentially responding to environmental change suggested that lncRNAs may play the roles as important as mRNAs in environmental adaptation. Two sub-networks containing 38 lncRNAs and 25 mRNAs with

lncRNA-mRNA connection degree more than 10 were presented in detail (Fig. 6a). Out of the two sub networks, 12 and 26 lncRNAs were found to be up-regulated and down-regulated respectively, while 16 and 9 mRNAs were up-regulated and down-regulated respectively. Our network showed that one lncRNA could co-express with multiple mRNAs and one mRNA could be correlated with multiple lncRNAs. The network implied a complex relationship between lncRNAs and mRNAs.

To reveal potential functions of the identified lncRNAs, we analyzed Gene Ontology (GO) terms and families of protein coding mRNA genes associated with the stress-responsive lncRNAs due to their regulating roles (Additional file 7). We detected significant enrichments of 17 families with more than 5 members and 8 GO terms in leaves under stressful environment. For examples, we found 5 categories of genes that Protein kinase domain, Ring finger domain, Cytochrome P450, WD domain, G-beta repeat had more than 12 members (Table 3). Based on the functional categories from the Pfam domain annotations, nine of the 25 protein coding mRNAs in the two sub-networks were found to be related with stress response (Fig. 6b).

Identification of key lncRNAs regulating water use efficiency (WUE)

Previous genome-wide association studies found hundreds of lncRNAs containing trait-associated SNPs, suggesting the putative contributions of lncRNAs to agricultural traits [16, 43], and we had identified 48 candidate genes whose expression were related with WUE in our previous study [32]. To find out and reveal the potential contribution of lncRNAs to WUE, we analyzed co-expression of lncRNAs and mRNAs using Spearman correlation coefficient for each candidate gene of WUE, and constructed the WUE-lncRNA-mRNA co-expression network (WUE-LMN). As a result, this co-expression network contained a total of 48 candidate genes, 3371 unique lncRNAs and 4277 edges (Additional file 8). The degree of the candidate genes ranged from 23 to 215, indicating that each of these genes associated with adaptation was regulated by

Table 2 Pairwise number of lncRNA-mRNA co-expression in the two environments

Parameter	Pairwise number	lncRNA number	mRNA number	Max nodes of lncRNA	Max nodes of mRNA	Implication
R(JH) > 0.9&R(QG) > 0.9	3086	1601	1431	33	8	Robust relationship between environment
R(JH) > 0.95&R(QG) > 0.95	929	598	553	14	7	
R(JH) > 0.95&R(QG) < 0.7	290705	6783	4180	396	241	Environmental response-related relationship
R(QG) > 0.95&R(JH) < 0.7	267035	7203	4245	317	179	
R(JH) < 0.1&R(QG) > 0.95	215251	6957	4096	308	169	
R(QG) < 0.1&R(JH) > 0.95	241459	6633	4059	374	227	

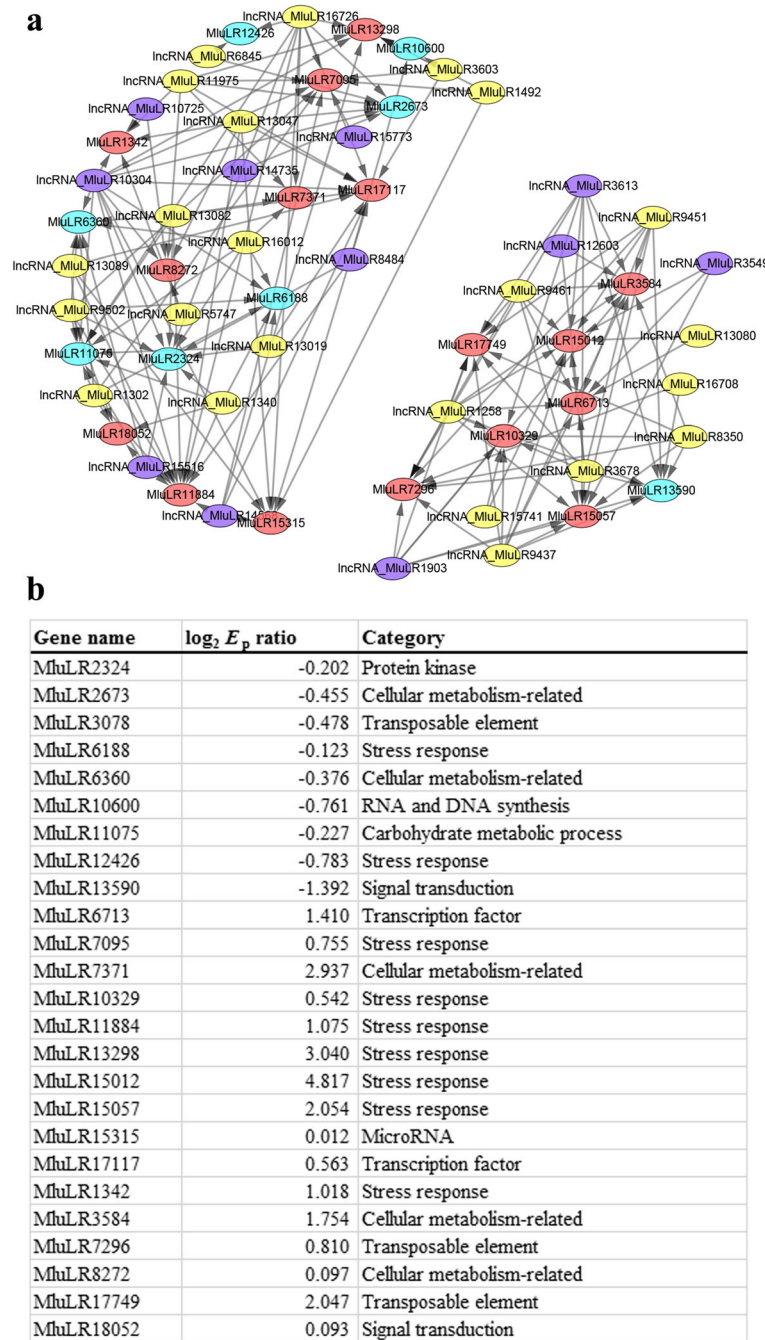


Fig. 6 **a** Visualization of environment responsive differential lncRNA-mRNA co-expression network. The nodes with red, blue, purple or green colors represent up-regulated mRNAs, down-regulated mRNAs, up-regulated lncRNAs and down-regulated lncRNAs. The size of node positively related with connection degree. **b** The fold change value and the functional categories of 25 mRNAs in the co-expression network

complex network involving multiple lncRNAs. We further filtered the co-expression network with more than five nodes for lncRNAs (Fig. 6a). Nine lncRNAs and 17 candidate genes for WUE were remained. And the remained 17 candidate genes for WUE were mainly functioned in abiotic stress responses, photosynthesis

and stomatal regulation (Fig. 7b). Of the 9 lncRNAs, 4 and 5 lncRNAs were up-regulated and down-regulated between two environments, respectively (Fig. 7c). We inferred that they may represent hub genes or regulators in WUE related pathway and played an important role in responding to environmental changes.

Table 3 The enrichments of families of the protein-coding genes identified by the environmental-specific lncRNA-mRNA co-expression network

Family	NO.
Protein kinase domain	13
Ring finger domain	13
Cytochrome P450	12
WD domain, G-beta repeat	12
RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	9
alpha/beta hydrolase fold	7
EF-hand domain pair	7
Alpha/beta hydrolase family	6
DnaJ domain	6
Protein phosphatase 2C	6
AMP-binding enzyme = Domain of unknown function (DUF4009)	5
Glutathione S-transferase, C-terminal domain = Glutathione S-transferase, N-terminal domain	5
GRAS family transcription factor	5
Myb-like DNA-binding domain	5
PPR repeat = PPR repeat family	5
Sugar (and other) transporter	5
Zinc finger, C3HC4 type (RING finger)	5

Discussion

The reliability of lncRNAs identification

Since *de novo* assembly is disadvantaged by its inability to account for incidents of structural alterations of mRNA transcripts, such as alternative splicing, *de novo* assembly of lncRNAs reference without a known genome faces even more huge difficulties due to lack of sequence conservation [2]. Here we reported population transcriptome-wide identification of lncRNAs with high reliability. We substantially improved the accuracy of the *M. lutarioriparius* lncRNAs assembly. First, we perform multiple assemblies with various *kmer* lengths, which could reach the best trade-off between the length and quantity of transcripts [44]. In addition, we used population transcriptome data to retain the best part of each one to form the final assembly [37]. The complementary effect of multiple individuals completed the transcriptome assembly in both transcript number and length [33]. Second, although lncRNAs often express in a low frequency, lncRNAs that expressed in less than 25% of individuals were filtered out. This procedure could greatly reduce the assembly errors and improve the reliability of the remained lncRNAs. Moreover, the sequences were further validated by PCR sequencing and the expression level of lncRNAs was proved by qPCR. The results suggested the sequence of lncRNAs was high-quality assembled and the expression level was

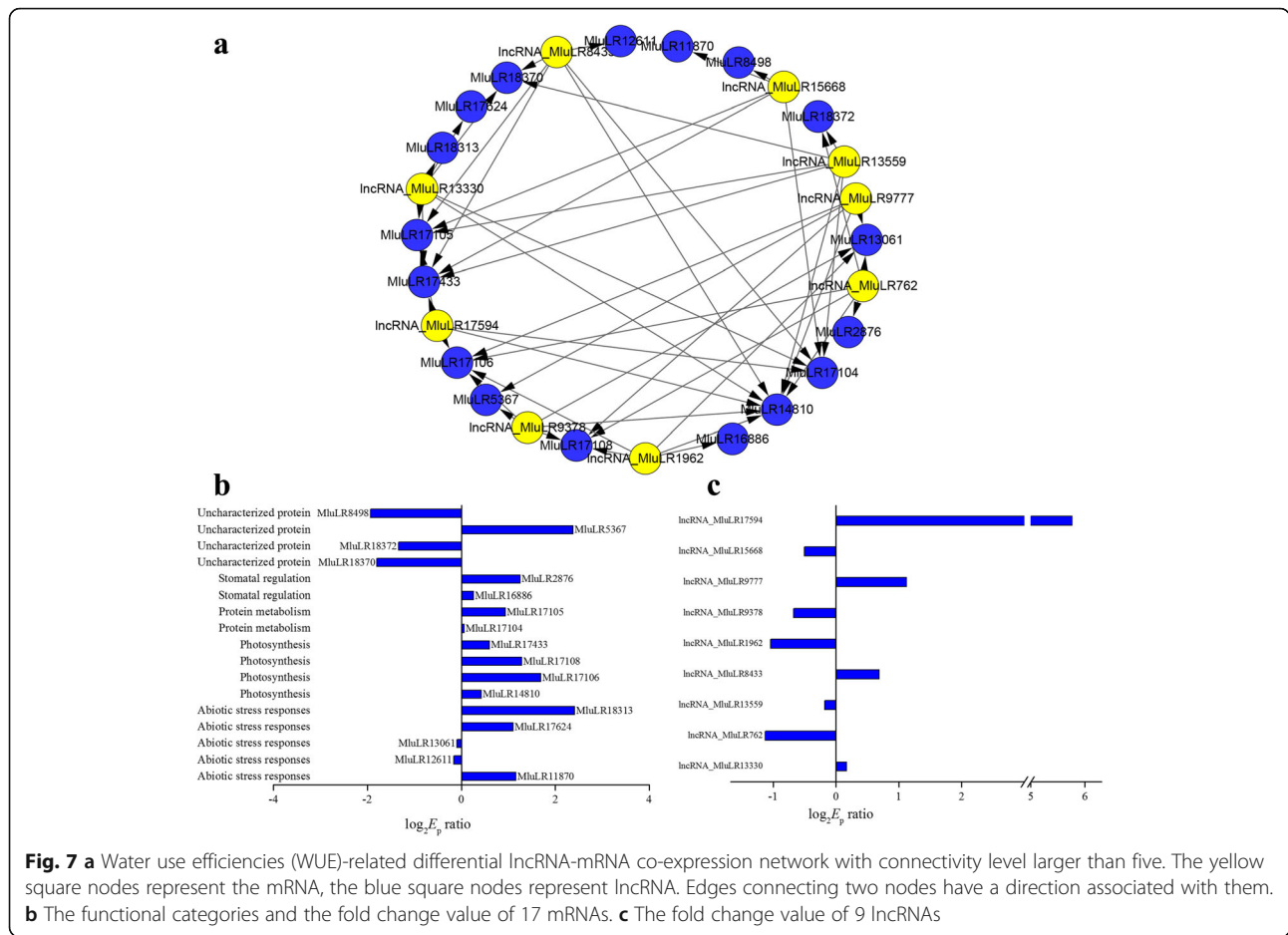
accurately evaluated. Nevertheless, we would demonstrate that a small proportion of the identified lncRNAs may be not real in our data, as non-coding transcripts derived from transposable elements were not filtered out due to the lack of a sequenced genome from *M. lutarioriparius*. In summary, the population-wide RNA-seq assembly approach for identifying lncRNAs was proven as an efficient method and could be applied for other species.

Relaxed expression regulation of lncRNAs and more sensitive to environment compared with protein coding mRNAs

In our study, we found the expression diversity of lncRNAs were significantly higher than that of protein coding RNAs in both JH and QG, the two independent strictly controlled experiments (Fig. 3a). This result may be explained by that population of *Miscanthus* harbored more substantial variation in noncoding regions of the genome, including extensive genetic variation within cis-regulatory elements and transcription factor binding sites. The results also implied that lncRNAs may be loosely regulated compared with protein coding mRNAs [45]. This may be due to that lncRNAs do not directly function and a large number of lncRNAs are nonfunctional, and thus changes in most lncRNA expression exact little fitness cost [5]. Higher-level expression diversity of lncRNAs may provide the indirect evidence for that lncRNAs experienced more relaxed evolution or divergent selection than protein coding mRNAs.

Previous results that lncRNA expression is typically more variable between tissues [5, 46, 47], which may be an indicator that the expression diversity of lncRNAs in population could be larger than mRNAs. Based on this inference, lncRNAs may be more variable among individuals, i.e. preferential expression in some individuals, but the majority of the mRNAs were consistently expressed in all individual [5]. This hypothesis could be illustrated by the comparison of expression frequency between lncRNAs and protein coding mRNAs (Fig. 4). Expression diversity provides raw materials for adaptation, because it represents a source of variation may enrich phenotypic variation at the level influencing more closely than genetic diversity and consequent fitness under natural selection [33]. Thus the lncRNAs and protein coding mRNAs with higher expression diversity had the higher potential as a variation factor in contribution to phenotypic variation [15]. Moreover, the lncRNAs had even more higher potential to be the variation factor than compared to protein coding mRNAs due to their higher variation of expression frequency (Fig. 4).

It intuitively seems that the range and average value of expression diversity ratio of lncRNAs between JH and QG are wider and higher than that of mRNAs, respectively



(Fig. 3). The widened ranges of expression levels and larger expression diversity suggest the lncRNAs are more sensitive to environmental change than mRNAs. As lncRNAs can epigenetically regulate the protein coding mRNAs by interacting with cis or trans elements [1, 17, 48], the increased expression diversity of lncRNAs might have triggered the mechanisms of enriching the expression diversity of genes in response to the environmental change [33]. Moreover, studies on adaptation in wild populations had revealed that ecologically important phenotypes changed due to the transcriptional regulation [49–51]. Thus, the increased diversity of lncRNAs may contribute to the change of ecologically important phenotypes such as WUE in our study and play an important role in improving the adaptation to new environment [43]. Thus, the lncRNAs with increased expression diversity could be the potential target contributing to adaptation.

Important roles of lncRNAs in regulation of stress response and agricultural trait

Previous studies have demonstrated that after stimulation with stress, expression level and expression diversity

for some genes or transcription factors related to stress response were elevated [33, 52]. However, the expression variation for lncRNAs had been little reported after stress stimuli. Here we found that the expression level of 807 lncRNAs (4.6%) and 995 mRNAs (5.4%) altered more than two-fold (Additional files 3 and 4), which suggested the proportion of lncRNAs with high expression fold change is comparable to mRNAs. Among the lncRNAs that we identified, we found that 2086 lncRNAs (11.8%) may be related with stress condition based on the E_d ratio, which reveal that extent of expression variation between two environments [33]. In particular, 76 lncRNAs had E_d ratio of more than 5-fold changes, which exhibited strong stress-responsive expression pattern [33, 53]. Moreover, 52 lncRNAs expressed at the level of nearly 100-fold increase from JH to QG. These lncRNAs with extreme change of expression could have the potential roles in adaptation to stress [17].

A large number of genes which were highly correlated with lncRNAs were identified using mRNA-lncRNA co-expression network analysis and their GO term enrichment analysis [54, 55]. And 5 categories of genes were

closely related to cellular stress response using GO term enrichment analysis (Table 2). Protein kinase domain played a role in a multitude of cellular processes, and Protein kinase pathway mediated drought and cold signaling [56]. Ring finger domain mainly involved in ubiquitination pathway and abiotic stress responses [57]. Cytochrome P450 mainly involved in biosynthetic reactions and biotic stress [58]. Functions of WD domain, G-beta repeat ranged from signal transduction and transcription regulation to cell cycle control, autophagy and apoptosis, and they may response to salt stress and nutrient stress [59–61]. These enrichment pathways could have been related with the cold and dry climates and poor soil conditions in the Loess Plateau. These results suggested that lncRNAs could have played the roles in many biological processes responding to stress through regulating gene network.

The higher WUE was likely to be one reason for the higher biomass production with much less precipitation in more stressful environment [30]. Using co-expression network analysis, we identified 3371 unique lncRNAs involving in the regulation of these candidate genes, which revealed a regulation pattern of lncRNAs in the manner of the accumulation of numerous minor lncRNAs for complex traits. Out of these lncRNAs, 9 lncRNAs regulated at least 5 candidate genes. This suggested that lncRNA has the potential to be the regulation center in the wide-range participation. Furthermore, the environmental change from JH to QG involves intricate natural condition, including soil water content, light condition, temperature etc. [30, 33]. Complex co-expression network regulation could be one of the most important molecular mechanisms underlying the adaptation to these conditional changes [9, 15, 22]. The evidence presented in this study indicates that only a small proportion of lncRNAs with large expression variation has an impact on phenotypic variation in WUE. It is reasonable to infer that much more lncRNAs may participate in the response to other conditional changes. Here we revealed the regulation pattern of lncRNAs for complex traits [43], which provided a good starting point towards understanding the role of lncRNAs in regulation of abiotic stress tolerance.

Although we identified a large number of lncRNAs related with many biological process, a limited number of lncRNAs had been speculated their functions using lncRNA-mRNA co-expression analysis. Further studies are necessary to understand the functional motifs of lncRNAs, and how specific lncRNAs seek out selective sites in the genome for lncRNA-mRNA and lncRNAs-lncRNAs interaction.

Conclusion

In this study, we identified population-wide lncRNAs in *M. lutarioriparius* under two different field conditions.

The expression level and the variation of expression of lncRNAs in population were systematically characterized and were compared with protein coding mRNAs. We proposed that lncRNAs may experience more relaxed evolution or more divergent selection compared with protein coding RNAs. In response to new environment, lncRNAs may be more sensitive to environmental change than protein coding RNAs. This study would provide insights into the roles of lncRNAs in plant stress responses. Such information can be useful in the process of adaptation of second generation of energy crop.

Methods

Data collection

We had collected populations of *M. lutarioriparius* across their distributional ranges along the Yangtze River in China in October and November 2008, and planted them in two experimental fields. One field site was in Jiangxia of Hubei Province (JH) which was near the native habitat and established by Wuhan Botanical Garden of Chinese Academy of Sciences, and the other field site was in Qingyang of Gansu Province (QG) which was near the domestic habitat and established by the Key Laboratory of Plant Resources and Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences. QG is colder, drier, and nutrient-poorer condition than its native habitat. The voucher specimen was deposited in the Wuhan botanical garden of Chinese Academy of Sciences (HIB). We then collected the same 14 populations of *M. lutarioriparius* in nearly native habitat JH and target domestic site QG in 2012 [33]. The growing stage was about one month later in QG than in JH, which was consistent with the temperature patterns between the two locations [29]. Therefore, samples which were listed in Additional file 9 were collected around noon on June 12th, 2012 in JH and on July 13th, 2012 in QG, respectively. When these plants were at the same growth stage between the two sites, 3 individuals for each population were randomly chosen and the fourth mature leaves of these plants were sampled for RNA-seq using Hiseq 2000 [33]. Considering the quality of reads, 2 individuals whose Q20 and genes expression level deviated from that of the majority of the individuals in both sites were discarded. Thus a total of 80 individuals were used for transcriptomic analysis ultimately. The raw data had been released at NCBI's Short Read Archive under three Bio Projects, PRJNA227191, PRJNA227195, and PRJNA226258. The transcriptome sequencing data, which was previously used in protein coding mRNAs expression analysis was proved to be a high quality of transcriptome reference [33].

Transcriptome assembly

We formed two assembly groups by randomly pooling 24 and 30 individuals from the 80 sampled individuals of *M. lutarioriparius*. *De novo* assembling was performed on SOAPdenovo-Trans for each group respectively [62], the average insert size was set at 230 bp, while the mergeLevel set at 0 to force perfect match. The coverage for contigs less than 3 was eliminated. The threshold value of output number of scaffolds was set at 15. The program GapCloser was used for filling the gaps of scaffolds with the sequence overlap length set at 25.

For each of the assembly group, *kmer* 41, *kmer* 51, and *kmer* 61 were used during assemble and all the resulting scaffolds were pooled [44]. The pooled scaffolds were merged when possible and joined together for extending sequence using the program CAP3 [63]. The minimal overlapping length was set at 45 bp and the overlapping identities of from 94% to 99% were tested. The scaffolds obtained from SOAPdenovo-Trans and merged scaffolds obtained from CAP3 for the two assembly groups were pooled together. The ORFs of all these sequences were predicted by the EMBOSS package [64], and only sequences with ORFs smaller than 150 bp were retained.

Identification and characterization of lncRNAs

All transcripts over 200 bp were filtered from raw assembled transcriptome for coding potential evaluation using Coding Potential Calculator software (http://cpc.cbi.pku.edu.cn/programs/run_cpc.jsp) [38], a sequence-based predictor for protein coding potential of transcripts and a widely used discoverer of lncRNAs. The negatively scored transcript was considered a noncoding transcript. To discriminate lncRNAs from previously annotated miRNA, rRNA, or other small noncoding RNA transcripts, we aligned all noncoding transcripts against the Rfam database [39]. The remaining transcripts were identified and were treated as candidate lncRNAs of *Miscanthus*, which was used for further functional analysis. The RNA sequencing reads from the 80 individuals were aligned independently against the lncRNA candidates. To gain enough sequencing data for lncRNA assembly, we combined all aligned results for each lncRNA candidate using SAMtools [65]. The mapped reads were then assembled into transcripts using Cufflinks [66]. The assembled transcripts with sequence length less than 200 bp were filtered to remove. To obtain a good comparison between two environments, we discarded the lncRNA candidates expressed in less than 20 individuals out of the 80 individuals. Under this criterion, 17610 lncRNA candidates were remained. The GC content as well as sequence length was compared between lncRNAs and protein coding mRNAs using both *t*-test and Wilcoxon test.

Differential expression of lncRNAs and mRNAs

Expression level of lncRNA was estimated based on transcript abundance calculated using FPKM. FPKM value for both genes and lncRNAs was calculated using Cufflinks. Reads of each individual were mapped to the reference transcriptome or lncRNA using Bowtie, TopHat, and Cufflinks [67, 68]. Reference sequence index was created using bowtie-build, and the short reads were aligned to the reference genes and lncRNAs sequence using TopHat using default settings. In this study, we used E_d which is the abbreviation of “expression diversity” as a way to estimate the Standard Deviation. Expression level and population expression diversity were calculated based on the formula

$$E_p = \frac{\sum_{i=1}^n E_i}{n} \quad \text{and} \quad E_d = \frac{\sum_{i=1}^n |E_i - E_p|}{(n-1)E_p},$$

here E_i represents the FPKM of a given gene or lncRNA of the i th individual in the population, n represents the number of individuals, and E_p represents the expression level of a given gene or given lncRNA [33]. E_p ratio and E_d ratio, which represents the fold change of E_p and E_d from JH to QG respectively, and E_d were compared between lncRNAs and protein coding mRNAs as well as between environments using both *t* test and Wilcoxon test. In order to more accurately detect candidate genes responding to new environmental stress, we used both significant test and a fold change ranking [69, 70]. The differentially expressed genes between the two sites were identified using the parametric *t*-test (normal bimodal distribution) or the non-parametric Wilcoxon-test (non-normal unimodal distribution). And gene expression change of lncRNA with FDRs less than 0.01 and larger than two-fold change was considered to be statistically significant.

Construction of the lncRNA-mRNA co-expression network

lncRNA-mRNA co-expression network was constructed following a method similar to previous study [9]. First of all, the expression values of the lncRNAs and mRNAs were obtained. Then, Spearman correlation coefficient was calculated between the expression values of each of the lncRNA-mRNA pairs across the near native samples JH and the transplanted site samples QG, respectively. The data were preprocessed without special treatment of the lncRNAs or mRNAs expression value. A *P*-value of less than 0.05 was considered statistically significant. The lncRNA-mRNA pairs whose Spearman correlation coefficients was larger than 0.7 in one environment (native environment JH or transplanted site QG), but smaller than 0.5 in the other environment were selected, as these parameters indicated that the lncRNA-mRNA pairs were differentially co-expressed between the two distinct environments [9]. Finally, the network was constructed, in which nodes were lncRNAs or mRNAs. A

total of 241 lncRNAs and mRNAs containing 334 relationships were visualized using Cytoscape [71]. The top 20 mRNA nodes which were those with the highest degree were shown. Two sub-networks containing 32 lncRNAs and mRNAs and 56 relationships with most lncRNA-mRNA interactions were presented in detail. Circle nodes represent lncRNAs and square nodes represent mRNAs.

Construction of water use efficiency (WUE) related differential lncRNA-mRNA co-expression network

By re-analyzing transcriptome expression level associated with water use efficiency (WUE), a WUE related differential lncRNA-mRNA co-expression network (WUE-LMN) was constructed in the current study. First of all, we conducted a matrix correlation between water use efficiency (WUE) and expression level of protein coding mRNAs. Before that, the genes with FPKM of the median value of 0 were dropped out, and 15,495 mRNAs were ultimately used for the further analysis. For each gene, the FPKM value of each individual in QG divided by the FPKM value in JH with all combinations, and then a matrix were obtained. The same process was also performed for water use efficiency (WUE). Following, we performed mantel test between the water use efficiency (WUE) matrix and FPKM matrix using Spearman's rank correlation method. Furthermore, a Spearman's rank correlation coefficient was calculated for each gene using a 10,000 permutation test to assess the statistical significance. The detail method was described in our previous work [32]. Using this method, 48 genes were identified at the $P < 0.01$ level as candidates genes for WUE alteration in the changing environment under stress [32]. The co-expression of lncRNAs and mRNAs as well as WUE was also visualized using Cytoscape [71]. The Pfam database was used to analyze the potential functions of lncRNAs based on their co-expressed genes.

Validation of sequence assembly and gene expression from RNA-Seq

In order to validate the sequence quality of this assembly, eight randomly selected lncRNAs were sequenced by PCR sequencing method. RNA sample from one of these individuals was used and reverse-transcribed into the first strand cDNAs for the PCR template. The primers were listed in Additional file 10. All the PCR products were obtained from the same sample. The clear PCR results were blasted against the assembly sequence from RNA-seq using BLASTN. RNA samples from the 18 individuals which passed the quality control were used and reverse-transcribed into the first strand cDNAs for the qPCR quantification. Eight lncRNAs were examined for the relative quantitation of expression level in

populations. The qPCR was performed using SYBR Premix Ex Taq (Takara) on ABI7500 (Applied Biosystems). The primers were listed in Additional file 11. According to previous method [32, 40–42], we performed three technical replicates for each template to calculate the average CT value, which were used to evaluate the relative expression level of lncRNAs. Relative expression levels of target lncRNAs among the individuals were determined using the $2^{-\Delta\Delta Ct}$ method to calculate the relative quantitation of expression levels with the normalization gene Actin as the endogenous control (The accession number was AT3G53750 from TAIR of the *Arabidopsis* homologue) [42]. The correlation of lncRNA expression estimated by qPCR and RNA-Seq were analyzed using Spearman test with R (Version 3.3.0).

Additional files

Additional file 1: The sequence of 17610 lncRNAs identified from the population of *M. luteoriparius*. (XLSX 3694 kb)

Additional file 2: The sequence of 8 lncRNAs obtained from population RNA-seq assembly and the PCR experiment. (XLSX 12 kb)

Additional file 3: The E_p ratios, E_d and E_d ratios for each lncRNA between two environments JH and QG. (XLSX 1967 kb)

Additional file 4: The E_p ratios, E_d and E_d ratios for each mRNA between two environments JH and QG. (XLSX 1784 kb)

Additional file 5: The expression frequency of each lncRNA in population and their different frequency between two environments JH and QG. (XLSX 1136 kb)

Additional file 6: The lncRNA-mRNA pairs in both environments with the correlation coefficient of more than 0.9. (XLSX 234 kb)

Additional file 7: The lncRNA-mRNA pairs with Spearman correlation coefficients larger than 0.7, including 1 052 mRNAs and 1 023 lncRNAs. (XLSX 44 kb)

Additional file 8: The lncRNA-mRNA pairs in the WUE-lncRNA-mRNA co-expression network. (XLSX 75 kb)

Additional file 9: Collection locations and experimental field sites for each of the sampled individuals of the *M. luteoriparius* species. (XLSX 10 kb)

Additional file 10: List of primer sequence of 8 lncRNAs for the PCR validation experiments. (XLSX 8 kb)

Additional file 11: List of primer sequence of 8 lncRNAs for qPCR validation experiments. (XLSX 9 kb)

Abbreviations

CT: Cycle threshold; E_d : Expression diversity; E_p : Average expression level; FPKM: Expected number of fragments per kilobase of transcript sequence per millions base pairs sequenced; GO: Gene ontology; indels (deletions and insertions); JH: Jiangxia in Hubei Province; lncRNA: Long non-coding RNA; QG: Qingyang in Gansu Province; qPCR: Quantitative real-time PCR; WUE: Water use efficiency

Acknowledgments

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.

Funding

The work was supported by grants from the National Natural Science Foundation of China (No. 31500186) in analysis and writing the manuscript and Key Program of the National Natural Science Foundation of China (No. 91131902) in the design of the study and collection.

Availability of data and materials

Sequence data from RNA-seq described in this article had been released at NCBI's Short Read Archive under three BioProjects, PRJNA227191 <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA227191>, PRJNA227195 <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA227195>, and PRJNA226258 <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA226258>. The accession numbers for the sequences of the 7 lncRNAs obtained by PCR were BankIt1973535: KY321443-KY321449. The data sets supporting the results of this article are included within the article and its additional files.

Authors' contributions

TS and QX conceived and designed the experiments. ZS and CZ performed the experiments. QX, ZS, CT, FH, JY, WL and LK participated in acquisition of data and data analysis. QX, ZS and TS wrote and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Since all *M. lutarioriparius* seed sample employed in this study are not endangered, no permission for collection of seed growing in public area is needed in China. All the leaf samples used in this study were collected from two field sites that were maintained in our own laboratory.

Author details

¹Key Laboratory of Plant Resources and Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. ²State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. ³University of Chinese Academy of Sciences, Beijing 100049, China. ⁴Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA. ⁵Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China.

Received: 23 November 2016 Accepted: 23 January 2017

Published online: 13 February 2017

References

- Lee JT. Epigenetic regulation by long noncoding RNAs. *Science*. 2012; 338(6113):1435–9.
- Johnsson P, Lipovich L, Grander D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta*. 2014;1840(3):1063–71.
- Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res*. 2007; 17(5):556–65.
- Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*. 2009;136(4):629–41.
- Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154(1):26–46.
- Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, et al. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nature Str Mol Biol*. 2013;20(7):908–13.
- Crea F, Watahiki A, Quagliata L, Xue H, Pikor L, Parolia A, et al. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. *Oncotarget*. 2014;5(3):764–74.
- Ounzain R, Micheletti R, Beckmann T, Schroen B, Alexanian M, Pezzuto I, et al. Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs. *Eur Heart J*. 2015;36(6):353–68.
- Wang P, Fu H, Cui J, Chen X. Differential lincRNA-mRNA co-expression network analysis revealing the potential regulatory roles of lincRNAs in myocardial infarction. *Mol Med Rep*. 2016;13(2):1195–203.
- Guo Q, Cheng Y, Liang T, He Y, Ren C, Sun L, et al. Comprehensive analysis of lincRNA-mRNA co-expression patterns identifies immune-associated lincRNA biomarkers in ovarian cancer malignant progression. *Sci Rep*. 2015;5:17683.
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458(7235):223–7.
- Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*. 2013;152(6):1308–23.
- Yarmishyn AA, Kurochkin IV. Long noncoding RNAs: a potential novel class of cancer biomarkers. *Front Genet*. 2015;6:145.
- He C, Hu H, Wilson KD, Wu H, Feng J, Xia S, et al. Systematic characterization of long noncoding RNAs reveals the contrasting coordination of cis-and trans-molecular regulation in human fetal and adult heart. *Circ Cardiovasc Genet*. 2016;9(2):110–8.
- Zhang W, Han Z, Guo Q, Liu Y, Zheng Y, Wu F, et al. Identification of maize long non-coding RNAs responsive to drought stress. *PLoS One*. 2014;9(6): e98958.
- Wang H, Niu QW, Wu HW, Liu J, Ye J, Yu N, et al. Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lincRNAs associated with agriculture traits. *Plant J*. 2015;84(2):404–16.
- Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, et al. Characterization of stress-responsive lincRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J*. 2014;80(5):848–61.
- Quan M, Tian J, Yang X, Du Q, Song Y, Wang Q, et al. Association studies reveal the effect of genetic variation in lincRNA UGTRL and its putative target PtoUGT88A1 on wood formation in *Populus tomentosa*. *Tree Genetics Genomes*. 2016;12(1):1–16.
- Fan C, Hao Z, Yan J, Li G. Genome-wide identification and functional analysis of lincRNAs acting as miRNA targets or decoys in maize. *BMC Genomics*. 2015;16(1):1.
- Muthusamy M, Uma S, Backiyarani S, Saraswathi M. Genome-wide screening for novel, drought stress-responsive long non-coding RNAs in drought-stressed leaf transcriptome of drought-tolerant and-susceptible banana (*Musa spp*) cultivars using Illumina high-throughput sequencing. *Plant Biotechnol Rep*. 2015;9(5):279–86.
- Megha S, Basu U, Rahman MH, Kav NN. The role of long non-coding RNAs in abiotic stress tolerance in plants. In: Pandey GK, editor, *Elucidation of abiotic stress signaling in plants*. New York: Springer; 2015. p. 93–106.
- Amor BB, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, et al. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res*. 2009;19(1):57–69.
- Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, et al. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol*. 2011;11(1):61.
- Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell*. 2012;24(11):4333–45.
- Liu W, Yan J, Li J, Sang T. Yield potential of *Miscanthus* energy crops in the Loess Plateau of China. *GCB Bioenergy*. 2012;4(5):545–54.
- Sang T. Toward the domestication of lignocellulosic energy crops: Learning from food crop domestication. *J Integr Plant Biol*. 2011;53(2):96–104.
- Liu W, Sang T. Potential productivity of the *Miscanthus* energy crop in the Loess Plateau of China under climate change. *Environ Res Lett*. 2013;8(4): 044003.
- Sang T, Zhu W. China's bioenergy potential. *GCB Bioenergy*. 2011;3(2):79–90.
- Yan J, Chen W, Luo F, Ma H, Meng A, Li X, et al. Variability and adaptability of *Miscanthus* species evaluated for energy crop domestication. *GCB Bioenergy*. 2012;4(1):49–60.
- Yan J, Zhu C, Liu W, Luo F, Mi J, Ren Y, et al. High photosynthetic rate and water use efficiency of *Miscanthus lutarioriparius* characterize an energy crop in the semiarid temperate region. *GCB Bioenergy*. 2015;7(2):207–18.
- Yan J, Zhu M, Liu W, Xu Q, Zhu C, Li 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*. 2016;8:764–76.
- Fan Y, Wang Q, Kang L, Liu W, Xu Q, Xing S, et al. Transcriptome-wide characterization of candidate genes for improving the water use efficiency of energy crops grown on semiarid land. *J Exp Bot*. 2015;66(20):6415–29.
- Xu Q, Xing S, Zhu C, Liu W, Fan Y, Wang Q, et al. Population transcriptomics reveals a potentially positive role of expression diversity in adaptation. *J Integr Plant Biol*. 2015;57:284–99.
- Liu W, Mi J, Song Z, Yan J, Li J, Sang T. Long-term water balance and sustainable production of *Miscanthus* energy crops in the Loess Plateau of China. *Biomass and Bioenergy*. 2014;62:47–57.

35. Mi J, Liu W, Yang W, Yan J, Li J, Sang T. Carbon sequestration by *Miscanthus* energy crops plantations in a broad range semi-arid marginal land in China. *Sci Total Environ*. 2014;496:373–80.
36. Xing S, Kang L, Xu Q, Fan Y, Liu W, Zhu C, et al. The Coordination of gene expression within photosynthesis pathway for acclimation of C4 energy crop *Miscanthus lutarioriparius*. *Front Plant Sci*. 2016;7:109.
37. Surget-Groba Y, Montoya-Burgos JI. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res*. 2010;20(10):1432–40.
38. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35 Suppl 2:W345–9.
39. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res*. 2003;31(1):439–41.
40. Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, et al. Mendelian and non-mendelian regulation of gene expression in Maize. *Plos Genet*. 2013;9(1):e1003202.
41. Schmittgen TD, Zakrjsek BA, Mills AG, Gorn V, Singer MJ, Reed MW. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: Comparison of endpoint and real-time methods. *Anal Biochem*. 2000;285(2):194–204.
42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*. 2001;25(4):402–8.
43. Shin SY, Shin C. Regulatory non-coding RNAs in plants: potential gene resources for the improvement of agricultural traits. *Plant Biotechnol Rep*. 2016;10(2):35–47.
44. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics*. 2014;30(1):31–7.
45. Billerey C, Boussaha M, Esquerré D, Rebours E, Djari A, Meersseman C, et al. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*. 2014;15(1):1.
46. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22(3):577–91.
47. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25(18):1915–27.
48. Matkovich SJ, Edwards JR, Grossenheider TC, Strong CD, Dorn GW. Epigenetic coordination of embryonic heart transcription by dynamically regulated long noncoding RNAs. *Proc Natl Acad Sci U S A*. 2014;111(33):12264–9.
49. Mitchell-Olds T, Schmitt J. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature*. 2006;441(7096):947–52.
50. Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T. Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol Ecol*. 2013;22(3):699–708.
51. Agren J, Oakley CG, McKay JK, Lovell JT, Schemske DW. Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 2013;110(52):21077–82.
52. Kidokoro S, Watanabe K, Ohori T, Moriwaki T, Maruyama K, Mizoi J, et al. Soybean DREB1/CBF-type transcription factors function in heat and drought as well as cold stress-responsive gene expression. *Plant J*. 2015;81(3):505–18.
53. Xu Q, Zhu CY, Fan YY, Song ZH, Xing SL, Liu W, et al. Population transcriptomics uncovers the regulation of gene expression variation in adaptation to changing environment. *Sci Rep*. 2016;6:25536.
54. Chen M, Wang CL, Bao H, Chen H, Wang YW. Genome-wide identification and characterization of novel lncRNAs in *Populus* under nitrogen deficiency. *Mol Gen Genomics*. 2016;291(4):1663–80.
55. Chang L, Qi H, Xiao Y, Li C, Wang Y, Guo T, et al. Integrated analysis of noncoding RNAs and mRNAs reveals their potential roles in the biological activities of the growth hormone receptor. *Growth Hormone IGF Res*. 2016;29:11–20.
56. Shinozaki K, Yamaguchi-Shinozaki K, Seki M. Regulatory network of gene expression in the drought and cold stress responses. *Curr Opin Plant Biol*. 2003;6(5):410–7.
57. Lyzenga WJ, Stone SL. Abiotic stress tolerance mediated by protein ubiquitination. *J Exp Bot*. 2012;63(2):599–616.
58. Morant M, Bak S, Møller BL, Werck-Reichardt D. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Curr Opin Biotechnol*. 2003;14(2):151–62.
59. Halder SK, Anumanthan G, Maddula R, Mann J, Chytil A, Gonzalez AL, et al. Oncogenic function of a novel WD-Domain protein, STRAP, in human carcinogenesis. *Cancer Res*. 2006;66(12):6156–66.
60. Chen R-H, Miettinen PJ, Maruoka EM, Choy L, Derynck R. A WD-domain protein that is associated with and phosphorylated by the type. *Nature*. 1995;377(6549):548–52.
61. Ford CE, Skiba NP, Bae H, Daaka Y, Reuveny E, Shekter LR, et al. Molecular basis for interactions of G Protein β subunits with effectors. *Science*. 1998;280(5367):1271–4.
62. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. 2014;30(12):1660–6.
63. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res*. 1999;9(9):868–77.
64. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 2000;16(6):276–7.
65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
66. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562–78.
67. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
68. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105–11.
69. Parker BL, Thaysen-Andersen M, Fazakerley DJ, Holliday M, Packer NH, James DE. Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF-Alpha-Induced insulin resistance in *Adipocytes*. *Mol Cell Proteomics*. 2016;15(1):141–53.
70. Lee ST, Xiao YY, Muench MO, Xiao JQ, Fomin ME, Wiencke JK, et al. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res*. 2012;40(22):11339–51.
71. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27(3):431–2.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

