# Chapter 6
# Transcriptome Dynamics in Rice Leaves Under Natural Field Conditions

**Takeshi Izawa**

**Abstract** Although crops have been domesticated and bred under natural field conditions, the majority of molecular genetic analyses have been performed under controlled artificial conditions, such as growth chambers. This restricts agricultural application of new findings on important crops based on molecular genetics. Recently, several transcriptome analyses to elucidate the dynamics of the transcriptome and several specific biological traits have been reported. These analyses made full use of cutting-edge methods of statistical modeling and Bayesian approaches. One critical finding of these studies was that thousands of genes expressed in rice leaves respond significantly to dynamic changes in ambient temperatures under natural fluctuating conditions. This should serve as a wake-up call for plant researchers using fixed-temperature conditions in growth chambers. This chapter discusses the processes involved and provides longitudinal perspectives on field transcriptome analysis.

**Keywords** Field transcriptome · Fluctuating environments · Statistical modeling · Crop science · Molecular biology

T. Izawa (✉)
Laboratory of Plant Breeding & Genetics, Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences,
The University of Tokyo, Tokyo, Japan
e-mail: a-izawa@mail.ecc.u-tokyo.ac.jp



T. Sasaki, M. Ashikari (eds.), *Rice Genomics, Genetics and Breeding*,
https://doi.org/10.1007/978-981-10-7461-5_6

## 6.1 Reasons Why Plant Molecular Genetic Researchers May Be Hesitant to Perform Experiments Under Field Conditions

### *6.1.1 Dealing with Fluctuating Data Obtained Under Natural Conditions Using Statistical Analysis*

Many researchers hesitate to use data obtained under natural field conditions because the reproducibility of experimental data would be severely affected under fluctuating ambient growth conditions. In contrast, experimental data are easily reproducible under artificially controlled growth conditions in the laboratory, so researchers do not have to repeat experiments to obtain statistically significant results. Molecular geneticists generally attempt to minimize the repetition of experiments, which is appropriate when working on mutant lines exhibiting very clear genetic phenotypes. In such cases, fold changes and differences between wild-type and mutant plants are normally high. However, when analyzing complex genetic traits, known as quantitative trait loci (QTL), experimental data contain both genetic variations due to segregation and environmental variations due to differences in growth settings under natural field conditions (Lander and Botstein 1989). Thus, QTL analyses include statistical evaluations, such as logarithm of odds (LOD) scores, to select the next approach or target locus. Furthermore, to evaluate agronomic traits of a new variety, many crop breeders normally grow a number of candidate lines for several years in at least a dozen areas. Therefore, upon next-genome sequencing era, not only plant breeders but also plant molecular biologists should become familiar with experiments on this scale. Statistically significant results obtained on a small scale with a small number of experiments are not suitable for practical use.

To examine the significance of transcriptome data, it has been recommended to use criteria based on multiple testing, such as the false discovery rate (FDR) (Yang et al. 2003). However, we often obtain only a few genes which can be beyond such criteria significantly with several repeats of transcriptome analyses. In our experience, the paired *t*-test using 40 pairs of rice leaf transcriptome data yielded several thousand genes that showed statistically significant changes in expression even after FDR correction (Izawa et al. unpublished). Interestingly, all of these paired samples were sampled at distinct timings, although they were obtained from two neighbor paddy fields with and without nutrition. Furthermore, most of the significant genes among the 40 pairs of transcriptome data exhibited very small fold changes of mean values in expression. Meanwhile, hundreds of genes showed significant changes in expression when we used 20 pairs of transcriptome data. These observations suggested that statistical analyses with more than a dozen transcriptome data would give us satisfactory results, even when there is a great

deal of experimental noise due to differences in environmental conditions. In addition, it might be very important that the fold changes should be taken into consideration rather than FDR-corrected *P*-values to select biologically important genes. That is, even the statistical significance obtained with larger amounts of data may have limits for identifying the biological importance.

### 6.1.2 Correlations Among Distinct Environmental Factors

Temporal correlations between distinct environmental factors in a few days, such as solar radiation and ambient temperature, are often so high and too difficult to distinguish them on which factors were relevant to the studies of plant biology. However, such correlations between distinct environmental data differ between seasons. Thus, we found previously that temporal correlations between the fluctuations in various environmental data over several months relatively become small (Izawa 2015). Therefore, when considering all of the crop seasons of target crops over several years, the correlations among environmental factors are thought not to be an obstacle for further biological analysis.

### 6.1.3 Costs of Gathering Field Transcriptome Data

When planning field transcriptome analysis, it is necessary to consider the costs associated with gathering data. As of 2016, the financial costs to perform transcriptome analysis are around 30,000 Japanese yen per sample. Here, we used a custom microarray with 180 K probes provided by Agilent Technologies. In a series of preliminary experiments, we examined the use of Cy5 and Cy3 labeling for the same RNA samples. After calibration to correct for characteristic trends between Cy5 and Cy3 signals using software provided by the manufacturer, we found that there were no differences in data according to which label was used (Izawa et al. data not shown). Thus, we do not care about the swapping effects between Cy5 and Cy3 for the cost efficiency in our field transcriptome analysis. Although the numbers of transcriptome data required for biologically significant field transcriptome analysis depend on the purpose of individual research, around 100 samples for one class of data would be sufficient for most purposes. We usually perform analyses using 12–13 samples per day to obtain the diurnal changes of transcriptome.

In addition, the major matters to be attended carefully to obtain reliable good quality data are those associated with sampling, RNA preparation, labeling, and microarray analysis. For sampling, we use standard tubes for all collaborating

teams and provide guidelines for sampling in the field. For microarray analysis, we use an open laboratory facility at the National Agriculture and Food Research Organization (NARO) in Tsukuba, Japan. In this open laboratory, experts in handling RNA and in operation of microarray equipment support us well to generate very reliable microarray data for all of our samples.

The RNA sequencing (RNA-Seq) technique provides a great deal of information for each transcript, such as its initiation site or splicing variants. We compared microarray analysis and RNA-Seq data for the same RNA samples and found that both yielded comparable data. Furthermore, we found that the reliability of transcriptome data between RNA-Seq and microarray data differed among target genes. Thus, it is difficult to say which of the two is better at this moment. In addition, the financial costs associated with both methods to obtain comparable data are currently almost the same. However, primary analysis of raw RNA-Seq data using a high-performance computer is still more laborious and complicated than microarray analysis. In addition, the conversion of legacy microarray data to the corresponding RNA-Seq data is largely dependent on the platform used and requires the setting of distinct rules for data regarding each gene. In the near future, direct RNA-Seq, such as the MinION nanopore sequencer technology, should be considered in place of the microarray method.

## 6.2 Environment-Driven Statistical Modeling of the Transcriptome

### *6.2.1 Lognormal Distribution Assumption for Transcriptome Data*

For statistical modeling, experimental noise should be distributed according to a normal distribution. In contrast, many raw data related to gene expression, including qRT-PCR data and transcriptome data obtained by microarray analysis, are not normally distributed. Empirically, it is known that qRT-PCR and transcriptome data obtained by microarray analysis have a lognormal distribution (Izawa 2012). Therefore, all transcriptome data should be transformed in a logarithmic manner before further statistical analysis. It is better to consider transforming data to a lognormal distribution when raw data obtained with larger values possesses larger experimental standard deviations. Experimental noise may not depend on the range of measured data.

### 6.2.2 A Model of Field Transcriptome Analysis in Rice

Determining the drivers of gene expression patterns is more straightforward under laboratory conditions than in the complex fluctuating environments seen in the field. Nagano et al. (2012) reported gathering 461 transcriptome data from the leaves of rice plants in a paddy field along with the corresponding meteorological data and developed statistical models for the endogenous and external influences on gene expression. In total, expression dynamics of more than 20,000 genes could be explained based only on information regarding environmental conditions, such as sampling date and time, transplantation date, and meteorological environmental data (solar radiation, temperature, atmosphere, wind, and precipitation). The results indicated that transcriptome dynamics were predominantly governed by endogenous diurnal rhythms, ambient temperature, plant age, and solar radiation. The data revealed diurnal gates for environmental stimuli to influence transcription and pointed to relative influences exerted by circadian and environmental factors on different metabolic genes.

#### 6.2.2.1 Pretreatment of Raw Environmental Data

Gene expression data are likely to be influenced by the dynamics of environmental factors. To integrate information regarding past dynamic changes in the environment, we pretreated environmental data with both a prefixed threshold for perception and a prefixed perception period. We prepared a series of pretreated environmental data with all possible combinations of several prefixed thresholds and several prefixed periods and tried to the best combination to explain the dynamic patterns of gene expression. Thus, we were able to test various perception patterns of dynamic data from each environment from a single set of temporal environmental data. With this approach, we can evaluate which environmental factors contribute to the dynamics of target genes.

#### 6.2.2.2 Grid Search Modeling of the Transcriptome

To select the best among various pretreated environment data, we developed a linear model that connects several terms, including bias, development, clock, and environment, to explain the dynamics of gene expression in the field. Among the candidate models, the best was selected by a grid search. Six environmental factors were considered: solar radiation, ambient temperature, humidity, atmosphere, wind, and precipitation for the entire crop season. Each environmental factor was considered on a distinct grid, with each grid location having prefixed thresholds and periods of perception. In addition, we considered gate effects, which reflect the diurnal changes of sensitivity for each environmental factor. Several hundred thousands of grid locations were evaluated to select the best model. In fact, we

developed models for 21,173 genes in which the dynamics of gene expression can be explained by the environmental conditions among 23,000 genes expressed in rice leaves under field conditions during the crop season. We considered a model successful when the residual values between observation and prediction (or estimation) were normally distributed. Several hundred genes expressed in rice leaves were still not modeled with our method, suggesting that other environmental factors, such as abiotic and biotic stress, are required to explain the dynamics of these genes.

#### 6.2.2.3 Major Findings from This Model

One of the most important findings in this modeling analysis was that more than 3000 genes are affected by the history of ambient temperature with specific thresholds and perception periods. As most plant researchers grow their research plant materials under artificial conditions with a fixed temperature, the findings of such studies are not necessarily reproducible at different temperatures. On the other hand, around 3000 genes are expressed very stably in rice leaves and are not affected by environmental or developmental factors. Of course, this model also generated predictions for the influence of changing temperature on transcriptome dynamics (Fig. 6.1). The models would also help to translate the knowledge amassed in laboratories to problems in agriculture, and our approach to deciphering the transcriptome fluctuations in complex environments will be applicable to other organisms.

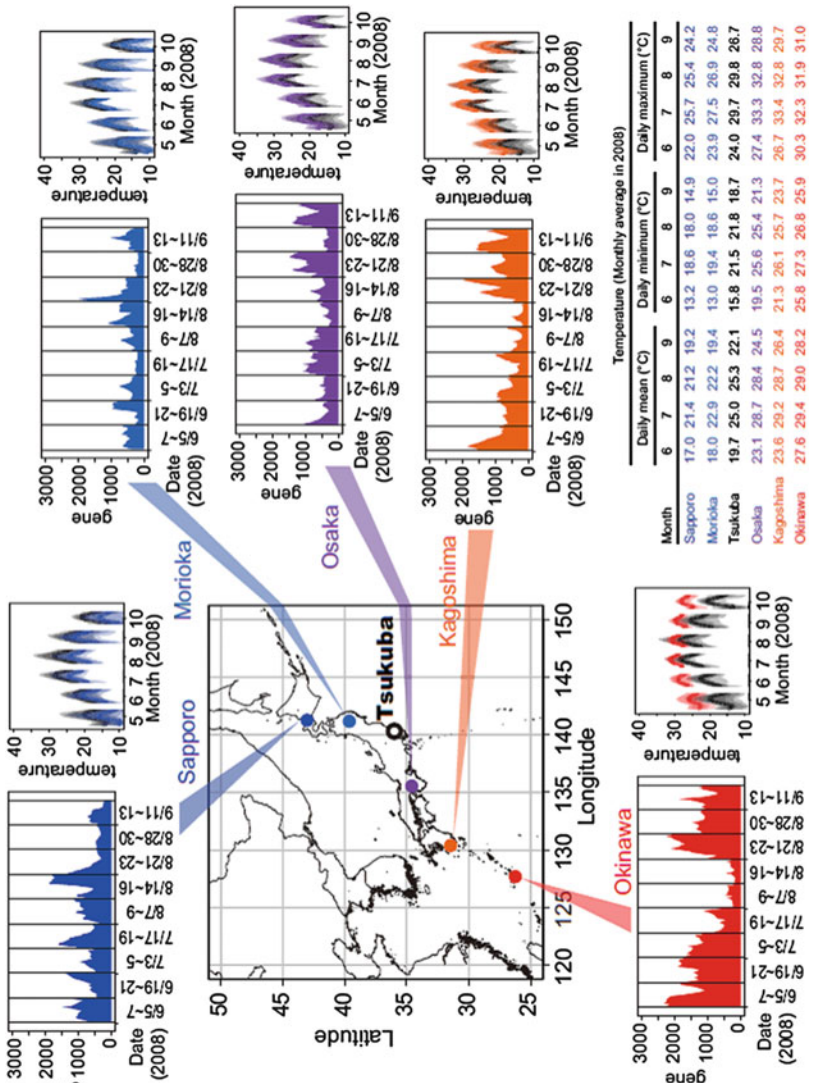
### *6.2.3 Differential Equation-Based Modeling of the Transcriptome*

In the above model (Nagano et al. 2012), the degree of generalizability to predict the entire transcriptome is quite high with Pearson's correlation coefficients of 0.95 between observations and predictions of the transcriptome based on environmental data. However, the generalizability for half of the genes was still not practical to predict gene expression based only on environmental data. One major reason for this is that we attempted to explain gene expression based on only one environmental factor. Many genes are likely to be regulated by several environmental factors, such as light and temperature signals. However, we selected the most effective environmental factor among the six tested—solar radiation, temperature, humidity, atmosphere, wind, and precipitation. To improve the generalizability of genes that did not show practical abilities, it is necessary to simultaneously integrate information from at least two independent environmental factors to explain the expression of the target gene. Therefore, we developed a new method to develop such models using a differential equation to input two temporal environmental

factors (Matsuzaki et al. 2015). From the data of the previous model, we focused on solar radiation and ambient temperature to explain the expression of a set of genes involved in the circadian clocks in rice. Here, we did not attempt to develop models at the transcriptome level as the calculation cost to determine the appropriate parameters would become much larger than the simple grid search models. In the previous case, we utilized the nonlinear least squares method to determine the parameters as algebraic solutions for each grid. We then attempted to select the grid to fit the gene expression data. In this case, we used the particle swarm optimization method, where several particles that form a vector of numerical parameter values are dispersed in the parameter space to search for optimal values of each parameter (Lu et al. 2002). The numerical values of each particle are updated after consideration of the inertia of each particle, the center of balance of all of the particles, and the direction to the optimal position. Therefore, they are not independent but are weakly connected to each other. With updated particle values, the differential equation is then solved again, and the values are evaluated for the next update. With this system, we can perform machine learning to determine the average values of all particles as the best parameters after iterated learning processes. The new model obtained in this way clearly improved the generalizability compared to the previous model selected by the grid search for most of the 20 circadian clock-related genes in rice examined in this study. We recently developed a fast algorithm to perform this modeling at the transcriptome level using an improved ABC (approximate Bayesian computation) method for parameter regression (Lenormand et al. 2013). We are currently preparing a two-environmental factor-driven model of the transcriptome with this algorithm.

### *6.2.4 Potential Use of the Neural Network Algorithm Concept for Transcriptome Modeling*

Although we have not yet evaluated the above two-environmental factor model, in the near future, we will consider the synergistic interactions of two environmental factors to explain gene expression for a special group of genes that are very sensitive and exhibit complicated responses to a given environment. To construct such a model, it could be reasonable to utilize the concept of multiple neural network algorithms (LeCun et al. 2015). This algorithm is known to be sufficiently flexible and rich to explain complicated interactions of various inputs. To maintain generalizability, we should use regularization terms in squared loss function and/or cross-validation. In addition, only a few pretreated inputs of environmental factors and pretreated time information should be used to make a model. Then, the best model among distinct combinations of pretreated data would be selected to obtain novel knowledge in biology. However, it is still not clear how many genes will be targets in a model with such complex interactions among environmental factors.




Temperature (Monthly average in 2008)

| Month | Daily mean (°C) | | | | Daily minimum (°C) | | | | Daily maximum (°C) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 6 | 7 | 8 | 9 | 6 | 7 | 8 | 9 | 6 | 7 | 8 | 9 |
| Sapporo | 17.0 | 21.4 | 21.2 | 19.2 | 13.2 | 18.6 | 18.0 | 14.9 | 22.0 | 25.7 | 25.4 | 24.2 |
| Morioka | 18.0 | 22.9 | 22.2 | 19.4 | 13.0 | 19.4 | 18.6 | 15.0 | 23.9 | 27.5 | 26.9 | 24.8 |
| Tsukuba | 19.7 | 25.0 | 25.3 | 22.1 | 15.8 | 21.5 | 21.8 | 18.7 | 24.0 | 29.7 | 29.8 | 26.7 |
| Osaka | 23.1 | 28.7 | 28.4 | 24.5 | 19.5 | 25.6 | 25.4 | 21.3 | 27.4 | 33.3 | 32.8 | 28.8 |
| Kagoshima | 23.6 | 29.2 | 28.7 | 26.4 | 21.3 | 26.1 | 25.7 | 23.7 | 26.7 | 33.4 | 32.8 | 29.7 |
| Okinawa | 27.6 | 29.4 | 29.0 | 28.2 | 25.8 | 27.3 | 26.8 | 25.9 | 30.3 | 32.3 | 31.9 | 31.0 |

**Fig. 6.1** Prediction of gene expression under different growing conditions. Predicted changes in gene expression in five locations at different latitudes in Japan based on environment-driven statistical modeling of the transcriptome (Nagano et al. 2012). For each location, the left plot indicates the number of genes predicted to differ in expression by more than twofold relative to corresponding data from Tsukuba, Japan. The *horizontal axis* represents the time points. The right plot indicates diurnal temperature data (00:00–24:00) for each month at that particular location (colored) and Tsukuba (black). Thick lines represent averages. The area densities represent ranges containing 90, 75, and 50% of data. The lower-right table shows monthly averages of the daily mean, minimum, and maximum temperature in 2008 (Cited from Fig. 7 in Nagano et al. 2012)

## 6.3 Transcriptome-Driven Sparse Modeling of Biological/Agricultural Traits

### *6.3.1 Bayesian Filtering Using Two-Dimensional Probability Distribution Heat Maps*

Transcriptome data obtained under field conditions is likely to contain enough information to explain dynamic changes in biological and agricultural traits (Yang et al. 2011). One simple way to extract such information on various traits is to determine the two-dimensional relationships of probability between the target trait and expression of a gene within a cluster of genes. According to Bayes' theorem, with information regarding gene expression in a set of selected genes, the target trait can be narrowed down as a probability distribution (Matsuzaki et al. 2015). In this way, we demonstrated that having gene expression data of only 16 related genes was sufficient to estimate the time of sampling with an accuracy of around 20 min (Figs. 6.2 and 6.3). Here, we examined 25 circadian clock-related genes and selected 16 genes. We searched all possible combinations of the 25 genes to select the best combination. This way works well for circadian clock-related traits since we were able to focus a dozen genes before selecting the best combination. However, this indicates that this way is not possible when we have transcriptome data, which includes more than ten thousand data, to select related genes.

### *6.3.2 Sparse Modeling of the Transcriptome with Regularization Terms, Such as LASSO Regression*

To find any relationship between gene expression of individual genes in transcriptome and any biological/agricultural traits, it would be a good way to develop a sparse modeling method to explain biological traits using gene expression data as the input. Here, it is important to extract as many genes related to biological traits as possible and establish statistical models to predict the traits with high generalizability. We examined the use of least absolute shrinkage and selection operator (LASSO) regression methods to develop such a model (Tibshirani 1996). Although we were unable to develop such models for various traits using transcriptome data artificially randomly mixed and connected to trait data, we were able to establish practical LASSO models for various traits with transcriptome data appropriately connected to trait data. Thus, the LASSO method is a very promising way. As we used glmnet, which provides the final model after tenfold cross-validation, the generalizability of the obtained model was sufficient. We used hundreds of microarray data containing 20–50 K data of gene expression to make

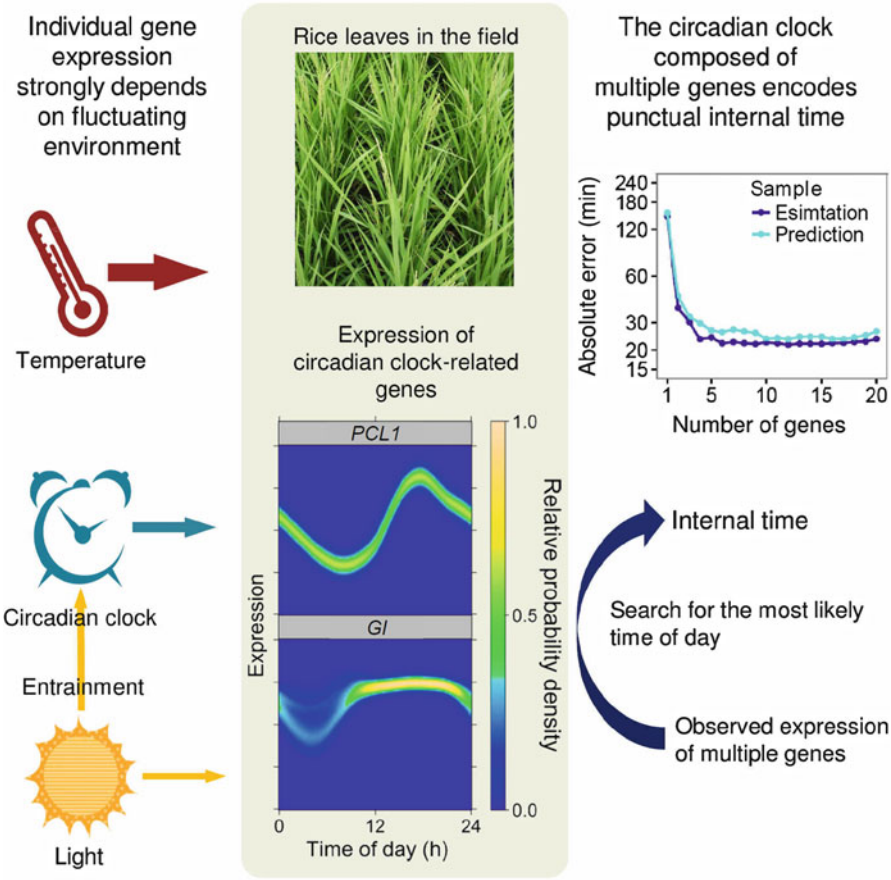



**Fig. 6.2** Construction of transcriptome-driven model for trait prediction**.** Left: a model of individual gene expression in wild-type rice growing in the field responding to solar radiation and ambient temperature was built for circadian clock-related genes and indicated a strong effect of temperature. Middle: prediction by the model was used to determine the relationship between physical time of day and expression. Right: using the relationship, we can infer internal time from expression of multiple genes and found accuracy to 22 min relative to physical time regardless of weather, day length, or plant developmental age (Cited from Supplemental Fig. 6.1 in Matsuzaki et al. 2015)

such LASSO models, and so we often obtained around 100 genes to make a LASSO model. Expression data of most of the selected genes showed clear correlations with the target trait values, but several genes did not. Thus, some synergistic interactions between gene expression values were integrated into the LASSO model. However, it should be noted that not all of the genes in the obtained LASSO model reflect all of the genes related to the trait. Thus, a new method is needed to extract all of the genes related to the target trait.

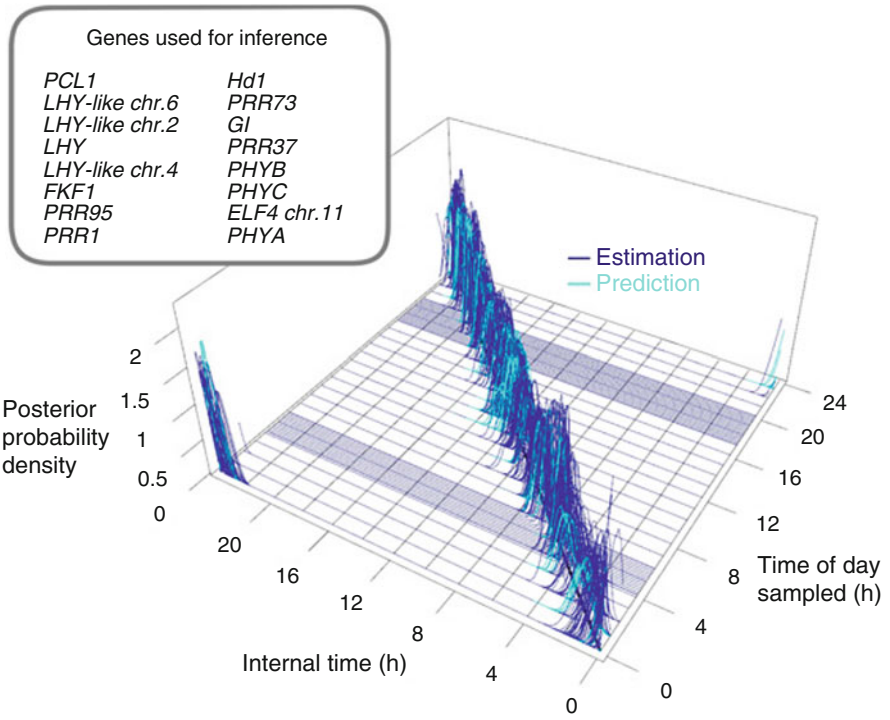



**Fig. 6.3** Prediction of sampling time based on expression data of 16 circadian clock-related genes. Estimation and prediction of sampling time using gene combination with the best estimation performance (i.e., the lowest mean absolute estimation). Posterior probability density of predicted sampling time is plotted against time of day sampled. Each blue (training sample for estimation, $n = 461$) and turquoise (validation sample for prediction, $n = 125$) line corresponds to a single sample. Among the training samples, those obtained at 10-min intervals at 04:00–06:00 and 17:00–20:00 are included. Ranges of predicted sampling time with zero posterior probability density for those samples are presented as areas with dense blue lines at the bottom of the three-dimensional (3D) space. The thick black diagonal line at the bottom of the 3D space indicates correspondence between internal time and time of day sampled. (Cited from Fig. 3 in Matsuzaki et al. 2015)

## 6.4 Potential Use of Field Transcriptome Data to Check Crops

In our preliminary LASSO model developed for prediction of heading date from sample RNAs (Izawa et al. data not shown), the accuracy of heading date prediction was around 2 days. Unlike the typical phenology model used to predict the target trait (Nakagawa et al. 2005), in which both real historical and predicted future meteorological data were needed, the LASSO model used only a single RNA sample to predict future traits. It is likely that such RNA may contain the developmental status of the sample and can therefore be used to predict the future from such

samples. This type of model would provide information regarding related genes, which can be used as biomarkers to predict biological traits. In addition, such RNA contains information on the health status of crops in the field, and LASSO modeling enables us to narrow the transcriptome data down to several key genes to find practical biomarkers to estimate the health status of crops. Applying this method to plant disease responses may provide information regarding the appropriate timing of herbicide treatments relying on these gene expression biomarkers. Several technologies will require further improvement to make this idea feasible. First, the cost of gene expression analysis for biomarker genes must be reduced. Considering the current cost of 30,000 Japanese yen for detection of approximately 180 K probes (or genes), it may be possible to reduce the cost to around 100 Japanese yen for 1000 genes. Second, we usually use liquid nitrogen for leaf sampling and storage, but this is an obstacle for wider use. Therefore, reagents for fixing RNA in rice leaves at normal temperature are needed. Finally, sampling under field conditions is laborious for farmers. As one example, automatic sampling machines, drones, and self-traveling vehicles would be useful for developing practical handling to check crops based on gene expression data. After efficient sampling, the users can send samples to the center for RNA and data analyses. The samples would be analyzed within a few days, and checkup data would be sent back to the users for diagnosis of their crops.

## 6.5 Potential Use of Field Transcriptome Data to Mine Genetic Resources Against Global Climate Change

Transcriptome data obtained under field conditions will be useful for breeding in the future. With global climate change, new cultivars with wide regional adaptability are required. Previously, a breeding method called shuttle breeding applied in the Green Revolution of wheat breeding was thought to be effective for developing new cultivars with wide adaptation (Hesser 2006). However, it is very laborious as the cultivars were selected in two distant areas with very different climates. Use of field transcriptome data from multiple fields with distinct climates would allow the selection of useful alleles to confer distinct responses depending on the given environments. In addition, we could evaluate which environments can give rise to distinct responses using statistical modeling. For example, a novel disease resistance gene allele that can exhibit a distinct temperature response would likely contribute to plant disease resistance within different temperature ranges. We can make use of this allele as a genetic resource to develop new cultivars with wide regional adaptability. In the process of maize domestication, ancient humans mainly would have used genetic changes in *cis*-regulatory elements of causal genes (Lemmon et al. 2012). The search for novel alleles with wide regional adaptability to improve crops would be reasonable from a historical perspective. There are three possible ways to identify candidates of such novel alleles. The first

would be the use of paired tests. As described above, we already have more than 40 paired samples in which significant detection would become very sensitive. The second would be to develop transcriptome-based models of specific traits, such as temperature responses, and compare with other models, such as LASSO models. The third is to develop environmental-driven models of gene expression. We can compare the selected values for specific parameters for each gene among tested cultivars and determine which environmental factors cause the differences of the gene expression. The third would require more than a hundred of RNA samples obtained under various cultivation conditions for each allele of tested cultivars for comparison. Known QTLs are included among candidate genes, and we would be able to identify novel useful alleles in the QTLs. If there is no available biological information on the candidate genes, the CRISPR/Cas9 method could have been applied to the candidate genes for molecular genetic analysis.

## 6.6 Future Crop Agronomic Performance Mediated by Field Transcriptome Data

### *6.6.1 Spatial and Developmental Regulation* In Planta *as Barriers for Field Transcriptome Analysis*

There are critical barriers when performing transcriptome analysis for plants cultured in the laboratory and/or field. One such barrier is that the samples for transcriptome analysis are mixtures of different plant organs. Although in the case of fully developed rice leaves, the ratios of mixing of tissues, such as vascular bundle cells and mesophyll cells, do not vary markedly, so we cannot discuss tissue specificity of expressed genes in rice leaves. In addition, when performing transcriptome analysis of differentiating tissues/organs, it would be very difficult to order the samples according to developmental stages among fluctuations in gene expression according to environmental factors. Therefore, at present we can only focus on fully differentiated tissues/organs. Furthermore, the damage due to sampling may affect future samples from the same plants. We usually try to harvest samples at intervals that are as short as possible in a sampling event. These flaws must be taken into account when designing biological experiments using field transcriptome analysis.

### *6.6.2 Understanding of the Effects of Ambient Environmental Conditions*

The final goal of field transcriptomics would be global integration between environment-driven models of gene expression and transcriptome-driven models of biological and agricultural traits (Fig. 6.4). Such integration would allow the

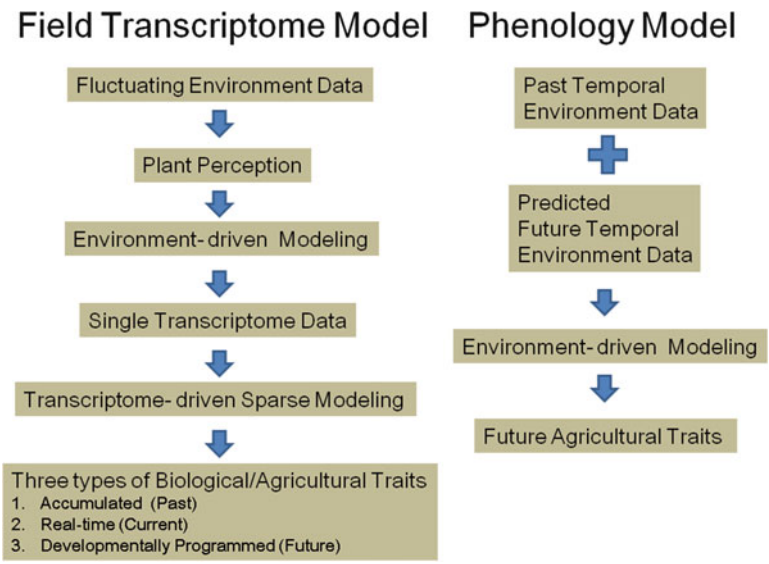



**Fig. 6.4** Comparison between field transcriptome and phenology models. In the field transcriptome model, prediction of single transcriptome data is a feature in the environment-driven model. Prediction of traits by single transcriptome data is a feature in the transcriptome-driven model

biological and agricultural traits, including some future traits, such as heading date or flowering time or yield-related traits, to be predicted based on historical environmental information. These analyses would also reveal the critical relationships between given environments and traits, including distinct responses in distinct developmental stages and gate effects of timing. Such integrated views from transcriptome data would provide dynamic responses of crops under naturally fluctuating environmental conditions.

**Acknowledgments** This work was supported by grants from MAFF, Japan (Genomics for Agricultural Innovation, RTR-1005; Genomics-based Technology for Agricultural Improvement, PFT-1001), to TI.

## References

Hesser L (2006) The man who fed the world: Nobel Peace Prize Laureate Norman Borlaug and his battle to end world hunger. Durban House, Dallas

Izawa T (2012) Physiological significance of the plant circadian clock in natural field conditions. Plant Cell Environ 35:1729–1741

Izawa T (2015) Deciphering and prediction of plant dynamics under field conditions. Curr Opin Plant Biol 24:87–92

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

LeCun Y, Bengio Y, Hinton GE (2015) Deep learning. Nature 521:436–444

Lemmon ZH, Bukowski R, Sun Q, Doebley JF (2012) PLoS Genet 10:E1004745

Lenormand M, Jabot F, Deffuant G (2013) Adaptive approximate Bayesian computation for complex models. Comput Stat 28:2777–2796

Lu WZ, Fan HY, Leung AY, Wong JC (2002) Analysis of pollutant levels in central Hong Kong applying neural network method with particle swarm optimization. Environ Monit Assess 79:217–230

Matsuzaki J, Kawahara Y, Izawa T (2015) Punctual transcriptional regulation by the rice circadian clock under fluctuating field conditions. Plant Cell 27:633–648

Nagano AJ, Sato Y, Mihara M, Antonio BA, Motoyama R, Itoh H, Nagamura Y, Izawa T (2012) Deciphering and prediction of transcriptome dynamics under fluctuating field conditions. Cell 151:1358–1369

Nakagawa H, Yamagishi J, Miyamoto N, Motoyama M, Yano M, Nemoto K (2005) Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. Theor Appl Genet 110:778–786

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol 58:267–288

Yang Y, Hoh J, Broger C, Neeb M, Edington J, Lindpaintner K, Ott J (2003) Statistical methods for analyzing microarray feature data with replications. J Comput Biol 10:157–169

Yang XS, Wu J, Ziegler TE, Yang X, Zayed A, Rajani MS, Zhou D, Basra AS, Schachtman DP, Peng M, Armstrong CL, Caldo RA, Morrell JA, Lacy M, Staub JM (2011) Gene expression biomarkers provide sensitive indicators of *in planta* nitrogen status in maize. Plant Physiol 157:1841–1852