Molecular phenology in plants: in natura systems biology for the comprehensive understanding of seasonal responses under natural environments

Hiroshi Kudoh
Center for Ecological Research, Kyoto University, Hirano 2-509-3, Otsu, Shiga 520-2113, Japan

Contents

Summary 399
I. Introduction 400
II. Definition and methodological advance of molecular phenology 401
III. in natura systems biology using high-resolution molecular phenology (HMP) data 402
IV. Predicting properties of a temperature-responding machinery (Step 1) 402
V. FLC as a candidate gene regulatory system for past temperature memory (Step 2) 404
VI. in natura quantification of past temperature memory by HMP (Step 3) 405
VII. Development of a molecular phenology-based mechanistic model (Step 4) 406
VIII. New insight into and analysis of the mechanism: digital repression of FLC (Step 5) 406
IX. New insight into and analysis of natural signals: phase lag calendar hypotheses (Step 6) 407
X. Transcriptomic HMP (revolution in Step 3) 409
XI. Perspectives: circles of in natura systems biology 409
Acknowledgements 410
References 410

Summary

Phenology refers to the study of seasonal schedules of organisms. Molecular phenology is defined here as the study of the seasonal patterns of organisms captured by molecular biology techniques. The history of molecular phenology is reviewed briefly in relation to advances in the quantification technology of gene expression. High-resolution molecular phenology (HMP) data have enabled us to study phenology with an approach of in natura systems biology. I review recent analyses of FLOWERING LOCUS C (FLC), a temperature-responsive repressor of flowering, along the six steps in the typical flow of in natura systems biology. The extensive studies of the regulation of FLC have made this example a successful case in which a comprehensive understanding of gene functions has been progressing. The FLC-mediated long-term memory of past temperatures creates time lags with other seasonal signals, such as photoperiod and short-term temperature. Major signals that control flowering time have a phase lag between them under natural conditions, and hypothetical phase lag calendars are proposed as mechanisms of season detection in plants. Transcriptomic HMP brings a novel strategy to the study of molecular phenology, because it provides a comprehensive representation of plant functions. I discuss future perspectives of molecular phenology from the standpoints of molecular biology, evolutionary biology and ecology.
I. Introduction

Flowering, bud break and leaf senescence occur at particular times of the year. In temperate regions, for example, seasonality of flowering time is common (Fitter & Peat, 1994; Tooke & Battey, 2010), and each plant species often has a peak of flowering time in the range $c. 10-20$ d (Ahas et al., 2000; Sparks et al., 2000). Such seasonal synchronization has attracted considerable attention in biology and agronomy, and ‘phenology’ refers to the study of seasonal schedules of organisms and how they are influenced by seasonal environmental changes.

Plant phenology shows year-to-year variation and its meteorological determinants have long been of interest in biometeorology (Schwartz, 2003). The prediction of plant phenology is an important task in agriculture and forestry, and phenology data over decades, especially flowering phenology, have accumulated in many localities (Fitter & Peat, 1994; Cleland et al., 2007; Tooke & Battey, 2010). Phenology has been modeled primarily in relation to temperature and with combinations of other environmental factors, such as photoperiod and precipitation (Schwartz, 2003; Donohue et al., 2015). Systematic shifts in flowering time, together with global warming, have been reported (Walther et al., 2002; Root et al., 2003), and the prediction of phenology has become an important task in monitoring climate changes.

Recently, the importance of the study of organisms in their natural habitats has been emphasized in the areas of biology that aim to understand gene function, as well as other molecular and cellular processes (Merquiol et al., 2002; Richards et al., 2009; Izawa, 2015), and this approach is referred to as $in~natura$ systems biology (Shimizu et al., 2011; Kudoh & Nagano, 2013). The term ‘$in~natura$’ has been coined to refer to the concept that combined knowledge from studies in the field ($in~natura$) and in the laboratory ($in~vivo$ and $in~vivo$) brings about a more comprehensive understanding of organismal functions (Fig. 1a). Molecular biology approaches have been introduced actively into ecology and evolutionary biology, but the recent emphasis on $in~natura$ studies stems from the idea that analysis in natural habitats is required for the understanding of gene and cell function, that is, the aims of

![Fig. 1 Schematic diagram explaining how a combined knowledge from studies in the field ($in~natura$) and in the laboratory ($in~vivo$ and $in~vivo$) provides a more comprehensive understanding of organismal functions (a), and a typical workflow of an $in~natura$ systems biological approach using high-resolution molecular phenology (HMP) data in the study of molecular phenology (b). Through the six steps, combined applications of molecular biology and statistical modeling in natural habitats can be achieved. First, one can predict the required properties of the machinery for seasonal responses solely by the patterns of environmental fluctuations (Step 1). However, as these steps make circles, it does not necessarily mean that Step 1 should come first. If one identifies a particular gene regulatory system of interest which shows an environmental response, the series of studies may be motivated by the knowledge of molecular mechanisms (Step 2). There are two circles in the series of steps in $in~natura$ systems biology, and the inner and outer circles tend to comprise mechanism-based studies using model species and field phenomenon-based studies using non-model plants, respectively.](image-url)
molecular genetics and cell biology (Fig. 1a). The ultimate reason for this requirement is attributable to the fact that all biological systems have been shaped by natural selection to function in natural habitats.

In an in natura study, different sets of questions concerning gene function are asked from those in laboratory studies. How are environmental signals perceived and how is noise filtered out by organisms in fluctuating environments? Is the regulatory system robust enough to function under the natural range of the whole set of fluctuations of abiotic and biotic factors? The analyses often aim to add novel interpretations, especially in terms of the robustness of the systems to existing noisy fluctuations of diverse factors in natural habitats. Phenology is the representative phenomenon in which underlying functions of molecular mechanisms are understood in the context of natural seasonal environments (Kudoh & Nagano, 2013).

The aim of this review is to introduce molecular phenology, a methodologically and conceptually new approach for the understanding of organismal seasonality. One of the recent advances in the study of plant phenology is the combined applications of molecular biology and statistical modeling in natural habitats.

II. Definition and methodological advance of molecular phenology

Here, I define ‘molecular phenology’ as ‘the study of the seasonal patterns of organisms captured by the techniques of molecular biology’. Therefore, molecular phenology data include seasonal patterns of gene expression, epigenetic modifications, and quantities of proteins, metabolites and other molecules. In particular, molecular phenology studies have been underpinned by the recent progress on the quantification methods of gene expression. An evaluation of the function and control of genes or gene networks in natural habitats represents an important conceptual advance of phenological studies. The yearly variation of phenological events is not merely a passive developmental or metabolic effect of preceding environments. The molecular phenology approach allows us to analyze phenology as an active response of plants that has evolved during environmental fluctuations in natural habitats (Kudoh & Nagano, 2013).

In earlier studies, seasonal changes in gene expression were examined semiquantitatively by the RNA gel blot (Northern blotting) or by semiquantitative reverse transcriptase PCR (Merquiol et al., 2002; Roca-Pérez et al., 2004; Anderson et al., 2005). As the earliest example, Merquiol et al. (2002) examined monthly changes in the transcription of heat-, drought- and oxidative stress-related genes for a desert legume, *Retama raetam*. They found the upregulation of genes encoding heat shock proteins (HSPs) during the summer, when plants often experience a temperature of 40°C. Summer accumulations of HSPs were also reported in *Solidago altissima* (Barua & Heckathorn, 2006) and *Iris pumila* (Manitašević et al., 2007). These studies provide good examples of molecular phenology at the protein level. Other examples using semiquantitative methods reported spring and summer upregulation of defense-related genes (cardenolide biosynthesis-related genes) in *Digitalis obscura* (Roca-Pérez et al., 2004), and seasonal changes in the transcription of dormancy- and carbohydrate metabolism-related genes in *Euphorbia esula* (Anderson et al., 2005).

Later, real-time quantitative (q)PCR became a standard method for the study of the molecular phenology of gene expression (Mayrhofer et al., 2005; Böhlenius et al., 2006; Aikawa et al., 2010). This allowed a more precise quantification of gene transcription for diverse plants by the determination of sequences of a target gene and the design of primers for qPCR. This method stimulated studies investigating the seasonal patterns of expression of genes with functions that were well characterized in model species. FLOWERING LOCUS T (FT) and its homolog were identified as florigens, mobile signals of flowering, in *Arabidopsis thaliana* (Corbesier et al., 2007) and rice (*Oryza sativa*, Tamaki et al., 2007). FT homologs and related genes were often examined in relation to the timing of initiation of flowering of diverse plants under natural conditions (*Populus trichocarpa*, Böhlenius et al., 2006; *Citrus unshiu*, Nishikawa et al., 2007; *Vitis vinifera*, Carmona et al., 2007; *Arabidopsis halleri*, Aikawa et al., 2010; Satake et al., 2013; Malus, Korda et al., 2010; a hybrid *Citrus*, Muñoz-Fambuena et al., 2011; a hybrid *Fortunella*, Nishikawa et al., 2011; Mangifera indica, Nakagawa et al., 2012; *Ficus carica*, Ikegami et al., 2013; a hybrid *Vitis*, Wang et al., 2014; *Brassica oleracea*, Ridge et al., 2014; *Fagus crenata*, Miyazaki et al., 2014), and all of these studies reported that FT homologs were upregulated at the associated timing of flowering. Other examples include the seasonal expression of the glutamine synthetase gene family, key regulators of ammonium metabolism and nitrogen flow in plants, in *Populus trichocarpa* (Castro-Rodriguez et al., 2011). Mayrhofer et al. (2005) showed that seasonal variation of isoprene biosynthesis-related genes in leaves of a hybrid *Populus* correlated with temperature and light environments.

Microarrays and RNA sequencing (RNA-Seq) are widespread and powerful transcriptomics technologies that simultaneously measure the genome-wide gene expression of multiple samples under natural conditions (Alvarez et al., 2015). These methods have made it possible to produce time-series transcriptomic data, and have been applied to molecular phenology studies (Nagano et al., 2012; Richards et al., 2012; Kobayashi et al., 2013). Microarray methods have been applied to the study of molecular phenology using model species, such as *A. thaliana* (Richards et al., 2012) and rice (*Oryza sativa*, Nagano et al., 2012; Matsuzaki et al., 2015). For example, Richards et al. (2012) grew two accessions of *A. thaliana* (Bay-0 and Sha) in an outside garden and obtained transcriptomic data of 15352 genes at nine time points. They reported that growth stage (either vegetative or reproductive) was the most effective component of transcriptional variance. By analyzing 8594 genes expressed during the vegetative stage, they identified temperature- and precipitation-associated major components of transcriptional variance that were over-represented by thermoregulatory and drought-responsive genes, respectively.

RNA-Seq brought the study of the molecular phenology of the transcriptome into non-model plants. Kobayashi et al. (2013) applied it to estimate a trigger of general flowering in South-East Asia. This phenomenon is spectacular in terms of the mass flowering that occurs at the community level. They examined the
transcriptomes of *Shorea beccariana* through the course of the phenomenon. They found that two flowering time genes, homologs of *FT* and a flowering repressor, *SHORT VEGETATIVE PHASE* (SVP), showed dynamic changes in gene expression, together with drought-responsive genes in response to drought. They concluded that drought is a trigger for general flowering (Kobayashi et al., 2013).

### III. In natura systems biology using high-resolution molecular phenology (HMP) data

Molecular phenology data represent a type of time-series observation made at multiple timings throughout the seasons. This contrasts with conventional phenology data, such as first flowering or bud break dates, that accumulate at the rate of once a year and take decades before conventional phenological models can be developed. The observation frequencies of typical molecular phenology data are 12 times yr⁻¹ (every month) or less, for example, four seasons, or before and after phenological events.

One important characteristic of molecular phenology is that one can increase the frequencies of observations as required. HMP data refer to measurements that occur frequently enough to give a high time resolution. The first weekly observations throughout the year (c. 50 times yr⁻¹) of molecular phenology were reported for a flowering time gene in a perennial *Arabidopsis* relative, *A. halleri* ssp. *gemmifera* (Aikawa et al., 2010). This 2-yr study produced 567 measurements for six individuals at 96 time points (Aikawa et al., 2010). Such HMP data allow us to develop phenological models that predict gene expression or phenology at phenotype levels from meteorological environments (Aikawa et al., 2010; Nagano et al., 2012; Satake et al., 2013). More importantly, in addition to simply correlating phenology with environmental factors, HMP is useful for the testing of more complex hypotheses with multiple parameters concerning gene regulatory systems under natural conditions.

A typical workflow of an in natura systems biological approach using HMP data is shown in Fig. 1(b). First, one can predict the required properties of the machinery for seasonal responses solely by the patterns of environmental fluctuations (Step 1 in Fig. 1b). Natural environments fluctuate enormously and seasonal cues contain a large amount of noise. Mechanisms that govern the seasonal responses should therefore have a property that is sensitive to the useful cues, but insensitive to other erroneous noises. As the next step (Step 2 in Fig. 1b), based on a knowledge of molecular biology, one needs to choose a candidate gene that may have predicted properties in its regulatory system. The accumulation of knowledge from laboratory studies is necessary for a successful choice of a gene to be analyzed in HMP. For example, flowering time control is one of the most extensively studied phenomena with regard to its underlying mechanisms, and it is relatively easy to choose target genes to be studied by HMP. Once targets are determined, the HMP data of the candidate genes are collected under natural conditions, and numerical analysis is applied to evaluate whether a particular regulatory system has the required properties (Step 3 in Fig. 1b). This process often brings about a novel interpretation or prediction of the function of the gene regulation under natural conditions.

HMP data allow us to develop a mechanistic model of phenology based on the gene regulation systems (Step 4 in Fig. 1b). The novel aspect in gene regulation requires new analyses for the underlying mechanisms (Step 5 in Fig. 1b), and further analyses of natural signals in a more complex context, for example, the processing and integration of multiple environmental cues (Step 6 in Fig. 1b). These new analyses are likely to stimulate the second round of HMP involving newly targeted genes or gene regulatory components.

In the following sections, as an example of an in natura systems biological approach to phenology studies, I review the analyses of *FLOWERING LOCUS C* (*FLC*), a temperature-responsive repressor of flowering. The extensive studies of the regulatory mechanisms of *FLC*, in both the laboratory and field, make this example a successful case in which a comprehensive understanding of gene functions is progressing at a considerable speed (Kudoh & Nagano, 2013; Berry & Dean, 2015).

### IV. Predicting properties of a temperature-responding machinery (Step 1)

The annual cycles of solar radiation create seasons (Trenberth, 1983; Jones et al., 1999), and amplitudes of seasonal changes in temperature are large enough to provide a major cue for plants to determine their annual schedules (Lang, 1952; Salisbury, 1963; Bernier, 1988; Battey, 2000). Seasonal patterns of temperature are often visualized as changes in monthly average temperature. However, when we draw a graph of instantaneous temperatures, we realize that the actual temperature in natural habitats fluctuates enormously at different time scales (Fig. 2a). The ranges of temperature changes within a single day often exceed the differences in the monthly average temperature over a couple of neighboring months. Although the spring and autumn transitions are the overall upward and downward directions of temperature measurements, it is common that colder days or weeks come later during the spring transition, and vice versa in the autumn transition. Therefore, the seasonal cue of temperature should be captured as long-term information that needs to be processed secondarily from the series of instantaneous temperatures.

The ease of detection of a seasonal pattern depends on the amount of short-term noise in the measurements. The relative magnitude of signal and noise is known as the signal-to-noise ratio (S : N ratio). A high ratio means that the signal is significantly larger than the noise, making the underlying trend of the measurements obvious. As we see from the instantaneous temperatures in Fig. 2(a), short-term noises of temperatures, such as day and night, day by day and week by week fluctuations, are greater than the seasonal signal. This means that seasonal temperature has low S : N ratio.

One way to level out large fluctuations in a set of time-series data is to use moving averages (Lewis, 1960). A simple moving average (SMA) is the unweighted mean of the previous data points. An SMA of temperature is formed by averaging past temperatures for a specific length of time. For example, 1-d SMA is the average...
temperature for the past 24 h, and 1-wk SMA is the average
temperature for the past 7 d (Fig. 2b and c, respectively). SMAs are
calculated at any timing, and therefore we can calculate
n-day SMAs for every day. One-, 2-, 4- and 6-wk SMAs are shown in
Fig. 2 (c, d, e and f, respectively). With longer reference periods of
SMA, the seasonal signal becomes conspicuous and the noise becomes less
remarkable. The temperature data (September 2006–August 2008) were
obtained from the meteorological station at Nishiwaki, Japan Meteorological Agency.

The description of past temperature trends using SMA has two
other effects. One is that the absolute magnitude of the signal
becomes smaller in longer day SMAs. Another is that SMA causes a
lag in the seasonal signal, and the lag becomes larger when the
averaging periods become longer. Principally, with longer averag-
ing periods for past temperature, the seasonal signal becomes easier
to detect. However, reference periods should have an upper limit
for three reasons; first, the signal needs to be larger than the
sensitivity of the putative temperature sensors of plants; second, the
lag of the signal should not obscure the timing of the response; and
third, the machinery should involve some type of memory to
average past temperature, but the error of memory itself is expected
to become larger when reference periods become longer. Based on the pattern of temperatures in central Japan, it has been predicted that the molecular machinery of flowering time control should use information of past temperatures for 4–6 wk (Kudoh & Nagano, 2013).

The earlier analyses suggest that a season-responding machinery using temperature cues is expected to have properties by which instantaneous temperatures are processed in a manner that is similar to calculating the SMA for the past 4–6 wk. Such machinery is expected to be characterized by a slow response that can filter out short-term temperature fluctuations and a reversible memory of temperatures that can store information for the past 4–6 wk. The memory is expected to be transferred through cell divisions that may occur during the 4–6-wk period and to be quantitative, at least, at a tissue level.

V. FLC as a candidate gene regulatory system for past temperature memory (Step 2)

One mechanism that is known to serve long-term memory at the cell level is based on the regulation of gene expression through chromatin remodeling (Turner, 2002; Ringrose & Paro, 2004). The gene expression is regulated by epigenetic modifications of histones through chromatin remodeling protein complexes, and the up- and downregulation of genes corresponds to facultative changes between euchromatin and heterochromatin states, respectively. The modification states can be transmitted through cell divisions, and the mechanism is referred to as ‘cellular memory’ (Lanzuolo & Orlando, 2012). This type of gene regulation has been reported in vernalization (Andrè, 2011). Expression of a gene encoding one of the PHD proteins, VERNALIZATION INSENSITIVE3 (VIN3), is induced by cold exposure for 3 wk and becomes highly active by 6-wk of cold (Sung & Amasino, 2004; Heo & Sung, 2011). The most prominent histone mark in the suppressed FLC region is the trimethylation of histone H3 at lysine 27 (H3K27me3). During the cold, the rising levels of H3K27me3 in the nucleation region show a quantitative change that correlates with an increase in expression of VIN3. Cold-induced H3K27me3 in the nucleation region accumulates to a maximum level after 4–6 wk exposure to cold (Angel et al., 2011).

After plants return to warm conditions, H3K27me3 spreads quantitatively across the rest of the FLC locus according to the length of the cold period (Angel et al., 2011). The quantitative increase in H3K27me3 levels correlates with the level of FLC suppression (Angel et al., 2011). A substantial increase in H3K27me3 across the whole gene is required for stable silencing throughout the rest of development, that is, completion of flowering and fruiting, of Arabidopsis (Finnegan & Dennis, 2007; De Lucia et al., 2008; Angel et al., 2011). Furthermore, in a perennial species, Arabidopsis alpina expression of PERPETUAL FLOWERING 1 (PEP1), an ortholog of FLC, is downregulated by prolonged cold, and this repression correlates with an increase in H3K27me3 levels within the locus, as in A. thaliana (Wang et al., 2009). PEP1 repression is not maintained stably after exposure of the plants to warm temperatures (Wang et al., 2009).

Overall, molecular genetic studies on FLC have characterized the following properties. (1) The multiple sequential processes allow gene regulation to respond slowly over the scale of 4–8 wk. (2) The response is quantitative, at least for the range of 2–6 wk of cold period, and is reversible for perennial species. (3) The mechanistic basis of the response is the state of epigenetic histone marks which is mitotically stable, and thus the response can be transmitted through cell divisions over multiple weeks, functioning as a quantitative memory at the tissue level. Therefore, FLC is a strong candidate of the molecular basis for past temperature memory.
VI. *In natura* quantification of past temperature memory by HMP (Step 3)

Whether or not this machinery provides a memory of seasonal temperature changes can be tested by HMP studies, using the *in natura* systems biological approach (Fig. 1b, “Step 3”). Assuming that *FLC* expression is controlled by past temperature memory, a quantitative estimation of memory length and its accuracy can be evaluated under natural conditions. We cannot measure the memory length of *FLC* suppression in most *Arabidopsis thaliana* accessions, because *FLC* silencing is maintained until the end of its annual life cycle. The measurement of memory length requires a phenomenon in which memory is wiped after it is maintained for a certain period. In some accessions of *A. thaliana*, the reactivation of *FLC* after vernalization has been reported when the length of cold is not long enough (Shindo *et al.*, 2006; Li *et al.*, 2014). In perennials in which suppression of *FLC* homologs is reversible even after long periods of cold, we can determine the length of memory in the *FLC* regulation.

Aikawa *et al.* (2010) conducted an HMP study of an *FLC* homolog, *AhgFLC*, in a natural population of *A. halleri* subsp. *gemmifera* located in central Japan. The species is a diploid perennial relative of *A. thaliana*, and its flowering peaks in late April in the study population. The expression of *AhgFLC* showed a clear seasonal pattern (Fig. 3a), although plants experienced seasonal temperature signals with enormous shorter term noise (Fig 3b). Marked changes in gene expression occurred in the autumn–spring seasons, and the expression decreased in autumn and increased in spring (Fig. 3a). The expression levels were kept high from May to November (Fig. 3a).

To quantitatively evaluate memory, the reference period of past temperature was estimated by applying cooling unit models to

![Fig. 3](image-url) Two-year high-resolution molecular phenology (HMP) of gene expression of *AhgFLC*, a *FLOWERING LOCUS C* (*FLC*) homolog, in a natural population of *Arabidopsis halleri* subsp. *gemmifera* examined at 1-wk intervals (a, average and SD for six individuals), hourly air temperatures during the HMP measurement period obtained from the nearest meteorological station to the study population (b), applications of cooling unit models to the gene expression data (c–f), and a schematic diagram representing the robust control of *AhgFLC* under natural conditions (g). To quantitatively evaluate the past temperature memory, the cooling degree-hour (CDH), the sum of temperatures lower than a certain threshold (*T*) over a certain past period (*L*), was calculated (c). Seasonal changes in CDH were calculated with the most likely *L* and *T* (d). Seasonal changes in *AhgFLC* expression correspond well with the past 6 wk (42 d) of CDH (e). Two-week weekly measurements of *AhgFLC* expression for six individuals were plotted against CDH with the most likely *L*, and solely the past 6 wk of temperature explained 83% of the variation in *AhgFLC* expression (f). The regulatory system of *AhgFLC* expression not only detects the seasonal temperature (past 6-wk temperatures, indicated by the orange arrow, g), but also provides robustness under complex natural conditions by ignoring short-term fluctuations of temperature and other environmental factors (indicated by strikethroughs and blue arrows, g). Original *AhgFLC* data were obtained from Aikawa *et al.* (2010).
HMP data (Fig. 3c). The cooling degree-hour (CDH) is the sum of temperatures lower than a certain threshold \( T \) over a certain past period \( L \) (Fig. 3c). HMP data of FLC expression were fitted against calculated CDH with diverse \( T \) and \( L \), and it was found that the most likely values for the threshold temperature and reference period were \( T = 10.5^\circ C \) and \( L = 42 \text{ d} \), respectively (Fig. 3d). Reference periods were important, and shorter and longer periods greatly reduced the explanatory power of the model.

The regulatory machinery of \( \text{AhgFLC} \) successfully extracted the seasonal pattern from noisy temperatures under natural conditions (Fig. 3e). Surprisingly, the selected model, that is, solely the past 6 wk of temperature, explained 83% of the 2-yr variation in \( \text{AhgFLC} \) expression (Fig. 3f). This estimate of a 6-wk reference period of past temperature corresponds well to what is expected for properties of season-detecting machinery solely by meteorological temperature data (Fig. 1b, Step 1). It also fits with the time scale of responses reported in the cellular memory mechanisms of FLC regulation (Fig. 1b, Step 2).

We should note that the study was conducted in a natural habitat in which factors other than temperature also fluctuate largely. The plants analyzed by Aikawa et al. (2010) experienced flooding, snow cover, desiccation stress and herbivory, but a large part of \( \text{AhgFLC} \) regulation was simply dependent on past temperatures (Fig. 3g). This indicates that the regulatory system of \( \text{AhgFLC} \) not only detects the seasonal temperature, but is also equipped with sufficient robustness to function under complex natural conditions (Fig. 3g).

**VII. Development of a molecular phenology-based mechanistic model (Step 4)**

The internal state of plants is measured all year around in molecular phenology, and HMP data have allowed us to develop a mechanistic model that predicts gene expression or phenology at phenotype levels from meteorological environments (Shimizu et al., 2011; Donohue et al., 2015; Izawa, 2015). HMP considerably accelerates the development of predictive models of phenology, and we acquired data for a year or two before developing phenology models (Satake et al., 2013). This is in contrast with the way in which conventional phenology models require data on phenology over decades (Schwartz, 2003).

Satake et al. (2013) modeled the timings of transitions between vegetative and reproductive growth, that is, the initiation and termination of the flowering period, in the yearly growth cycle of \( A. \text{halleri} \) by incorporating a gene network consisting of two key genes, \( \text{AhgFLC} \) and \( \text{AhgFT} \). In \( A. \text{thaliana} \), upregulation of FLC strongly suppresses \( F7 \) expression, and thus floral induction. Because Aikawa et al. (2010) reported that \( \text{AhgFLC} \) has the property to monitor long-term temperature changes, Satake et al. (2013) aimed to predict phenology at the phenotype level by a model with a minimally sized gene network. The dynamics of \( \text{AhgFLC} \) and \( \text{AhgFT} \) were represented by differential equations that included temperature-dependent production and degradation processes of the two genes. The dissociation constant of \( \text{AhgFT} \) suppression by \( \text{AhgFLC} \) was assumed to be temperature dependent. The lag in temperature rise and \( \text{AhgFLC} \) upregulation after cold was also incorporated into the model by assuming a certain level of irreversibility of \( \text{AhgFLC} \) expression at the initial stage after vernalization.

Plants from two populations located in different climatic regions (north and south populations in cool and warm temperate regions, respectively) were used (Satake et al., 2013). The laboratory experiments revealed that there was genetic variation in the response of gene expression. For example, \( \text{AhgFLC} \) decreased in response to cold of \( \leq 12.5^\circ C \) and \( \leq 10.0^\circ C \) for plants from the north and south populations, respectively. They determined parameters in the model base from the data obtained in controlled environments in the laboratory.

Four sets of HMP data were obtained for plants that were reciprocally transplanted into gardens located within the two climatic regions (Satake et al., 2013). These HMP data validated that the model successfully predicted the expression of the two genes, as well as the initiation and termination of the flowering periods in the four cases (two population origins \( \times \) two transplantation sites). The laboratory-based model was shown to predict flowering phenology accurately under natural fluctuating environments. The model further predicted that future global warming would cause significant decreases in the length of the flowering period of \( A. \text{halleri} \) (Satake et al., 2013). This example demonstrates clearly that laboratory-based studies on regulatory mechanisms can contribute to the prediction of plant phenology, especially when the model is validated by in natura HMP data.

Another set of recent phenology models that incorporate gene regulation networks are those characterized by the use of mutants of flowering time genes in \( A. \text{thaliana} \) (Wilczek et al., 2009; Chew et al., 2012; Valentim et al., 2015). The models were validated by comparing predictions and actual responses of the mutants. These models are also good candidates for evaluation by HMP data under natural conditions.

**VIII. New insight into and analysis of the mechanism: digital repression of FLC (Step 5)**

The novel insights into the function of a gene regulatory system under natural conditions invoke new analyses in the laboratory and also theoretical studies (Fig. 1b, Step 5). The fluctuating nature of temperature means that seasonal responses require a quantitative memory of long-term past temperatures that can be filtered from short-term noise (Aikawa et al., 2010; Kudoh & Nagano, 2013). Because this context is expected to be shared between temperate plant species, one expects that the machinery of the vernalization process at the tissue level, but a large number of cells exist in the sample. One of the questions is whether quantitative memory is stored at the cell level or tissue level (Angel et al., 2011). The former requires a quantitative memory mechanism of FLC regulation at a single cell, but the latter can be achieved by changing the number of cells that are in either one of the two states, that is, an ON state and an OFF state. Using a \( \beta \)-glucuronidase (GUS) reporter gene inserted in \( \text{FLC} \), Angel et al. (2011) demonstrated that FLC expression is
actually ON or OFF in individual cells. After longer cold treatment, an increasing number of cells are switched off, and the quantitative response of FLC expression at a tissue level is achieved as the ratio of cells in an ON and OFF state. Angel et al. (2011) showed that, by mathematical simulation, cell-autonomous ON/OFF switching generates the quantitative nature of vernalization.

The next question is whether this digital system functions as a mechanism for the response of FLC expression to interrupted cold under fluctuating natural temperatures. Using mathematical modeling, Angel et al. (2015) showed that, by assuming a digital process with a low but constant probability of transition from the ON to OFF states and irreversibility of the transition, cell-autonomous digital repression serves as an interruption-buffering system. They further exposed A. thaliana plants to c. 2–8 wk of cold that included one to three occasions of 4-d warm breaks, and found that plants responded similarly to interrupted and non-interrupted cold exposure. Overall, the studies indicated that digital repression is a mechanism by which plants can respond quantitatively to interrupted cold exposure without the need for individual cells to store complex quantitative information.

IX. New insight into and analysis of natural signals: phase lag calendar hypotheses (Step 6)

Studies introduced in this review indicate that plants can respond to seasonal temperatures by monitoring the long-term temperature change and filtering short-term fluctuations. However, there is one inherent problem for seasonal detection from an oscillating signal: the existence of the same signal level twice in a year, except for the maximum and minimum points. This problem can be solved by the use of multiple signals which differ in their oscillation phases. It is possible to identify seasonal timing all year round if plants utilize phase lags of multiple seasonal signals. Here, I term the function as the ‘phase lag calendar’. Although the use of multiple signals for seasonal detection has been suggested in previous reports (reviewed in Srikanth & Schmid, 2011; Andrés & Coupland, 2012; Franklin et al., 2014), there has been little reference to phase lags in these candidate signals under natural conditions. Here, using HMP data on AhgFLC of A. balleri (Aikawa et al., 2010), I illustrate hypothetical phase lag calendars by examining the patterns in candidate signals under natural environments. I discuss two phase lag calendars that are likely to have major roles in the seasonal control of plants, namely the photoperiod–temperature phase lag calendar and the short- and long-term temperature phase lag calendar.

1. Photoperiod–temperature phase lag calendar

One general but less explored context in natural conditions is the phase lag between photoperiod and temperature changes across the seasons (Fig. 4a). At the model study site (Aikawa et al., 2010), the hottest and coldest days (August and February, respectively) occur c. 1.5 months later than the longest and shortest days (c. June 22 and December 22, respectively; Fig. 4a). This phase lag between photoperiod and seasonal temperature is relatively long near the coastal region compared with that inland (Prescott & Collins, 1951; El-Hussainy & Essa, 1997). It has been reported that the phase lag is in the range 38–48 d at the eastern margins of the continents, 45–65 d at the western margins of the continents, and 12–21 d at the inland basins and highlands (Prescott & Collins, 1951).

Photoperiod provides a reliable signal of seasons in terms of predicting calendar dates (Farner, 1964). There is no yearly variation in photoperiod pattern and, with a given calendar date and a latitude, one can calculate the theoretical photoperiod precisely (Withrow, 1959; Lee, 1970). In A. thaliana, FT expression is upregulated by the accumulation of CONSTANS (CO) protein towards dusk (Fig. 4b; Yanovsky & Kay, 2002; Y. H. Song et al., 2012). CO gene transcription is regulated by the circadian clock, and the transcribed CO protein is stabilized in the light, but rapidly degrades in the dark (Valverde et al., 2004). Therefore, only when the peak timing of CO gene transcription coincides with daytime (i.e. long day) does the CO protein accumulate, thus leading to FT upregulation (Andrés & Coupland, 2012). It is likely that photoperiod is measured and revised every day in A. thaliana, and the experimental transfer of plants between long and short days has indicated that the levels of FT expression are adjusted to the corresponding photoperiods within 3 d (Corbier et al., 2007).

Because of the memory of long-term past temperature in plants, the phase lag between photoperiod and temperature signals is expected to become even greater in plant cells. Assuming that A. balleri has a similar system of daily measurements of photoperiod as A. thaliana, we can visualize a hypothetical photoperiod–temperature phase lag calendar by plotting 1-yr records of AhgFLC expression against photoperiod (Fig. 4c). In this two-dimensional phase space, the timing of floral initiation (observed as bolting in A. thaliana), as well as successive flowering period (terminated by reversion to vegetative growth), are well separated from other seasons.

2. Short- and long-term temperature phase lag calendar

As a property of past temperature memory, the lag between phases of long past memory and current temperature becomes larger when periods of memory become longer. This property provides phase lag information between short- and long-term temperatures. In A. thaliana, the temperature dependence of flowering time is not only regulated by FLC-mediated vernalization, but also by other mechanisms that respond to ambient temperature (Wigge, 2013; Verhage et al., 2014; Capovilla et al., 2015).

In A. thaliana, earlier flowering at 23°C than at 16°C has been reported under long-day conditions (Blázquez et al., 2003). Furthermore, a moderate temperature increase from 23 to 27°C has been shown to induce flowering under a short-day photoperiod (Balasubramanian et al., 2006). In contrast with the detailed understanding of the vernalization mechanism, responses to warm ambient temperatures have been studied recently (Samach & Wigge, 2005; Capovilla et al., 2015). As temperature increases under long-day photoperiods, SVP protein, one of the components of a repressive complex, is actively degraded, allowing for the activation of FT and SOC1 expression (Lee et al., 2013). Temperature-dependent alternative splicing of FLOWERING LOCUS M
(FLM) also affects the activity of the SVP–FLM complex (Lee et al., 2013; Posé et al., 2013). Recently, microRNAs (miRNAs) have been suggested to be involved in the temperature-dependent regulation of flowering (Lee et al., 2010; Kim et al., 2012).

Assuming that short-term warm temperatures enhance flowering, I propose that there is a hypothetical phase lag calendar between short- and long-term temperatures. Such a calendar can be visualized by plotting 1-yr records of AhgFLC expression against photoperiod (Fig. 4c). The above two combinations of phase lag calendar hypotheses are not mutually exclusive. If one considers the reported mechanisms of flowering time control in A. thaliana (Pajoro et al., 2014), the three signals in Fig. 4c are likely to be integrated to function as a three-dimensional phase lag calendar. Here, it is most noteworthy to point out that three representative environmental signals of flowering time control have lags between the phases of their seasonal patterns under natural conditions. Photoperiod has the most advanced phase, followed by the short-term temperature signal with a 4–6-wk phase difference, and then by the long-term temperature signal with an additional 4–6-wk phase difference. We need to take the lags into account when we interpret the role of signal perceptions of these seasonal signals. Growth chamber experiments in which phase lags are manipulated between different signals will be revealing. Obviously, we require in natura studies to understand how multiple seasonal signals are coordinated in a plant cell, and how plants can robustly determine their phenology under natural fluctuating environments.
X. Transcriptomic HMP (revolution in Step 3)

As an example of recent advances in molecular phenology, I have reviewed studies on FLC in both the laboratory and field by locating them on the six steps of in natura systems biology (Fig. 1b). Through these six steps, combined applications of molecular biology and statistical modeling in natural habitats can be achieved. As these steps make circles, it does not necessarily mean that Step 1 should come first. If one identifies a particular gene regulatory system of interest which shows an environmental response, the series of studies may be motivated by the knowledge of the molecular mechanisms (Step 2 in Fig. 1b).

Transcriptomic technologies, such as microarrays and RNA-Seq, allow us to simultaneously measure genome-wide gene expression on large numbers of time-series samples under natural conditions (Alvarez et al., 2015). HMP of the transcriptome brings a novel strategy to the study of molecular phenology. Now, it is possible to begin studies using in natura systems biology from Step 3 in Fig. 1(b). Transcriptomes provide a comprehensive functional perspective of plant function, and therefore more exploratory questions can be asked, for example, ‘which genes show seasonality?’ Seasonal transcriptome analyses allow us to identify candidate genes that associate with phenology at the phenotype level and that respond to particular biotic and abiotic environmental signals.

The first transcriptomic HMP data were obtained under natural fluctuating environments on field-grown rice (Oryza sativa, Nagano et al., 2012; Matsuzaki et al., 2015). The data were time-series transcriptomes of leaf tissues obtained by 461 microarrays at distinct time points during rice cultivation from May to October (Nagano et al., 2012; Matsuzaki et al., 2015). Nagano et al. (2012) successfully modeled the expression patterns of 97% of expressed genes by single environmental factors, even though they sampled plants outside, where multiple environmental factors fluctuated with large complexity. Their results indicated that robustness might be common in gene regulation, and that the regulatory mechanisms of certain groups of genes are generally sensitive only to a few particular environmental factors and are robust against noises that exist in natural conditions.

In the analyses of rice HMP, they evaluated two properties that are required for the photoperiodic response, that is, punctual time measurement provided by the circadian clock and precise light/dark discrimination initiated by sensitive photoreceptors (Troein et al., 2009; Nagano et al., 2012; Matsuzaki et al., 2015). Rice is a typical short-day plant and the timing of ear emergence is under strict controls of differences in day length of <30 min (Itoh et al., 2010). Obviously, the two properties are required for rice to control flowering time in response to small changes in photoperiod, and both machineries need to be robust enough not to be disturbed by changes in the external environment in natural conditions (Izawa, 2012).

The robustness of the circadian clock has been evaluated in rice by constructing a molecular timetable (Matsuzaki et al., 2015). The molecular timetable is a systems biological approach in which a predictive model for physical time in a day is developed using expression data of multiple circadian-regulated genes (Ueda et al., 2004; Kerwin et al., 2011). Matsuzaki et al. (2015) quantified the punctuality of the internal circadian clock and evaluated how the punctuality is affected by environmental fluctuations caused by weather. Interestingly, the expression of individual genes was strongly affected by temperature, but the time table estimated by 16 genes was punctual to 22 min. The plant clock system as a whole was robust against changes in external conditions, and had the predicted property that provided punctual internal time under natural fluctuating environments.

Light/dark discrimination of plants in natural conditions has been evaluated using HMP data of the rice transcriptome (Nagano et al., 2012). It has been predicted that photosensing organisms should have low thresholds of light sensitivity because their effective photoperiods are not dependent on the weather (Withrow, 1959). By calculating the S : N ratio for seasonal changes in day length, Nagano et al. (2012) predicted that light/dark discrimination of rice plants should be initiated by specific photoreceptors with a low threshold of solar radiation. For the 130 genes whose expression was successfully modeled using solar radiation with time gating, they found that the gating times of many genes corresponded to the center of seasonal transition of dawn and dusk when they assumed light/dark discrimination at 0.3 kJ m^{-2} min^{-1} solar radiation (Nagano et al., 2012). They concluded that a highly sensitive photoreceptor, such as phyA (Shinomura et al., 1996), should contribute to the seasonal responses to photoperiod.

By transcriptomic HMP, the modeling of phenology, not only for flowering time control genes, but also for genes responsible for other functions, such as photosynthesis, stress tolerance, defense, dormancy and senescence, becomes possible. These models will provide us with a unique opportunity to discover novel functions of genes that cannot be granted through controlled conditions in laboratory studies (Alvarez et al., 2015). Furthermore, the molecular phenology of the transcriptome may identify unexpected cross-talk between distinct regulatory systems.

XI. Perspectives: circles of in natura systems biology

There are two circles in the series of steps in in natura systems biology (Fig. 1b). The inner circle tends to comprise mechanism-based studies using model species. HMP analyses will allow them to identify novel transcripts or novel functions of known transcripts. The outer circle tends to be followed by field phenomenon-based studies using naturally growing, non-model plants. Predictive models can be developed using HMP data. The recognition of these two circles in in natura systems biology provides a reinforcing toolset to accelerate the inner and outer circles, respectively; that is, detailed ecology of model species and transgenic technologies for non-model species. Here, I wish to make three points concerning the future perspectives of molecular phenology from the standpoints of (1) molecular biology, (2) evolutionary biology and (3) ecology.

1. Molecular biology

Rapid advances in the knowledge of the control of gene expression should be incorporated into in natura systems biology. Epigenetic
mechanisms, such as DNA methylation and histone modifications, are expected to become particularly important for the understanding of the robust control of gene expression under fluctuating complex environments (Huff & Zilberman, 2012; Baulcombe & Dean, 2014; Berry & Dean, 2015). In particular, functions such as a memory of past environments and a high-pass filter for noise reduction are expected. Gene expression controls through ncRNAs, such as miRNA and short interference RNA (siRNA), should also be taken into account (Fahlgren et al., 2007; Leung & Sharp, 2010). Genome-wide analyses of DNA methylation, histone modifications and small RNAs along seasonal time series will be revealing.

2. Evolutionary biology

I expect that transcriptomic HMP data will increase using diverse plants because the RNA-seq method is widely applicable to non-model species. The comparative methods between different lineages will be revealing to understand how diversification and convergence have shaped the response machinery during the evolution of the seasonal response in plants at the macroevolutionary level. For example, advances in geographic distributions from the tropics to temperate regions have occurred repeatedly in multiple plant lineages, and vernalization responses in cereals (monocot) and Arabidopsis (dicot) are likely to have evolved independently. In wheat, the FLC equivalent is a MADS-box gene that is non-homologous, but shares a histone modification-mediated regulatory mechanism with FLC (Greenup et al., 2009; Hemming & Trevaskis, 2011). HMP analyses of these genes are required to evaluate whether this is the result of convergence via adaptation to the shared signal/noise context of seasonal temperature. At the microevolutionary level, HMP from common garden experiments and reciprocal transplant experiments will detect genetic variation and local adaptation in seasonal responses.

3. Ecology

HMP data for multiple species at a community level can be obtained. In a community, diverse types of interactions between species are taking place along seasonal schedules under a shared seasonal signal. For example, reproductive isolation of closely related plant species is sometimes achieved by phenological isolation (Franks & Weis, 2009). Seasonal synchronization is often observed in plant–animal interactions, such as pollination, seed dispersions and herbivory (Elzinga et al., 2007). These phenomena should be analyzed in terms of molecular phenology to reveal what mechanisms underlie the responses to shared environmental signals. Furthermore, predictions of perturbations in plant–plant and plant–animal interactions should be made in response to global environmental changes.

Acknowledgements

The author thanks Dr A. N. Dodd and an anonymous referee for comments on an earlier version of the manuscript; Drs K. K. Shimizu, A. Satake, D. Buzas and A. J. Nagano for discussions during the collaboration on molecular phenology studies; and T. Kato for support during manuscript preparation. This study was supported by a Grant in Aid for Scientific Research (S) 26221106, Japan Society for the Promotion of Science (JSPS), and by Research Grants in Natural Science (26118), the Mitsubishi Foundation.

References


Angel A, Song J, Yang H, Quest JI, Dean C, Howard M. 2015. Vernalizing cold is registered digitally at FLC. *Proceeding of the National Academy of Sciences, USA* 112: 4146–4151.


