Demographic history and genetic factors associated with flowering time variation in Japanese *Lotus japonicus*

Tomomi WAKABAYASHI

要旨

種子植物にとって、開花時期は繁殖成功度に関わる大きな要素の一つである。温度や日長は開花時期決定の外的要因として知られており、植物はこれらのシグナルを受けて適切な時期に栄養成長から繁殖成長へと移行する。広域分布種では、異なる環境下にある生育地間で開花時期の違いが観察される場合があり、開花時期の種内多型はそれぞれの地域への環境適応を反映すると考えられる。日本列島には約 25 度の幅広い緯度差があり、地域間で温度や日長に大きな違いが存在する。そのため国内に広域分布する植物種では、各々の生育環境に適応して最適な開花時期を維持しており、これが遺伝的に固定されている可能性が想定される。マメ科のミヤコグサは、本邦で琉球列島から利尻島までの広い分布域を持ち、研究用モデル植物のバイオリソースとして多数の野生系統の種子が維持されている。これら野生系統の種子では、高緯度の由来産地ほど発芽から開花開始までの所要日数が長期化すること、ならびにこの性質は遺伝的に固定されていることが知られている。各野生系統の全ゲノム塩基配列が既知であることから、本研究では、ミヤコグサの野生系統における開花時期多型に関与する遺伝的要因を解析して、日本列島の幅広い環境への適応機構を解明することを目的にして、以下の研究を行った。

広域分布種では分布拡大に伴って、各地域の環境要因による自然選択の影響を受け、生育環境に適した開花時期を呈する個体が選択されたと考えられる。そこで第一章では、ミヤコグサの日本列島における分布形成の歴史的背景を明らかにした。国内の分布を網羅する 136 系統の野生系統を対象に、全ゲノム塩基多型情報を用いた集団遺伝学的解析を行った。fastSTRUCTURE によるクラスタリング解析では、日本のミヤコグサは地理的なまとまりを持つ 2-6 の集団に分けられ、特に K=3 の場合には緯度に沿って明瞭な地理的構造があることが示された。PSMC 解析による分岐年代推定では、クラスタリング解析と同様な地理的なまとまりを持つ集団構造の存在や、対馬の系統が他の全ての系統との分岐が浅く、その年代は 3,000-10,000 年前であると推定された。先行研究においても、対馬と他の系統との遺伝的距離が大きいことが示されているため、これらの知見を併せると、国内のミヤコグサは対馬周辺から最終氷期後に日本列島に広く分布拡大したと考えられる。

次に第二章では、既知の開花遺伝子に焦点を絞り、ミヤコグサにおけるその相同遺伝子が自然選択の影響を受けているかどうかを分子進化学的解析により明らかにした。対象の遺伝子は、同じマメ科のダイズで開花時期決定に特に重要である *E1*, *GI* 遺伝子で、本章ではそれぞれ 18 地点 49 個体、8 地点 23 個体のミヤコグサについて、相同遺

伝子 *LjE1*, *LjGI* の配列を決定し、MK 検定と Tajima's *D* 検定によって自然選択の有無を検定した。その結果、*LjGI* については 2,930bp の配列にほとんど変異が見つからず、純化選択の影響を受けた可能性が示唆された。一方、*LjE1* については 552bp の CDS の配列において、塩基置換パターンに地理的な傾向を持った 5 サイトが確認された。しかもこれらは非同義的置換であった。これらの遺伝子型はそれぞれ、一部を除く北海道と青森、岐阜以北の本州と九州の一部と琉球列島の一部、鳥取、九州東部、琉球列島の一部に由来する系統に見られ、MK 検定の結果、正の自然選択を受けたことが明らかにされた。このことから、ミヤコグサにおいてもダイズと同様に、*E1* が緯度の違いに応じた開花時期の調節に関与している可能性が考えられる。

さらに、第三章ではダイズの *E1* とは異なる本種特有の開花時期多型に関わる遺伝子を探索した。全ゲノム関連解析によるゲノム網羅的な解析による候補遺伝子の探索を行った。132 系統の国内野生系統を用いた同一条件下による開花時期多型の計測では、生育地の緯度と種子発芽から開花までの所要日数の間に相関があることを確認した。従って、緯度に沿って変わる温度や日長などの環境要素が本種の開花時期決定に関連していると考えられた。この開花時期多型データをもとに、121 系統について混合モデルを利用した全ゲノム関連解析を行ったところ、シロイヌナズナで開花時期制御に関わる *SmD3b-like* の相同遺伝子など含む 16 遺伝子が検出された。また、全ゲノム関連解析において開花時期多型データと特に高い関連を検出された 10 の塩基多型を用いて、本種が呈する開花時期変動のモデル予測を行ったところ、上位 2 つの塩基多型で約 56%の開花時期変動が説明されることを示した。このうち開花時期多型と最も関連の高かった塩基多型の周辺領域には、多種で開花時期制御に関連する *SmD3b-like* の相同遺伝子などが位置していた。このことから、本種の開花時期多型は *SmD3b-like* の相同遺伝子などを含む少数の遺伝的要因によってその大部分を制御されている可能性が考えられる。

これらの研究から、本種は日本列島各地の環境に対して開花のタイミングを適応させながら、最終氷期後の温暖化な時期に対馬周辺から日本列島各地へ分布を拡大したと考えられる。本種で観察される開花時期多型は、*LjE1* に見られたように自然選択の影響を受けて種内分化が促進され、また、*LjE1* や *SmD3b-like* の相同遺伝子などを含む少数の遺伝的要因によってその大部分を制御されている可能性が考えられた。

Contents

Contents …… **1**

Summary …… **2**

General Introduction …… **4**

Chapter 1 …… **6**

Demographic history of Japanese *Lotus japonicus* accessions based on genome-wide nucleotide polymorphisms

Chapter 2 …… **14**

Polymorphisms of *E1* and *GIGANTEA* in wild populations of *Lotus japonicus*

Chapter 3 …… **23**

Detecting loci that contribute to flowering time variation in *Lotus japonicus*

General Discussion …… **43**

Acknowledgements …… **45**

References …… **46**

Summary

In angiosperms, the flowering time is an important factor in reproductive success and is subjected to local adaptation. The Japanese archipelago lies across a wide latitudinal range of approximately 25 degrees and exhibits wide temperature and daylength differences from north to south. It is therefore likely that local adaptation may cause a variation in the flowering time of any species distributed throughout Japan. To understand the mechanism of local adaptation, it is important to determine a species' demographic history, as climatic change and natural selection effects may have caused its members to undergo repetitive microevolutions, the traces of which remain in its demographic history.

Lotus japonicus is a leguminous plant with a wide distribution along the length of the Japanese archipelago, from northern Hokkaido to the southern Ryukyus; it exhibits remarkable differences in flowering time. In this study, I aimed to reveal the demographic history of *L. japonicus* in Japan and detect the genetic factors affected by natural selection that cause the flowering time variation in this species, using genome-wide single-nucleotide polymorphism (SNP) data.

Chapter 1 reveals the demographic history of *L. japonicus* in Japan. Population clustering analysis and divergence time estimation were performed with genome-wide nucleotide polymorphism maintained among wild accessions. The population clustering analysis, with 136 accessions, showed the existence of subpopulations with geographical units. Considering the composition of each group and the geographical units, a large fraction of the variation could be attributed to three subpopulations distributed in the southern, middle, and northern parts of the Japanese archipelago. Some of the accessions originated in the Kyushu area, while Tsushima Island harboured three mixed clusters. For divergence time estimation, pairwise sequentially Markovian coalescent (PSMC) analyses based on pairwise pseudo-diploid sequences of the 65 accessions were conducted. The results showed that the Tsushima accessions appear to have diverged recently from all other accessions, suggesting that the northern subpopulation spread from the southern part of Japan. In addition, according to these results and the genetic diversity analysis performed by Shah *et al*. (2020), all the Japanese *L. japonicus* accessions may have originated in Tsushima Island and colonised the Japanese archipelago from there over the past ~20,000 years.

In Chapter 2, molecular biological analyses were conducted for two orthologues of the known flowering time causal genes in soybean (*Glycine max*). *E1* and *GIGANTEA* (*GI*) play a central role in transmitting signals to *FLOWERING LOCUS T* (*FT*). In order to reveal whether the orthologues in *L. japonicus* have been affected by natural selection, I analysed polymorphisms of *LjE1* and *LjGI* among wild populations covering the entire distribution range of this species in Japan. *LjE1* harboured five nucleic acid changes in a 552-bp CDS, all of which were nonsynonymous; four of the changes were located in the core function area. *LjE1* alleles exhibited partial north-south differentiation and non-neutrality. In contrast, *LjGI* harboured one synonymous and one nonsynonymous change. Thus, my results suggest that *LjE1* may be involved in the control of flowering times, whereas *LjGI* may be under the effect of strong purifying selection.

In Chapter 3, I conducted genome-wide association (GWA) analyses for flowering time variation in wild accessions of *L. japonicus* and for two environmental factors based on the information originating from the accessions. The flowering time variation in 132 wild accessions showed that there were remarkable flowering time differences along the latitude in which each accession originated. Considering the population structure, I adopted a mixed model that takes into account population structure for GWA analysis. In the analysis of flowering time variation, 16 protein-coding genes were detected based on the Bonferroni-corrected 5% significance threshold. For the 16 candidate genes, two orthologues of known flowering time genes were contained based on the results of gene annotation. For the environmental factors, there were few SNPs beyond the threshold; therefore, I focused on the top 100 SNPs. In regions around the SNPs detected for annual average temperature and integrated annual sunshine duration, 157 and 216 protein-coding genes were located, respectively. The 216 genes for integrated annual sunshine duration contained two orthologues of known flowering time genes. Based on the results of the GWA analysis for flowering time variation, correlation tests were conducted between genotypes of highly detected SNPs and the flowering time variation. Adding the non-linked top SNPs in order from one to 10, the coefficient determinations levelled off at approximately 0.6 in cases with more than two SNPs; this result suggests that the flowering time variation could be mainly explained by two SNPs in the candidate genes.

A comprehensive consideration of these results suggests that this species has expanded its distribution from Tsushima Island by adapting to each environment

of the various regions in Japan after the last ice age. It could be suggested that during this process *L. japonicus* could have been affected by natural selection in terms of its flowering time, which is one of its adaptive traits; a few of the genetic factors, including *LjE1* and orthologues of *SmD3b-like* related to the control of flowering time may contribute significantly to the local adaptation and distribution of the species on the various environments.

General Introduction

Since Schemske (1978), many botanists have been interested in the relationship between flowering time and fitness. In the wild, flowering time is important for a plant species because it is one of the factors influencing its reproductive success (e.g., Hall and Willis, 2006; Izawa, 2007). As the appropriate flowering time phenotype could be selected in each environment, phenotype differences may arise among ecotypes of different environments. Accordingly, it is important to reveal the genetic factors that cause flowering-time variation in a species in order to understand the mechanism of local adaptation. Recent studies have focused on the intraspecific polymorphism of flowering time as an adaptive trait (e.g., Franks *et al*., 2007; Ashworth *et al*., 2016). Considering that flowering time is controlled by several environmental factors (e.g., temperature and daylength; Andrés and Coupland, 2012), flowering-time differentiation tends to be clear, especially in areas such as the Japanese archipelago with its latitudinal spread, which has a wide range of abiotic environments. The Japanese archipelago extends from approximately 20° N to 46° N, covers more than 27,000 km, and has great temperature and daylength differences among its different regions. It is therefore likely that a species distributed throughout Japan would exhibit flowering-time variations as a result of adapting to the particular environmental characteristics of each habitat. Local adaptation studies should aim to determine the demographic history of a species, in addition to identifying the genes responsible for flowering-time variation. This is especially true for species distributed over a wide range of environments, as climatic change and natural selection effects have caused its members to undergo repetitive microevolutions, the traces of which remain in the species' demographic history.

The signalling pathways controlling flowering time have been studied extensively in the field of physiology, with several model plants. In *Arabidopsis thaliana*, more than 100 genes are involved in flowering-time control (Srikanth and Schmid, 2011). These studies have shown that four principal pathways (photoperiod, temperature, gibberellin, and autonomous) converge on *FLOWERING LOCUS T* (*FT*). *FT* is the common functional gene of most angiosperms and triggers the initiation of a flower bud on the meristem of *A. thaliana* (Kardailsky *et al*., 1999; Kobayashi *et al*., 1999). The flowering-time variation in wild plants could be ascribed to the molecular evolution of several genes, such as *PHYTOCHROME C* (*PHYC*; Balasubramanian *et al*., 2006), *PHYTOCHROME B2* (*PHYB2*; Ingvarsson *et al*., 2006), *FLOWERING LOCUS C* (*FLC*; Deng *et al*., 2011), and *FT* (Liu *et al*., 2014). In addition, *E1*, a leguminous plant species-specific gene, is known as one of the main flowering-time genes in soybean.

Recent developments in sequencing technology have enabled us to determine the nucleic sequences of many organisms and identify genome-wide nucleotide polymorphisms. This high-density information helps reveal demographic history details and detect the causal genes of various traits in the whole genome. As far as revealing demographic history using genome-wide nucleotide polymorphism is concerned, several studies describe the dynamics of the effective population size and of the divergence process from past to present population structures (Li and Durbin, 2011; Durvasula *et al*., 2017; Fulgione *et al*., 2018). Genome-wide association studies (GWAS) are especially useful for the detection of causal genes, because species polymorphisms are widely available. In fact, many studies on several plant species have adopted GWA analyses for various traits, including flowering time (*Oryza sativa*, Zhao *et al*., 2011; *A. thaliana,* Atwel *et al*., 2010; soybean, Zhang *et al*., 2015). In addition, some of the studies involved the use of GWA analyses to detect loci with adaptive traits and abiotic environmental factors, such as temperature and daylength.

Lotus japonicus is a leguminous plant that is widely distributed throughout the Japanese archipelago from the northern Hokkaido Prefecture to the southern Ryukyus; it exhibits remarkable differences in flowering time. As a model plant, its whole-genome information is available (Sato *et al*., 2008). Moreover, wild accessions collected from wild populations have been deposited in the National BioResource Project Legume Base (University of Miyazaki, Japan) and over 150 accessions are available for genetic study worldwide. With this available information, it should be possible to reveal this species' demographic history and mechanism of local adaptation through flowering time, from a genetic perspective.

In this study, I aimed to reveal the mechanisms of local adaptation of *L. japonicus* by examining its flowering time and the preservation of its flowering-time variation. To achieve this, I employed three types of analysis. First, I performed population clustering analysis and an estimation of the effective population size, based on the whole-genome nucleotide polymorphism of Japanese wild accessions (Chapter 1), in order to reveal the population dynamics of *L. japonicus*. Secondly, I conducted molecular biological

analyses for two orthologues of the known causal flowering-time genes in soybeans (Chapter 2). Finally, I conducted genome-wide association (GWA) analyses for flowering-time variation in wild accessions and for environmental factors, based on the original accession information (Chapter 3). The results of the population clustering analysis showed that Japanese *L. japonicus* accessions could be divided into three clusters according to the geographical units within Japan and Tsushima Island, which harbours a group composed of three mixed clusters. The effective population size estimation suggests that the Tsushima accessions have diverged recently from all the other accessions. Considering these results along with the genetic distance analysis results of Shah *et al*., (2020), the *L. japonicus* populations in Japan probably originated in Tsushima Island and colonised the Japanese archipelago from there after the last glacial period. The molecular biological analysis revealed that the *E1* gene in this species was affected by positive natural selection and that the allele distribution exhibited geographical differentiation. In addition, several candidate genes, including orthologues of known flowering-time genes, were detected by GWA analysis for flowering-time variation. A correlation test between the genotypes and the flowering-time variation showed that approximately 60% of the flowering-time variation could be explained by the top two single-nucleotide polymorphisms (SNPs) detected by GWA analysis. These results can provide valuable information to studies of local adaptation and biodiversity.

Chapter 1

Demographic history of Japanese *Lotus japonicus* accessions based on genome-wide nucleotide polymorphisms

Abstract

The migration and colonisation history of a species can contribute to the understanding of mechanisms of local adaptation and biodiversity, especially in a region with very wide environmental ranges, such as the Japanese archipelago. In addition, knowledge of the colonisation process and existence of population clusters helps identify candidate genes by association analysis with natural nucleotide variations for adaptive traits, such as flowering time and overwintering. *Lotus japonicus* is a plant distributed across the Japanese archipelago, which has various environments formed by location and topography. Here, I tracked the colonisation process whereby Japanese *L. japonicus* took up this long-range distribution across various environments; I used whole-genome nucleotide polymorphisms based on genome re-sequencing. First, I conducted a fastSTRUCTURE analysis to confirm the subpopulations involved in *L. japonicus* in Japan. The results showed that the wild accessions can be divided into several subpopulations with a geographical tendency. Next, pairwise sequentially Markovian coalescent (PSMC) analysis was conducted using pseudo-diploid sequences for all 65 accession pairs; the pairwise divergence time was then estimated. According to the estimation, several clusters aligned with the topography and the major clusters overlapped with the subpopulations detected by the fastSTRUCTURE analysis. In addition, the PSMC analysis results showed that two accessions that originated in Tsushima Island appear to have diverged recently from other accessions. Considering these results and the genetic distance analysis results obtained in a previous study, it is probable that the Japanese *L. japonicus* populations spread from Tsushima Island after the last ice age.

Introduction

Knowledge of the population history of a species is instrumental in understanding its mechanisms of local adaptation and intraspecific differentiation. This is especially true for a region with a variety of environments, such as the Japanese archipelago, in which an explanation of the process whereby a species spreads and colonises would contribute to the inference of the factors or mechanisms related to adaptive traits that are observable today.

Lotus japonicus is a leguminous plant that is distributed across the entire Japanese archipelago; it is reasonable to expect that the wild populations would adapt to each environment. Owing to its wide range, the habitat has broad environmental ranges affected by properties such as daylength, temperature, and precipitation (Hashiguchi *et al*., 2011). The Japanese archipelago extends across a latitudinal range of approximately 25 degrees and its climate ranges from subtropical to cold-temperate. In addition, three quarters of Japan are covered by mountains, the topography and climatic gradients of which can limit dispersal and form barriers. This can result in strong selection pressures and promote population differentiation driven by local adaptation (Kaweki and Elvert, 2004). Moreover, Japan is surrounded by sea and affected by seasonal winds from Eurasia and the ocean. These factors would create a variety of habitats, and each wild population would be optimised to maintain a certain fitness level, thereby accelerating the intraspecific differentiation in *L. japonicus*. This species is a model plant with a relatively small genome size of ~472 Mb (Handberg and Stougaard, 1992). Wild *L. japonicus* accessions from natural populations over the entire Japanese archipelago have already been collected and are maintained by self-reproduction (Hashiguchi *et al*., 2012). The use of these accessions enables us to reveal the population history that led to the huge distribution of the species, and inquire into the mechanisms of local adaptation to various environments.

A recent development of genome sequencing technology, the use of genome-wide nucleotide polymorphisms, has improved population genetics. Genome-wide nucleotide polymorphisms are available for hundreds of accessions, and the data offer a large amount of information. With high-density polymorphism information, it is possible to reveal the dynamics of the effective population size and the divergence process from past to present population structures (Li and Durbin, 2011; Durvasula *et al*., 2017; Fulgione *et al*., 2018). In this study, population clustering analysis and divergence time estimations were conducted with whole-genome nucleotide polymorphisms by using developed materials from

wild accessions of Japanese *L. japonicus* populations. This work enables us to explain the process of spreading over and colonising various environments in Japan.

Understanding the mechanisms of local adaptation is an important theme for evolutionary biology. Genome sequencing technology has also enabled us to use genome-wide association studies (GWAS) to identify the genes responsible for contributing to local adaptation (Atwell *et al*., 2010; Fournier-Level *et al*., 2011; Shah *et al*., 2020; Yu *et al*., 2016). Revealing a population's demographic characteristics helps us understand the local adaptation process and could contribute to improving GWA analyses with nucleotide polymorphisms maintained among wild populations. Population structure is recognised as having a major confounding effect in this kind of analysis (e.g., Korte and Farlow, 2013), thus demonstrating that it would help in the selection of statistical models with population structure correction.

Here, I identified subpopulations and estimated the accession-pairwise divergence time in Japanese *L. japonicus* accessions. The results show the demographic history of *L. japonicus* and its colonisation history in Japan.

Materials and Methods

Population structure analysis

To infer the population structure, a Bayesian model-based clustering analysis was conducted with fastSTRUCTURE version 1.0 (Raj *et al*., 2014) for 201,694 non-repetitive SNP markers with minor allele count > 5 in 136 wild accessions (Table 1-1). The SNP markers were determined by Shah *et al*. (2020) based on whole-genome re-sequencing data that used *Lotus japonicus* genome version 3.0 (http://www.kazusa.or.jp/lotus/). The analysis was run with a number of clusters (*K*) ranging from 1 to 8 with default parameter settings. QGIS 2.18.2 was used to visualise the results on a map of Japan with coastline data released by the Japan Meteorological Agency. A line plot of marginal likelihood for each number of clusters was drawn using R package 3.1.2 (R Core Team, 2015).

PSMC analysis

For the estimation of the population size history, the pairwise sequentially Markovian coalescent (PSMC) model was used (Li and Durbin, 2011). For this analysis, I used Illumina paired-end reads from the accessions (ENA accession PRJEB27969). The analysis was conducted for pseudodiploid sequences of each pair of 65 accessions. Accessions with more than 10× coverage were subsampled to 10× coverage. Diploid consensus sequences were created with SAMtools v1.3 and BCFtools v1.3. SNPs with mapping quality > 50 and read depths between 5 and 100 were included in the analysis. Using the recommended settings (Li and Durbin, 2011), the PSMC analysis was run with each pseudodiploid consensus sequence. The mutation rate per nucleotide in *A. thaliana* (6.5×10^{-9} year/site) (Ossowski *et al*., 2010) was used in the visualisation step.

Results

The results of the clustering analysis showed geographic structures of subpopulations in Japanese *L. japonicus* accessions.

The fastSTRUCTURE method groups accessions according to their allele frequencies (Raj *et al*., 2014). According to Raj *et al*. (2014), it is likely that the analysis will lead to better inference when the marginal likelihood is increasing. Fig. 1B shows the marginal likelihood for each number of clusters. The likelihood increased until the likelihood for six clusters was reached, indicating that the clustering analysis worked well when the accessions were grouped into 2–6 subpopulations. Considering the composition and geographical tendency of each group, a large fraction of the variation in Japanese *L. japonicus* accessions could be attributed to three subpopulations with a geographical structure (Fig. 1-1 and 1-2). The subpopulations pop1, pop2, and pop3, were located in the southern, middle, and northern parts of the Japanese archipelago, respectively, (Fig. 1-1). Accessions of pop1 originated in areas from the southernmost part of Kyushu to Miyako Island, the southern boundary of distribution in Japan. The area covered by pop2 extended from the Kinki area to the northeast Kyushu area. Finally, pop3 was distributed from the northernmost part of Japan to the Chubu area. The genotypic compositions of accessions around borders between these subpopulations showed a tendency towards several mixed components. Several accessions in the border between pop2 and pop3 had a genetic component in which elements of pop2 and pop3 commanded a majority. In the Kyushu area encompassing the border between pop1 and pop2, although most accessions with mixed components had genetic elements of both pop1 and pop2, several, especially in the central Kyushu area, showed components in which the pop3 element accounted for the majority (Fig. 1).

Table 1-1 Sampling points of *Lotus japonicus* wild accessions and genetic component rates of three subpopulations based on the fastSTRUCTURE analysis.

Accession	Latitude	Longitude	Trivial name	Prefecture	pop1	pop2	pop3
Gifu*	35.72	137.27	Gifu	Gifu	0.00	0.36	0.64
MG001*	32.72	133.02	Ashizuri	Kochi	0.00	0.96	0.03
MG003*	35.40	139.52	Totsuka	Kanagawa	0.12	0.88	0.00
MG004	31.44	130.31	Kaseda	Kagoshima	0.89	0.00	0.11
MG005*	35.30	139.27	Ninomiya	Kanagawa	0.01	0.10	0.89
MG007*	39.32	141.34	Towa	Iwate	0.00	0.00	1.00
MG008	35.02	135.56	Kameoka	Kyoto	0.00	1.00	0.00
MG009*	39.93	139.74	Oga	Akita	0.00	0.22	0.78
MG010*	39.32	141.48	Tono	Iwate	0.00	0.00	1.00
MG011	39.33	140.81	Yuda	Iwate	0.00	0.00	1.00
MG012	35.26	139.60	Hayama	Kanagawa	0.26	0.05	0.69
MG013*	35.21	139.61	Arasaki-1	Kanagawa	0.04	0.14	0.82
MG014	35.21	139.61	Arasaki-2	Kanagawa	0.05	0.18	0.77
MG016	35.14	139.68	Bishamon	Kanagawa	0.06	0.19	0.76
MG017*	35.13	138.96	Mishima-1	Shizuoka	0.00	0.00	1.00
MG018*	35.13	138.96	Mishima-2	Shizuoka	0.00	0.00	1.00
MG019*	35.13	138.96	Mishima-3	Shizuoka	0.00	0.00	1.00
MG020	24.72	125.47	Miyakojima-1	Okinawa	0.76	0.23	0.01
MG021	41.46	140.09	Matsumae	Hokkaido	0.01	0.13	0.86
MG022*	31.19	130.56	Kaimon	Kagoshima	1.00	0.00	0.00
MG023*	40.82	140.76	Aomori	Kagoshima	0.00	0.00	1.00
MG024*	24.72	125.47	Miyakojima-2	Okinawa	1.00	0.00	0.00
MG025	30.33	130.65	Ambo	Kagoshima	1.00	0.00	0.00
MG026	35.70	139.76	Kokyo	Tokyo	0.02	0.04	0.93
MG027*	26.40	127.74	Yomitan	Okinawa	1.00	0.00	0.00
MG028*	31.16	130.59	Nagasakibana	Kagoshima	1.00	0.00	0.00
MG030	42.85	143.17	Obihiro	Hokkaido	0.00	0.00	1.00
MG032	42.55	143.48	Taiki-1	Hokkaido	0.00	0.00	1.00
MG033	42.55	143.48	Taiki-2	Hokkaido	0.00	0.00	1.00
MG034*	42.61	143.55	Toyokoro	Hokkaido	0.00	0.00	1.00
MG035*	42.45	141.16	Noboribetsu	Hokkaido	0.00	0.00	1.00
MG036*	42.41	140.31	Oshamambe-1	Hokkaido	0.00	0.13	0.87
MG038*	42.11	140.57	Mori	Hokkaido	0.00	0.00	1.00
MG039*	40.93	141.02	Hiranai	Aomori	0.00	0.00	1.00
MG040*	40.59	141.25	Towada	Aomori	0.00	0.00	1.00
MG041	39.76	140.16	Asahigawa	Akita	0.00	0.09	0.91
MG042*	39.75	140.71	Tazawako	Akita	0.00	0.00	1.00
MG044	38.39	141.39	Ishinomaki	Miyagi	0.00	0.00	1.00
MG045*	37.59	139.92	Shiokawa-1	Fukushima	0.00	0.00	1.00
MG046	37.59	139.91	Shiokawa-2	Fukushima	0.00	0.00	1.00
MG049*	35.64	134.62	Kasumi	Hyogo	0.00	0.99	0.01
MG050	35.50	133.73	Tohaku	Tottori	0.08	0.88	0.04
MG051	35.50	133.85	Hojyo	Tottori	0.01	0.42	0.57
MG052	35.39	132.79	Hikawa	Tottori	0.00	1.00	0.00
MG053*	34.65	131.83	Masuda	Tottori	0.20	0.74	0.06
MG056*	31.58	131.42	Nichinan-1	Miyazaki	0.07	0.93	0.00
MG057	31.70	131.46	Nichinan-2	Miyazaki	0.06	0.94	0.00
MG058	31.70	131.46	Nichinan-3	Miyazaki	0.50	0.50	0.00
MG059	31.70	131.46	Nichinan-4	Miyazaki	1.00	0.00	0.00
MG060	31.26	130.29	Makurazaki	Kagoshima	1.00	0.00	0.00
MG061	31.37	130.99	Kushira	Kagoshima	0.64	0.36	0.00
MG062	NA	NA	NA	NA	0.12	0.88	0.00
MG063*	31.74	131.47	Noshima	Miyazaki	0.14	0.86	0.00
MG065	31.83	131.42	Miyadai	Miyazaki	0.00	1.00	0.00
MG066*	32.13	131.41	Saito	Miyazaki	0.00	1.00	0.00
MG067	30.77	131.01	Tanegashima	Kagoshima	1.00	0.00	0.00
MG068*	30.23	130.56	Yakushima	Kagoshima	1.00	0.00	0.00
MG069*	34.59	135.51	Asakayama	Osaka	0.01	0.94	0.06
MG070	33.87	135.23	Inami	Wakayama	0.12	0.83	0.05
MG071	32.95	131.10	Aso	Kumamoto	0.42	0.51	0.06
MG072	32.85	131.05	Hakusui	Kumamoto	0.09	0.32	0.60
MG073*	33.04	131.27	Kujyu	Ohita	0.41	0.51	0.08
MG074*	33.39	132.54	Uwa-1	Ehime	0.00	1.00	0.00
MG075	33.34	132.54	Uwa-2	Ehime	0.00	1.00	0.00
MG076*	34.26	131.59	Yamaguchi	Yamaguchi	0.06	0.28	0.66
MG077	35.12	138.66	Fuji	Shizuoka	0.09	0.33	0.58
MG078*	35.90	136.72	Kuzuryuko	Fukui	0.00	0.00	1.00
MG080*	36.65	137.19	Toyama	Toyama	0.00	0.06	0.94

Table 1-1. Continued.

MG081*	37.04	137.85	Omi	Niigata	0.00	0.12	0.88
MG082*	37.76	139.21	Gosen	Niigata	0.00	0.00	1.00
MG083	36.54	138.36	Sugahira	Nagano	0.00	0.00	1.00
MG084*	37.60	139.65	Nishiaizu	Fukushima	0.00	0.00	1.00
MG085*	37.15	140.41	Ishikawa	Fukushima	0.00	0.00	1.00
MG086*	38.76	140.76	Naruko	Miyagi	0.00	0.00	1.00
MG088	36.65	139.98	Ujiie	Tochigi	0.00	0.00	1.00
MG089*	39.76	141.14	Morioka-1	Iwate	0.00	0.00	1.00
MG090	43.00	141.43	Sapporo	Hokkaido	0.00	0.00	1.00
MG091	41.90	140.64	Ono	Hokkaido	0.00	0.00	1.00
MG092	44.15	143.94	Tokoro	Hokkaido	0.00	0.00	1.00
MG093	42.03	140.67	Komagatake	Hokkaido	0.00	0.00	1.00
MG094	42.42	140.90	Date	Hokkaido	0.00	0.00	1.00
MG095	42.48	141.22	Shiraoi	Hokkaido	0.00	0.00	1.00
MG096*	42.60	141.50	Tomakomai	Hokkaido	0.00	0.00	1.00
MG097	42.45	140.33	Oshamambe-2	Hokkaido	0.00	0.00	1.00
MG098	41.92	140.73	Nanae	Hokkaido	0.00	0.00	1.00
MG099	41.76	140.72	Hakodate	Hokkaido	0.00	0.00	1.00
MG100	42.22	140.39	Yakumo	Hokkaido	0.00	0.00	1.00
MG101*	42.00	140.65	Mori-2	Hokkaido	0.00	0.00	1.00
MG102	45.30	141.03	Rebunto	Hokkaido	0.00	0.00	1.00
MG103	42.99	143.20	Otofuke	Hokkaido	0.00	0.00	1.00
MG104	32.08	131.41	Saito-2	Miyazaki	0.00	1.00	0.00
MG105	32.12	131.42	Saito-3	Miyazaki	0.00	1.00	0.00
MG106*	32.09	131.38	Saito-4	Miyazaki	0.00	1.00	0.00
MG107*	34.82	135.52	Suita	Osaka	0.00	0.97	0.03
MG109	39.71	141.11	Morioka	Iwate	0.00	0.02	0.98
MG110*	32.83	131.13	Takamori-1	Kumamoto	0.41	0.54	0.04
MG111*	32.86	131.00	Choyo	Kumamoto	0.43	0.56	0.02
MG112*	32.93	131.23	Namino	Kumamoto	0.00	0.00	1.00
MG113*	32.86	131.16	Takamori-2	Kumamoto	0.01	0.09	0.89
MG115	NA	NA	Ibigawa	Gifu	0.00	0.84	0.16
MG116	35.90	136.65	kuzuryu	Fukui	0.10	0.30	0.60
MG117	36.06	136.50	Katsuyama	Fukui	0.12	0.30	0.58
MG118	35.97	136.13	Asahi	Fukui	0.00	0.00	1.00
MG119	NA	NA	Sninanogawa	Niigata	0.00	0.01	0.99
MG120*	36.99	138.58	Sakae	Nagano	0.00	0.00	1.00
MG121*	36.85	138.37	Iiyama	Nagano	0.00	0.04	0.96
MG123	36.00	138.16	Chino	Nagano	0.00	0.01	0.99
MG124*	37.56	139.82	Aizubange	Fukushima	0.00	0.00	1.00
MG125*	37.51	139.92	Aizuwakamatsu	Fukushima	0.00	0.00	1.00
MG126*	37.13	140.21	Shirakawa	Fukushima	0.00	0.00	1.00
MG127	38.53	141.22	Eaigawa	Miyagi	0.00	0.00	1.00
MG128*	36.78	140.13	Houkigawa	Tochigi	0.00	0.00	1.00
MG129*	39.55	140.37	Nishisenboku	Akita	0.00	0.00	1.00
MG130*	35.42	136.41	Ibuki	Shiga	0.05	0.66	0.29
MG131	NA	NA	Ymatogawa	Osaka	0.00	1.00	0.00
MG132	28.18	129.37	Amami	Kagoshima	1.00	0.00	0.00
MG133	31.61	131.43	Kazeda	Miyazaki	0.96	0.04	0.00
MG134	31.50	131.38	Meotoura	Miyazaki	0.00	1.00	0.00
MG135	31.47	131.37	Fuji	Miyazaki	0.87	0.13	0.00
MG136	31.46	131.36	Kamiishinami	Miyazaki	0.79	0.21	0.00
MG138*	31.43	131.35	Nadani	Miyazaki	0.00	1.00	0.00
MG139	31.82	131.42	Nakajima	Miyazaki	0.00	1.00	0.00
MG140*	31.85	131.40	Kihara	Miyazaki	0.00	1.00	0.00
MG141*	31.95	131.31	Gecchibai	Miyazaki	0.00	1.00	0.00
MG142	31.64	131.46	Udo-1	Miyazaki	0.30	0.70	0.00
MG143	31.65	131.47	Udo-2	Miyazaki	0.30	0.70	0.00
MG144*	31.64	131.47	Udo-3	Miyazaki	0.30	0.70	0.00
MG145	31.63	131.46	Oura	Miyazaki	0.71	0.29	0.00
MG146*	33.55	129.93	Yobuko	Saga	0.43	0.54	0.03
MG149	34.10	129.17	Tsutsusaki-1	Nagasaki	0.64	0.28	0.08
MG150	34.11	129.17	Tsutsusaki-2	Nagasaki	0.64	0.30	0.06
MG151	34.11	129.21	Azamo-1	Nagasaki	0.79	0.21	0.00
MG152*	34.11	129.21	Azamo-2	Nagasaki	0.65	0.29	0.07
MG154*	34.66	129.49	Tonosaki	Nagasaki	0.58	0.33	0.09
MG155	34.68	129.49	Shitazaki	Nagasaki	0.63	0.31	0.05
MG156	34.68	129.47	Izumi	Nagasaki	0.59	0.35	0.06

*Accessions used for PSMC analysis

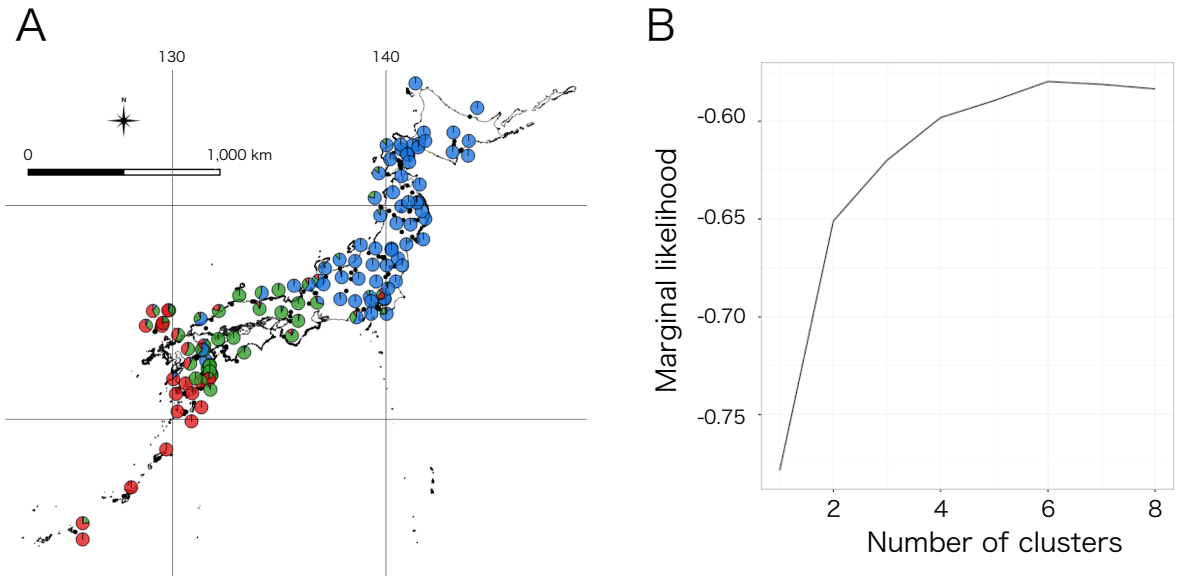

Fig. 1-1. Results of the fastSTRUCTURE analysis

A: Pie charts of clustering elements with locality information for each accession (*K*=3). Each colour indicates the ratio of the clustering elements. **B**: Marginal likelihood for each number of population clusters. The vertical and horizontal axes represent the marginal likelihood and number of clusters, respectively.

Effective population size and population differentiation

To examine the migration history of this species, I carried out a PSMC analysis (Li and Durbin, 2011). In this analysis, population size history was inferred from a diploid sequence based on the extent of heterozygous regions in the genome sequence. *Lotus japonicus* is a self-reproducing plant and most accessions have low heterozygosity levels (Shah *et al.*, 2020), therefore the individual accessions were not suitable for PSMC analysis. I generated pseudo-diploids by merging read alignments from pairs of selected accessions with at least 7× read coverage and identified a pseudo-diploid consensus sequence based on merged alignments. These pseudo-diploids were used for PSMC analysis instead of individual accessions. Inference with pseudo-diploids can show the time of the last contact of two accessions as the time of the last coalescence events, shown by an increase in the estimated effective population size. This feature was adopted to make rough estimates of the divergence times for pseudo-diploid pairs, assuming a mutation rate of 6.5×10^{-9} per year (Ossowski *et al.*, 2010). The estimated divergence time for all accession pairs showed the same clustering tendency as the fastSTRUCTURE analysis did (Fig. 3). Particularly for pop1, pop3, and the Southeast Kyushu area, the estimated divergence time was low among each cluster and showed tight clustering. The results for accession pairs from pop1, pop2, and pop3 showed that the accessions from southern and northern Japan last came in contact 10–18 thousand years ago, a result consistent with the colonisation after the last ice age theory. The accessions from Tsushima (MG152, MG154) seem to have diverged more recently than all the other accessions (Fig. 3).

Discussion

Here, I inferred the population structure and divergence time of Japanese *L. japonicus* accessions using genome-wide nucleotide polymorphisms. The results of fastSTRUCTURE revealed that Japanese *L. japonicus* accessions involved 2–6 clusters. In the case of cluster numbers 2–3, the accessions were geographically unified as one cluster, while the accessions in several clusters, grouped into 4–6 subpopulations, were scattered across several areas (Fig. 1A and 2). Considering the composition of each group in each accession and geographical unit, the accessions could be roughly divided into three subpopulations ranging in the southern, middle, and northern parts of the Japanese archipelago, respectively (Fig. 1A and 2). According to the results for cluster numbers 4–6, more small clusters are likely to exist within pop2 and pop3 (Fig. 2B). These geographical units indicate that the climatic differences across Japan could be strong barriers to colonisation. On the other hand, some accessions harboured multiple group compositions on the borders between the distributions of subpopulations (Fig. 1), particularly in the Kyushu area, where all the

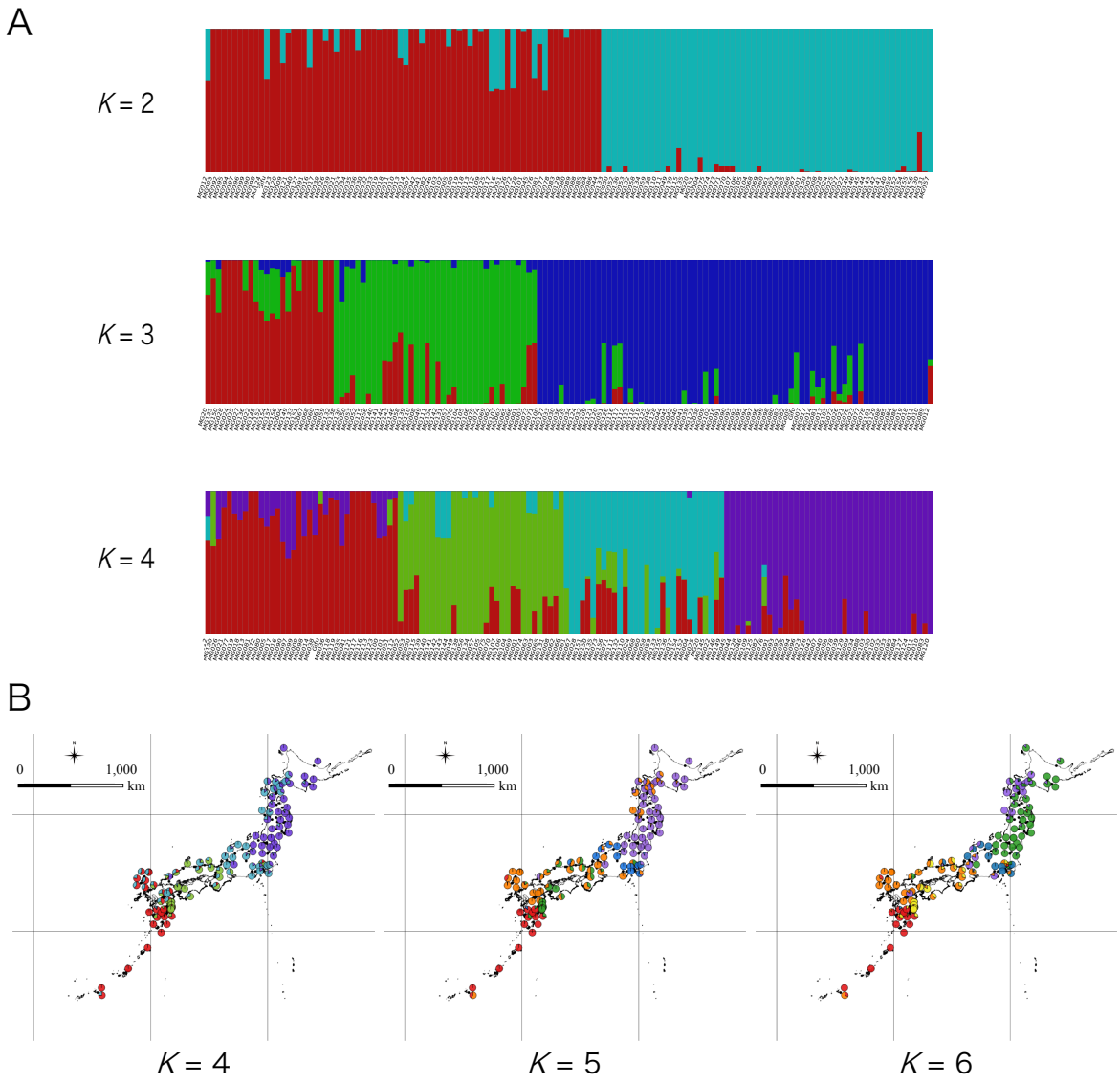

Fig. 1-2. Results of the fastSTRUCTURE analyses

Each colour indicates the ratio of clustering elements for each accession. **A**: Component bar charts of clustering elements for each accession with cluster numbers 2–4. Accession IDs are shown under bar charts. **B**: Pie charts of clustering elements with locality information of each accession for cluster numbers 4–6.

compositions were involved, implying that the area has high genetic diversity.

The pairwise divergence time estimation resulting from the PSMC analysis revealed several clusters that were aligned with the topography; the major clusters overlapped with the subpopulations detected by fastSTRUCTURE analysis (Fig. 3). This result reflects the robust clustering of the accessions. While some accession groups had belonged to the same subpopulation, pop2 (Fig. 3). This result also implies that more small clusters would exist within this group as suggested by the fastSTRUCTURE analysis results (K=4–6; Fig. 2B). In addition, two accessions that originated in Tsushima Island appear to have diverged recently from other accessions (Fig. 3).

The tight clustering of northern accessions in pop3 that resulted from both the fastSTRUCTURE and PSMC analyses, indicated that accessions in this subpopulation could have lower genetic diversity. In fact, pop1 and pop3 had the highest and lowest genetic diversity (Pi) in the three subpopulations, respectively, as the pairwise genetic distance increases along with the given geographic distance (Shah *et al*., 2020).

Therefore, it could be suggested that this species spread from South to North Japan, losing genetic

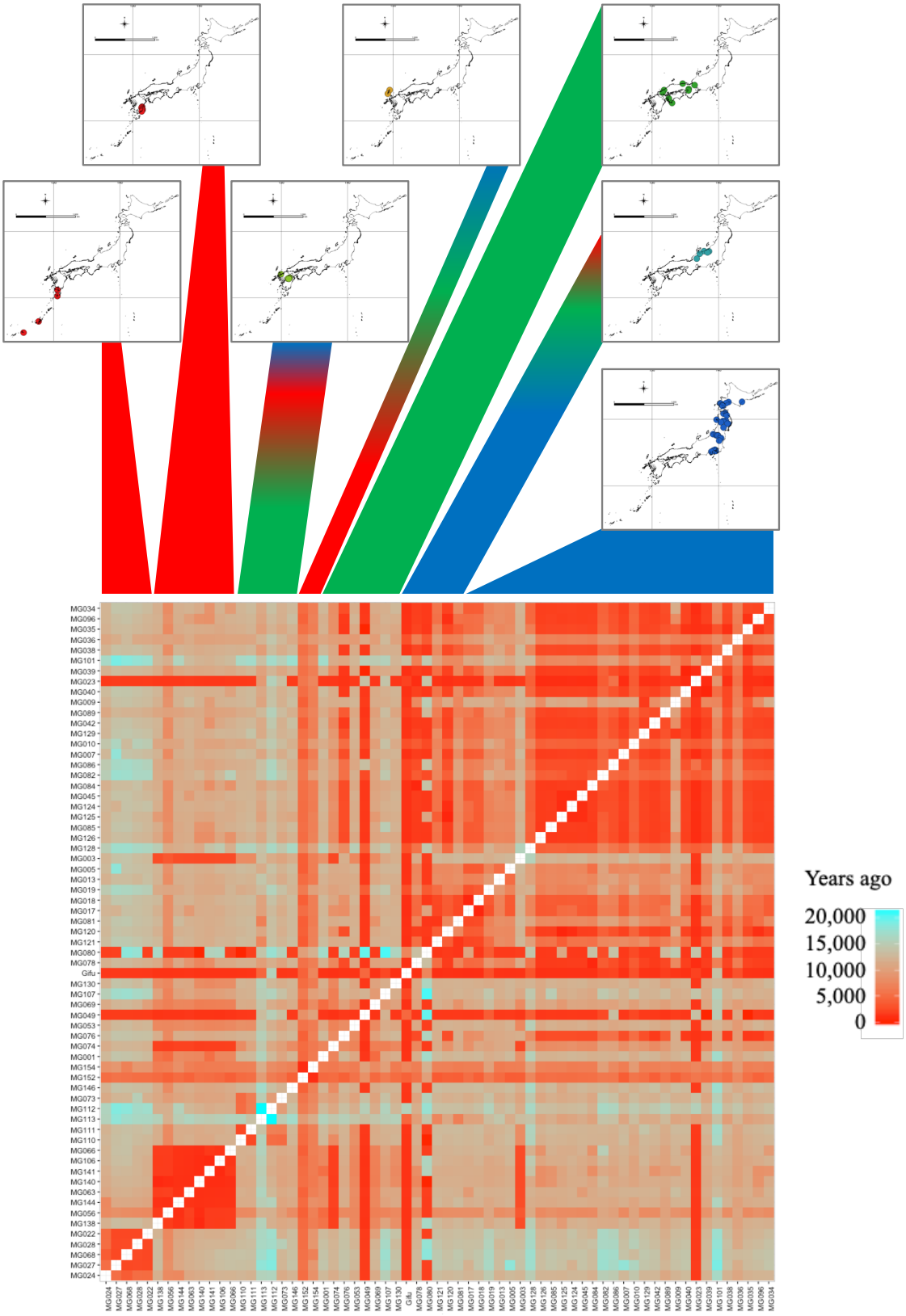

Fig. 1-3. Result of the pairwise sequentially Markovian coalescent (PSMC) analysis

The heatmap shows the esti mated divergence time for 65 accession pairs. Red and light blue indicate recent and past divergence times, respectively. Accession IDs are shown beside and under the heat map. Each plot on the map of Japan indicates the locality of the accessions.

strong ties, several accessions in the Kinki, Shikoku, and Chugoku areas were not united, although they diversity. In addition, the high genetic diversity of accessions in the Kyushu area implies the expansion of *L. japonicus* from this area. Moreover, in contrast to the results of the genetic components and the estimated divergence time, the results of the genetic distance analysis for the same 65 accession pairs showed that Tsushima accessions had wide genetic distances from all other accessions, with no strong association to any other population cluster (Shah *et al*., 2020). These results suggest that Tsushima Island may be a point of origin for all Japanese *L. japonicus* accessions. Considering my results and the genetic diversity shown by Shah *et al*. (2020), *L. japonicus* reached southern Japan before reaching the northern part, with long-distance gene flow post colonisation, and it is considered that this species colonised Japan from Tsushima Island after the last ice age.

Knowledge of the demographic history of a species is necessary for understanding the mechanisms of its local adaptation and intraspecific differentiation. This is especially true for regions surrounded by sea or having a wide range of environmental gradients, such as the Japanese archipelago, where local adaptation and intraspecific differentiation would be accelerated and revealing a species' history could be more important for evolutionary biology. *Lotus japonicus* is a suitable model for providing a pattern of this kind of research, as this species has a wide distribution in the Japanese archipelago and may adapt to each habitat. In order to understand the mechanisms of local adaptation and intraspecific differentiation from the perspective of the responsible genes, it is important to exploit the population history of a species as well as reveal its demographic history. Although association analysis (e.g., genome-wide association analysis) with nucleotide polymorphisms maintained in wild populations has high statistical power and could detect genetic factor regions directly, the existence of population clusters would complicate this kind of association analysis (Korte and Farlow, 2013). Therefore, in order to use appropriate statistical models for analysis and avoid detecting pseudo-positive candidates, it is necessary to confirm the existence and tendency of population clusters before this kind of association analysis is conducted. The migration and colonisation history of this species would contribute not only to an understanding of the biodiversity in the Japanese archipelago, but also to identifying the candidate genes for adaptive traits.

Chapter 2

Polymorphisms of *E1* and *GIGANTEA* in wild populations of *Lotus japonicus*

Abstract

In plants, timing of flowering is an essential factor that controls the survival rates of descendants. The circadian clock genes *E1* and *GIGANTEA* (*GI*) play a central role in transmitting signals to *Flowering Locus T* (*FT*) in leguminous plants. *Lotus japonicus* is a wild Japanese species that ranges from northern Hokkaido to the southern Ryukyus and exhibits a wide range in terms of the time between seeding and first flowering. In this study, I first identified *LjGI* and analysed polymorphisms of *LjE1* and *LjGI* among wild populations covering the entire distribution range of this species in Japan. *LjGI* had a coding sequence (CDS) length of 3,528 bp and included 14 exons. The homologies of DNA and amino acid sequences between *LjGI* and *GmGI* were 89% and 88% (positive rate was 92%), respectively. *LjE1* harbored five nucleic acid changes in a 552-bp CDS, all of which were nonsynonymous; four of the changes were located in the core function area. *LjE1* alleles exhibited partial north–south differentiation and non-neutrality. In contrast, the *LjGI* harbored one synonymous and one nonsynonymous change. Thus, my study suggests that *LjE1* may be involved in the control of flowering times, whereas *LjGI* may be under strong purifying selection.

Introduction

In the plant life cycle, the shift from growth to the reproductive stage is critical in ensuring successful seed production. Flowering time has an essential influence on the survival rate of descendants and the persistence of the plant species. Since Garner and Allard's (1920) discovery of the photoperiodic response in soybean and cotton, the flowering response of a number of plant taxa have been shown to be sensitive to photoperiod. Various factors are involved in the mechanisms that control flowering time, including hormones and genetic regulation (internal factors), and photoperiod, light quality, and temperature (external factors). Plants regulate flowering time by receiving environmental cues and by expressing functional genes that control phenotype. Genotypic variation, including single nucleotide polymorphisms (SNPs), can also be involved in functional gene expression (e.g., Levy and Dean 1998).

Over 80 genes involved in the regulation of flowering time have been identified in the model plant *Arabidopsis thaliana*, and these form a large genetic network (Mouradov et al. 2002; Komeda 2004; Bäurle and Dean 2006). The photoperiodic network consists of three major pathways: light-sensing and circadian clock pathways, and a developmental pathway that causes flowering at shoot apical meristems (Mouradov et al. 2002; Komeda 2004; Bäurle and Dean 2006). In the circadian clock pathway in *A. thaliana*, *CONSTANS* (*CO*), *GIGANTEA* (*GI*), *PRRfamily*, *LHY*, and *TOC1* have been identified as major genes (Fankhauser and Chory 1997; Whitelam and Devlin 1997; Fowler et al. 1999; Alabadí et al. 2001; Suárez-López et al. 2001; Nakamichi et al. 2005). These genes are involved in regulating flowering in response to environmental changes. In addition, previous studies have suggested that circadian clock genes affect both the flowering pathway and plant growth including hypocotyl elongation (Farré 2012). The circadian clock network comprises three loops (Locke 2006) that, when activated, transfer signals to *GI* and *CO* as end-core circadian clock genes. Thus, *GI* and *CO* are considered to be among the most important genes in the circadian clock pathway in relation to flowering-time regulation (Fowler et al. 1999; Samach et al. 2000).

The photoperiodic flowering response in soybean (*Glycine max* [L.] Merr.) is among the most sensitive of that of any plant species (Garner and Allard 1920). Because of its agricultural importance, the genome of *G. max* has been investigated extensively as a model leguminous plant, as have those of *Medicago truncatula* Gaertn. and *Lotus japonicus* (Regel) Larsen (e.g., Sato et al. 2001; Schmutz et al. 2010; Young et al. 2011). Among these legumes, *G. max* has been investigated extensively in efforts to identify the genes responsible for flowering-time regulation. Regulation of flowering and maturity in soybean is associated with the *E* loci. Classic genetics

techniques have shown that *E1* and *E2* (Bernard 1971), *E3* (Buzzell 1971), *E4* (Buzzell and Voldeng 1980), *E5* (McBlain and Bernard 1987), *E6* (Bonato and Vello 1999), *E7* (Cober and Voldeng 2001), and *E8* (Cober et al. 2010) are involved in flowering-time regulation. Analyses of association suggested that four of the eight *E* loci (*E1–E4*) account for 62–66% of the flowering-time response (Tsubokura et al. 2014). A recent study showed that *G. max* lacks *CO* but that *E1* of *G. max* exhibits a function similar to that of *CO* (Xia et al. 2012). In addition, *E2*, *E3*, and *E4* were identified as *GI* (*GmGI*) (Watanabe et al. 2011), *PHYA3* (*GmPHYA3*) (Watanabe et al. 2009), and *PHYA2* (*GmPHYA2*) (Liu et al. 2008), respectively. Soybean cultivars exhibited several genotypes as polymorphisms in these four genes, and each genotype was associated with either the early- or late-flowering phenotype (Xia et al. 2012; Tsubokura et al. 2014).

E1 is a relatively small protein (174 aa); ~67% of *E1* comprises the B3 domain, which may bind to DNA and repress expression of *GmFTs* in *G. max* (Xia et al. 2012). The B3 domain is a DNA-binding motif that is specific to plants and several restriction enzymes (Yamasaki et al. 2004; Zhou et al. 2004). *GmGI*, responsible for the *E2* locus, harbors polymorphisms across Japanese cultivars that control the early- or late-flowering phenotype (Watanabe et al. 2011).

Lotus japonicus, a model leguminous plant, is distributed across a wide range of latitudes in Japan, from ~24ºN (Miyakojima) to 45ºN (Hokkaido). Previous studies have shown that the flowering time of wild *L. japonicus* correlates significantly with latitude; plants from southern areas tend to have shorter flowering periods (Kawaguchi 2000; Kai et al. 2010). These observations suggest that a natural variation in genes regulating flowering time is involved in local adaptation. Although the presence of *E1* and the lack of *CO* have been reported in *L. japonicus* (Xia et al. 2012), *GI* has not yet been identified (Ono et al. 2010). Because the entire genome sequence is available in public databases (e.g., Sato et al. 2008), comparative genomic analysis of *LjE1* and *LjGI* as representative circadian clock genes can be performed.

In this study, I aimed to address whether polymorphisms of genes thought to be involved in flowering time had genetic structures that were consistent with the differences in flowering time in *L. japonicas* according to latitude, and whether these genes were affected by natural selection. I identified the location and genomic structure of *GI* in the *L. japonicus* genome. I then sequenced *LjE1* and *LjGI* from wild accessions of *L. japonicus* collected from individuals across its distribution range in Japan. Based on these sequences, I clarified the evolutionary history and spatial patterns of their polymorphisms.

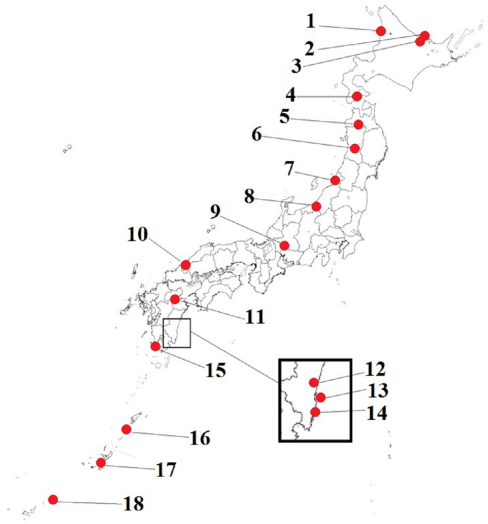

Fig. 2-1. Collection sites of *Lotus japonicus* used in this study

Table 2-2 The geographic information of *Lotus japonicus* used in this study.

No.	Prefecture	Locality/ Accession No.	Latitude/Longitude	No. of samples	
				E1	*GI*
1	Hokkaido	Tomamae	44°28'N/141°69'E	4	4
2	Hokkaido	Wakka	44°14'N/143°95'E	3	
3	Hokkaido	Kimuneappu	44°10'N/143°90'E	4	
4	Hokkaido	Hakodate	41°45'N/140°36'E	4	
5	Aomori	Hakkouda	40°38'N/140°51'E	4	4
6	Akita	MG-42*	39°44'N/140°42'E	1	
7	Niigata	Kitsunezaki	38°27'N/139°29'E	4	4
8	Niigata	Kakuda	37°00'N/138°49'E	4	
9	Gifu	Gifu B-129*	35°43'N/136°76'E	1	1
10	Shimane	MG-53	34°39'N/131°49'E	1	
11	Oita	MG-73	33°02'N/131°16'E	1	
12	Miyazaki	MG-63	31°44'N/131°28'E	1	
13	Miyazaki	MG-66	30°44'N/131°16'E	1	
14	Miyazaki	Udo	31°65'N/131°46'E	3	
15	Kagoshima	MG-22*	31°11'N/130°33'E	1	1
16	Kagoshima	Tokunoshima	27°75'N/129°02'E	4	4
17	Okinawa	Nishihara	26°23'N/127°76'E	4	4
18	Okinawa	Miyakojima	24°44'N/125°28'E	4	1

Materials and methods

Sampling and DNA extraction

I analysed 49 samples from 18 *Lotus japonicus*

accessions for *LjE1* and 23 samples from 8 *L. japonicus* accessions for *LjGI*. Seeds were collected from the field or were provided by the National Bioresource Project (Legume Base), Japan. Samples covered the species' distribution range within Japan, ranging from Tomamae, Hokkaido (44º28'N/141º69'E) to Miyakojima, Okinawa (24º44'N, 125º28'E; Fig. 2-1). Sample details are summarized in Table 2-1. Because of the length of *LjGI* (~3534 bp of CDS in *Glycine max*), I restricted the number of samples as necessary to adequately cover the species distribution range in Japan. The seed coats were partially scarified with sandpaper; seeds were sown on 1% agar and planted in pots after germination.

For *LjE1* gene sequencing, mature leaves were subjected to DNA extraction. Leaves were frozen in liquid nitrogen and pulverized with a mortar and pestle. After removal of polysaccharides from powdered leaves using HEPES buffer (pH 8.0; Setoguchi and Ohba 1995), DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle 1990). The extracted DNA was dissolved in 100-µL Tris–EDTA buffer and used for subsequent amplification by polymerase chain reaction (PCR), as described below.

RNA was extracted from mature leaves for *LjGI* sequencing. Because gene expression was highest in ZT 8–12 (Flower et al. 1999), the fresh leaves were sampled accordingly. The samples were frozen in liquid nitrogen and RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol, and dissolved in 50-mL RNase-free water. I prepared cDNA from the RNA using QuantiTect Reverse Transcription (Qiagen) according to the manufacturer's protocol. Before reverse transcription, genomic DNA (gDNA) wipeout was conducted in a 14.0-µL reaction volume containing 11.0-µL autoclaved ion-exchanged water, 2.0-µL gDNA wipeout buffer, and 1.0-µL template RNA.

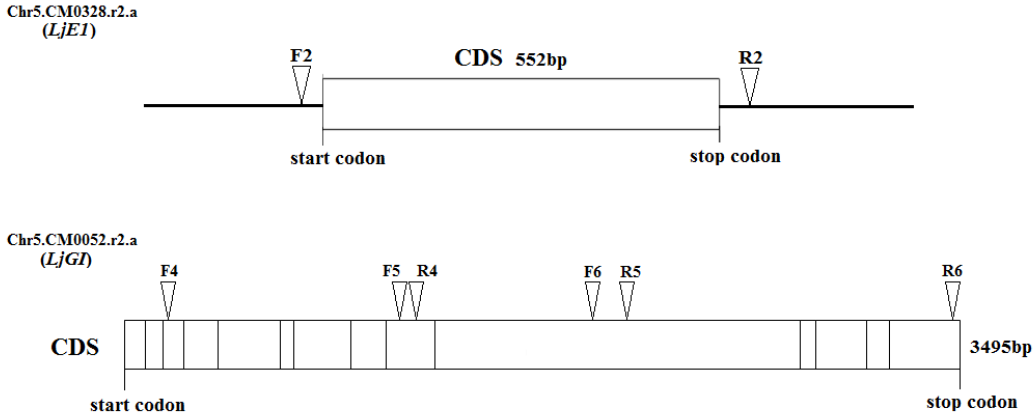

Fig. 2-2. Binding sites of PCR primers used to amplify the *LjE1* and *LjGI* genes of *Lotus japonicus*

Closed and open triangles indicate forward and reverse primers, respectively. The number of primers is noted adjacent to the triangles.

Table 2-1 PCR primer sequences used for *E1* and *GIGANTEA* genes in *Lotus japonicus*.

Gene	Type of primer	Primer name	Abbreviation	primer sequence (5'-3')
E1	Forward	LjE1-F2	F2	ACACTACCATAATAGCCTTTGATCT
		LjE1-F3	F3	ACCGTCACAACCCTGCATGTCGATCA
	Reverse	LjE1-R2	R2	ACCGTCACAACCCTGCATGT
		LjE1-R3	R3	TGGCTATGAAAACGATTAAGGAAGCCCA
GIGANTEA	Forward	LjGItr-F4	F4	CCGTAACCGCTATCCAGCCAAAG
		LjGItr-F5	F5	CTTGCGTGCAATAGGGACTGC
		LjGItr-F6	F6	GCCTGTTAGTCTCAACGGGTCTT
		LjGItr-F7	F7	TGCTGCAGGAGAACTCAAGCCTCCGTGA
		LjGItr-F8	F8	ACCTTGCTGTGGCAGAACTTCGTACTATGG
	Reverse	LjGItr-R4	R4	CCGTCAACTTGCCTAAGTGGAGG
		LjGItr-R5	R5	GCATGCAAGCTTTTGACCGTCTG
		LjGItr-R6	R6	GACAACCCATTGGAAAGTCGTGTG
		LjGItr-R7	R7	ACGAGAGCATCAAGTGCATCCGGAGCCCA
		LjGItr-R8	R8	AGAGGGGCTGAATCTTCCAATGGTGCA

The gDNA wipeout was performed by incubation for 5 min at 42 ºC. Reverse transcription was conducted in a 20.0-μL reaction volume containing 14.0-μL template RNA (the same as used for gDNA wipeout), 4.0-μL RT buffer, 1.0-μL RT primer mix, and 1.0-μL Quantiscript reverse transcriptase (Quantiscript Inc., Quebec, Canada) Reverse transcription was performed with an extension for 15 min at 42ºC and inactivation of reverse transcriptase for 3 min at 95ºC. Reverse transcription products were used as described below. After verification that there were no problems using the transcriptome for *LjGI* sequencing, I used gDNA from each sample or the parent plant. All samples used in this study contained homoalleles at the two SNP sites.

PCR and sequencing

PCR primers were designed for amplification and sequencing of the exons of each gene (Fig 2-2). The primer information was based on sequences found in the *Lotus japonicus* Genome Database (http://www.kazusa.or.jp/lotus/). PCR amplification was conducted in a 10.0-μL reaction volume containing 5.8-μL autoclaved ion-exchanged water, 0.8-μL dNTP mixture (2.5 mM each), 2.0-μL 5× PrimeSTAR GXL Buffer (Takara Bio, Kusatsu, Japan), 0.2-μL PrimeSTAR GXL DNA polymerase (Takara Bio), 0.1 μL of each primer (10 pmol), and 1.0-μL template DNA. Amplification was performed with initial denaturation for 2 min at 94ºC followed by 45 cycles of denaturation for 45 s at 94ºC, annealing for 45 s at 50–54ºC, and extension for 1.5 min at 72ºC. PCR products were visualized on 1.0% TAE–agarose gels stained with ethidium bromide and were purified with glass powder using GeneCleanII (Bio 101, Vista, CA, USA). Products were sequenced directly using the standard methods of the BigDye™ Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA, USA) using all listed primers (Table 2-1) and sequenced using an ABI 3130 Genetic Analyzer (Applied Biosystems).

Sequence alignment and analysis

All sequences were aligned using the AutoAssembler software (Applied Biosystems) and the *Lotus japonicus* Genome Database (http://www.kazusa.or.jp/lotus/). No dual peaks were found in any analysed sequence, and polymorphisms of all sequences were determined directly. DnaSP version 5.10 (Librado and Rozas 2009) was used to analyse DNA polymorphisms. Tajima's *D* test (Tajima 1989) and the McDonald–Kreitman (MK) test (McDonald and Kreitman 1991) were

performed to determine statistical neutrality for each sequence of *LjE1* or *LjGI*. To compare the patterns of substitutions of these and other loci, the MK test was applied to two loci (*PHYA*, *CRY1*), whose polymorphisms were identified in a previous study (Oh et al. unpublished). Nucleotide variation was estimated as nucleotide diversity π (Tajima 1989) for each sequence. The protein structure based on *LjE1* nucleic acid sequence data was estimated using Phyre (http://www.ncbi.nlm.nih.gov) and the genomic structure of *G. max* reported in the literature (Xia et al. 2012).

Identification of *GIGANTEA* and *E1* orthologs in *L. japonicus*

I identified candidate genes that had high homology with *GmGI* by NCBI BLAST searching (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The genetic structure of the most promising *LjGI* gene was identified by comparison with *L. japonicus* genomic data from the Kazusa DNA Research Institute (http://www.kazusa.or.jp/lotus/). Genomic information for *LjE1* was referenced in Xia et al. (2012).

Results

Identification of the *GIGANTEA* Gene in *L. japonicus*

I identified *LjGI* at chr5.CM0052.430.r2.a as the *GI* gene in *L. japonicus*, located ~6 cM above *LjE1* (chr5.CM0328.440.r2.d) (Fig. 2-3). The length of *LjGI* was 3,528 bp, the total length of the coding sequence (CDS). The homology of the DNA and amino acid sequences between *LjGI* and *GmGI* was 89% and 88% (positive rate was 92%), respectively. *LjGI* and *GmGI* each included 14 exons (Fig. 2-4).

Detailed structure of *LjE1* protein and *LjE1* polymorphisms

The structure of the *LjE1* protein was compared to that of the *GmE1* protein, and *LjE1* was confirmed to also contain the DNA-binding B3 domain (Fig.

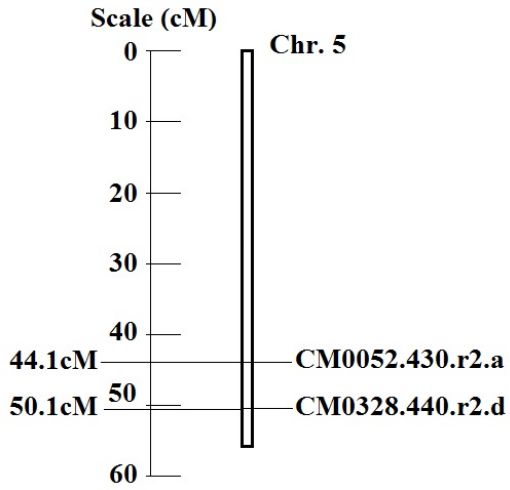

Fig. 2-3. Genomic locations of *LjE1* (CM0328.440.r2.d) and *LjGI* (CM0052.430.r2.a) on chromosome 5
The scale on the left side indicates length in centimorgans (cM).

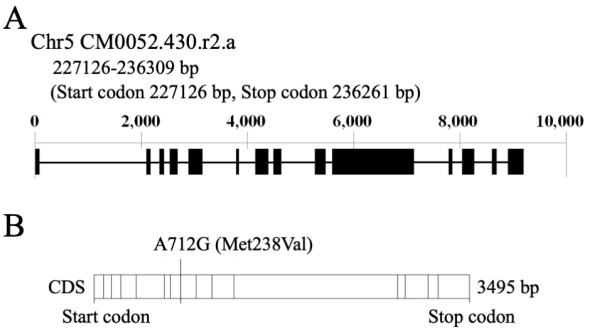

Fig. 2-4. Structure of *GI* of *Lotus japonicus*
A: Closed boxes and thin lines indicate exons and introns, respectively. Scale is shown above as base pairs of nucleic acid sequences. **B**: Structure of *LjGI* coding sequence. Nonsynonymous SNP sites are indicated by vertical lines.

2-5B). This domain of *LjE1* occupied ~66% of the length of *E1* (366 of 552 bp) and comprised 122 amino acids. *E1* in *L. japonicus* consisted of a continuous exon (introns were absent), the same as *GmE1*. I was unable to determine the 5'-UTR of *LjE1*.

Five substitutions were found among 49 samples from 18 wild accessions, and each SNP was specific to a particular geographic region (Table 2-3). Substitution A of *G65A* was most frequent in the northern area (7 of 12 samples from collection sites 1 and 3-5 in Fig. 2-1), substitution A of *A197G* was most frequent in the southern area (16 of 19 samples from collection sites 16-18), substitutions T of *A205T* and G of *A329G* were detected only in southeastern Kyushu, and substitution T of *G488T* was found only in MG-53, Shimane (Table 2-3; Fig. 2-5). All of the SNPs were nonsynonymous site changes.

In *LjE1* and *GmE1*, one of the SNPs (*G65A*) was located in the N-terminal region, while *A197G*, *A205T*, *A329G* and *G480T* were located in the B3 DNA-binding domain. Among the SNPs in the putative DNA-binding motif (expected to contain the B3 DNA-binding domain group), *A329G* and *G480T* were located at a blunt end, while *A197G* and *A205T* were located in a sensitive region in which coded amino acids whose residues with backbone chemical shifts were affected by the DNA binding.

The genetic diversity of *LjE1* from the 49 wild samples was low (π = 0.00175) compared to that of 116 functional genes in *Glycine soja* (π_{total} = 0.00217; Hyten et al. 2006). Tajima's *D* test showed that *LjE1* did not deviate from neutral equilibrium (*D* = -0.11, *P* >0.10; Table 2-4). In contrast, the MK test showed that *LjE1* did not evolve under neutral selection (*P* < 0.05): *LjE1* harbored 54 synonymous and 34 nonsynonymous fixed differences between *G. max* (outgroup) and *L. japonicus*. All five SNPs within the 49 wild samples of *L. japonicus* were nonsynonymous substitutions (Table 2-3). Three of the five amino acid replacements (*Met69Leu* and *His110Arg*), and (*Arg163Leu*) were detected exclusively in accessions MG-63, 66, 73 and accession MG-53, respectively. The other two replacements (*Gly22Ser*, *Asn66Ser*) were found in the northern and southern populations, respectively (Table 2-3).

Polymorphism of *GIGANTEA*

Approximately 2,930 bp of *LjGI* CDS (~83% of sequences between exons 4 and 14) were analysed in 23 samples from eight wild accessions. All of the CDS in two SNPs were specific to a geographic location. One of the two SNPs was a synonymous change (*T2152A* in Gifu GifuB-129), while *A748G* was nonsynonymous and specific to Tomamae (Hokkaido). All samples from Tomamae had an identical genotype (Table 2-5). The genetic diversity of *LjGI* was low (π = 0.0001) compared to that of 116 functional genes in *G. soja* (π_{total} = 0.00217; Hyten et al. 2006). Tajima's *D* test and the MK test suggested that *LjGI* did not deviate from neutral equilibrium (*D* = –0.034, *G*-value = 0.156, *p* = 0.69330; Tables 2-4 and 2-6). *LjGI* harbored 216 synonymous substitutions and 123 nonsynonymous substitutions between *G. max* and *L. japonicus*.

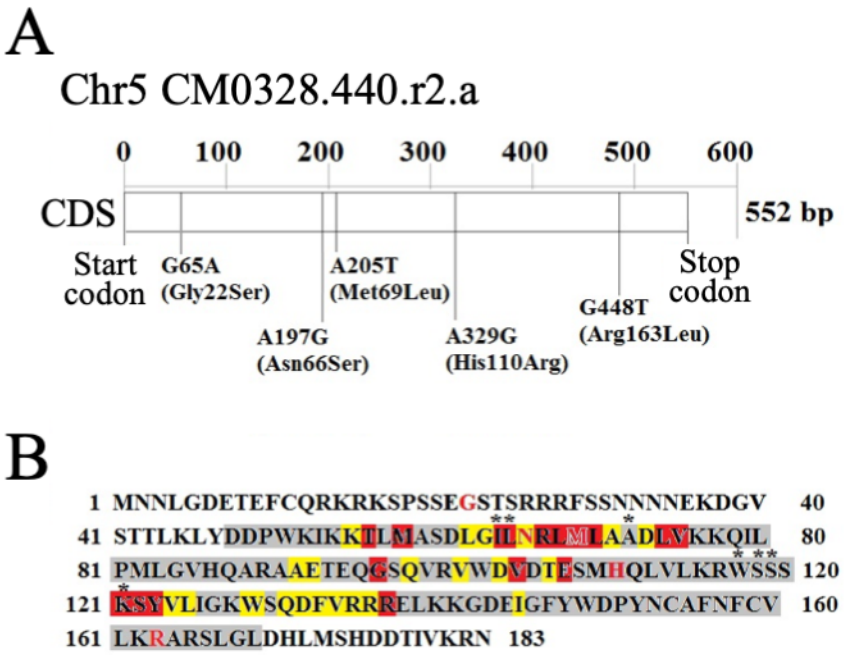

Fig. 2-5. Structure of *E1* of *Lotus japonicus*

A: Structure of the *LjE1* sequence. Scale is shown above as nucleic acid base pairs. SNP sites are indicated by vertical lines below. **B**: Amino acid sequence of *LjE1* protein (MG-20). Amino acids in red represent the polymorphic amino acids identified in this study. Regions marked in gray, red, and yellow represent the B3-domain structure, amino acids likely affected by DNA-binding, and amino acids strongly affected by DNA-binding, respectively.

Discussion

Polymorphisms in *E1* suggest possible flowering-time regulation in *L. japonicus*

MK tests detected non-neutral evolutionary selection in *LjE1*, while other genes lacked non-neutral evolution tendencies (*LjPHYA* and *LjCRY1*; $p > 0.05$; Table 2-6). These results suggest that *LjE1* may have a role in adaptation to local habitats. According to the domain database of the National Center for Biological Information (http://www.ncbi.nlm.nih.gov/), part of the *LjE1* protein forms the B3 domain, which is involved in DNA binding. The B3-domain structure of *LjE1* was almost equivalent to that of the AtRAV1 protein based on a Phyre search (http://www.imperial.ac.uk/phyre). The B3 DNA-binding domain is found in transcription factor proteins. Consequently, this domain probably has a critical function in *L. japonicus* in relation to flowering time regulation. In *G. max*, *E1* directly represses the expression of the *FT* gene (Xia et al. 2012). Comparison of amino acid sequences between the *LjE1* protein and the *E1* protein (Xia et al. 2012) revealed that one of the SNPs was located in the N-terminal while the other four SNPs were in the putative DNA-binding motif (Table 2-3; Fig. 2-5). Note that samples with three of the four SNPs (*A197G*, *A205T* and *A329G*) were from southern populations in Japan that tend to exhibit early flowering (Kawaguchi 2000; Kai et al. 2010). In addition, two of the four SNPs in this domain (A197G and A205T) were located at adjacent amino acids, with backbone chemical shifts affected by DNA binding (Fig. 2-5B; Yamasaki et al. 2004). Thus, these SNPs may affect the function of the *LjE1* protein and cause early flowering in these plants. Considering the structure of this gene, all or some of these five substitutions may be involved in the control of flowering time.

The *L. japonicus LjFTa* gene likely has a role in flowering activation as a downstream reaction (Yamashino et al., 2013), and Xia et al. (2012) reported that *E1* likely controls the expression of *GmFT*s in *G. max*. Therefore, *LjE1* may play a role in the direct control of *LjFTa*, and comparative analysis of *LjE1* and *LjFTa* expression between wild populations should be undertaken to evaluate whether *LjE1* polymorphisms among the alleles contribute to flowering-time regulation.

Table 2-3 SNPs and amino acid changes identified in *E1* gene in *Lotus japonicus*.

No.#	Location or Accession No. of Legume Base	SNP information	*E1*				
		Position	65*	197*	205*	329*	488*
		Domain	N-terminal	B3domain	B3domain	B3domain	B3domain
		Amino Acid	22	66	69	110	163
		SNP*	G/A	A/G	A/T	A/G	G/T
		Codon change	GGT AGT	AAC AGC	ATG TTG	CAC CGC	CGC CTC
		Amino acid change	Gly/Ser	Asn/Ser	Met/Leu	His/Arg	Arg/Leu
1	Tomamae 1		A	G	A	A	G
	Tomamae 2		A	G	A	A	G
	Tomamae 3		A	G	A	A	G
	Tomamae 4		A	G	A	A	G
2	Wakka 1		G	G	A	A	G
	Wakka 2		G	G	A	A	G
	Wakka 3		G	G	A	A	G
3	Kimuneappu 1		A	G	A	A	G
	Kimuneappu 2		A	G	A	A	G
	Kimuneappu 3		A	G	A	A	G
	Kimuneappu 4		A	G	A	A	G
4	Hakodate 1		A	G	A	A	G
	Hakodate 2		A	G	A	A	G
	Hakodate 3		A	G	A	A	G
	Hakodate 4		A	G	A	A	G
5	Hakkouda 1		A	G	A	A	G
	Hakkouda 2		A	G	A	A	G
	Hakkouda 3		A	G	A	A	G
	Hakkouda 4		A	G	A	A	G
6	Kitsunezaki 1		G	G	A	A	G
	Kitsunezaki 2		G	G	A	A	G
	Kitsunezaki 3		G	G	A	A	G
	Kitsunezaki 4		G	G	A	A	G
7	MG-42		G	G	A	A	G
8	Kakuda 1		G	G	A	A	G
	Kakuda 2		G	G	A	A	G
	Kakuda 3		G	G	A	A	G
	Kakuda 4		G	G	A	A	G
9	Gifu B-129		G	G	A	A	G
10	MG-53		G	G	A	A	T
11	MG-63		G	G	T	G	G
12	MG-66		G	G	T	G	G
13	MG-73		G	G	T	G	G
14	Udo 1		G	G	A	A	G
	Udo 2		G	G	A	A	G
	Udo 3		G	G	A	A	G
15	MG-22		G	G	A	A	G
16	Tokunoshima 1		G	G	A	A	G
	Tokunoshima 2		G	A	A	A	G
	Tokunoshima 3		G	G	A	A	G
	Tokunoshima 4		G	A	A	A	G
17	Nishihara 1		G	G	A	A	G
	Nishihara 2		G	G	A	A	G
	Nishihara 3		G	A	A	A	G
	Nishihara 4		G	G	A	A	G
18	Miyakojima 1		G	A	A	A	G
	Miyakojima 2		G	A	A	A	G
	Miyakojima 3		G	A	A	A	G
	Miyakojima 4		G	A	A	A	G

#: No. is corresponded to those in Fig. 2-1 and Table 2-2.

* SNP position that causes nonsynonymous change

Table 2-4 Results of neutral test using Tajima's *D* test and nucleotide diversity in each *E1* and *GIGANTEA* genes of *Lotus japonicus*.

Locus	π	Tajima's *D*	*P-value*
E1	0.00175	-0.01124	>0.10
GIGANTEA	0.00011	-0.63623	>0.10

Tajima's *D* and nucleotide diversity (π) are shown. Tajima's *D* and nucleotide diversity were estimated for the entire sequence.

Table 2-5 SNPs and amino acid changes identified in *GIGANTEA* gene in *Lotus japonicus*.

No.#	Location or Accession No. of Legume Base	SNP information	***GIGANTEA***	
		Position	712*	2115
		SNP	A/G	T/A
		Codon change	ATG GTG	ATT ATA
		Amino acid change	Met/Val	Lle/Lle
1	Tomamae 1		A	T
	Tomamae 2		A	T
	Tomamae 3		A	T
	Tomamae 4		A	T
5	Hakkouda 1		G	T
	Hakkouda 2		G	T
	Hakkouda 3		G	T
	Hakkouda 4		G	T
7	Kitsunezaki 1		G	T
	Kitsunezaki 2		G	T
	Kitsunezaki 3		G	T
	Kitsunezaki 4		G	T
9	Gifu B-129		G	A
15	MG-22		G	T
16	Tokunoshima 1		G	T
	Tokunoshima 2		G	T
	Tokunoshima 3		G	T
	Tokunoshima 3		G	T
17	Nishihara 1		G	T
	Nishihara 2		G	T
	Nishihara 3		G	T
	Nishihara 4		G	T
18	MG-20		G	T

#: No. is corresponded to those in Fig. 2-1 and Table 2-2.

* SNP position that causes nonsynonymous change

Table 2-6 Results of neutral test using Mcdonald and Kreitman test

Locus	Type of substitutions	Synonymous	Nonsynonymous	Fisher's exact test	*G*-test
E1	Polymorphisms	0	5	*P*=0.011078	not be performed
	Fixed difference	54	34		
GI	Polymorphisms	1	1	*P*=1.000000	*P*=0.69330
	Fixed difference	216	123		
PHYA	Polymorphisms	2	5	*P*=0.462901	*P*=0.35784
	Fixed difference	164	195		
CRY1	Polymorphisms	0	1	*P*=1.000000	not be performed
	Fixed difference	20	127		

The number of polymorphisms and fixed differences indicate the number of substitutions within *L. japonicus* and those fixed between *L. japonicus* and *G. max*.

Identification of the circadian clock gene *GIGANTEA* and its intraspecific polymorphisms

In this study, the location of *GI* was identified as chr5.CM0052.430.r2.a because the existing DNA database for *L. japonicus* included unidentified sequences around this region. I obtained the complete DNA sequence of *LjGI* by sequencing cDNA and then compared these sequence data to a revised database provided by the Kazusa DNA Research Institute. Despite the relatively high homology of DNA and amino acid sequences between *LjGI* and *GmGI*, the 5'-UTR region could not be determined because of the insertion of a retrotransposon. Thus, further studies are needed to clarify the complete structure of *LjGI*.

Polymorphisms of *LjGI* may be restricted by purifying selection (Table 2-5), indicating strong constraints on its functions. In general, circadian clock genes are important for the plant life cycle (Yano et al. 2000; Hecht et al. 2005) and *GI* plays an important role in *A. thaliana* and in *G. max* (Fowler et al. 1999; Watanabe et al. 2011). Nevertheless, mutations of circadian clock genes can alter the phenotype of *A. thaliana* (e.g., slow growth rate of vegetative parts and low survival rate of mutants; Dodd et al. 2005). Therefore, I cannot exclude the possibility that the single SNP resulting in an amino acid replacement (*A748G*, *Met249Val*) in the northernmost samples at Tomamae (Accession 1 in Fig. 2-1 and Table 2-2) could affect flowering time. In fact, *L. japonicus* plants from this locality exhibited remarkably late flowering (data not shown). Thus, despite the few polymorphisms in *LjGI*, I cannot exclude the possibility that the SNP is involved in the regulation of flowering time. Further investigations of the functions of *LjGI* alleles should not exclude the possibility that the SNP is involved in the regulation of flowering time.

Chapter 3

Detecting loci that contribute to flowering time variation in *Lotus japonicus*

Abstract

Flowering time is an important factor for fitness and local adaptation in plants. Genome-wide association (GWA) studies have allowed the identification of candidate genes for various traits, and GWA studies have been successfully used to detect loci associated with regulation of flowering time in some plant species. *Lotus japonicus* can be widely found in the entire Japanese archipelago and shows intra-specific flowering time variations. Flowering time data were collected for 132 wild accessions originating from various points across the Japanese archipelago. Flowering time measurements showed that this species had flowering time variation along with latitude, with southern accessions showing earlier flowering times. This result suggested that there is a possibility that environmental factors, e.g. temperature and photoperiod, might affect flowering time control and contribute to local adaptation in this species. Comparing with another flowering time data under different conditions, it is considerable that day length would be important for flowering time control in this species. GWA analyses were conducted on flowering time variation and two environmental variables, and detected several single nucleotide polymorphisms (SNPs) and candidate genes, including orthologues of known flowering time genes. The analysis for flowering time variation detected 16 protein coding genes that contain two orthologues of known flowering time genes. Correlation tests between flowering time and SNPs strongly detected in the GWA analysis suggested that approximately 60 % of flowering time variation can be explained by the top two SNPs. This result suggested that most part of the variation could be explained with small number of genetic factors. One of these SNPs was associated with an orthologue of known flowering time genes, *SmD3b-like* and its gene expressions had been observed in previous study. Considering the strong association with flowering time variation, these candidates might cause the variation and be related to local adaptation in this species.

Introduction

Plants have four stages in their life cycles; the timings of germination and flowering are dictated by environmental factors. In particular, flowering time is critical for wild plants, as it is directly linked to fitness through reproductive success (Schemske 1978). There are several studies that focused on flowering time to investigate local adaptation of plant species (Dittmar et al. 2014; Hall and Willis 2006; Keller et al. 2012; Leinonen et al. 2013). Many major causal genes for flowering time have been detected in genetic and physiological studies in several model plants, including *Arabidopsis*, *Populus* sp. and *Glycine max*. It is well known that flowering time is controlled by four major genetic pathways (temperature, photoperiod, autonomous, and GA pathways), with more than 100 genes identified in these pathways (Srikanth and Schmid 2011). Of the four factors, temperature and photoperiod are important for plants to determine flowering time in wild environments (Henderson and Dean 2004).

Improvements in sequencing technology have enabled the identification of many genes associated with various phenotypes in many species by using genome-wide association (GWA) studies. GWA studies use a large number of shared alleles, and therefore, could detect genetic factors that account for natural variation among wild populations, including adaptive phenotypic variations. Indeed, many genes responsible for various phenotypes have been identified by GWA studies, such as in *Arabidopsis thaliana* (Atwell et al. 2010), rice (Zhao et al. 2011), soybean (Zhang et al. 2015), and *Lotus japonicus* (Shah et al. 2020). While previous GWA studies have placed emphasis on agricultural varieties, several recent studies have attempted to find adaptive loci with GWA by using high density SNPs in the conserved alleles of wild accessions (Atwell et al. 2010; Fournier-Level et al. 2011; Shah et al. 2020; Yu et al. 2016). Some of these studies aimed to determine possible candidate genes that are key to flowering time, as it is an important trait for plant adaptation and reproductive success (Burgarella et al. 2016; Sasaki et al. 2015; Shah et al. 2020). While, it is known that GWA analysis for flowering time is not easy because of its population structure within

species. Flowering time is mostly correlated with population structure, which further complicates GWA analysis. Therefore, in this study, a mixed-models which takes population structure into account was adopted (AMM; Kang et al. 2010; Seren et al. 2012). Aside from flowering time variation, other environmental variables were also included to detect genetic factors that contribute to local adaptation, and to provide further insights into relationships between SNPs and flowering time/local adaptation. In fact, GWA analyses for environmental variables enabled the detection of candidate genes in several previous studies (e.g. Hancock et al. 2011). Some of these studies have used environmental factors to detect associations between local adaptations and genotypes. In plant species that have huge representations in various environments, this method is effective in identifying local adaptation loci.

L. japonicus is a legume model plant and is distributed across the entire Japanese archipelago. The Japanese archipelago extends across a large latitudinal range (approximately 20 degrees), which includes many different environments. For instance, day length is extremely different between the north-most and south-most points, by up to 4 h. In addition, other environmental factors, including temperature and precipitation, vary greatly throughout the archipelago due to its mountainous topology, the prevailing westerlies, and seasonal winds. Given its presence in various environments in Japan, it is conceivable that *L. japonicus* could adapt to many environments. Previous studies have shown that this species has intraspecific flowering time variation (Kai et al. 2010; Kawaguchi 2000); wild *L. japonicus* accessions from natural populations in Japan has been previously collected (Hashiguchi et al. 2012), and may be used to reflect genome-wide polymorphisms in GWA studies. In this study, I tracked flowering time days under natural light condition in a greenhouse in the middle of Japan, because this species is one of the long-day plants and the natural light condition from early spring would make the flowering time difference reflect the response to natural daylength. Therefore, the measurement was started in early spring in accordance with the natural life cycle of this plant.

GWA analyses for several traits of this species, such as flowering time, have already been performed on species in greenhouses at Aarhus, Denmark, and in a field in Miyagi; several candidate genes were revealed in the variation analysis in Miyagi (Shah et al. 2020). In this study, we conducted GWA analyses for flowering time and environmental factors. For the flowering time data, as compared with the data at Aarhus and Miyagi, my flowering time data showed much extremer differences between accessions because of the natural light condition. The result of GWA analysis for more prominent difference of flowering time in different condition should be reflect the different genetic factor contributing to flowering time differences in this species. In addition, although it is almost the same, the used accessions were partially different between these studies and it might cause differences in flowering time variation and GWA analyses results. In fact, my GWA study results differed from that of the Aarhus and Miyagi data because of the different reactions of *L. japonicus* to various environments. By combining the genetic information and variations in the flowering time in these wild accessions, I can determine genetic factors that are associated with local adaptation. In addition, GWA analyses for other environmental factors may help me to detect other candidate genes that contribute to local adaptation. This approach would contribute to studies for local adaptation and biodiversity, especially in species that are widely distributed across a wide range of latitudes, such as in the Japanese archipelago.

Furthermore, flowering time variations are generally a gradual trait and the flowering time in this species was not an exception. However, there are few studies focusing on the combination of effects by multiple genetic factors. Besides GWA analysis, to estimate how much flowering time variation could be explained by genotypes of SNPs detected by GWA analysis, I conducted correlation tests between flowering time variation and genotypes of top detected SNPs considering the effect of a combination of genetic factors. Because GWA analysis is a "SNP by SNP" method, adopting both of these analyses would be effective to reveal mechanisms of the intraspecific differentiation. Moreover, my results would contribute to understanding the local adaptation through flowering time in *L. japonicus*.

Materials and Methods

Plant material and measurement of flowering time

Flowering time phenotypes of 132 wild Japanese *L. japonicus* lines (Fig. 3-1, Table 3-1) were collected. Seeds from 100 and 28 lines were

provided by NBRP (Legume Base) and Aarhus University, respectively. In addition, four lines (Tomamae, Hakkoda, Kitsunezaki, and Tokunoshima) previously collected and cultivated were used after two generations of inbreeding by self-pollination. Fig. 3-2 shows the locality information of accessions with collection site information, drawn by QGIS 2.18.2.

Before sowing the seeds, the seed coat was partially scrubbed away with sandpaper; 15 seeds from each line were sown on 1.0 % agar in a 90 mm-diameter, 15 mm-deep Petri dish, and were germinated in a growth chamber (Biotron LPH410S; NK Systems, Osaka, Japan) under a 16 h day: 8 h night cycle at 25 °C for 1 week. Plants from each line were planted in three 9.0 cm-diameter, 20 cm-high polyethylene pots filled with potting compost, expanded vermiculite, and soil. The potted plants were placed in a greenhouse (Okazaki, Aichi, 34°57′N, 137°09′E) on 31 January 2014 and were grown under natural light conditions: approximately 10.5–14.5 h of daylight in the experimental period, considering that *L. japonicus* behaves as a long-day plant in the wild. After 1 week, the plants in the pot were thinned, leaving only the largest individual. At this stage, each plant had three to five compound leaves.

The date on which the first flower was fully opened was recorded, and the date for each line was calculated as the mean of three individuals. In addition, flowering days from germination to opening of the first flower were calculated, and the correlation between the latitude of collection sites of wild lines and average flowering days in each accession was evaluated using the R package 3.1.2 (R Core Team 2015).

Genome-wide association mapping for flowering time and environmental factors

I conducted genome-wide association analyses on the genotype data (Shah et al. 2020) with an accelerated mixed model (AMM; Kang et al. 2010; Seren et al. 2012) on the website https://lotus.au.dk/gwas/. The variables were transformed in square root (sqrt), natural logarithm (log), and boxcox and used for the analysis besides not-transformed data. For flowering time data gathered under the same condition, not-flowered individuals were removed and 121 out of 132 wild accessions were used for GWA analyses (Sampling points and flowering days of used accessions shown in Table 3-1). GWA analysis for a tentative data set in which not-flowering individuals were treated as 250 flowering days was also conducted; however, a (quantile–quantile) Q–Q plot for tentative data showed an overestimated result compared to that without not-flowered individuals (Fig. 3-3, 3-4). Therefore, I adopted the flowering time result without not-flowered individuals in this study. Genes around the SNPs beyond Bonferroni-corrected 5 % significant threshold were selected as possible candidates.

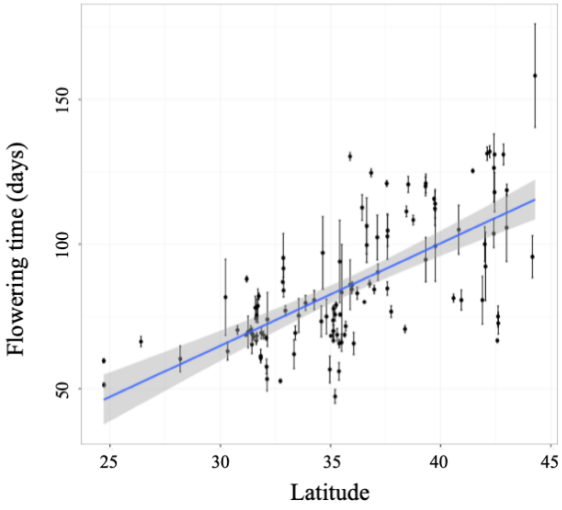

Fig. 3-1. Variations in flowering time of 127 wild accessions in Japan across latitudes

Flowering time is shown on the vertical axis (days), and latitude where accessions originated from are shown on the horizontal axis. Black bars denote standard deviations of flowering days in each accession. The blue line and grey area represent the regression line and the 95 % confidence interval, respectively.

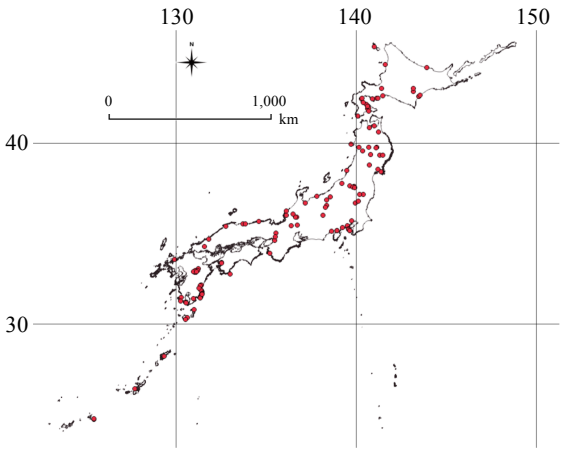

Fig. 3-2. Locality of accessions with collection site information

Red circles denote locations at which accessions originated from.

Table 3-1 Sampling points of *Lotus japonicus* wild lines using in this study and results of flowering time measurement.

Accession ID	north latitude	west longitude	Flowering time (Days)			Average	GWA analysis
MG-001	32°43'13"	133°01'03"	52	54	52	52.67	FT, Temp, Sunshine
MG-002	42°34'19"	143°30'11"	68	66	66	66.67	FT, Temp, Sunshine
MG-003	35°23'54"	139°30'59"	66	63	68	65.67	FT, Temp, Sunshine
MG-004	31°26'10"	130°18'49"	na	na	na	na	Temp, Sunshine
MG-005	35°18'08"	139°16'01"	66	73	67	68.67	FT, Temp, Sunshine
MG-007	39°18'56"	141°20'16"	119	126	115	120.00	FT, Temp, Sunshine
MG-008	35°01'06"	135°33'27"	68	68	69	68.33	FT, Temp, Sunshine
MG-009	39°55'36"	139°44'37"	196	not flowered	192	194.00	FT, Temp, Sunshine
MG-010	39°19'28"	141°29'05"	110	87	87	94.67	FT, Temp, Sunshine
MG-011	39°19'38"	140°48'32"	126	115	122	121.00	FT, Temp, Sunshine
MG-012	35°15'25"	139°35'50"	80	77	80	79.00	FT, Temp, Sunshine
MG-013	35°12'36"	139°36'20"	80	73	74	75.67	FT, Temp, Sunshine
MG-014	35°12'36"	139°36'20"	46	52	44	47.33	FT, Temp, Sunshine
MG-015	35°08'07"	139°40'35"	76	80	77	77.67	-
MG-016	35°08'07"	139°40'35"	na	na	na	na	Temp, Sunshine
MG-017	35°07'59"	138°57'32"	68	66	66	66.67	FT, Temp, Sunshine
MG-018	35°07'59"	138°57'32"	68	68	74	70.00	FT, Temp, Sunshine
MG-019	35°07'59"	138°57'32"	70	74	66	70.00	FT, Temp, Sunshine
MG-020	24°43'07"	125°28'59"	59	59	61	59.67	FT, Temp, Sunshine
MG-021	41°27'38"	140°05'36"	126	124	126	125.33	FT, Temp, Sunshine
MG-022	31°11'30"	130°33'45"	87	90	87	88.00	FT, Temp, Sunshine
MG-023	40°49'25"	140°45'43"	122	96	97	105.00	FT, Temp, Sunshine
MG-024	24°43'57"	125°28'11"	52	52	50	51.33	FT, Temp, Sunshine
MG-025	30°19'44"	130°39'00"	59	61	69	63.00	FT, Temp, Sunshine
MG-026	35°41'47"	139°45'29"	75	70	70	71.67	FT, Temp, Sunshine
MG-027	26°24'07"	127°44'10"	63	67	69	66.33	FT, Temp, Sunshine
MG-028	31°09'25"	130°35'15"	69	68	69	68.67	FT, Temp, Sunshine
MG-030	42°50'51"	143°10'22"	138	126	129	131.00	FT, Temp, Sunshine
MG-032	42°32'47"	143°28'37"	143	not flowered	not flowered	143.00	FT, Temp, Sunshine
MG-033	unkown	unkown	124	129	122	125.00	FT
MG-034	42°36'51"	143°32'54"	68	70	80	72.67	FT, Temp, Sunshine
MG-035	42°26'44"	141°09'38"	108	115	131	118.00	FT, Temp, Sunshine
MG-036	42°24'34"	140°18'21"	94	108	109	103.67	FT, Temp, Sunshine
MG-038	42°06'29"	140°34'18"	129	129	136	131.33	FT, Temp, Sunshine
MG-039	40°56'39"	141°01'08"	80	87	75	80.67	FT, Temp, Sunshine
MG-040	40°35'13"	141°14'57"	80	84	80	81.33	FT, Temp, Sunshine
MG-041	39°45'20"	140°09'42"	122	105	115	114.00	FT, Temp, Sunshine
MG-042	39°44'43"	140°42'28"	124	108	105	112.33	FT, Temp, Sunshine
MG-044	38°23'12"	141°23'18"	70	73	69	70.67	FT, Temp, Sunshine
MG-045	37°35'19"	139°54'54"	111	110	87	102.67	FT, Temp, Sunshine
MG-046	37°35'15"	139°54'51"	87	87	80	84.67	FT, Temp, Sunshine
MG-049	35°38'48"	134°37'08"	68	68	70	68.67	FT, Temp, Sunshine
MG-050	35°30'06"	133°43'54"	63	63	72	66.00	FT, Temp, Sunshine
MG-051	35°30'37"	133°51'15"	76	115	59	83.33	FT, Temp, Sunshine

Table 3-1. Continued

MG-052	35°23'03"	132°47'20"	54	52	62	56.00	FT, Temp, Sunshine
MG-053	34°39'03"	131°49'35"	72	112	107	97.00	FT, Temp, Sunshine
MG-056	31°34'51"	131°24'58"	82	82	70	78.00	FT, Temp, Sunshine
MG-057	31°41'59"	131°28'56"	80	68	88	78.67	FT, Temp, Sunshine
MG-058	unkown	unkown	80	68	62	70.00	FT
MG-059	unkown	unkown	76	70	94	80.00	FT
MG-060	31°15'14"	130°17'36"	73	77	59	69.67	FT, Temp, Sunshine
MG-061	31°22'28"	130°59'18"	70	68	73	70.33	FT, Temp, Sunshine
MG-062	unkown	unkown	96	87	73	85.33	-
MG-063	31°44'23"	131°28'15"	80	82	84	82.00	FT, Temp, Sunshine
MG-064	31°49'46"	131°24'51"	59	61	62	60.67	-
MG-065	31°49'46"	131°24'55"	59	59	66	61.33	FT, Temp, Sunshine
MG-066	32°07'54"	131°24'34"	91	72	59	74.00	FT, Temp, Sunshine
MG-067	30°46'25"	131°00'31"	68	70	73	70.33	FT, Temp, Sunshine
MG-068	30°14'03"	130°33'22"	68	108	69	81.67	FT, Temp, Sunshine
MG-069	34°35'14"	135°30'27"	68	68	84	73.33	FT, Temp, Sunshine
MG-070	33°52'27"	135°13'41"	80	84	75	79.67	FT, Temp, Sunshine
MG-071	32°56'45"	131°05'55"	74	77	80	77.00	FT, Temp, Sunshine
MG-072	32°51'06"	131°03'11"	87	87	112	95.33	FT, Temp, Sunshine
MG-073	33°02'14"	131°16'01"	na	na	na	na	Temp, Sunshine
MG-074	33°23'41"	132°32'07"	74	66	68	69.33	FT, Temp, Sunshine
MG-075	33°20'19"	132°32'38"	66	52	68	62.00	FT, Temp, Sunshine
MG-076	34°15'48"	131°35'28"	74	84	84	80.67	FT, Temp, Sunshine
MG-077	35°07'09"	138°39'45"	68	70	83	73.67	FT, Temp, Sunshine
MG-078	35°53'45"	136°43'23"	129	133	129	130.33	FT, Temp, Sunshine
MG-079	36°13'13"	136°08'29"	82	87	80	83.00	-
MG-080	36°38'58"	137°11'26"	87	115	117	106.33	FT, Temp, Sunshine
MG-081	37°02'07"	137°51'00"	na	na	na	na	Temp, Sunshine
MG-082	37°45'38"	139°12'31"	80	77	73	76.67	FT, Temp, Sunshine
MG-083	36°32'38"	138°21'48"	80	80	80	80.00	FT, Temp, Sunshine
MG-084	37°36'09"	139°39'01"	108	112	94	104.67	FT, Temp, Sunshine
MG-085	37°09'14"	140°24'49"	88	87	96	90.33	FT, Temp, Sunshine
MG-086	38°45'39"	140°45'35"	105	110	110	108.33	FT, Temp, Sunshine
MG-088	36°39'07"	139°58'38"	108	103	88	99.67	FT, Temp, Sunshine
MG-089	39°45'47"	141°08'27"	75	112	111	99.33	FT, Temp, Sunshine
MG-090	43°00'16"	141°25'57"	122	115	119	118.67	FT, Temp, Sunshine
MG-091	41°53'44"	140°38'21"	75	70	97	80.67	FT, Temp, Sunshine
MG-092	44°09'36"	143°56'10"	94	84	109	95.67	FT
MG-093	42°02'01"	140°40'13"	75	115	87	92.33	FT, Temp, Sunshine
MG-094	42°25'07"	140°54'11"	140	136	103	126.33	FT, Temp, Sunshine
MG-095	42°28'31"	141°13'13"	138	not flowered	not flowered	138.00	FT, Temp, Sunshine
MG-096	42°36'10"	141°30'10"	68	80	77	75.00	FT, Temp, Sunshine
MG-097	42°26'59"	140°20'35"	131	131	131	131.00	FT, Temp, Sunshine
MG-098	41°55'15"	140°43'41"	not flowered	147	not flowered	147.00	FT, Temp, Sunshine
MG-099	41°45'25"	140°43'30"	147	not flowered	not flowered	147.00	FT, Temp, Sunshine
MG-100	42°13'44"	140°24'07"	136	129	131	132.00	FT, Temp, Sunshine

Table 3-1. Continued

MG-101	42°00'11"	140°39'03"	94	94	112	100.00	FT, Temp, Sunshine
MG-102	45°17'46"	141°01'54"	not flowered	182	131	156.50	FT, Temp, Sunshine
MG-103	42°59'37"	143°11'52"	94	94	129	105.67	FT, Temp, Sunshine
MG-104	32°05'02"	131°24'44"	68	66	69	67.67	FT, Temp, Sunshine
MG-105	32°07'05"	131°25'02"	47	52	61	53.33	FT, Temp, Sunshine
MG-106	32°05'41"	131°22'46"	53	58	62	57.67	FT, Temp, Sunshine
MG-107	34°49'04"	135°31'20"	69	69	87	75.00	FT, Temp, Sunshine
MG-109	39°42'21"	141°06'51"	117	115	115	115.67	FT, Temp, Sunshine
MG-110	32°49'40"	131°08'01"	87	87	87	87.00	FT, Temp, Sunshine
MG-111	32°51'36"	131°00'03"	88	89	98	91.67	FT, Temp, Sunshine
MG-112	32°55'58"	131°13'45"	na	na	na	na	Temp, Sunshine
MG-113	32°51'47"	131°09'18"	80	84	88	84.00	FT, Temp, Sunshine
MG-116	35°53'60"	136°39'12"	80	75	103	86.00	-
MG-117	36°03'41"	136°30'02"	70	58	69	65.67	FT, Temp, Sunshine
MG-118	35°58'29"	136°07'47"	82	84	87	84.33	FT, Temp, Sunshine
MG-119	unkown	unkown	87	87	87	87.00	FT
MG-120	36°59'16"	138°34'38"	87	82	84	84.33	FT, Temp, Sunshine
MG-121	36°51'06"	138°21'58"	122	126	126	124.67	FT, Temp, Sunshine
MG-122	36°26'25"	138°18'14"	104	119	115	112.67	-
MG-123	35°59'44"	138°09'34"	84	87	87	86.00	FT, Temp, Sunshine
MG-124	37°33'41"	139°49'20"	122	122	119	121.00	FT, Temp, Sunshine
MG-125	37°30'51"	139°55'15"	not flowered	82	108	95.00	FT, Temp, Sunshine
MG-126	37°07'33"	140°12'39"	87	112	108	102.33	FT, Temp, Sunshine
MG-127	38°31'53"	141°13'20"	117	126	119	120.67	FT, Temp, Sunshine
MG-128	36°46'60"	140°07'41"	88	87	84	86.33	FT
MG-129	39°32'43"	140°22'10"	108	dead	117	112.50	FT, Temp, Sunshine
MG-130	35°25'05"	136°24'23"	68	117	97	94.00	FT, Temp, Sunshine
MG-131	34°34'48"	135°37'12"	52	66	52	56.67	FT
MG-132	28°10'59"	129°22'22"	54	69	58	60.33	FT, Temp, Sunshine
MG-133	31°36'46"	131°26'02"	77	73	73	74.33	FT, Temp, Sunshine
MG-134	31°29'54"	131°23'03"	68	66	70	68.00	FT, Temp, Sunshine
MG-135	31°28'23"	131°21'57"	66	dead	62	64.00	FT, Temp, Sunshine
MG-136	31°27'31"	131°21'23"	na	na	na	na	Temp, Sunshine
MG-137	31°26'51"	131°22'14"	66	68	73	69.00	-
MG-138	31°25'33"	131°21'16"	68	59	69	65.33	FT, Temp, Sunshine
MG-139	31°49'15"	131°25'14"	68	66	dead	67.00	FT, Temp, Sunshine
MG-140	31°51'07"	131°23'58"	69	73	66	69.33	FT, Temp, Sunshine
MG-141	31°57'04"	131°18'29"	70	70	66	68.67	FT, Temp, Sunshine
MG-142	31°38'41"	131°27'52"	80	73	82	78.33	FT, Temp, Sunshine
MG-143	31°38'43"	131°28'00"	66	69	70	68.33	FT, Temp, Sunshine
MG-144	31°38'41"	131°27'58"	73	73	80	75.33	FT, Temp, Sunshine
MG-145	31°37'37"	131°27'27"	68	64	68	66.67	FT, Temp, Sunshine
MG-146	33°32'59"	129°55'43"	87	70	69	75.33	FT, Temp, Sunshine
Gifu B-129	35°43'46"	136°76'	75	75	77	75.67	-
Tokunoshima	27°45'	129°01'	dead	82	dead	82.00	-
Kitsunezaki	38°27'	139°29'	115	109	110	111.33	-
Hakkoda	40°38'	140°51'	187	not flowered	not flowered	187.00	-
Tomamae	44°28'	141°69'	194	138	143	158.33	-

In addition to flowering time, environmental factors released by the Japan Meteorological Agency were used to detect local adaptation loci. In this data set, the predicted values for each factor (on a 1 km^2 mesh) were selected based on values obtained at meteorological observatories from 1981 to 2010. In this study, the annual average temperature and integrated annual sunshine duration were adopted as part of the environmental factors and were included in analyses of the 119 wild accessions when collection site information was available. Each mesh point on which an accession was located was used as the value for the accession (sampling points of used accessions are shown in Table 3-1). Because there were few SNPs beyond Bonferroni-corrected 5 % significant threshold, I focused on genes around the top 100 SNPs as possible candidates. To draw Manhattan and Q–Q plots for every analysis, the qqman program (Turner 2018) was used.

Annotation of candidate genes

Based on the position of SNPs detected in GWA analysis, amino acid sequences of candidate genes were listed from the *L. japonicus* genome assembly build 3.0 using bedtools2 v2.26.0 (Quinlan and Hall 2010). Genes located 10 kilo base pairs (kbp) upstream and downstream (20 kbp in total) of SNPs were detected and were considered in linkage disequilibrium (Shah et al. 2020). To further annotate candidate genes in detail, the blastp function in BLAST 2.2.26 (Altschul et al. 1990) on the NCBI BLAST web site (https://blast.ncbi.nlm.nih.gov/Blast.cgi). For the candidate genes of annual average temperature and integrated annual sunshine duration, I collated

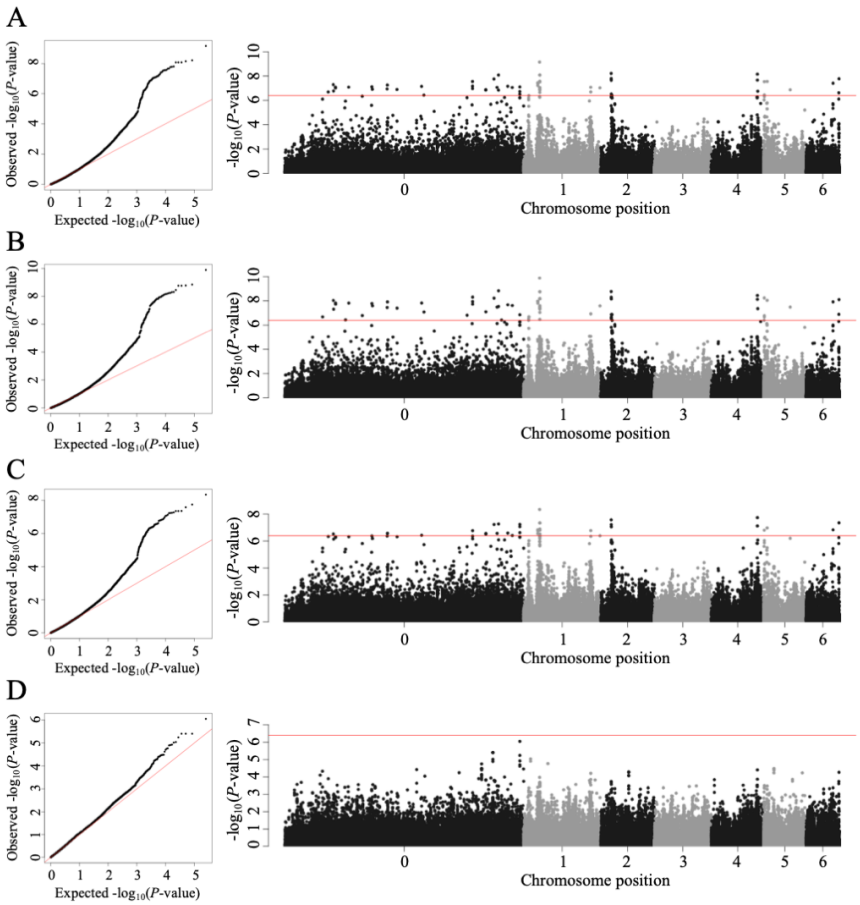

Fig. 3-3. Manhattan and *Q*–*Q* plots of genome-wide association analyses of tentative flowering time variation

A–D show the plots for the results with sqrt, not-transformed, log, and boxcox flowering time data, respectively. The (quantile–quantile) *Q*–*Q* plots are shown in left side. The vertical axis and the horizontal axis show observed $-\log_{10}$(*P*-values) and expected $-\log_{10}$(*P*-values), respectively. Right side figures show each Manhattan plot. The vertical axis and the horizontal axis show $-\log_{10}$(*P*-values) and chromosome positions, respectively. The position and $-\log_{10}$(*P*-values) for each single nucleotide polymorphism were plotted. Red line indicates Bonferroni-corrected 5 % significant threshold.

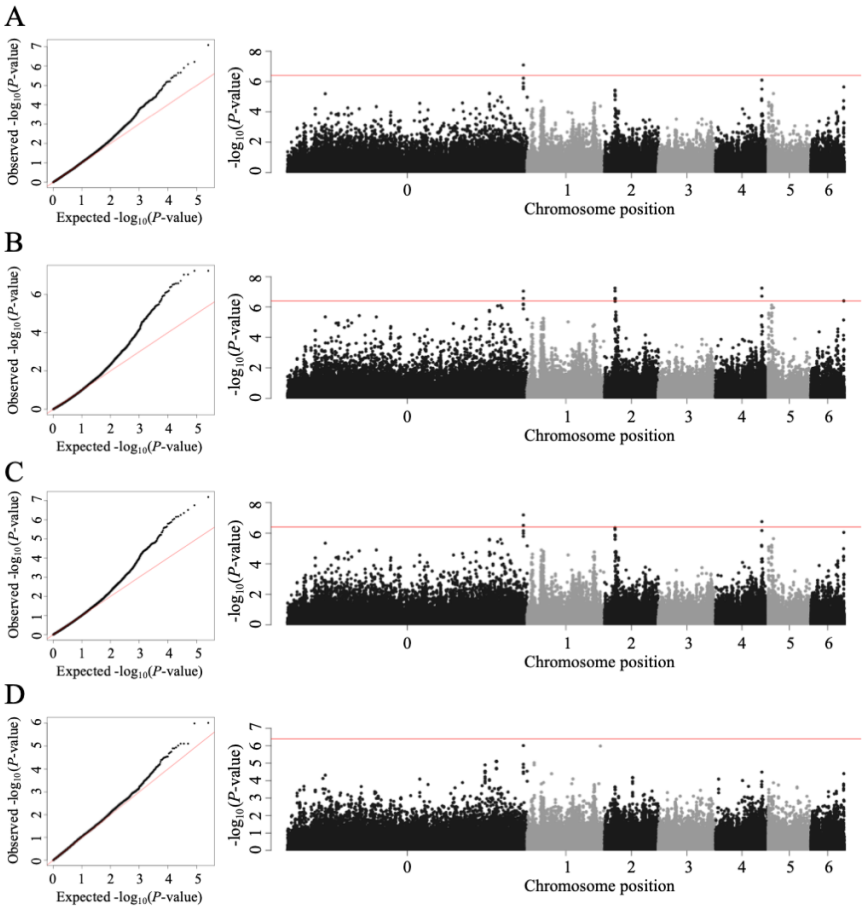

Fig. 3-4. Manhattan and *Q*–*Q* plots of genome-wide association analyses of flowering time variation

A–D show the plots for the results with sqrt, not-transformed, log, and boxcox flowering time data, respectively. The (quantile–quantile) *Q*–*Q* plots are shown in left side. The vertical axis and the horizontal axis show observed $-\log_{10}$(*P*-values) and expected $-\log_{10}$(*P*-values), respectively. Right side figures show each Manhattan plot. The vertical axis and the horizontal axis show $-\log_{10}$(*P*-values) and chromosome positions, respectively. The position and $-\log_{10}$(*P*-values) for each single nucleotide polymorphism were plotted. Red line indicates Bonferroni-corrected 5 % significant threshold.

the gene IDs with a priori to check whether the candidates contained orthologues of known flowering time genes. According to several previous studies and the *Arabidopsis* flowering genes list available on the website of the Max Planck Institute for Plant Breeding Research (Jung et al. 2012; Lu et al. 2013; Murtas et al. 2003; Stroud et al. 2013; Liu et al. 2012; Held et al. 2011; https://www.mpipz.mpg.de/14637/Arabidopsis_flowering_genes), a priori flowering time genes were listed based on known genes in *A. thaliana* and *G. max* (Table 3-2). Top hit genes in BLAST with the amino acid sequences coded by known flowering time genes were used as a priori in *L. japonicus*.

Flowering time prediction with candidate SNPs

I determined the correlation between flowering time and genotype combinations in 121 accessions to evaluate the amount of flowering time variation contributed by genotypes of detected SNPs. To impute missing data in the SNP set, BEAGLE 5.1 (Browning et al. 2018) was used in default settings. To select candidate SNPs, SNPs on chromosome 0 were first removed because this group contains short genome sequence fragments (< 10 kbp) that were not assembled in chromosomes 1 to 6. Next, regarding linkage SNPs, one SNP with the lowest *P*-value in the GWA analysis of each linkage SNP set was selected based on linkage disequilibrium (LD) decay estimation (< 0.5). The top ten non-linkage SNPs were used for flowering time prediction. Genotypes of the top SNPs with low *P*-values in the GWA analysis were added as possible

candidates in the linear regression models; each coefficient of determination value (R^2) for multiple regression analyses was calculated using R version 3.2.0. The formulae were also estimated in the linear regression analysis and predicted flowering time was calculated by the formulae shown in Fig. 3-5B. The norms of flowering time in each combination of genotypes were shown using one and two SNPs. Fig. 3-6 shows the genotype distributions of the top SNPs with collection site information, drawn by QGIS 2.18.2 with coastline data released by the Japan Meteorological Agency.

Results

Flowering time variation in *L. japonicus* in Japan

The flowering times of 132 Japanese wild lines are summarised in Table 3-1 with their original data. The first flowering was observed in a MG-105 (Miyazaki; 32°07′05″N, 131°25′02″E) on 12 March 2014, 47 days after germination. In eight lines (MG-9, 32, 35, 94, 95, 98, 99, and 125), no flowering occurred within 200 days (22 August 2014). However, those plants continued to grow in size, and the shoots were large enough for flowering; some were even larger than those of the early flowering lines. No further studies were conducted on these plants, and they were categorised as "not flowered" in Table 3-1. The average flowering time of the three individual plants in each line was calculated, and further analyses were conducted.

Based on 120 accessions, of which flowering time data could be gathered, a broad-sense heritability of 0.89 was calculated using ANOVA. In 115 of the 120 wild lines (the locality information for five accessions was lost), a strong correlation was observed between latitude and flowering time, as shown in Fig. 3-1 ($R = 0.67$; $P < 0.01$). Accessions from southern Japan tended to flower earlier. Compared to the flowering time variation reported in a greenhouse in Denmark (Shah et al. 2020), my data showed a more remarkable difference in flowering time among accessions and a clear correlation between flowering time and latitude.

GWA analyses for flowering time and environmental factors

GWA analyses for flowering time variation showed several peaks in Manhattan plots for all of variable forms. Several peaks were located in the same regions among these analyses (Fig. 3-4) and the result for not-transformed flowering time data had a largest number of detected SNPs. Fig. 3-7 shows the result for not-transformed data of 124,921 SNPs (MAF $\geq$ 0.1) and there were 10 SNPs beyond the threshold in the result. These regions were also detected or showed peaks in the results for transformed variables (Fig. 3-4A, C, D). Five, two, and one out of the 10 SNPs were located on chromosome 2, 4, and 6, respectively and the same regions had peaks in the Manhattan-plots for variables in sqrt and log conspicuously. Regarding the rest two SNPs on chromosome 0, the same region was also detected in results for sqrt and log transformed variables (Fig. 3-4A, C) and had a peak in the result for boxcox transformed data (Fig. 3-4D). Based on linkage disequilibrium of 10 kbp, 28 genes were found around the 10 SNPs; 12 were NULL or repeat genes that serve no function, and 16 protein coding genes were identified by BLAST (Table 3-3).

Next, GWA analyses for the two environmental factors, the annual average temperature and integrated annual sunshine duration, were conducted with 119 accessions. Fig. 3-8 shows the Manhattan-plots for 128,736 SNPs (MAF $\geq$ 0.1) in the results for not-transformed variables. Each manhattan-plot shows several peaks, however, there were few SNPs beyond Bonferroni-corrected 5 % significant threshold. For both of the variables, there were not big differences among these results with the data in different transformations (Fig. 3-9, 3-10). Focusing on each top 100 SNPs, 270 and 373 genes were located in linked regions for the annual average temperature and integrated annual sunshine duration, respectively. 270 genes for the annual average temperature contained 157 protein coding genes and 113 NULL or repeat genes. While, for integrated annual sunshine duration, 216 out of 373 genes were protein coding and 157 genes were NULL or repeat. As with the candidates for flowering time variation, these protein coding genes were identified by BLAST.

Gene annotation of candidate genes

Gene annotation with peptide sequences was conducted for protein coding candidate genes that were associated with flowering time and two environmental factors, respectively. Table 3-3 shows the result for candidate genes associated

Table 3-2 A priori of known flowering time genes listed based on known genes in *A . thaliana* and *G . max* .

Flowering time gene name	*Arabidopsis* ID	Corresponding Lj ID
ABF1	AT1G49720	Lj4g3v0619920.1
ABF2	AT1G45249	Lj1g3v1650270.1
ABF3	AT4G34000	Lj4g3v0619920.1
ABF4	AT3G19290	Lj4g3v0619920.1
ABI5	AT2G36270	Lj5g3v0769560.1
AGL17 (AGAMOUS-LIKE 17)	AT2G22630	Lj1g3v0175780.1
AGL19 (AGAMOUS-LIKE 19)	AT4G22950	Lj4g3v1327620.2
AKT2/3	AT4G25760	Lj6g3v2274490.1
AP1 (APETALA1)	AT1G69120	Lj0g3v0012409.1
AP2	AT4G36920	Lj0g3v0092269.2
APRR5 (PRR5)	AT5G24310	Lj1g3v1076690.1
APRR9	AT2G46790	Lj1g3v1076690.1
AT1G51120	AT1G50680	Lj5g3v2300330.1
AT2G25920	AT2G25920	Lj5g3v1601830.1
AT5G27220	AT5G27220	Lj3g3v3639220.2
AT5G62040	AT5G62040	Lj0g3v0236979.1
ATC	AT2G27550	Lj3g3v1169160.1 Lj1g3v4716210.1
ATH1 (ARABIDOPSIS THALIANA HOMEOBOX 1)	AT4G32980	Lj0g3v0086419.1
ATSWC6	AT5G37055	Lj0g3v0346189.1
ATX1 (ARABIDOPSIS TRITHORAX 1)	AT2G31650	Lj1g3v3139990.1
BRI1 (BRASSINOSTEROID INSENSITIVE 1)	AT4G39400	Lj1g3v1222370.1
CDF2 (CYCLING DOF FACTOR 2)	AT5G39660	Lj1g3v2975990.1
CDF3 (CYCLING DOF FACTOR 3)	AT3G47500	Lj1g3v0513130.1
CKA1	AT5G67460	Lj2g3v1963030.2
CKA2	AT3G44180	Lj2g3v1963030.2
CKA3	AT2G23080	Lj2g3v1963030.2
CKA4	AT2G23070	Lj4g3v0768420.1
CLF (CURLY LEAF)	AT2G23380	Lj2g3v1890940.2
CO (CONSTANS)	AT5G15840	Lj1g3v3102660.1
COL4	AT5G24830	Lj4g3v2046140.1
COL5 (CONSTANS LIKE 5)	AT5G57660	Lj4g3v2046140.1
COP1 (CONSTITUTIVE PHOTOMORPHOGENIC 1)	AT2G32950	Lj0g3v0114209.2
CPS1/GA1 (ENT-COPALYL DIPHOSPHATE SYNTHETASE 1)	AT4G02780	Lj1g3v4434270.1
CRY1 (CRYPTOCHROME 1)	AT4G08920	Lj0g3v0286829.3
CRY2 (CRYPTOCHROME 2)	AT1G04400	Lj5g3v1203330.1
Cstf64 ((CLEAVAGE STIMULATION FACTOR 64)	AT1G71800	Lj5g3v1888380.1
Cstf77 (CLEAVAGE STIMULATION FACTOR 77)	AT1G17760	Lj0g3v0121319.1
CUL4 (CULLIN 4)	AT5G46210	Lj6g3v1946230.1
E1	GLYMA_06G207800	Lj5g3v2221340.1
E12A11	AT1G18100	Lj4g3v3114560.1 Lj2g3v0690130.1
EFS (EARLY FLOWERING IN SHORT DAYS) also called SDG8 (SET DOMAIN GROUP 8)	AT1G77300	Lj1g3v0887590.1
ELF3 (EARLY FLOWERING 3)	AT2G25930	Lj1g3v1785530.1
ELF6 (EARLY FLOWERING 6)	AT5G04240	Lj0g3v0106299.1
ELF7 (EARLY FLOWERING 7)	AT1G79730	Lj3g3v2318150.2
EMF2 (EMBRYONIC FLOWER 2)	AT5G51230	Lj2g3v1989150.2
ESD1 (EARLY IN SHORT DAYS 1) also called ACTIN RELATED PROTEIN 6 (ARP6)	At3g33520	Lj0g3v0068039.1
ESD4 (EARLY IN SHORT DAYS 4)	AT4G15880	Lj6g3v1449680.1
FCA (FLOWERING TIME CONTROL PROTEIN ALPHA)	AT4G16280	Lj4g3v2263920.1
FIE1	AT3G20740	Lj3g3v1543510.1

Table 3-2. Continued.

FIO1 (FIONA1)	AT2G21070	Lj4g3v0217740.1
FKF1 (FLAVIN BINDING, KELCH REPEAT, F-BOX 1)	AT1G68050	Lj4g3v3113210.1
FLD (FLOWERING LOCUS D)	AT3G10390	Lj0g3v0244869.1
FLK (FLOWERING LATE KH MOTIF)	AT3G04610	Lj1g3v4997240.1
FPA	AT2G43410	Lj3g3v3315960.1
FRI (FRIGIDA)	AT4G00650	Lj1g3v1386480.1
FT (FLOWERING LOCUS T)	AT1G65480	Lj1g3v3944940.1
FTIP1	GLYMA_11G024300	Lj2g3v2017630.1
FUL (FRUITFUL)	AT5G60910	Lj0g3v0012409.1
FVE	AT2G19520	Lj3g3v0309020.5
FY	AT5G13480	Lj3g3v1876080.1
GA2/KS (GA REQUIRING 2/ENT-KAURENE SYNTHASE)	AT1G79460	Lj3g3v0948400.1
GA20ox2 (GIBBERELLIN 20-OXIDASE 2)	AT5G51810	Lj0g3v0071479.1
GA2ox1 (GIBBERELLIN 2-OXIDASE 1)	AT1G78440	Lj6g3v1945170.1
GA2ox2 (GIBBERELLIN 2-OXIDASE 2)	AT1G30040	Lj6g3v1945170.1
GA2ox3 (GIBBERELLIN 2-OXIDASE 3)	AT2G34555	Lj6g3v1945170.1
GA2ox4 (GIBBERELLIN 2-OXIDASE 4)	AT1G47990	Lj0g3v0322199.1
GA2ox6 (GIBBERELLIN 2-OXIDASE 6)	AT1G02400	Lj4g3v2253520.1
GA2ox7 (GIBBERELLIN 2-OXIDASE 7)	AT1G50960	Lj0g3v0081939.2
GA3ox1 (GIBBERELLIN 3-OXIDASE 1)	AT1G15550	Lj3g3v2247040.1
GA3ox2 (GIBBERELLIN 3-OXIDASE 2)	AT1G80340	Lj3g3v2247040.1
GAI (GIBBERELLIC ACID INSENSITIVE)	AT1G14920	Lj6g3v0433880.1
GI (GIGANTEA)	AT1G22770	Lj5g3v1914110.1
GID1A (GA INSENSITIVE DWARF1A)	AT3G05120	Lj0g3v0154849.1
GID1B (GA INSENSITIVE DWARF1B)	AT3G63010	Lj1g3v1686830.1
GID1C (GA INSENSITIVE DWARF1C)	AT5G27320	Lj0g3v0154849.1
GRF1	AT2G22840	Lj5g3v0324520.1
GRF12	AT1G26480	Lj2g3v1252820.1
GRF2	AT4G37740	Lj5g3v0324520.1
GRF4	AT3G52910	Lj1g3v4699740.1
GRF5	AT3G13960	Lj3g3v2591390.1
HAP5A	AT3G48590	Lj1g3v1502590.1
HAP5B	AC009894.2:4058..4471	Lj1g3v1502590.1
HUA2 (HUA is the full name, means flower in Chinese)	AT5G23150	Lj3g3v3513720.2
HUB2 (HISTONE MONOUBIQUITINATION 2)	AT1G55250	Lj2g3v2876160.2
LD (LUMINIDEPENDENS)	AT4G02560	Lj1g3v4915280.2
LHY (LATE ELONGATED HYPOCOTYL)	AT1G01060	Lj3g3v2518940.1
LIP1 (LIGHT INSENSITIVE PERIOD 1)	AT5G64813	Lj4g3v2787000.1
LKP2 (LOV KELCH PROTEIN 2)	AT2G18915	Lj6g3v1368660.3
LUX/PCL1 (LUX ARRHYTHMO/PHYTOCLOCK 1)	AT3G46640	Lj3g3v3235380.1
LWD1/ATAN11 (LIGHT-REGULATED WD 1/ANTHOCYANIN11)	AT1G12910	Lj1g3v3943610.1
LWD2 (LIGHT-REGULATED WD 2)	AT3G26640	Lj1g3v3943610.1
MSI1	AT5G58230	Lj4g3v0244120.1
MYB33 (MYB DOMAIN PROTEIN 33)	AT5G06100	Lj3g3v1855350.1
MYB65	AT3G11440	Lj3g3v1855350.1
NFYA1/HAP2A (NUCLEAR FACTOR Y, SUBUNIT A1/HEME ACTIVATED PROTEIN 2A)	AT5G12840	Lj6g3v0647470.1
NFYB3/HAP3C (NUCLEAR FACTOR Y, SUBUNIT B3/HEME ACTIVATED PROTEIN 3C)	AT4G14540	Lj6g3v1720580.3
NFYC1/HAP5A (NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT C-1/HEME ACTIVATEI	AT3G48590	Lj1g3v1502590.1
NFYC3/HAP5C (NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT C-3/HEME ACTIVATEI	AT1G08970	Lj1g3v4931770.1
NFYC4	AT5G63470	Lj1g3v1502590.1
NFYC9 (NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT C-9)	AT1G54830	Lj1g3v4931770.1
PAF1	AT5G40700	Lj5g3v0108300.1

Table 3-2. Continued.

PAF2	AT1G41200	Lj5g3v0108300.1
PEP	AT4G30000	Lj4g3v2139160.1
PFT1 (PHYTOCHROME AND FLOWERING TIME 1)	AT1G25540	Lj2g3v1168790.3
PHYA (PHYTOCHROME A)	AT1G09570	Lj5g3v1533480.1
PHYB (PHYTOCHROME B)	AT2G18790	Lj6g3v1915810.1
PHYC (PHYTOCHROME C)	AT5G35840	Lj6g3v1915810.1
PHYD (PHYTOCHROME D)	AT4G16250	Lj6g3v1915810.1
PHYE (PHYTOCHROME E)	AT4G18130	Lj6g3v1093490.1
PI	AT5G20240	Lj1g3v0886560.1
PIE1 (PHOTOPERIOD INDEPENDENT EARLY FLOWERING 1)	AT3G12810	Lj5g3v0921010.1
PIF3 (PHYTOCHROME INTERACTING FACTOR 3)	AT1G09530	Lj5g3v1533330.1
PLDalpha1	AT3G15730	Lj1g3v1911660.1
PNF (POUND-FOOLISH)	AT2G27990	Lj3g3v3082350.1
PNY (PENNYWISE)	AT5G02030	Lj1g3v3443690.1
PRR1/TOC1	AT5G61380	Lj4g3v1658890.2
PRR5 (PSEUDO-RESPONSE REGULATOR 5)	AT5G24470	Lj1g3v1076690.1
PRR7 (PSEUDO-RESPONSE REGULATOR 7)	AT5G02810	Lj5g3v0615690.2
PRR9 (PSEUDO-RESPONSE REGULATOR 9)	AT2G46790	Lj1g3v1076690.1
REF6 (RELATIVE OF EARLY FLOWERING 6)	AT3G48430	Lj0g3v0351959.1
RFI2 (RED AND FAR-RED INSENSITIVE 2)	AT2G47700	Lj5g3v1498020.1
RGA (REPRESSOR OF GA1-3 1)	AT2G01570	Lj6g3v0433880.1
SEF (SERRATED LEAVES AND EARLY FLOWERING)	AT5G37055	Lj0g3v0346189.1
SEP3 (SEPALLATA 3)	AT1G24260	Lj4g3v2573630.2
SOC1 (SUPPRESSOR OF OVEREXPRESSION OF CONSTANS)	AT2G45660	Lj1g3v2608040.2
SPA1 (SUPPRESSOR OF PHYA-105)	AT2G46340	Lj3g3v2414270.2
SPA2 (SPA1-RELATED 2)	AT4G11110	Lj2g3v2088010.1
SPY	AT3G11470	Lj5g3v0803050.1
STM	AT1G62360	Lj6g3v1812030.1
SUMO1-2	AT2G45695	Lj1g3v3150030.1
SUVR2	GLYMA_19G167900	Lj1g3v4516130.1
SVP (SHORT VEGETATIVE PHASE)	AT2G22540	Lj2g3v1450690.1
TEM1 (TEMPRANILLO 1)	AT1G25560	Lj2g3v1171110.1
TEM2 (TEMPRANILLO 2)	AT1G68840	Lj2g3v1171110.1
TFL1 (TERMINAL FLOWER 1)	AT5G03840	Lj1g3v4716210.1
TIC (TIME FOR COFFEE)	AT3G22380	Lj5g3v2056280.1
TOE1 (TARGET OF EAT 1)	AT2G28550	Lj3g3v3107620.3
TOE2 (TARGET OF EAT 2)	AT5G60120	Lj0g3v0037529.1
TOE3	AT5G67180	Lj0g3v0092269.2
TSF (TWIN SISTER OF FT)	AT4G20370	Lj1g3v3944940.1
UBC1 (UBIQUITIN CONJUGATING ENZYME 1)	AT1G14400	Lj1g3v0465660.1
UBC2 (UBIQUITIN CONJUGATING ENZYME 2)	AT2G02760	Lj1g3v0465660.1
UBP26 (UBIQUITIN-SPECIFIC PROTEASE 26)	AT3G49600	Lj5g3v0574430.1
UFO	AT1G30950	Lj4g3v0189930.1
VEL1	AT4G30200	Lj4g3v2120360.1
VEL2	AT2G18880	Lj4g3v2120360.1
VIL1 (VERNALIZATION INSENSTIVE 3 LIKE 1) also called VRN5 (VERNALIZATION 5)	At3g24440	Lj4g3v3093980.2
VIN3 (VERNALIZATION INSENSITIVE 3)	AT5G57380	Lj4g3v2120360.1
VIP4 (VERNALIZATION INDEPENDENCE 4)	AT5G61150	Lj0g3v0107339.1
VIP5 (VERNALIZATION INDEPENDENCE 5)	AT1G61040	Lj4g3v0450520.1
VIP6 (VERNALIZATION INDEPENDENCE 6) also called EARLY FLOWERING 8 (ELF8)	AT2G06210	Lj6g3v1201200.1
VRN1 (VERNALIZATION 1)	AT3G18990	Lj3g3v3336240.2
VRN2 (VERNALIZATION 2)	AT4G16845	Lj2g3v1989150.2
WNK1 ((WITH NO LYSINE (K) 1))	AT3G04910	Lj5g3v1415400.1
ZTL (ZEITLUPE)	AT5G57360	Lj6g3v1368660.3

with flowering time variation. Although candidate genes for flowering time variation had several orthologues of unknown genes or genes with no hit results, there were also two orthologues of known flowering time genes (Table 3-3). The known flowering time genes in the results coded for the following protein sequences: Small nuclear ribonucleoprotein -like (SmD3b-like; Swaraz et al. 2011), and potassium channel AKT2/3 (AKT2/3; Held et al. 2011). Each known flowering time gene showed altered flowering time phenotypes in their mutant lines in each previous study (Swaraz et al. 2011 and Held et al. 2011, respectively). In addition, although there were no references which showed direct evidence, candidate genes contained another one orthologue that might be related to flowering time regulation: *importin-4* (Kevei et al. 2007). Besides these genes, *Mediator of RNA polymerase II transcription subunit 28* (*MED28*) was one of the gene family which contained several flowering time-related genes (e.g. *MED8*, *MED12*, *MED13*, *MED15*, *MED16*, *MED17*, *MED18*, *MED20a*, and *MED25*; Buendia-Monreal & Gillmor, 2016; Yao et al. 2019). These mediator subunit gene mutants showed a late flowering phenotype. All of detected protein coding genes on chromosome 4, including the orthologues of *SmD3b-like* and *MED28*, were located in the 20 Kbp region around the SNPs of 37,258,719 bp and 37,258,674 bp which were most strongly associated with the flowering time variation. Especially for *MED28*, both of the SNPs were located on its gene region. *AKT2/3* and *importin-4* were located in a detected region around the SNP of 26,476,140 bp on chromosome 6. For ten out of the 16 protein coding genes, the gene expressions had been observed by RNA-sequencing in two experimental strains, Gifu B-129 and MG-20, and produced hits on a gene expression database (Table 3-3; Verdier et al. 2013; https://lotus.au.dk/).

While there were no orthologues of known flowering time genes in the candidates for annual average temperature, there were two orthologues of known flowering time genes in the detected genes for integrated annual sunshine duration; *Gibberellin 2-beta-dioxygenase 1*(*GA2ox1*; Rieu et al. 2008), Lj3g3v0421660, and *Target of EGR1 protein 1*, Lj3g3v3107620 (*TOE1*; e.g. Aukerman and Sakai, 2003). These orthologues were in regions around the SNPs of 3,655,438 bp and 38,916,742 bp on chromosome 3 and its gene expressions have been also observed in Verdier et al. 2013.

Table 3-3 The result of gene annotation for the candidate genes on the basis of the GWA analysis for flowering time.

Linkage group	Detected SNP	Gene ID	Description	Length	Identity (%)	Reference
Chr0	189095050	Lj0g3v0359209	Spastin [Glycine soja]	425	92.8	
Chr2	8439786	Lj2g3v0621090	YTH domain-containing protein 1-like isoform X1 [Abrus precatorius]	305	66.9	
Chr2	8482295	Lj2g3v0621130	No hit			
Chr2		Lj2g3v0621140	No hit			
Chr2	8569194	Lj2g3v0621330	No hit			
Chr2	8569194, 8580319	Lj2g3v0621350	kinase-interacting family protein-like isoform X2 [Abrus precatorius]	329	81.4	
Chr2		Lj2g3v0621360	No hit			
Chr4	37258674, 37258719	Lj4g3v2785700	PREDICTED: small nuclear ribonucleoprotein SmD3b-like [Lupinus angustifolius]	131	96.2	Swaraz et al. 2011
Chr4		Lj4g3v2785710	UDP-rhamnose/UDP-galactose transporter 4 [Glycine max]	336	92.4	
Chr4		Lj4g3v2785730	mediator of RNA polymerase II transcription subunit 28 [Abrus precatorius]	141	90.5	
Chr4		Lj4g3v2785740	Alkaline/neutral invertase E, chloroplastic, partial [Mucuna pruriens]	658	86.5	
Chr6	26476140	Lj6g3v2274460	importin-4 isoform X1 [Abrus precatorius]	1,048	93.4	Kevei et al., 2007*
Chr6		Lj6g3v2274470	ceramide metabolic process [Spatholobus suberectus]	288	86.5	
Chr6		Lj6g3v2274490	potassium channel AKT2/3 isoform X1 [Abrus precatorius]	824	85.3	Held et al., 2011
Chr6		Lj6g3v2274500	BYPASS-related protein [Spatholobus suberectus]	314	88.9	
Chr6		Lj6g3v2274510	BYPASS-related protein [Spatholobus suberectus]	314	88.0	

The E-values for all of these results were under 1.0e-4.

* Candidates that had no direct references but might have a possibility to be related to flowering time.

Table 3-4 SNPs that have higher -$\log_{10}$(p-values), ≧ 5, and adjusted R^2 values of models made with the ranked factors.

Rank	-$\log_{10}$(P-value) for GWA	Linkage group	SNP position (bp)	*R^2
1	7.24	4	37258719	0.391
2	7.24	2	8439786	0.567
3	6.41	6	26476140	0.573
4	6.12	5	2772060	0.608
5	5.95	5	4184207	0.663
6	5.61	5	2941168	0.665
7	5.60	5	448242	0.669
8	5.43	4	37118847	0.672
9	5.39	5	2666733	0.677
10	5.26	1	12947890	0.677

*The R^2 values of models made with the factors added to the order.

Prediction of flowering time with SNPs that had lower *P*-values in genome-wide association analysis for flowering time

The ten selected SNPs are shown in Table 3-4. With the most strongly detected SNP, 37,258,719 bp on chromosome 4, the R^2-value was 0.391. When the top two SNPs were added, the value reached 0.567 (Fig. 3-5 and Table 3-4). In both cases, all added SNPs were significantly selected explanatory variables for flowering time variation ($P < 0.001$). Fig. 3-5B shows the plots for flowering time and predicted flowering time with the top one and two genotypes, respectively. Predicted flowering time was calculated by each estimated formula shown in Fig. 3-5B, respectively. Further, R^2-value reached around 0.66 with the top five SNPs (Fig. 3-5A) and the values levelled off at 0.66-0.67 when more than five SNPs were used. With more than five SNPs, 1st, 2nd, and 5th SNPs were selected significantly as explanatory variables. Genotype distributions for each of 1st and 2nd SNPs are shown in Fig. 3-6B, C and the accumulation of both of the top two SNPs is shown in Fig. 3-6A; they indicate that there were geographic tendencies in each SNP distribution. In addition, there were four and one protein coding genes around these SNPs, respectively (Table 3-3), and one of the genes around 1st SNP was *SmD3b-like* orthologue which is one of known flowering time-related genes.

Discussion

To introduce genome-wide methods that utilise next generation sequencing, candidate genes for various phenotypes have been identified in several model species and cultivars (Atwell et al. 2010; Zhang et al. 2015; Zhao et al. 2011); number of these types of research could further increase in the next few years. The development of sequencing technologies has enabled construction of *de novo* genome sequences, and these technologies have been used to detect candidate genes associated with local adaptation in wild plants (Shah et al. 2020; Burgarella et al. 2016; Hancock et al. 2011; Sasaki et al. 2015). Although these types of studies are mostly conducted with cultivars or typical model plant species, it is necessary to determine genetic factors in a variety of species to understand the mechanisms that underlie local adaptation and the development of biodiversity. As one of the pioneering studies, my study could contribute to future studies of the identification of genes that regulate legume flowering time.

Flowering time measurements revealed extreme differences in *L. japonicus* flowering time. Compared to the results in Shah et al. (2020), the flowering time variation in my greenhouse data showed different tendencies. Especially, the variation measured in this study had a larger variance among accessions than that in greenhouse data at Denmark; for example, the differences in averages between the earliest and latest flowering accessions were 146.67 days and 45.5 days in the data shown in this study and measured in Denmark, respectively. Considering that *L. japonicus* is a long-day plant, the variance difference suggested that the extremely long-day condition in Denmark uniformly promoted flowering in this species, even in late flowering accessions under shorter day lengths, and the flowering time and latitude were not correlated with the flowering variation in Denmark. Regarding the field flowering time data in Miyagi, although there was a correlation between the flowering time and latitude, the difference in average flowering time between the earliest and latest flowering accessions was 52.4 days; this is much shorter than that shown in this study.

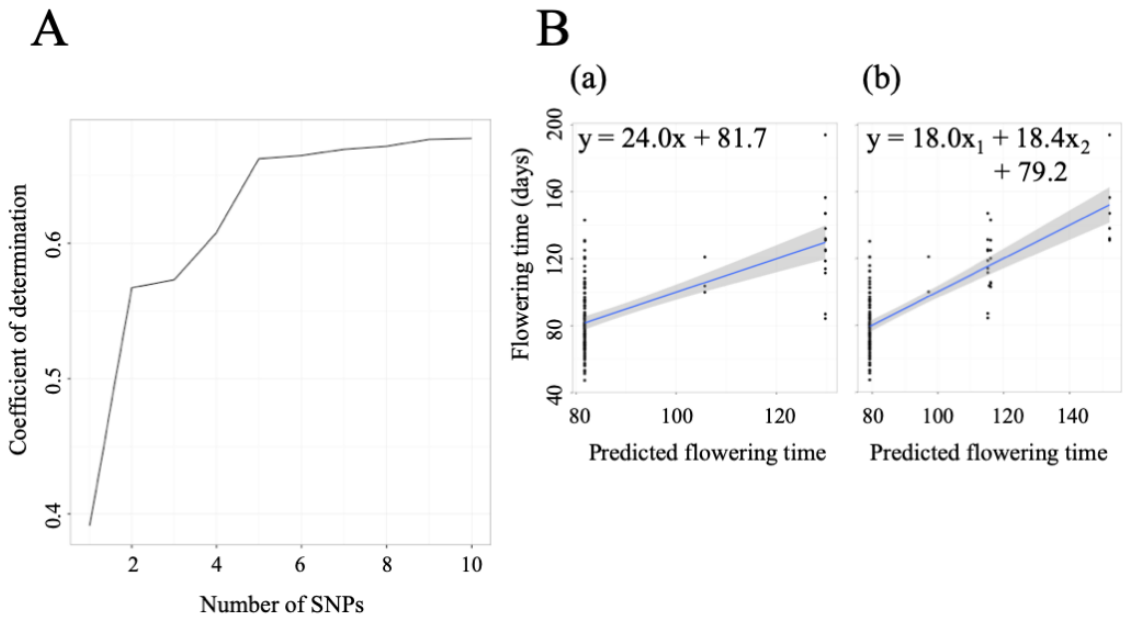

Fig. 3-5. Correlation between flowering time variation and combination of detected *SNPs* by genome-wide association analysis for flowering time

(A) The vertical axis and the horizontal axis show the coefficient of determination and the number of single nucleotide polymorphisms (*SNPs*) contained in the *SNP* combinations, respectively. (B) The vertical axis and the horizontal axis show flowering time (days) and predicted flowering time for cases with the top one and two genotypes in (a) and (b), respectively. Plots denote flowering time and predicted flowering time for each accession. predicted flowering time was calculated by the estimated formula shown in each figure.

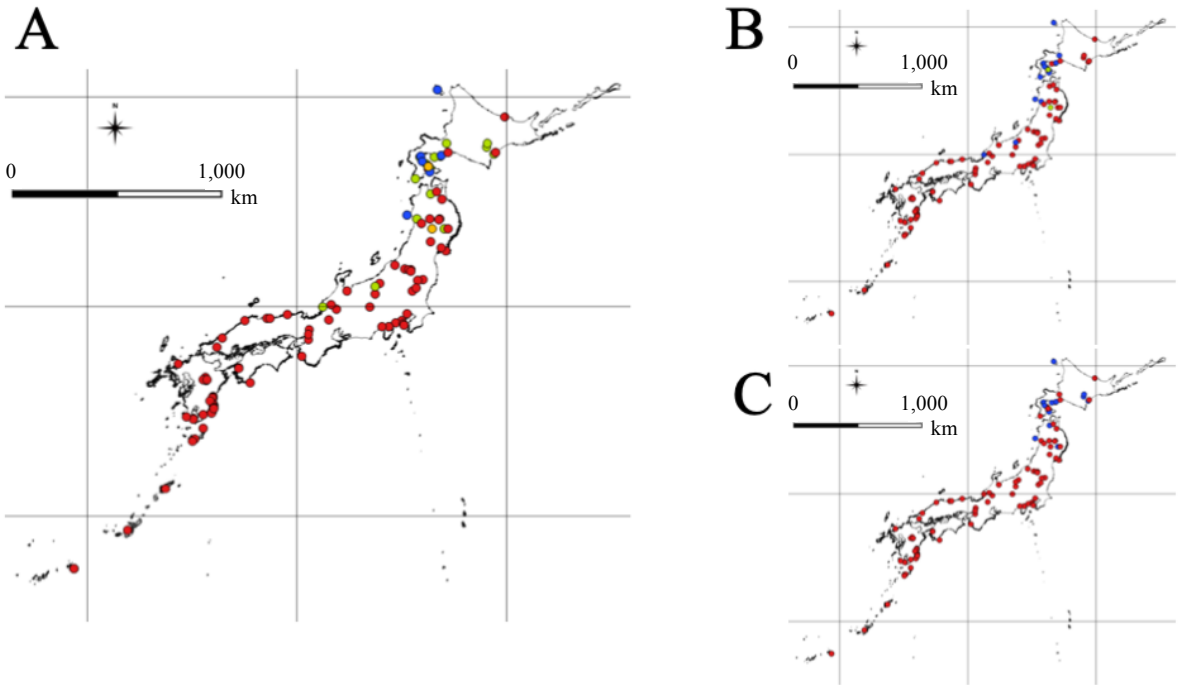

Fig. 3-6. Genotype distribution for selected *SNPs* in correlation test with flowering time variation

Each color of circles denotes the locations of accessions with genotypes of single nucleotide polymorphisms (*SNPs*) selected by correlation with flowering time variation. The genotype distribution in which top two SNPs were taken into account is shown in A. Red and blue circles show early- and late-flowering genotype, respectively. Green circles show accessions in which one of two *SNPs* fixed in the early-flowering and the other had the late-flowering genotype. Yellow circles show accessions in which one of two *SNPs* had the hetero and the other had the early-flowering genotype. The genotype distributions of the top one and two SNPs, Chr4 37258719 and Chr2 8439786, are shown in B and C, respectively. Red and blue circles show early- and late-flowering genotypes, respectively. Green circles show hetero genotypes.

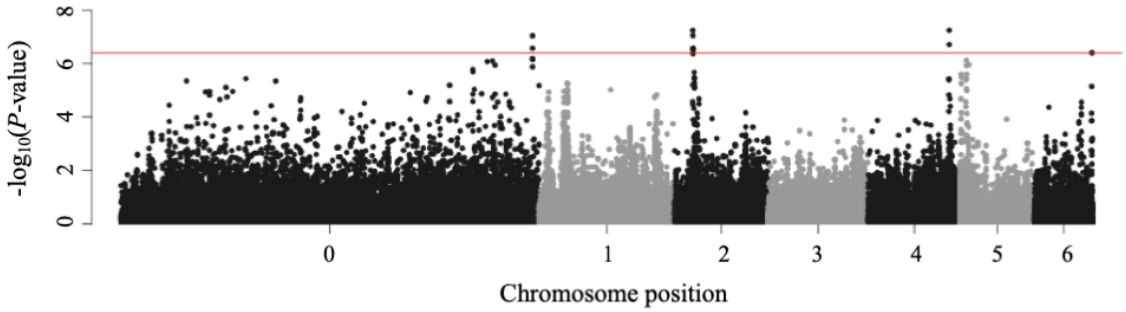

Fig. 3-7. Manhattan plots of genome-wide association analyses for flowering time variation

The vertical axis and the horizontal axis show $-\log_{10}$(*P*-values) and chromosome positions, respectively. The position and $-\log_{10}$(*P*-values) for each single nucleotide polymorphisms (*SNPs*) were plotted. Red line indicates Bonferroni-corrected 5 % significant threshold.

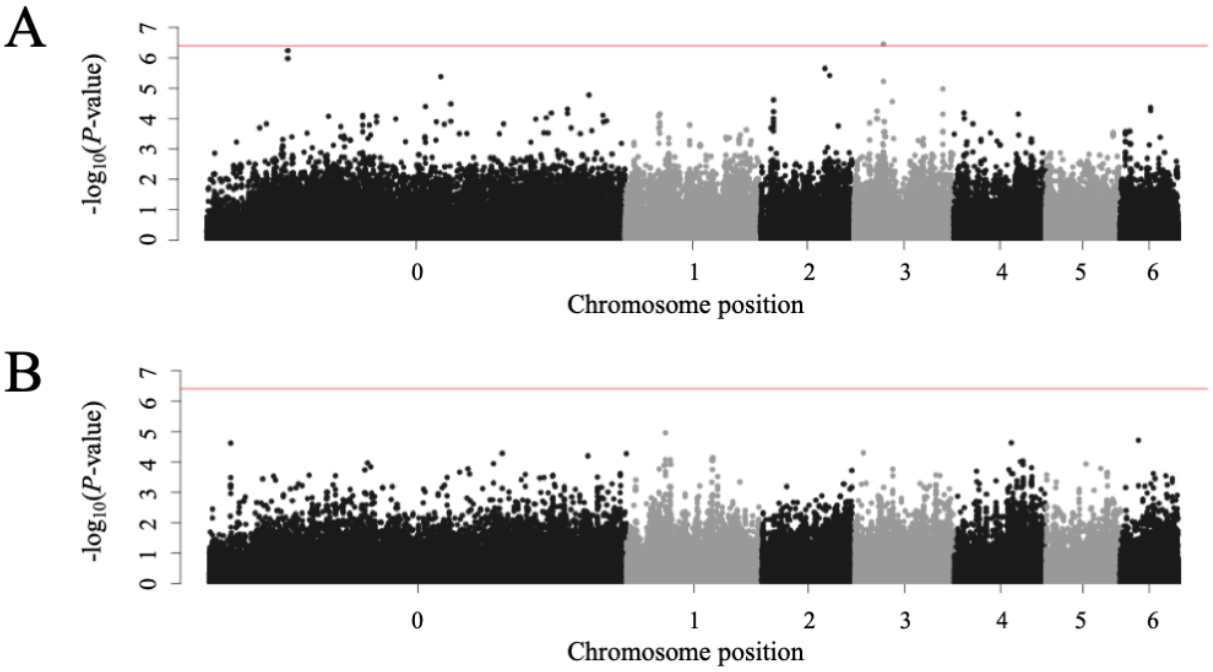

Fig. 3-8. Manhattan plots of genome-wide association analyses for two environmental factors

Manhattan plots for annual average temperature and integrated annual sunshine duration are shown in A and B, respectively. The vertical axis and the horizontal axis show $-\log_{10}$(*P*-values) and chromosome positions, respectively. Plots show the position and $-\log_{10}$(*P*-values) for each single nucleotide polymorphisms (*SNPs*). Red line indicates Bonferroni-corrected 5 % significant threshold.

Moreover, the flowering time differences among accessions were emphasised in my data, and clearly showed the correlation with latitude. This suggested that day length was one of the most important signals for *L. japonicus* flowering time and contributed to the flowering time variation being correlated with latitude. The broad sense heritability for flowering time variation in this study reached 0.89 for all flowered individuals in 120 out of 132 accessions, as calculated by analysis of variance; this variation was considered to be largely due to genetic factors. Besides large geographic distances between wild populations, and taking into account the large variation in flowering time and the occurrence of non-flowering accessions for more than 200 days, it is possible that crossing opportunities may be lost between some ecotypes of this species, even if they have some opportunities of secondary contact. The correlation between flowering time and latitude suggested that flowering time variation may be associated with environmental factors that varies by latitude, such as temperature and day length, and that accessions can adapt to each habitat through flowering time. As previous studies have revealed, flowering time is controlled by several external factors, such as day length and temperature (Andrés and Coupland 2012; Song et

al. 2013). Therefore, although I mentioned that day length would be an important factor controlling flowering time above, the differences in flowering time might be reflected by complex combinations of these environmental factors. Due to the complex nature of flowering time, genetic factors that induce flowering time variation would contain genetic heterogeneities or allelic heterogeneities. In fact, the flowering time variation showed continuous values, which suggested that it is contributed by several factors. The result of GWA analyses for flowering time variation showed several peaks, which also implies that there may be genetic or allelic heterogeneities in the flowering time pathway in this species.

GWA analyses for flowering time could detect candidate genes that contain orthologues of known flowering time genes. Compared to GWA analysis results for several flowering time variations in Shah et al. (2020), my GWA analysis detected different candidate genes. This might have been caused by the differences in flowering time tendencies and accessions used. Phenotypes tend to be correlated with genetic structures such as flowering time, and analysis by GWA is likely to detect false positives. Therefore, in this study, I adopted a model that takes population cluster and structure into account. There were several peaks in the Manhattan plots and most of them were common among the data forms (Fig. 3-4). 16 protein coding genes were detected with the not-transformed variable on the basis of Bonferroni-corrected 5 % significant threshold and two of them were orthologues of known flowering time genes (Fig. 3-7 and Table 3-3). The detected orthologues of known flowering time genes were genes that coded SmD3b-like and AKT2/3 (Held et al. 2011); these genes have been shown to have functions associated with flowering time phenotypes in other species. Besides these two genes, there were two genes which might be related to flowering time control considering its or paralogue's functions. One of the genes, Lj6g3v2274460.1 is an orthologue of *Importin-4* (Table 3-3); this gene is known to be involved tin targeting photoreceptors (Kevei et al. 2007) and may regulate flowering time as well. The other is Lj4g3v2785730, an orthologue of *MED28*; this gene coded one of mediator subunit and several mediator subunit mutants of *Arabidopsis thaliana* showed a late flowering phenotype (Buendia-Monreal & Gillmor, 2016; Yao et al. 2019). Candidate genes, especially the above two gene and two orthologues of known flowering time genes, could contribute to the regulation of flowering time in this species and may even promote intraspecific differentiation and local adaptation via changes in flowering time. In future studies, I am planning to conduct several experiments with mutant lines or transformation lines to reveal the relationships between these genes and flowering time control in this species.

GWA analyses for two environmental factors showed that few SNPs strongly associated with each variable with all four data forms (Fig. 3-9, 3-10). Although several peaks were shown in Manhattan plots for annual average temperature, there were no orthologues of known flowering time-related genes in the regions around top 100 SNPs. While, two orthologues of known flowering time-related genes were detected in top 100 SNPs for integrated annual sunshine duration. Although these top 100 SNP sets did not include the detected SNPs in the analysis for flowering time variation, two detected orthologues of known flowering time genes might contribute for the variation partially, because it is considerable that day length is one of the most important external factors for flowering time control in this species. Moreover, these two orthologues may play roles of genetic factors for local adaptation via flowering time variation.

In *A. thaliana*, *FRIGIDA* played a large role in latitudinal decline of flowering time (Stinchcombe et al. 2004). However, it is possible that not only one, but several or many genes contribute to flowering time variation in an additive manner in a single species. In addition, genetic factors for intraspecific flowering time variation should be species-specific. The assessment of correlations between flowering time variation and combinations of top detected SNPs in GWA analysis showed that adding the 1st to 2nd SNPs resulted in a large increase in the R^2-value (Fig. 3-5A). Especially the top two SNPs, 37,258,719 bp on chromosome 4 and 8,439,786 bp on chromosome 2, were selected as significant variables for flowering time variation. This result suggests that the variation in the flowering time of *L. japonicus* could be mainly explained by a small number of genetic factors. The genotype distributions of each SNP of the genetic factors were along the latitude (Fig. 3-6B, C). Furthermore, the genotype of a combination of two SNPs had a gradual distribution along the latitude (Fig. 3-6A). This implies that the clinal flowering time variation in this species might be

caused by a combination of multiple responsible genes. The genes around these two SNPs included an orthologue of known flowering time genes, *SmD3b-like* (Table 3-3). The gene expression was observed in a previous study by RNA-sequencing and produced hits on a gene expression database (Verdier et al. 2013; https://lotus.au.dk/); this implies that the gene have some sort of function. Although, on the protein coding genes located around these two SNPs, there was no non-synonymous changes listed in the SNP dataset used in this study, because the SNP dataset was used for markers to detect associated regions for variables, there is a possibility that this SNP set did not contain all of the mutations conserved among these accessions. That is to say, the observed gene expressions and my result suggested that some of these genes could be related to flowering time variation in this species. Moreover, flowering time variation may manifest due to the accumulation of small effects from many genetic factors. To understand the mechanisms of local adaptation and intraspecific differentiation in this species, it is also necessary to identify the genetic factors that induces small effects in future studies.

The general mechanisms that underlie local adaptation through flowering time is complicated, even with recent advancements in technologies, because of the complex genetic pathways and population structure. However, several previous studies and this study have offered several candidates that could contribute to the regulation of flowering time; all or some of these candidates may be associated with local adaptation in the wild population. Especially in widely distributed species such as *L. japonicus*, local adaptation processes are important for the emergence of biodiversity. In the case of *L. japonicus*, it is distributed across latitudes that span the entire country of Japan. Therefore, flowering time variation along the latitude may be important for local adaptation to each environment; the identification of genetic factors that affects flowering time variation in this species would contribute to my understanding of *L. japonicus* biodiversity in Japan. In addition, compared with other model plant species (e.g. *A. thaliana*), *L. japonicus* is distributed in different environmental ranges, and therefore, different factors that are unique to this species may be involved in flowering time variation. Revealing common and specific mechanisms for flowering time variation between plant species should elucidate the mechanisms of local adaptation.

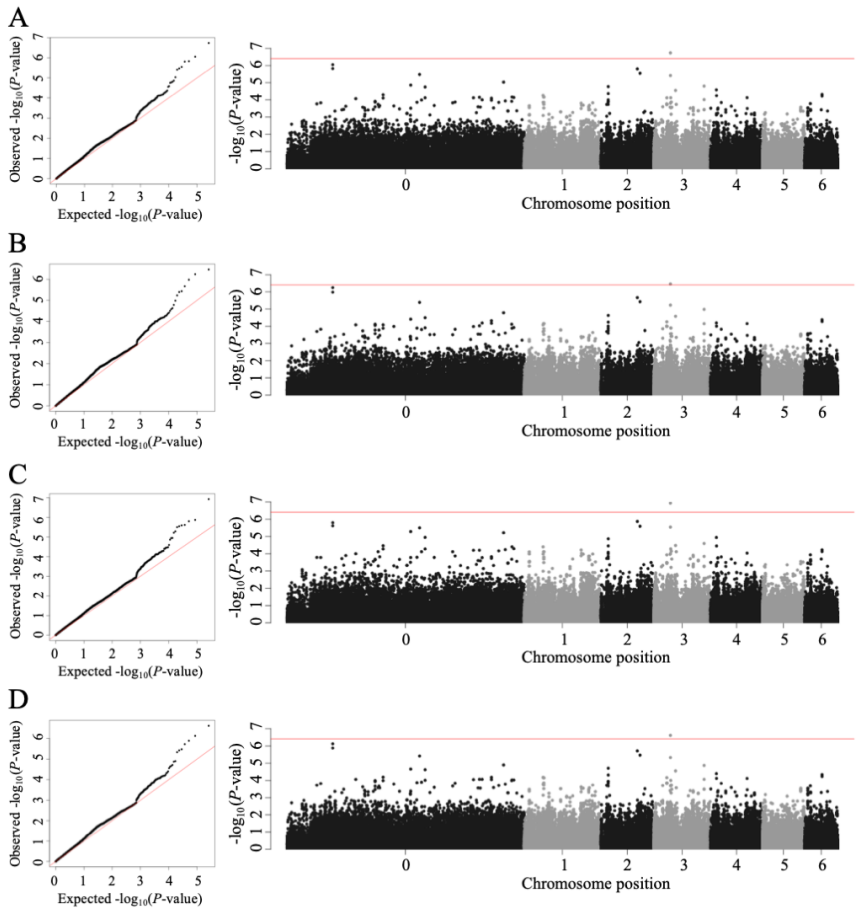

Fig. 3-9. Manhattan plots and *Q*–*Q* plots of genome-wide association analyses for annual average temperature

A–D show the plots for the results with sqrt, not-transformed, log, and boxcox flowering time data, respectively. The (quantile–quantile) *Q*–*Q* plots are shown in left side. The vertical axis and the horizontal axis show observed $-\log_{10}$(*P*-values) and expected $-\log_{10}$(*P*-values), respectively. Right side figures show each Manhattan plot. The vertical axis and the horizontal axis show $-\log_{10}$(*P*-values) and chromosome positions, respectively. The position and $-\log_{10}$(*P*-values) for each single nucleotide polymorphism were plotted. Red line indicates Bonferroni-corrected 5 % significant threshold.

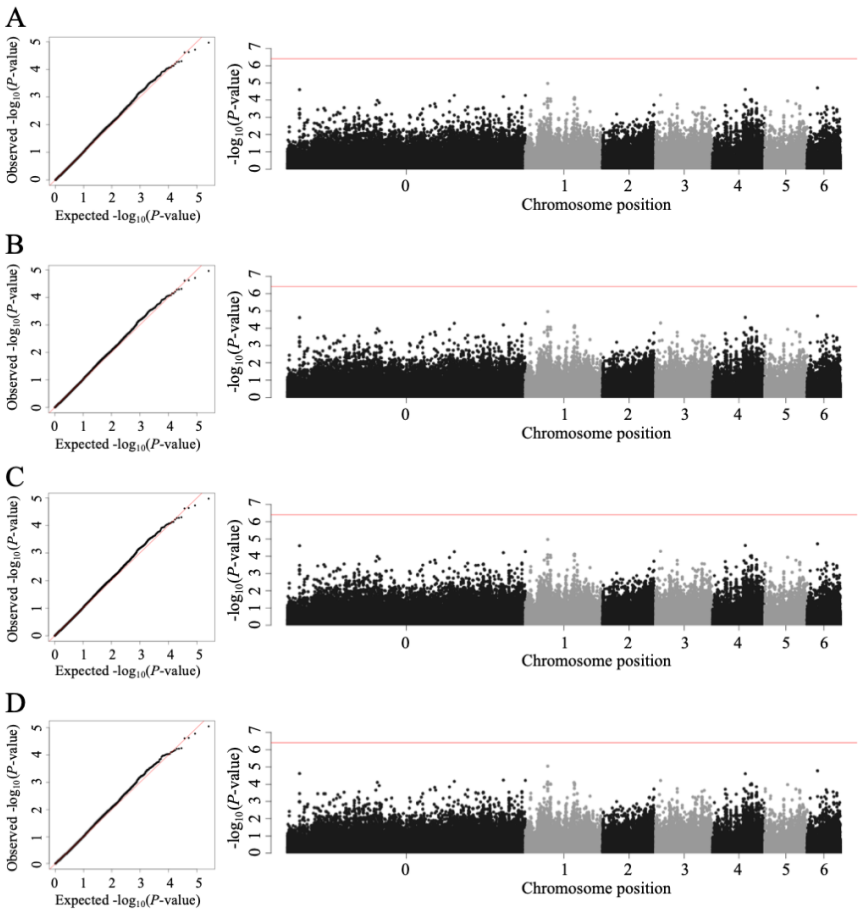

Fig. 3-10. Manhattan plots and *Q*–*Q* plots of genome-wide association analyses for integrated annual sunshine duration

A–D show the plots for the results with sqrt, not-transformed, log, and boxcox flowering time data, respectively. The (quantile–quantile) *Q*–*Q* plots are shown in left side. The vertical axis and the horizontal axis show observed $-\log_{10}$(*P*-values) and expected $-\log_{10}$(*P*-values), respectively. Right side figures show each Manhattan plot. The vertical axis and the horizontal axis show $-\log_{10}$(*P*-values) and chromosome positions, respectively. The position and $-\log_{10}$(*P*-values) for each single nucleotide polymorphism were plotted. Red line indicates Bonferroni-corrected 5 % significant threshold.

General Discussion

These studies focused on the demographic history and flowering-time variation of *L. japonicus* in Japan with the objective of understanding the mechanisms of its local adaptation. Many botanists have been concerned about the relationship between flowering time and reproductive fitness because of its critical effect on reproductive success. Recent developments in sequencing technology have enabled us to gather genome-wide information from many species. Based on the genome-wide sequencing data, genome-wide nucleotide polymorphisms can be identified; these help reveal details on the dynamics of the effective population size and the divergence process from past to present population structures (Li and Durbin, 2011; Durvasula *et al*., 2017; Fulgione *et al*., 2018). In addition to population dynamics, genome-wide nucleotide polymorphisms enable the detection of causal genes for various traits, including flowering time (e.g., Sasaki *et al*., 2015; Burgarella *et al*., 2016). In this study, I tried to reveal the demographic history of Japanese *L. japonicus* accessions and detect the genetic factors affected by natural selection that cause intraspecific flowering-time variation in this species. In pursuit of these aims, I employed several types of analyses, including molecular biological methods and analyses using genome-wide nucleotide polymorphisms.

First, population clustering analyses with genome-wide nucleotide polymorphisms were conducted to infer the divergence times of Japanese *L. japonicus* accessions. To analyse population clustering, fastSTRUCTURE analysis was conducted. The results of this analysis revealed that the distribution of Japanese *L. japonicus* accessions involved several clusters with geographical tendencies and that the accessions could be roughly divided into three subpopulations distributed in the southern, middle, and northern parts of the Japanese archipelago. The divergence time was inferred by PSMC analysis; the results showed that the major clusters overlapped with the subpopulations detected by the fastSTRUCTURE analysis. In addition, two accessions that had originated in Tsushima Island appeared to have recently diverged from other accessions, approximately 5,000 and 7,000 years ago, respectively. This period is within the postglacial climatic optimum when Tsushima Island was divided from the main land of Japan by sea. In addition, seeds of this species sink in liquid; therefore, it can be suggested that the Tsushima population diverged from other Japanese *L. japonicus* populations by geographical reproductive isolation. Based on these results, it could be suggested that this species could have spread from South to North Japan and its distribution could have expanded from the Kyushu area. Apart from these results, a consideration of the genetic distance analysis for the same accession sets suggests that Tsushima Island may be a point of origin for all Japanese *L. japonicus* accessions and that the species colonised the Japanese archipelago from there after the last ice age.

Next, the results of the molecular biological analyses using Tajima's D test (Tajima, 1989) and the McDonald-Kreitman test (McDonald and Kreitman, 1991) showed that the *LjE1* gene was affected by positive natural selection; the sequencing results showed that *LjE1* had several mutations even in the 552 bp sequence. In addition, there were some mutations in the functional region, and its allele pattern had a geographic structure. Considering the function of *E1* in *G. max*, these results suggest that *LjE1* may be related to flowering time and its polymorphism may cause flowering time variation in this species. *E1* is known as the most effective factor for flowering time regulation in *G. max*. Therefore, the results indicated the importance of this gene in the wild population of *L. japonicus* as well. *LjGI* harboured one synonymous and one nonsynonymous change in the 2,930 bp of the sequences. Therefore, it could be suggested that the *LJGI* gene may be under the effect of strong purifying selection. In addition, the results of the demographic history analyses suggest that the polymorphisms harboured in *LjE1* would have been maintained in the colonisation process as standing genetic variation. Regarding *LjGI*, this gene may have been affected by purifying selection during the time of colonisation into Tsushima because it is unlikely that the genotypes could be independently fixed in such a small variation in allopatric subpopulations.

Finally, a genome-wide association study was conducted for flowering time variation in 121 wild accessions collected from various habitats covering the distribution of this species. The flowering time variation had a continuous distribution along the latitude and several or many genetic factors could be related to this intraspecific variation. In addition, the analyses for two environmental factors, the annual average temperature and the integrated annual sunshine

duration, were conducted with 119 wild accessions. Considering the structure of Japanese *L. japonicus* populations, a mixed model that takes population structure into account was adopted (AMM; Kang *et al*., 2010; Seren *et al*., 2012). The GWA analyses for flowering time showed several clear peaks and detected 16 protein-coding candidate genes that contained two orthologues of known flowering time genes. The correlation test between flowering time variation and the genotypes of the detected SNPs showed that two out of the genotypes could explain about 60% of the flowering time variation in this species. This result suggests that the flowering time variation may be caused by a small number of genetic factors, and that there could be genetic factors with weak effects. In addition, four and one candidate genes were detected by these two SNPs, respectively, and the candidates around the first SNP included a known flowering time gene, *SmD3b-like*. As the result of continuous flowering time variation suggested, there could be several or many genetic factors affecting flowering time variation; therefore, in order to understand local adaptation through flowering time, it is necessary to identify multiple genetic factors, which would be mainly related to the variation. Although the GWA analysis enables the detection of candidates associated with the flowering time data, it is necessary to detect candidate genetic factors for flowering time data under different conditions and take all the candidates, including *LjE1* affected by natural selection, into consideration. In a future study, I plan to conduct GWA analysis with multi-trait models using several flowering time data and correlation tests with several candidate genetic factors detected by plural GWA analyses.

Throughout this study, I focused on the demographic history of and the flowering-time variation in Japanese *L. japonicus* accessions. A comprehensive consideration of the results of these studies suggests that this species has been adapted to each environment of the various regions in Japan and has expanded its distribution into the entire Japanese archipelago from Tsushima Island, while being affected by natural selection in terms of flowering time, one of its adaptive traits. In the process of local adaptation through proper flowering time, it is considerable that *LjE1,* which has a nucleotide polymorphism with geographical units affected by natural selection and may be related to the flowering-time variation in this species, is an orthologue of important flowering-time genes in *G. max*. Moreover, it could be suggested that the combinations of a few of the genetic factors related to the control of flowering time, possibly including *LjE1* and the detected candidate genes (e.g., orthologues of *SmD3b-like* and *MED28*), have been affected by natural selection and may have contributed significantly to local adaptation through flowering time as the species spread over the various environments.

Acknowledgements

First, I really appreciate my super visor Prof. H. Setoguchi (Kyoto University) to have extremely supported me and all of my research with advises, to have discussed from the various sights, and to have given me precious opportunities to work on my research with many collaborators. Also, I am especially grateful to Prof. M. Kawaguchi (National Institute for Basic Biology) and Dr. S. Sato (Tohoku University) for helping me always to have given me many chances to learn several kinds of methods though the collaborations for about eight years. In addition, I am deeply grateful to Dr. S. U. Andersen (Aarhus University) for giving me the chance to learn about population genetics and GWA analysis in Aarhus University, Denmark and for having discussed with me about the analyses in many times. And, I appreciate Dr. H. Ikeda (Okayama University) to teach me several analyses and to have discussed contents of my research with me deeply. In addition, I am deeply grateful to Dr. K. Harada for his advices especially for the research about *LjE1* and the discussion about the research with me. Also, I am proudly grateful to Dr. S. Tanaka (National Institute for Basic Biology) for having supported me in the several experiments through collaboration. And, I appreciate Dr. N. Shah to give me technical advices of population genomics methods while I was studying at Aarhus. In addition to them, I'd like to express my gratitude for National BioResource Project, University of Miyazaki, Kazusa DNA Research Institute, and Aarhus University for supplying the seeds of wild accessions.

Special thanks go to Dr. T. Ohtsuki and H. Higashi (Kyoto University) in the deep discussions about plans and several analyses related to my research. And I appreciate Mr. N. Ishibashi and Mrs. H. Oh to teach me several analyses and experiments and field works and to help me during the beginning of my laboratory life. Also, I'm very grateful to Mrs. M. Ohki, Mr. H. Yagi, and Mr. S. Yamada to help me to start my laboratory life. In addition, I thank all of the members of Setoguchi laboratory, Kawaguchi laboratory, and Stougaard laboratory. Finally, I'd like to express my obligation to my parents and my husband for having respected my intension and having supported me always. Finally, this study was supported by the Japan Society for the Promotion of Science KAKENHI (grant numbers JP24247013, JP23128508, JP25128708, and JP15J00897).

References

Alabadí D, Oyama T, Yanovsky MJ, Harmon FG, Más P, Kay SA. 2001. Reciprocal Regulation Between *TOC1* and *LHY* / *CCA1* Within the *Arabidopsis* Circadian Clock. *Science* 293: 880–883.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Moleculer Biology* 215: 403–10.

Andrés F, Coupland G. 2012. The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics* 13: 627–639.

Andrés F, Coupland G. 2012. The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics* 13: 627–639.

Ashworth MB, Walsh MJ, Flower KC, Vila-Aiub MM, Powles SB. 2016. Directional selection for flowering time leads to adaptive evolution in *Raphanus raphanistrum* (Wild radish). *Evolutionary Applications* 9: 619–629.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, TT H, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M. 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.

Aukerman MJ, Sakai H. 2003. Regulation of flowering time and floral organ identity by a MicroRNA and its *APETALA2*-Like target genes. *The Plant Cell* 15: 2730–2741.

Balasubramanian S, Sureshkumar S, Agrawal M, Todd P, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, Weigel D. 2006. The *PHYTOCHROME C* photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nature Genetics* 38: 711–715.

Bäurle I, Dean C. 2006. The Timing of Developmental Transitions in Plants. *Cell* 125: 655–664.

Bernard RL. 1971. Two major genes for time of flowering and maturity in soybeans. *Crop Science* 11: 242–244.

Bonato ER, Vello NA. 1999. E6, a dominant gene conditioning early flowering and maturity in soybeans. *Genetics and Molecular Biology* 22: 229–232.

Buendía-Monreal M, Gillmor CS. 2016. Mediator: A key regulator of plant development. *Developmental Biology* 419: 7–18.

Burgarella C, Chantret N, Gay L, Prosperi J-M, Bonhomme M, Tiffin P, Young ND, Ronfort J. 2016. Adaptation to climate through flowering phenology: a case study in *Medicago truncatula*. *Molecular Ecology* 25: 3397–3415.

Buzzell RI. 1971. Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Canadian Journal of Genetics and Cytology* 13: 703-707.

Buzzell RI, Voldeng HD. 1980. Inheritance of Insensitivity to Long Daylength. *Soybean Genetics Newsletter* 7: 26–29.

Cober ER, Voldeng HD. 2001. A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Science* 41: 698–701.

Cober ER, Molnar SJ, Charette M, Voldeng HD. 2010. A new locus for early maturity in soybean. *Crop Science* 50: 524–527.

Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES. 2011. FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. *Proceedings of the National Academy of Sciences* 108: 6680–6685.

Dittmar EL, Oakley CG, Ågren J, Schemske DW. 2014. Flowering time QTL in natural populations of *Arabidopsis thaliana* and implications for their adaptive value. *Molecular Ecology* 23: 4291–4303.

Dodd AN, Salathia N, Hall A, Kévei E, Réka T, Nagy F, Hibberd JM, Millar AJ, Webb AAR. 2005. Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. *Science* 309: 630–633.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15

Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Xavier Pico F, Alonso-Blanco C, Hancock AM. 2017. African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 114: 5213–5218.

Fankhauser C, Chory J. 1997. Light Control of Plant Development. *Annual Review of Cell and Developmental Biology* 13: 203–229.

Farré EM. 2012. The regulation of plant growth by the circadian clock. *Plant Biology* 14: 401–410.
Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek a. M. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* 334: 86–89.
Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J. 1999. *GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *The EMBO Journal* 18: 4679–88.
Franks SJ, Sim S, Weis AE. 2007. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 1278–82.
Fulgione A, Koornneef M, Roux F, Hermisson J, Hancock AM. 2018. Madeiran *Arabidopsis thaliana* reveals ancient long-range colonization and clarifies demography in eurasia. *Molecular Biology and Evolution* 35: 564–574.
Garner WW, Allard HA. 1920. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *Journal of Agricultural Research* 18: 553-606.
Hall MC, Willis JH. 2006. Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. *Evolution* 60: 2466–2477.
Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* Genome. *Science* 334: 83–86.
Handberg K, Stougaard J. 1992. *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *The Plant Journal* 2: 487–496.
Hashiguchi M, Tsuruta S-I, and Akashi R. 2011. Morphological traits of *Lotus japonicus* (Regal) ecotypes collected in Japan. *Interdisciplinary Bio Central* 3: 1–8
Hashiguchi M, Abe J, Aoki T, Anai T, Suzuki A, Akashi R. 2012. The National BioResource Project (NBRP) *Lotus* and *Glycine* in Japan. *Breeding Science* 61: 453–461.
Hecht V, Foucher F, Ferrándiz C, Macknight R, Navarro C, Morin J, Vardy ME, Ellis N, Beltrán JP, Rameau C, Weller JL. 2005. Conservation of *Arabidopsis* Flowering Genes in Model Legumes. *Plant Physiology* 137: 1420–1434.
Held K, Pascaud F, Eckert C, Gajdanowicz P, Hashimoto K, Corratgé-Faillie C, Offenborn JN, Lacombe B, Dreyer I, Thibaud JB, Kudla J. 2011. Calcium-dependent modulation and plasma membrane targeting of the AKT2 potassium channel by the CBL4/CIPK6 calcium sensor/protein kinase complex. *Cell Research* 21: 1116–1130.
Henderson IR, Dean C. 2004. Control of *Arabidopsis* flowering: the chill before the bloom. *Development* 131: 3829–3838.
Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335–2352.
Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences* 103: 16666–16671.
Ingvarsson PK, García MV, Hall D, Luquez V, Jansson S. 2006. Clinal variation in *phyB2*, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* 172: 1845–1853.
Izawa, T. 2007. Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *Journal of Experimental Botany* 58: 3091–3097.
Jung CH, Wong CE, Singh MB, Bhalla PL. 2012. Comparative genomic analysis of soybean flowering genes. *PLoS ONE* 7.
Kai S, Tanaka H, Hashiguchi M, Iwata H, Akashi R. 2010. Analysis of genetic diversity and morphological traits of Japanese *Lotus japonicus* for establishment of a core collection. *Breeding Science* 60: 436–446.
Kang HMH, Sul JHJ, Zaitlen NN a, Kong S-YY, Service SK, Zaitlen NN a, Kong S-YY, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42: 348–354.
Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D. 1999. Activation tagging of the floral inducer *FT*. *Science* 286: 1962–1965.
Kawaguchi M. 2000. *Lotus japonicus* `Miyakojima' MG-20: An Early-Flowering

Accession Suitable for Indoor Handling. *Journal of Plant Research* 113: 507–509.
Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7: 1225–1241.
Keller SR, Levsen N, Olson MS, Tiffin P. 2012. Local adaptation in the flowering-time gene network of balsam poplar, *populus balsamifera* L. *Molecular Biology and Evolution* 29: 3143–3152.
Kevei E, Schafer E, Nagy F. 2007. Light-regulated nucleo-cytoplasmic partitioning of phytochromes. *Journal of Experimental Botany* 58: 3113–3124.
Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T. 1999. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* 286: 1960–1962.
Komeda Y. 2004. Genetic regulation of time to flower in *Arabidopsis Thaliana*. *Annual Review of Plant Biology* 55: 521–535.
Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9: 1.
Leinonen PH, Remington DL, Leppälä J, Savolainen O. 2013. Genetic basis of local adaptation and flowering time variation in *Arabidopsis lyrata*. *Molecular Ecology* 22: 709–723.
Levy YY, Dean C. 1998. The Transition to Flowering. *the Plant Cell Online* 10: 1973–1990.
Li H, Durbin R. 2011. Inference of Human Population History From Whole Genome Sequence of A Single Individual. *Nature* 475: 493–496.
Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J. 2008. Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 180: 995–1007.
Liu L, Adrian J, Pankin A, Hu J, Dong X, Von Korff M, Turck F. 2014. Induced and natural variation of promoter length modulates the photoperiodic response of *FLOWERING LOCUS T*. *Nature Communications* 5: 1–9.
Liu L, Liu C, Hou X, Xi W, Shen L, Tao Z, Wang Y, Yu H. 2012. FTIP1 is an essential regulator required for florigen transport. *PLoS Biology* 10.
Locke JCW, Kozma-Bognár L, Gould PD, Fehér B, Kevei É, Nagy F, Turner MS, Hall A, Millar AJ. 2006. Experimental validation of a predicted feedback loop in the multi-oscillator clock of *Arabidopsis thaliana*. *Molecular Systems Biology* 2: 1–6.
Lu S, Bahn SC, Qu G, Qin H, Hong Y, Xu Q, Zhou Y, Hong Y, Wang X. 2013. Increased expression of phospholipase Dα1 in guard cells decreases water loss with improved seed production under drought in *Brassica napus*. *Plant Biotechnology Journal* 11: 380–389.
Mcblain BA, Bernard RL. 1987. A new gene affecting the time of flowering and maturity in soybeans. *Journal of Heredity* 78: 160–162.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351: 652–654.
Mouradov A, Cremer F, Coupland G. 2002. Control of flowering time: interacting pathways as a basis for diversity. *The Plant cell* 14 Suppl: S111–S130.
Murtas G, Reeves PH, Fu Y-F, Bancroft I, Dean C, Coupland G. 2003. A nuclear protease required for flowering-time regulation in *Arabidopsis* reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. *the Plant Cell Online* 15: 2308–2319.
Nakamichi N, Kita M, Ito S, Yamashino T, Mizuno T. 2005. PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, Together play essential roles close to the circadian clock of *Arabidopsis thaliana*. *Plant and Cell Physiology* 46: 686–698.
Ono N, Ishida K, Yamashino T, Nakanishi H, Sato S, Tabata S, Mizuno T. 2010. Genomewide characterization of the light-responsive and clock-controlled output pathways in *Lotus japonicus* with special emphasis of its uniqueness. *Plant and Cell Physiology* 51: 1800–1814.
Ossowski S, Schneeberger K, Lucas-lledó JI, Clark RM, Shaw RG, Weigel D, Lynch M, Warthmann N. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana* Supporting Online Material. 24073: 92–95.
Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
Raj A, Stephens M, Pritchard JK. 2014. FastSTRUCTURE: Variational inference of

population structure in large SNP data sets. *Genetics* 197: 573–589.
Rieu I, Eriksson S, Powers SJ, Gong F, Griffiths J, Woolley L, Benlloch R, Nilsson O, Thomas SG, Hedden P, Phillips A. 2008. Genetic analysis reveals that C19-GA 2-oxidation is a major gibberellin inactivation pathway in *Arabidopsis*. *Plant Cell* 20: 2420–2436.
Samach A, Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-sommer Z, Yanofsky MF, Coupland G. 2000. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. *Science* 288: 1613–1616.
Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M. 2015. "Missing" G x E Variation Controls Flowering Time in *Arabidopsis thaliana*. *PLoS Genetics* 11: 1–18.
Sato S, Kaneko T, Nakamura Y, Asamizu E, Kato T, Tabata S. 2001. Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 mb regions of the genome. *DNA Research* 8: 311–318.
Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, FujishiroT, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S. 2008. Genome structure of the legume, *Lotus japonicus*. *DNA Research* 15: 227–239.
Schemske DW, Willson MF, Melampy MN, Miller LJ, Verner L, Schemske KM, Best LB. 1978. Flowering Ecology of Some Spring Woodland Herbs. *Ecology* 59: 351–366.
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.
Seren U, Vilhjalmsson BJ, Horton MW, Meng D, Forai P, Huang YS, Long Q, Segura V, Nordborg M. 2012. GWAPP: A Web Application for Genome-Wide Association Mapping in *Arabidopsis*. *The Plant Cell* 24: 4793–4805.
Setoguchi H, Ohba H. 1995. Phylogenetic relationships in *Crossostylis* (Rhizophoraceae) inferred from restriction site variation of chloroplast DNA. *Journal of Plant Research* 108: 87-92
Shah N, Wakabayashi T, Kawamura Y, Skovbjerg CK, Wang M-Z, Mustamin Y, Isomura Y, Gupta V, Jin H, Mun T, Sandal N, Azuma F, Fukai E, Seren Ü, Kusakabe S, Kikuchi Y, Nitanda S, Kumaki T, Hashiguchi M, Tanaka H, Hayashi A, Sønderkær M, Nielsen KL, Schneeberger K, Vilhjalmsson B, Akashi R, Stougaard J, Sato S, Schierup MH, Andersen SU. 2020. Extreme genetic signatures of local adaptation during *Lotus japonicus* colonization of Japan. *Nature Communications* 11.
Song YH, Ito S, Imaizumi T. 2013. Flowering time regulation: Photoperiod- and temperature-sensing in leaves. *Trends in Plant Science* 18: 575–583.
Srikanth A, Schmid M. 2011. Regulation of flowering time: All roads lead to Rome. *Cellular and Molecular Life Sciences* 68: 2013–2037.
Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences* 101: 4712–4717.
Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* 152: 352–364.
Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. 2001. *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* 410: 1116–1120.
Swaraz AM, Park YD, Hur Y. 2011. Knock-out mutations of *Arabidopsis SmD3-b* induce pleotropic phenotypes through altered transcript splicing. *Plant Science* 180: 661–671.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.
Tsubokura Y, Watanabe S, Xia Z, Kanamori H, Yamagata H, Kaga A, Katayose Y, Abe J, Ishimoto M, Harada K. 2014. Natural variation

in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. *Annals of Botany* 113: 429–441.

Turner SD. 2018. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *The Journal of Open Source Software* 3: 25.

Verdier J, Torres-Jerez I, Wang M, Andriankaja A, Allen SN, He J, Tang Y, Murray JD, Udvardi MK. 2013. Establishment of the *Lotus japonicus* Gene Expression Atlas (LjGEA) and its use to explore legume seed maturation. *Plant Journal* 74: 351–362.

Watanabe S, Hideshima R, Zhengjun X, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T, Tabata S, Harada K. 2009. Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* 182: 1251–1262.

Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, Harada K. 2011. A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* 188: 395–407.

Whitelam GC, Devlin PF. 1997. Roles of different phytochromes in *Arabidopsis* photomorphogenesis. *Plant, Cell and Environment* 20: 752–758.

Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lü S, Wu Hongyan, Tabata S, Harada K. 2012. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proceedings of the National Academy of Sciences* 109: E2155–E2164.

Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S. 2004. Solution structure of the B3 DNA binding domain of the *Arabidopsis* cold-responsive transcription factor RAV1. *The Plant cell* 16: 3448–3459.

Yamashino T, Yamawaki S, Hagui E, Ueoka-Nakanishi H, Nakamichi N, Ito S, Mizuno T. 2013. Clock-controlled and FLOWERING LOCUS T (FT)-Dependent Photoperiodic Pathway in *Lotus japonicus* I: Verification of the Flowering-Associated Function of an FT Homolog. *Bioscience, Biotechnology, and Biochemistry* 77: 747–753.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *The Plant Cell* 12: 2473–2484.

Yao T, Park BS, Mao HZ, Seo JS, Ohama N, Li Y, Yu N, Mustafa NFB, Huang CH, Chua NH. 2019. Regulation of flowering time by SPL10/MED25 module in *Arabidopsis*. *New Phytologist*: 493–504.

Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, Peer YV, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, Mita SD, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen, B, Silverstein KAT, Tang H, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Bergès H, Bidwell S, Bisseling T, Choisne N, Couloux A, Denny R, Deshpande S, Dai X, Doyle J, Dudez A-M, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, González AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray S, Jeong D-H, Jing Y, Jöcker A, Kenton SM, Kim D-J, Klee K, Lai H, Lang C, Lin S, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun J-H, Najar FZ, Nicholson C, Noirot C, O'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Saurat ISO, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang B-B, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing Y, Yang L, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Dénarié J, Dixon RA, May GD, Schwartz DC, Rogers J, Quétier F, Town CD, Roe BA. 2011. The *Medicago* Genome Provides Insight into the Evolution of Rhizobial Symbioses. *Nature* 480: 520–524.

Yu L-X, Liu X, Boge W, Liu X-P. 2016. Genome-Wide Association Study Identifies Loci for Salt Tolerance during Germination in Autotetraploid Alfalfa (*Medicago sativa* L.) Using Genotyping-by-Sequencing. *Frontiers in Plant Science* 7: 1–12.

Zhang J, Song Q, Cregan PB, Nelson RL, Wang X, Wu J, Jiang GL. 2015. Genome-wide

association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16: 1–11.

Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications* 2: 467.

Zhou XE, Wang Y, Reuter M, Mücke M, Krüger DH, Meehan EJ, Chen L. 2004. Crystal structure of type IIE restriction endonuclease EcoRII reveals an autoinhibition mechanism by a novel effector-binding fold. *Journal of Molecular Biology* 335: 307–319.