Maternal phylogeographic patterns and coalescent times of *Arabidopsis thaliana* based on chloroplast DNA analyses

Ningyu Sun¹, Yang Zhong^{1†} and Takahiro Yonezawa^{1,2*}

¹*School of Life Science, Fudan University, SongHu Road 2005, Shanghai 200438, China*

²*Faculty of Agriculture, Tokyo University of Agriculture, 1737 Funako, Atsugi, Kanagawa 243-0034, Japan*

(Received 24 August 2018, accepted 6 May 2019; J-STAGE Advance published date: 6 September 2019)

Arabidopsis thaliana, one of the most important model plants, has played an essential role in every biological field including evolutionary biology. Recent population genomic studies have gradually clarified the origin and evolution of this species. Nevertheless, incongruent patterns among gene trees based on cytogenetic data have highlighted the importance of understanding the life history evolution and landscape biogeography of extant *A. thaliana* populations. Here, we focus on the maternally inherited chloroplast genome in *A. thaliana* and carry out phylogeographic analyses and coalescent time estimations. The maternal lineage of *A. thaliana* originated in the European to West and Central Asian regions in the Early Pleistocene. Relicts, the ancient lineages suggested by population genomic data, are not ancestral maternal lineages, but are derived from the European population. Part of the European population then dispersed eastward and spread to the Indian region, and finally extended to the Yangtze River region. The branching patterns and evolutionary time scales of the maternal genealogy are significantly different from those estimated from analyses of autosomal genes, and these cannot be explained by incomplete lineage sorting of the ancestral polymorphisms during the coalescent process due to large differences in the evolutionary time scale involved.

Key words: ancestral state, *Arabidopsis thaliana*, Bayesian timetree, chloroplast genome, evolutionary relationship

INTRODUCTION

Arabidopsis thaliana is one of the most important model plants and has played an essential role across all biological fields, especially genetics (Arabidopsis Genome Initiative, 2000) and developmental biology (Tsukaya, 2006; Kawade et al., 2013). This species is also attractive for studies of evolutionary biology, in part because its reproduction system includes self-fertilization (selfing). Because selfing changes the polymorphic patterns of populations (Shimizu and Tsuchimatsu, 2015), it is important to elucidate the phylogeographic history of this species. The origin of *A. thaliana* has been gradually clarified by recent population genomic studies (1001 Genomes Consortium, 2016; Durvasula et al., 2017; Zeng

et al., 2017). Selfing in this species is thought to have originated in Africa and then subsequently migrated to Eurasia via the Iberian Peninsula (Durvasula et al., 2017). Therefore, several old lineages among non-African populations exist in the Iberian Peninsula as well as its peripheral region. These lineages are called “relicts” (1001 Genomes Consortium, 2016).

Cytogenetic analyses of various organisms have been carried out extensively over the past two decades, and incongruent patterns of species (or population) trees and gene trees have been reported (Roca et al., 2005; Nakagome et al., 2008; Yonezawa et al., 2009; Akihito et al., 2016). These studies suggested that incongruent patterns of gene trees are related to the life histories of the organisms. In particular, discrepancies between the evolutionary histories of maternally inherited genes (e.g., mitochondrial genes or chloroplast (cp) genes) and autosomal genes often reflect complex breeding systems or different behaviors between the sexes (Roca et al., 2005; Nakagome et al., 2008). Therefore, it is important to

Edited by Atsushi Tajima

* Corresponding author. E-mail: cyclotis@gmail.com

† Deceased, 25th September, 2017.

DOI: <http://doi.org/10.1266/ggs.18-00044>

compare the different genealogies among cp gene trees and nuclear gene trees to improve our understanding of the complete historical evolutionary landscape of *A. thaliana*. Nevertheless, while several previous studies (e.g., Yin et al., 2010; Tyagi et al., 2015) reported extensive phylogeographic and population genetic analyses of local populations of this species, the maternal history of *A. thaliana* across its geographic distribution areas is little known. Moreover, the evolutionary histories of the chloroplast genomes and the nuclear genomes have seldom been compared. Here, we carry out phylogeographic studies of the maternal history of *A. thaliana* based on cpDNA data, and discuss the different evolutionary histories of the species tree and the maternal lineages.

MATERIALS AND METHODS

Data matrix We retrieved 195 complete cp genomes of *A. thaliana* from previous studies (1001 Genomes Consortium, 2016; Zeng et al., 2017). These data comprise all of the reported cp genomic data from relicts (25 individuals) and the Asian population (81 individuals) as well as randomly selected subsamples of the European population (89 individuals). The accession information for these complete cp genome data is shown in Supplementary Table S1.

Yin et al. (2010) and Tyagi et al. (2015) carried out phylogenetic and population genetic studies of Chinese

and Western Himalayan populations based on 11 cpDNA markers (GenBank accession numbers are GU293237–GU293316, GU293317–GU293396, GU293397–GU293476, GU293477–GU293556, GU293557–GU293636, GU293637–GU293716, GU293717–GU293796, GU293797–GU293876, GU293877–GU293956, GU293957–GU294036, GU294037–GU294116). Therefore, we also retrieved the nucleotide sequence data of 68 *A. thaliana* individuals with these 11 cpDNA markers. The accession information for these cp data is shown in Supplementary Table S2. Finally, these 68 individuals were merged with 195 complete cp genomes, and the full alignment consists of 263 *A. thaliana* individuals with 157,192 bp in total length. Since 68 individuals have only 11 cpDNA markers (~8,742 bp), most of the aligned sites of these individuals were treated as missing data. This complete data matrix consists of sequences from Relicts (25 individuals), the Asian population (149 individuals) and the European population (89 individuals). As discussed below, the individuals distributed in the Western Himalayan to Yangtze River regions form a distinct clade (clade 7) in the phylogenetic tree (Fig. 2). This China–India clade is derived from the European population instead of the Remaining Asia population, which is geographically closer. Therefore, the Asian population was further divided into a “China–India” population (67 individuals) and a “non China–India” population (82 individuals). Hereafter, we refer to this non China–India population as the “Remain-

Fig. 1. Geographic origins of all the accessions (accurate to the country). According to their geographic locations and previous *A. thaliana* population genomic research (Yin et al., 2010; Tyagi et al., 2015; 1001 Genomes Consortium, 2016), all the populations are divided into four groups: Remaining Asia region (green), European region (pink), China–India region (hyacinthine), Relicts (yellow). Clades 1–7 (Fig. 2) are marked with seven different symbols (circle, cross, triangle, rectangle, pentagram, diamond and asterisk).

Fig. 2. Timetree of *Arabidopsis thaliana* and its related species. The timetree is divided into seven clades marked with different symbols (circle, cross, triangle, rectangle, pentagram, diamond and asterisk). The branch lengths are proportional to the time scale and the concentric circles are at intervals of 1.6 million years. *Arabidopsis lyrata*, *A. halleri*, *A. suecica* and *Capsella grandiflora* were used as the outgroups, and these are not colored.

ing Asia population”, and this population is mainly distributed in Central Asia and its peripheral regions, including Russia, Ukraine, Uzbekistan, Tajikistan, Armenia and Xinjiang (China) (Fig. 1). *Arabidopsis lyrata*, *A. halleri*, *A. suecica* and *Capsella grandiflora* were also used as outgroups.

Demographic analysis Tajima’s *D* (Tajima, 1989) was estimated by MEGA version 6 (Tamura et al., 2013). Since there are many missing data in our matrix, we conducted the analysis with (1) complete deletion, (2) pairwise deletion and (3) partial deletion (50%, 60%, 70%, 80% and 90%).

Haplotype network A haplotype network of the 263 *A. thaliana* individuals was constructed by the MJ (median-joining) method (Bandelt et al., 1999) using Network version 5.0.0.3 (<http://www.fluxus-engineering.com/sharenetwork.htm>). In this analysis, only 11 chloroplast DNA markers were used for inferring the phylogenetic network. The MJ network was further refined using the maximum parsimony criterion, and a clear tree-like network structure was obtained.

Phylogenetic inference and coalescent time estimation To estimate detailed phylogenetic relationships and coalescent times for *A. thaliana*, we used 195 complete cp genomes as well as 68 accessions that shared the same 11 cp DNA markers. A phylogenetic tree was inferred using the RAxML program v8.2.10 (Stamatakis, 2014) with the GTR+I+ Γ model. The bootstrap method was applied to evaluate the confidence of the internal branches with 1,000 replications using the rapid bootstrap algorithm (Stamatakis et al., 2008). Coalescent times were estimated by the Bayesian method using BEAST v1.8.0 (Drummond et al., 2005). For taking account of rate variation among lineages, coalescent times were estimated under a relaxed molecular clock using the hierarchical Bayesian method (Thorne et al., 1998).

MCMC was conducted with 2×10^8 generations, and trees were sampled every 20,000 generations. Four coalescent demographic models (constant size, exponential growth, logistic growth and expansion growth) were compared with AICM using Tracer 1.6.0. (<http://beast.community/tracer>), and the logistic growth model was selected as the most suitable model for this data matrix. The initial 1,000 trees were discarded as burn-in, and the remaining 9,000 trees were used for reconstructing the Bayesian timetree using TreeAnnotator 1.6.1 (<http://beast.community/treeannotator>). The timetree with the maximum node credibility was visualized using FigTree 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). We repeated these analyses three times with different MCMC conditions and confirmed the stability of our estimates.

To minimize the computational burden, we selected the simplest nucleotide model using the following procedures. Our saturation plot analysis shows that the effect of homoplasy is limited within the framework of the intra- and interspecies analyses (Supplementary Fig. S1). Although the selection of different nucleotide substitution models (e.g., JC69 model, K80 model, TN93 model) had nearly no effect on the estimation of genetic distances, the selection of models for rate heterogeneity among sites (with and without Γ model) resulted in a small effect. Therefore, we applied the simpler nucleotide substitution model, but took rate heterogeneity among sites into account; subsequently, the HYK+I+ Γ model (Hasegawa et al., 1985; Yang, 1994) was employed using the empirical base frequency data.

The following two divergence times were applied for the calibration points: first, 15 million years ago (Mya) for the divergence time between the genus *Capsella* and the genus *Arabidopsis*; second, 9.2 Mya for the divergence time between *A. lyrata* and *A. thaliana*. These divergence times were estimated within the comprehensive framework of angiosperm evolution based on 334 orthologous single-copy genes examined in our previous studies (Zeng et al., 2017). The uncorrelated log normal relaxed clock was used for the clock model.

Ancestral reconstruction of geographic distribution Ancestral geographic distributions were reconstructed by parsimony and likelihood methods using MESQUITE software version 3.40 (<http://mesquiteproject.org>). Tree topology of the timetree was inferred by the BEAST program. The following three sets of geographic classifications were applied: (1) Relicts, Europe, Remaining Asia and China–India division; (2) Europe to Remaining Asia division; and (3) Relicts to Non-Relicts division.

RESULTS

Evolutionary relationships among chloroplast haplotypes A total of 118 haplotypes were detected from 263 individuals. The China–India population spanning from the Western Himalayan to the Yangtze River regions shows the highest haplotype diversity, having 51 haplotypes out of 70 individuals.

The Bayesian timetree, the MJ network and the Maximum Likelihood tree are shown in Fig. 2, Fig. 3 and Supplementary Fig. S2, respectively. These results are essentially consistent, and the China–India population is directly derived from the European population instead of the Remaining Asia population.

Twenty-five Relicts accessions have 14 haplotypes and are nested within the Europe population, which covers wide areas of the European continent spanning Sweden, Germany, Spain and Italy. Forty-three haplotypes were detected out of 74 individuals.

Fig. 3. MJ network of chloroplast haplotypes of *Arabidopsis thaliana*. MJ network of 118 haplotypes detected in 263 *A. thaliana* individuals based on the concatenated sequence of 11 chloroplast markers. Each haplotype is indicated by a circle, and the sizes of these circles are proportional to the number of individuals. Evolutionary relationships of haplotypes are indicated by gray lines, and red bars on the lines indicate the number of substitutions. All the individuals were divided into four populations (Europe, Remaining Asia, China-India, Relicts).

Compared with other groups, the Remaining Asia population shows the lowest haplotype diversity: 80 Asian individuals only have 15 haplotypes. European populations and the Remaining Asia population are partially intermingled on the phylogenetic tree and several haplotypes are shared by these populations.

Bayesian timetree The Bayesian timetree of *A. thaliana* and its related species, based on cpDNA data and estimated by the BEAST program, is shown in Fig. 2, and the detailed result of tMRCA (time of the most recent common ancestor) determinations is summarized in Table 1.

The tMRCA of *A. thaliana* is estimated to be 2.41 Mya (HPD = 0.42–5.39 Mya). The timetree (Fig. 2) shows that the oldest lineages of *A. thaliana* are from North Europe. Subsequently, most of the Asian population, the Relicts and most of the European population diverged in this order, and finally the China-India population

Table 1. Divergence and coalescent times of *A. thaliana* accessions

Evolutionary event	Estimated time (Mya)	95% HPD (Mya)
Split of <i>Arabidopsis</i> – <i>Capsella</i>	16.5	10.00–27.24
tMRCA of the genus <i>Arabidopsis</i>	8.49	1.61–18.07
tMRCA of <i>Arabidopsis thaliana</i>	2.41	0.42–5.38
tMRCA of Asian population	2.41	0.43–5.39
tMRCA of European population	2.41	0.42–5.37
tMRCA of Indian population	0.93	0.17–2.07
tMRCA of Chinese population	0.69	0.09–1.63
tMRCA of Relicts	1.17	0.19–2.66

emerged.

The tMRCA of each population are as follows: Remaining Asian population (2.41 Mya, HPD = 0.43–5.39 Mya),

Fig. 4. Ancestral geographic distribution areas of *Arabidopsis thaliana* reconstructed by the maximum parsimony method. This figure indicates the phylogenetic relationships of the seven major clades shown in Fig. 1. Relationships within each clade are compressed and the branch lengths are proportional to the time scale. Clades are colored in accordance with their populations.

European population (2.41 Mya, HPD = 0.42–5.37 Mya), Relicts (1.17 Mya, HPD = 0.19–2.66 Mya), and China–India population (0.93 Mya, HPD = 0.17–2.07 Mya).

Ancestral reconstruction of the geographic distribution Based on the topology of the Bayesian timetree (Fig. 2), the ancestral geographic distribution areas of *A. thaliana* were reconstructed. Nodes weakly supported (<50% posterior probability) were treated as multifurcations. Ancestral geographic distributions of important nodes estimated by the maximum parsimony method are shown in Fig. 4. Ancestral reconstruction by the maximum likelihood method was consistent with this result (data not shown). Our analysis shows that the common ancestor of *A. thaliana* was distributed in Central Asia (Remaining Asia) and Europe. The China–India region and the region of the Relicts (Iberian Peninsula and its peripheral region) were not assigned to the ancestral distribution area.

DISCUSSION

Our phylogeographic analysis suggests that the ancestral distribution areas of *A. thaliana* were in Central Asia (Remaining Asia) or Europe excluding Relicts (Fig. 4). Since larger numbers of geographic categories seem to have lower resolution for reconstructing ancestral geo-

graphic distribution areas (Yonezawa et al., 2017), we also carried out reconstruction of ancestral geographic distribution areas based on two geographic categories. The results of these analyses using the maximum parsimony method are shown in Supplementary Fig. S3 (Relicts vs. Non-Relicts) and S4 (Europe including Relicts vs. Asia including Remaining Asia and China–India). Ancestral reconstruction by the maximum likelihood method was consistent with these results (data not shown). These results also suggest that the Relicts are not included in the ancestral populations of the extant *A. thaliana* populations. The comparison of two geographic categories (Europe including Relicts vs. Asia including Remaining Asia and China–India) has no power to further resolve the detailed ancestral geographic distribution areas of this species (Supplementary Fig. S4).

Taking these results together, we can use these ancestral geographic distribution areas and the coalescent (and divergence) times to suggest the following evolutionary scenario for the maternal history of extant *A. thaliana* populations. The first evolutionary event was the emergence of the earliest extant *A. thaliana* population itself. After the branching from other *Arabidopsis* species during the Late Miocene (8.49 Mya), the most recent common ancestor of the maternal lineage of the modern population emerged at 2.41 (0.42–5.38) Mya. The emergence of this species corresponds to the Gelasian stage of

the Early Pleistocene, and, if HPD is taken into account, it corresponds to the Pliocene to Middle Pleistocene. Ice sheets are known to have expanded in the northern hemisphere during this time (Zachos et al., 2001). It has been suggested that global cooling events resulted in increased terrestrial aridity, which can cause the spread of savannah and open scrublands (Yonezawa et al., 2007). It is possible that *A. thaliana*, seen in undisturbed rocky sites and forest openings in natural habitats (Mitchell-Olds, 2001), emerged in such an arid environment. The geographic origin of this species seems to be Central and West Asia (the distribution areas of the Remaining Asia population) and Europe. A more detailed resolution of the region of origin could not be determined based on the phylogenetic ancestral reconstructions in this study (Fig. 4 and Supplementary Fig. S3 and S4). Therefore, we subsequently undertook the coalescent time-based approach (Zeng et al., 2017). If there are two populations (A and B), and population A is derived from population B, then the tMRCA of population B will be equal to the tMRCA of (population A + population B). In addition, the tMRCA of population A will be younger than the tMRCA of (population A + population B). The tMRCA of the Remaining Asia population (West Asia and Central Asia) and the European population were estimated to be 2.41 (0.43–5.39) Mya and 2.41 (0.42–5.37) Mya, respectively (Table 1). These estimates are consistent with the estimate for the entire extant *A. thaliana* population itself. This finding suggests that the European population and the Remaining Asia population are not ancestrally related by descent, but rather that they differentiated by vicariance.

The second evolutionary event was the emergence of the Relicts. The Relicts were derived from the European population (Fig. 4), and their tMRCA was estimated to be 1.17 (0.19–2.66) Mya (Table 1). This timing corresponds to the Calabrian stage of the Early Pleistocene (Fig. 2).

The third evolutionary event was the emergence of the China–India population. This population is also derived from the European population in the Calabrian stage (Fig. 2, Fig. 4). Although this population is the youngest population, it has the highest haplotype diversity. This is most likely a signature of recent population expansion. If this population is further divided into the Chinese subpopulation and the Indian subpopulation, the tMRCA of the Chinese subpopulation is 0.69 (0.09–1.63) Mya and that of the Indian subpopulation is 0.93 (0.17–2.07) Mya (Table 1), suggesting that the Chinese subpopulation is derived from the Indian subpopulation. Thus, the ancestors of the China–India population most likely migrated from the European region, and then some portions of this population continued to further migrate eastward and finally reached the Yangtze River region.

The maternal evolutionary scenario described above largely contradicts the scenario suggested by recent popu-

lation genomic studies regarding historical biogeographic patterns (origin and migration routes) (1001 Genomes Consortium, 2016; Durvasula et al., 2017; Zeng et al., 2017) as well as the evolutionary time scale (Durvasula et al., 2017; Zeng et al., 2017). Previous nuclear genomic analyses suggested that the African population or the Tibetan population was the most ancestral lineage (Durvasula et al., 2017; Zeng et al., 2017), but these populations were not included in the present study. However, we show that the Relicts (1001 Genomes Consortium, 2016) clearly have a derived maternal lineage in this study (Fig. 2 and Fig. 4). There is a considerable difference in the tMRCA estimated from whole genomic data (about 140 thousand years ago (Kya) by Durvasula et al. and 126–149 Kya by Zeng et al.) (Durvasula et al., 2017; Zeng et al., 2017) and from cpDNA data (2.41 Mya: this study). If there are considerable numbers of multiple substitutions per site in comparisons between the cp genomes of *A. thaliana* and those of other species, this would result in the overestimation of the coalescent times among *A. thaliana* accessions because the lengths of the internal branches near the root were underestimated and the nodes near the root were used as calibration points. However, since there are very limited numbers of multiple substitutions per site among *Arabidopsis* species or between genus *Arabidopsis* species and genus *Capsella* (Supplementary Fig. S1), the old tMRCA of the cp genome of *A. thaliana* is not an analytical artifact caused by saturation of the genetic distances.

This remarkable discrepancy of the time scale cannot be explained by incomplete lineage sorting of ancestral polymorphism. Zeng et al. (2017) and Durvasula et al. (2017) estimated the tMRCA of the nuclear genomes, and obtained very similar ages (about 150 Kya). Since these estimates are based on multiple nuclear genes at the whole-genome level, about 150 Kya can be interpreted as the mean tMRCA of the nuclear genes. Because the expected tMRCA and their standard deviations are $4N_e \pm 4N_e$ generations ago, tMRCA of more than 99% of the genes are principally younger than 600 Kya. Moreover, considering that the effective population size of cp genes is 1/4 of the autosomal genes, the expected tMRCA of cp genes should be about 1/4 of that of autosomal genes, which is about 37.5 Kya. Even taking account of the standard deviations, the tMRCA of the cp genome should be younger than 150 Kya. However, the tMRCA of the cp genome was estimated to be 2,410 Kya. Thus, the estimated tMRCA of cpDNA in this study is clearly outside this range even taking account of the large standard deviation of the tMRCA.

Accordingly, it can be interpreted that this discrepancy of the time scale is caused by the different evolutionary history of the cp genome and the nuclear genome. One possible explanation is a balancing selection of the cp genome. However, since the Tajima's D of the cp genome

is negative (−2.65 to −2.85), balancing selection of the cp genome is unlikely. Another possible explanation is a recent very strong selective sweep of the nuclear genome but not of organelle genomes in the modern *A. thaliana* population. A recent origin for self-compatibility in *A. thaliana* has been suggested (Shimizu and Tsuchimatsu, 2015). It is possible that the maternal lineage of *A. thaliana* has retained older phylogeographic structures that existed prior to when the capacity for self-compatibility spread in this species, thereby revealing a maternal history suggested by this study that contradicts the evolutionary scenario suggested by recent population genomic studies (1001 Genomes Consortium, 2016; Durvasula et al., 2017; Zeng et al., 2017). To examine this hypothesis further, we have initiated extensive population genomic analyses that are now in progress.

This paper is dedicated to the memory of Prof. Yang Zhong, who passed away on September 25th, 2017. We thank Dr. Zhuoya Gu for assistance with data analysis, Dr. Andrew M. Shedlock for improving our manuscript, and Dr. Atsushi Tajima as well as three anonymous reviewers for helpful comments.

REFERENCES

- 1001 Genomes Consortium (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491.
- Akihito, Akishinonmiya, F., Ikeda, Y., Aizawa, M., Nakagawa, S., Umehara, Y., Yonezawa, T., Mano, S., Hasegawa, M., Nakabo, T., et al. (2016) Speciation of two gobioid species, *Pterogobius elapoides* and *Pterogobius zonoleucus* revealed by multi-locus nuclear and mitochondrial DNA analyses. *Gene* **576**, 593–602.
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Bandelt, H. J., Forster, P., and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48.
- Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192.
- Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F. X., Alonso-Blanco, C., et al. (2017) African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **114**, 5213–5218.
- Hasegawa, M., Kishino, H., and Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174.
- Kawade, K., Horiguchi, G., Usami, T., Hirai, M. Y., and Tsukaya, H. (2013) ANGUSTIFOLIA3 signaling coordinates proliferation between clonally distinct cells in leaves. *Curr. Biol.* **23**, 788–792.
- Mitchell-Olds, T. (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* **16**, 693–700.
- Nakagome, S., Pecon-Slattery, J., and Masuda, R. (2008) Unequal rates of Y chromosome gene divergence during speciation of the family Ursidae. *Mol. Biol. Evol.* **25**, 1344–1356.
- Roca, A. L., Georgiadis, N., and O'Brien, S. J. (2005) Cytonuclear genomic dissociation in African elephant species. *Nat. Genet.* **37**, 96–100.
- Shimizu, K. K., and Tsuchimatsu, T. (2015) Evolution of selfing: recurrent patterns in molecular adaptation. *Annu. Rev. Ecol. Evol. Syst.* **46**, 593–622.
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758–771.
- Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.
- Thorne, J. L., Kishino, H., and Painter, I. S. (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**, 1647–1657.
- Tsukaya, H. (2006) Mechanism of leaf-shape determination. *Annu. Rev. Plant Biol.* **57**, 477–496.
- Tyagi, A., Singh, S., Mishra, P., Singh, A., Tripathi, A. M., Jena, S. N., and Roy, S. (2015) Genetic diversity and population structure of *Arabidopsis thaliana* along an altitudinal gradient. *AoB Plants* **8**, plv145.
- Yang, Z. (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314.
- Yin, P., Kang, J., He, F., Qu, L. J., and Gu, H. (2010) The origin of populations of *Arabidopsis thaliana* in China, based on the chloroplast DNA sequences. *BMC Plant Biol.* **10**, 22.
- Yonezawa, T., Fukui, D., Nakamura, R., Bando, G., Endo, H., Sasaki, M., and Hasegawa, M. (2009) A case study of the molecular genetic diagnosis of a small African elephant (*Loxodonta* sp.) “Nana” kept at Asahiyama Zoo. *Mammal Study* **34**, 171–177.
- Yonezawa, T., Nikaido, M., Kohno, N., Fukumoto, Y., Okada, N., and Hasegawa, M. (2007) Molecular phylogenetic study on the origin and evolution of Mustelidae. *Gene* **396**, 1–12.
- Yonezawa, T., Segawa, T., Mori, H., Campos, P. F., Hongoh, Y., Endo, H., Akiyoshi, A., Kohno, N., Nishida, S., Wu, J., et al. (2017) Phylogenomics and morphology of extinct *paleognaths* reveal the origin and evolution of the ratites. *Curr. Biol.* **27**, 68–77.
- Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693.
- Zeng, L., Gu, Z., Xu, M., Zhao, N., Zhu, W., Yonezawa, T., Liu, T., Qiong, L., Tersing, T., Xu, L., et al. (2017) Discovery of A high-altitude ecotype and ancient lineage of *Arabidopsis thaliana* from Tibet. *Sci. Bull.* **62**, 1628–1630.