**Materials and Methods:**

**Plant material and treatments**

The seeds and leaves of Tibet-0 were collected in the wild forest of Duilongdeqing County (N29.6903, E90.9338, altitude: 4200m asl), Tibet in 2013 when it had blossomed and borne fruit. The leaves were immediately dried with silica gel and stored in sealed bags for the following molecular identification. We then grew the Tibet-0 under a 16h light (22℃)/8 h dark (18℃) photoperiod regime and collected the flowers and leaves for further tests and sequencing.

**Molecular Identification by gene barcoding**

In order to prove Tibet-0 to be *A. thaliana*, we compared the nuclear internal transcribed spacer (ITS), four chloroplast genes (matK, rbcL, rpoB, rps16) and two chloroplast intergenic spacers (IGS) (trnL-trnF, trnT-trnL) between *A. thaliana* and A. lyrata (Online, Simon, Trajanoski et al. 2012). The total DNA was extracted by a TIANGEN Plant Genomic DNA Kit. We used the universal primers of each barcoding sequence in PCR amplification. The PCR reaction system contained 10ng DNA template, 2ul (10μM) forward-reverse primers, 25 μl 2x Taq PCR Master Mix (Tiangen Biotech). The PCR procedure consisted of one cycle of 94℃ 5min, 32 cycles of 94℃ 1min, Ta℃ (Supplementary Table 1) 1 min, and 72℃ 1.5 min, and one cycle of 72℃ 7min and 10℃10min. The PCR products were sequenced using an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, California, USA). We then compared all the 7 sequences of the Tibetan sample with the homologous sequence of Col-0 and Arabidopsis lyrata downloaded from NCBI by alignment using CLUSTALX (Thompson, Gibson et al. 1997). We also tested coverage and identity of those genes in the Tibetan sample, Col-0, and A. lyrata using the Blast tool of NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

**Polyploidy determination**

To determine the polyploidy of the new *Arabidopsis thaliana*, the seeds of *A. thaliana* ecotypes Columbia and Tibetan were grown on filter paper with distilled water at 22℃ with 16hr light/8hr dark cycles. The seedlings (4-5 days old, about 1cm long) were transferred on filter paper with 0.002M hydroxyquinoline solution for 2 hours at room temperature and for 2 hours at 4℃ in the dark. Roots of seedlings were submerged in hydroxyquinoline solution. The root tips were isolated and fixed in ethanol/glacial acetic acid (3:1) for one hour at room temperature and overnight at 4℃. The root tissue was then incubated in carbolfuchsin solution for 2-3 hours at room temperature. The root tips were transferred onto a slide, and covered by a cover slip. Most of the residual staining solution was removed by once strongly pressing a flat hand on the slide (with cover slip) wrapped with filter paper. The preparations were viewed in a light microscope.

The pollen of Tibet-0 was also observed. Seeds of wild Tibet-0 were grown under a 16h light (22℃)/8 h dark (18℃) photoperiod regime until blossoming occurred. The flower bud of Tibet-0 (about 0.8–1mm) was collected at 10 am, fixed with Carony's fixative (ethanol: acetic acid =3:1) for 24 h and then stored in 70% alcohol at 4℃. Following cleaning with distilled water, the pollen mother cell was then squeezed out from anther to a slide, with one drop of mixed enzyme containing 0.3% cellulase, 0.3% pectinase, and 0.5% snailase at 37℃ for 10 min. Using filter paper to remove the excess enzyme solution, we then added a drop of dyeing liquid (1.5ug/ml DAPI, H-1500.vector) and covered the sample with coverslips, pressing it slightly. After staining in dark for 10 min, we observed and photographed the sample under a fluorescence microscope.

**Genome-wide resequencing**

The total DNA was extracted by a TIANGEN Plant Genomic DNA Kit, and genome-wide sequenced with a mean coverage of 40x the reference genomes Col-0, TAIR10 which are available at ftp://ftp.arabidopsis.org/Genes/TAIR10\_genome\_release/, by an Illumina HiSeq 2000 system.

**Sequence preparation and orthologus identification**

The gene annotation of the reference genome of *A. thaliana* was downloaded from [https://www.arabidopsis.org](https://www.arabidopsis.org/download_files/Genes/TAIR10_genome_release/TAIR10_gff3/TAIR10_GFF3_genes.gff). The CDs sequence of other 47 *A. thaliana* ecotypes including Col-0, Bur-0, Ct-1, Edi-0, Hi-0, Kn-0, Ler-0, Mt-0, No-0, Po-0, Oy-0, Rsch-4, Sf-2, Tsu-0, Wil-2, Ws-0, Wu-0, Zu-0, Kas-1, Kas-2, Altai-5 and 26 relicts including Qar-8a, Etna-2, IP-Alm-0, IP-Cor-0, IP-Sne-0, IP-Con-0, Don-0, IP-Per-0, IP-Gra-0, IP-Moj-0, IP-Her-12, IP-Lso-0, IP-Cat-0, IP-Mar-1, IP-Vis-0, IP-Cem-0, IP-Vim-0, IP-Hum-2, Ped-0, IP-Nac-0, IP-Pun-0, IP-Ven-0, IP-Fun-0, Can-0, Cvi-0, IP-Iso-4 (Consortium 2016) were downloaded from http://1001genomes.org/. The A. lyrata is an outcrossing perennial relative of *A. thaliana*. The sequence of A. lyrata was downloaded from Phytozome v9.1 (http://www.phytozome.net/) as outgroup. We first concatenated the CDS sequence of each gene according to its direction. We then searched the orthologous genes of the Tibet-0, 47 other *A. thaliana* ecotypes (26 relicts, Col-0, Bur-0, Can-0, Ct-1, Edi-0 Hi-0, Kn-0, Ler-0, Mt-0, No-0, Po-0, Oy-0, Rsch-4, Sf-2, Tsu-0, Wil-2, Ws-0, Wu-0, and Zu-0) and A. lyrata by a bidirectional-best-hit method in BLAST, and obtained a total of 5741 single-copy orthologous genes. The genes of which less than half of the total CDS length were covered were then eliminated. Finally, we aligned 5611 orthologous genes in 48 *A. thaliana* ecotypes.

**Nucleotide Diversity**

The scaled mutation rate θω = 4Neμ was estimated by using the proportion of segregation sites θs (Watterson 1975) and the average pairwise nucleotide diversity θπ (Tajima 1989), which derived to the Tajima’s D to demonstrate the distribution of the segregating sites. Putative genes were included in the analyses only if resequence data covered more than 50% of the putative coding sequences from all the accessions. As a result, a total of 5611 single copy orthologous genes of 48 *A. thaliana* ecotypes were applied in the calculation. Among which, 103 genes containing no segregating sites were eliminated in the Tajima’s D deduction. Analyses were conducted of the tajima\_d function provided from DendroPy Python library (Sukumaran and Holder 2010), available R codes, or custom R or Python scripts.

**Phylogenetic analysis**

We aligned each of the 5611 orthologues genes of 48 *A. thaliana* ecotypes and concatenated them to long orthologous alignments to construct the maximum-likelihood (ML) tree, using RAxML v8.0 (Stamatakis 2014) . The ML tree applied a partition model to take account of the different evolution rates of the first, second and third position of the codon. GTRGAMMA model was used for nucleotide substitution model. The confidence of the tree topology was evaluated by the bootstrap method with 100 replications. The time of the most recent common ancestor (tMRCA) was calculated by the Bayesian coalescent method with the BP&P program (Rannala and Yang 2003, Yang and Rannala 2010). We then verified the accuracy of the phylogenetic tree by calculating the time of the most recent common ancestor (tMRCA) by the coalescent method (Song, Liu et al. 2012, Nakagome, Nakajima et al. 2013). Single-copy orthologous genes in 47 species and longer than 1000bp were extracted, and the tMRCA48 value of each gene was obtained using Bpp software (Yang 2015). We then removed each ecotype to form 48 new datasets and calculated the tMRCA47 values. For each dataset, the difference between each tMRCA47 and tMRCA48 was detected by using one tail distribution pairwise t-test.

**Divergence time estimation**

We used two steps to estimate the divergence time of Tibet-0 and other ecotypes. First, we downloaded the published genome sequence of 23 species including 21 flowering, 1 fern (Selaginella moellendorffii) and 1 moss (Physcomitrella patens) from Phytozome v9.1 (<http://www.phytozome.net/>) (Goodstein, Shu et al. 2012) before March. 24th, 2014, and downloaded the whole genome sequence of Norway spruce (Picea abies) from Spruce Genome Project (<ftp://congenie.org/congenie/>) (Supplementary Table 2). After bidirectional blast between the sequence of these 24 species and the reference genome of Col-0, and removing paralogue contamination, we obtained 334 single-copy orthologous genes shared by 22 species or more. We concatenated the aligned 334 genes in each species and used ‘?’ to fill the deletion. Based on the nucleotide and amino acid sequence super matrix, the phylogenetic ML tree was then constructed using RAxML v8.0 (Stamatakis 2014). The nucleotide ML tree applied a partition model to normalize the difference between evolution rates of the first, second and third position of the codon, and the model for nucleotide substitution rate was GTR+I+G, while the amino acid ML tree was constructed using a GAMMA+LG4XF amino acid substitution rate model. Both trees were bootstrapped for 1000 times. The divergence time was estimated by using MCMCTREE package in PAML software (Yang and Rannala 2006, Rannala and Yang 2007). The model for nucleotide sequence is a GTR substitution rate model, while the model for amino acid sequence is a F84 substitution rate model. We used the relative rate test and the likelihood ratio test to detect each branch of the phylogenetic tree. In this study, 7 fossil calibrations were used to correct the divergence time of each branch (Supplementary Table 3), thus obtain the divergence time between *A. thaliana* and A. lyrata.

Secondly, we estimated the time of common ancestor between Tibet-0 and the other 47 ecotypes using BP&P software based on Bayesian coalescent models (Yang 2002, Rannala and Yang 2003, Burgess and Yang 2008, Yang and Rannala 2010). By using ratio of the tau of Tibet-0 minus other thaliana accessions and the tau of the *A. thaliana* minus A. lyrata, we obtained the ratio of the divergence time between the *A. thaliana* and the A. lyrata and the divergence time between Tibet-0 and other *A. thaliana* accessions. Since we already obtained the divergence time between *A. thaliana* and A. lyrata, the divergence time between the Tibet-0 and other accessions could be calculated by multiplying the ratio and the divergence time between *A. thaliana* and A. lyrata.

**Supplementary Reference**

Burgess, R. and Z. Yang (2008). "Estimation of hominoid ancestral population sizes under bayesian coalescent models incorporating mutation rate variation and sequencing errors." Mol Biol Evol **25**(9): 1979-1994.

Consortium, G. (2016). "1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana." Cell **166**(2): 481-491.

Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam and D. S. Rokhsar (2012). "Phytozome: a comparative platform for green plant genomics." Nucleic Acids Res **40**(Database issue): D1178-1186.

Nakagome, S., Y. Nakajima and S. Mano (2013). "Biogeography revealed by mariner-like transposable element sequences via a Bayesian coalescent approach." J Mol Evol **77**(3): 64-69.

Online, S. s. m. o. S.

Rannala, B. and Z. Yang (2003). "Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci." Genetics **164**(4): 1645-1656.

Rannala, B. and Z. Yang (2007). "Inferring speciation times under an episodic molecular clock." Syst Biol **56**(3): 453-466.

Simon, U. K., S. Trajanoski, T. Kroneis, P. Sedlmayr, C. Guelly and H. Guttenberger (2012). "Accession-specific haplotypes of the internal transcribed spacer region in Arabidopsis thaliana--a means for barcoding populations." Mol Biol Evol **29**(9): 2231-2239.

Song, S., L. Liu, S. V. Edwards and S. Wu (2012). "Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model." Proc Natl Acad Sci U S A **109**(37): 14942-14947.

Stamatakis, A. (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." Bioinformatics **30**(9): 1312-1313.

Sukumaran, J. and M. T. Holder (2010). "DendroPy: a Python library for phylogenetic computing." Bioinformatics **26**(12): 1569-1571.

Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-595.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins (1997). "The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-4882.

Watterson, G. A. (1975). "On the number of segregating sites in genetical models without recombination." Theor Popul Biol **7**(2): 256-276.

Yang, Z. (2002). "Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci." Genetics **162**(4): 1811-1823.

Yang, Z. (2015). "A tutorial of BPP for species tree estimation and species delimitation " Current Zoology **61**: 854-865.

Yang, Z. and B. Rannala (2006). "Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds." Mol Biol Evol **23**(1): 212-226.

Yang, Z. and B. Rannala (2010). "Bayesian species delimitation using multilocus sequence data." Proc Natl Acad Sci U S A **107**(20): 9264-9269.