



# Genetic Adaptation of Giant Lobelias (*Lobelia aberdarica* and *Lobelia telekii*) to Different Altitudes in East African Mountains

Shu-Ying Zhao<sup>1,2</sup>, Ling-Yun Chen<sup>1,2\*</sup>, John K. Muchuku<sup>1,2</sup>, Guang-Wan Hu<sup>2</sup> and Qing-Feng Wang<sup>1,2\*</sup>

<sup>1</sup> Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China, <sup>2</sup> Sino-Africa Joint Research Centre, Chinese Academy of Sciences, Wuhan, China

## OPEN ACCESS

### Edited by:

Frederic J. J. Chain,  
McGill University, Canada

### Reviewed by:

Longjiang Fan,  
Zhejiang University, China  
Xue-jun Ge,  
South China Botanical Garden, China  
Baojun Wu,  
Wayne State University, USA

### \*Correspondence:

Ling-Yun Chen  
lychen83@qq.com;  
Qing-Feng Wang  
qfwang@wbgcas.cn

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 25 January 2016

**Accepted:** 25 March 2016

**Published:** 12 April 2016

### Citation:

Zhao SY, Chen LY, Muchuku JK,  
Hu GW and Wang QF (2016) Genetic  
Adaptation of Giant Lobelias (*Lobelia  
aberdarica* and *Lobelia telekii*)  
to Different Altitudes in East African  
Mountains. *Front. Plant Sci.* 7:488.  
doi: 10.3389/fpls.2016.00488

The giant lobelias in East African mountains are good models for studying molecular mechanisms of adaptation to different altitudes. In this study, we generated RNA-seq data of a middle-altitude species *Lobelia aberdarica* and a high-altitude species *L. telekii*, followed by selective pressure estimation of their orthologous genes. Our aim was to explore the important genes potentially involved in adaptation to different altitudes. About 9.3 Gb of clean nucleotides, 167,929–170,534 unigenes with total lengths of 159,762,099–171,138,936 bp for each of the two species were generated. OrthoMCL method identified 3,049 1:1 orthologous genes (each species was represented by one ortholog). Estimations of non-synonymous to synonymous rate were performed using an approximate method and a maximum likelihood method in PAML. Eighty-five orthologous genes were under positive selection. At least 8 of these genes are possibly involved in DNA repair, response to DNA damage and temperature stimulus, and regulation of gene expression, which hints on how giant lobelias adapt to high altitudinal environment that characterized by cold, low oxygen, and strong ultraviolet radiation. The negatively selected genes are over-represented in Gene Ontology terms of hydrolase, macromolecular complex assembly among others. This study sheds light on understanding the molecular mechanism of adaptation to different altitudes, and provides genomic resources for further studies of giant lobelias.

**Keywords:** *dN/dS* ratio ( $\omega$ ), giant lobelias, natural selection, RNA-Seq, high altitude

## INTRODUCTION

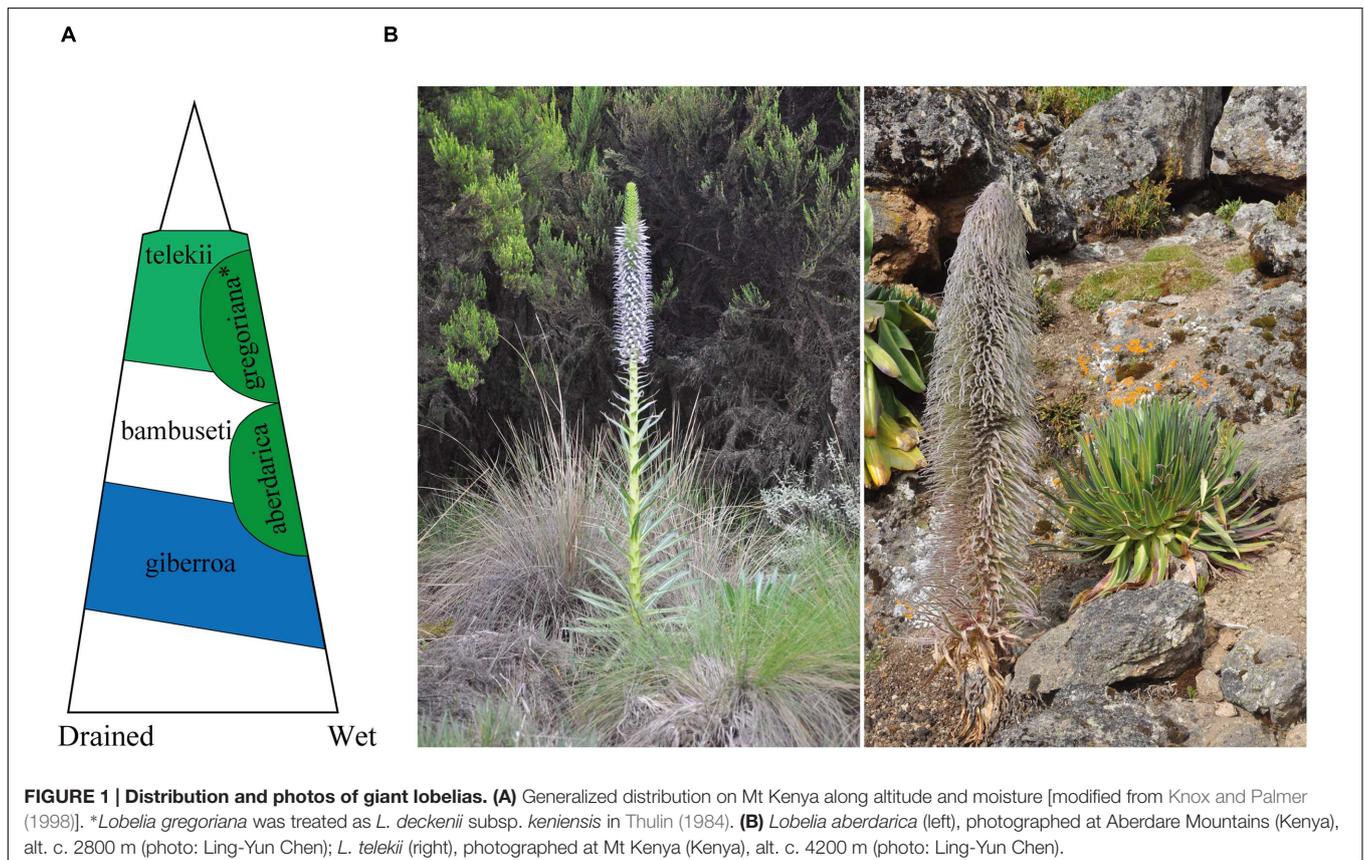
The upland East Africa is characterized by isolated mountains that reach alt. of 4000 m or higher. Vegetation in these mountains displays a conspicuous altitudinal zonation, starting with a montane forest belt, followed by an (subalpine) ericaceous belt, and finally an afro-alpine belt above 3500–4000 m alt. (Hedberg, 1951, 1970). The climate of montane forest is relatively temperate and seasonal, with temperatures falling below 10°C in cold season and rising to above 30°C in warm season. The belt contains moderate levels of species richness, which is higher than the surrounding lowlands (Agnew and Shirley-Agnew, 1994). Typical plants include bamboo, *Hagelia*, *Podocarpus*, etc. Species richness decreases with increase in altitude and fluctuating temperature (Hedberg, 1969). The afro-alpine belt is characterized by an extreme weather pattern with “summer every

day and winter every night” [intense insolation in daytime and heavy frost at night; Hedberg (1957, 1964)]. The number of vascular plants at afro-alpine belt is significantly reduced, with only 70–150 species at each of the mountains (Hedberg, 1957). Typical plants in afro-alpine zone include the well-known giant senecios, giant lobelias among others.

Giant lobelias (Lobeliaceae) in East African mountains are good models for studying plant adaptation to different altitudes. Giant lobelias are perennial, rosette forming herbs and gradiently occur at different ecological belts of East African mountains (Thulin, 1984). The group represents an iconic example of plant adaptation to alpine conditions (Hedberg, 1964, 1969) and a conspicuous landscape of East African mountains. Five species of giant lobelias occur in Kenya and northern Tanzania (according to our observation at Mt Elgon, Cherangani hills, Aberdare mountains, Mt Kenya, Mt Meru, and Mt Kilimanjaro; **Figure 1** illustrated the general distribution of giant lobelias at Mt Kenya and photos of *L. aberdarica* and *L. telekii*). *L. telekii* Schweinf. occurs in the afro-alpine zone from alt. 3400 m to a hostile environment at high altitude (alt. 4640 m, Mt Kenya). This species is also considered to reach the highest distribution in altitude of giant lobelias in Africa. *L. aberdarica* R.E.Fr. & T.C.E.Fr. commonly occurs in moorland, high lands along streams, surrounding marshy area or mountain bogs, and montane forest edges from 2360 m to 3300 m (**Figure 1**). *L. gibberroa*

Hemsl. and *L. bambuseti* R.E.Fr. & T.C.E.Fr. occur in montane forest belt. *L. deckenii* (Asch.) Hemsl. occurs in the afro-alpine belt from ericaceous zone at lower altitude to lower edge of upper alpine zone (3300–4380 m). Recent studies suggested that these species are closely related, and the alpine ones were recently derived from the low altitude relatives during the Pliocene and Pleistocene (4.0–0.8 million years ago [Ma]) (Knox and Palmer, 1998; Chen et al., 2016). Chromosome number for the five species are  $2n = 28$  (Knox and Kowal, 1993).

Hedberg (1964) and Beck et al. (1982) investigated the adaptive trends in the afro-alpine flora which included *L. telekii*, *L. deckenii*, and giant senecios. These rosette plants have evolved to present a conspicuous structure good for temperature insulation. In daytime, most of their leaves unfold for photosynthesis, whereas at night they are folded up and become firmly compressed, forming a compact cabbage-like head, which maintains temperature above freezing (Hedberg, 1964). Flowers of *L. telekii* are concealed among long, hairy bracts, which can buffer vigorous daily temperature fluctuations in hostile alpine environment (Hedberg, 1964). These features were not observed in the mountain forest species such as *L. aberdarica*. The progressive adaptation of giant lobelias to afro-alpine conditions might have been facilitated by extensive volcanism through creating new habitats (Hedberg, 1970), and by induced mutations in flower buds through radiant heat



shocks (Pettersson, 1961). Although previous works shed light on understanding the adaptive evolution of giant lobelias to different altitudes (Hedberg, 1951, 1957, 1964, 1969, 1970; Beck et al., 1982; Knox and Kowal, 1993; Knox and Palmer, 1998), the genes that could be involved in the adaptation remain unknown.

Acquisition of advantageous mutations by positive selection has been associated with adaptation to differentiated environments (Clark et al., 2007; Zhang et al., 2013; Poppe et al., 2015). Negative (purifying) selection plays important roles in maintaining the stability of biological structures by removal of alleles that are deleterious (Loewe, 2008). Positive and negative selection can be inferred by estimating the ratio of non-synonymous substitution rate to synonymous substitution rate ( $dN/dS$ , equivalent to  $\omega$ ) (Yang, 1998). Facilitated by next generation sequencing technology, the genetic basis of human and animal adaptation to different altitudes has been largely investigated by genome comparison (Yi et al., 2010) and assessing the selective pressure of orthologous genes (Simonson et al., 2010; Qiu et al., 2012; Qu et al., 2013). However, the genetic basis of plant adaptation to different altitudes has been poorly studied (but see Chapman et al., 2013; Zhang et al., 2013). Zhang et al. (2013) compared the RNA-seq data of *Primula poissonii* and *P. wilsonii*. Nevertheless, the distribution altitudes of the two species are similar<sup>1</sup> (eFlora of China).

As part of a suite of works to explore the molecular mechanism of plant adaptation to high altitude, we here generated RNA-seq data for *L. aberdarica* and its closest alpine relative *L. telekii* (Knox and Palmer, 1998; Chen et al., 2016), and tested the selective pressure in orthologs of the two species. Our aims were to (1) increase the limited genetic resources of African mountain plants, and (2) identify candidate genes involved in adaptation to different altitudes by analyzing functions of the positively selected genes (PSGs) and environmental differences of the two species.

## MATERIALS AND METHODS

### Materials and Sequencing

Seeds of *L. aberdarica* and *L. telekii* were collected from Aberdare mountains (00°31'27.27" S; 36°43'17.13 E; 2925 m alt.) and Mt Kenya (00° 08' 12. 65" S; 037°21'16.60" E; 4214 m alt.) respectively, in July 2014. The corresponding specimens (SAJIT-P.P1 and SAJIT-002116) were deposited at Botanische Staatssammlung München (M) and herbarium of Wuhan Botanical Garden (HIB), respectively. Tissues preserved in RNAlater storage solution did not yield high quality total RNA in our preliminary study. Therefore, seeds were grown in a plant growth incubator for 4 months with day temperature of c. 15°C and night temperature of c. 10°C. Whole plant of one individual for each species was used for total RNA extraction using RNaiso TMPlus (Takara, Qingdao, China) and then treated with RNase-free DNase I (Takara, Qingdao, China) for 45 min. Quality of RNA was checked using 2% agarose gel

electrophoresis. Double stranded cDNA was sequenced using the Illumina HiSeq<sup>TM</sup> 2000 sequencer (90 bp paired-end) in Beijing Genomics Institute (Wuhan, China) following the methodology in Chen et al. (2015).

### Assembling and Functional Annotation

Raw reads were cleaned by removing adaptor sequences, reads with unknown base calls (N) more than 5%, and low quality reads (>20% of the bases with a quality score  $\leq 10$ ) using Filterfq (an internal program of Beijing Genomics Institute). *De novo* assembly was carried out with the program Trinity v. 20130225 (Grabherr et al., 2011). Contigs were assembled to unigenes by Trinity using pair-end information. The unigenes were further processed by the TGI Clustering Tool (TGICL) v. 2.1 (Perteza et al., 2003) to remove redundancies, and assembled to acquire non-redundant unigenes as long as possible. Overlaps of at least 40 bp, and maximum length of unmatched overhangs of 20 bp were used in parameters of TGICL.

The non-redundant unigenes of the two species were annotated to NCBI non-redundant protein database (NR), Swiss-Prot protein database<sup>2</sup> (Swiss-Prot), Kyoto Encyclopedia of Genes and Genomes<sup>3</sup> (KEGG), Cluster of Orthologous Groups database<sup>4</sup> (COG), and Gene Ontology (GO) using BLASTX ( $E$ -value  $< 10^{-5}$ ). The unigenes were also annotated to NCBI nucleotide database (NT) by BLASTN ( $E$ -value  $< 10^{-5}$ ). Sequence direction of the unigenes was determined using the best aligning results between the unigenes and the protein databases. Incongruent results from different databases were settled by a priority order of NR, Swiss-Prot, KEGG, and COG. Coding sequences (CDSs) of the unigenes were predicted by firstly aligning to NR, then Swiss-Prot, then KEGG and finally COG with BLASTX. CDSs and protein sequences were predicted by using ESTScan v. 1.1 (Iseli et al., 1999).

### Identification of Orthologous Genes

The orthologous genes between *L. aberdarica* and *L. telekii* were identified using the program OrthoMCL v.1.4 (Li et al., 2003) with all-against-all BLASTP comparisons of the predicted protein sequence. Putative orthologous relationships were identified between pairs of genomes by reciprocal best similarity pairs (Li et al., 2003). A third species can increase efficiency in computational screening for orthologs (Lee et al., 2002; Li et al., 2003; Wu et al., 2006). Therefore, protein sequences of *Vitis vinifera*, which show high similarity to the two lobelia species, were downloaded from Genoscope<sup>5</sup>. Redundant sequences of *V. vinifera* were removed, and used for ortholog identification. Only orthologous clusters with a single gene for each of the three species and CDS length longer than 150 bp were kept. One strategy was further used to exclude possible paralogs: a local BLAST database was constructed using protein sequences accessed from NCBI (August 2015) with the software NCBI blast+ v. 2.2.31 (Camacho et al., 2009). Protein sequences of 19

<sup>2</sup><http://www.expasy.ch/sprot>

<sup>3</sup><http://www.genome.jp/kegg/>

<sup>4</sup><http://www.ncbi.nlm.nih.gov/COG/>

<sup>5</sup><http://www.genoscope.cns.fr/spip/>

<sup>1</sup>[http://www.efloras.org/flora\\_page.aspx?flora\\_id=2](http://www.efloras.org/flora_page.aspx?flora_id=2)

species, which showed high similarity to sequences of lobelias in our preliminary analyses, were incorporated, viz. *Amborella trichopoda*, *Arabidopsis thaliana*, *Brassica napus*, *Camelina sativa*, *Citrus sinensis*, *Cucumis sativus*, *Elaeis guineensis*, *Fragaria vesca*, *Glycine max*, *Gossypium raimondii*, *Malus domestica*, *Medicago truncatula*, *Nelumbo nucifera*, *Populus trichocarpa*, *Prunus mume*, *Ricinus communis*, *Sesamum indicum*, *V. vinifera*, and *Zea mays*. Orthologs between *L. aberdarica* and *L. telekii* were used as queries to search the local database using BLASTX with  $E$ -value =  $10^{-5}$  and default settings. The matched sequence with the highest score for each gene was kept. If two orthologs within one pair were matched to different sequences, they were excluded from further analyses. After removing the sequences of *V. vinifera*, CDS of the putative orthologs between *L. aberdarica* and *L. telekii* were aligned by MUSCLE (Edgar, 2004) with default parameters. Finally, the aligned sequences were inspected in BioEdit v. 7.1.3 (Hall, 1999). All gaps and codons with mismatches were deleted. Orthologs with mismatches more than 20 bp or stop codons were excluded from further analyses.

### Estimation of Selective Pressure

Pairwise comparison was implemented to test the selective pressure of each gene. An approximate method proposed by Yang and Nielsen (2000) and a maximum likelihood (ML) method were used. The approximate analysis was conducted using yn00 in PAML toolkit v. 4.8 (Yang, 2007). The ML analysis was conducted using codeml in PAML toolkit with seqtype = 1, codonfreq = 2, runmode = -2. After preliminary analyses, orthologs with  $dS > 0.1$  (Bustamante et al., 2005), or  $\omega > 98$  caused by extremely low  $dS$  value, were excluded to avoid potential paralogs and bias on results of selective pressure estimation.

### Functional Annotations

All orthologs were annotated to a local protein database using BLASTX in NCBI blast+ v. 2.2.31. The local database was constructed by blast+ v. 2.2.31 using all protein sequences of *Arabidopsis thaliana* accessed from NCBI (August 2015).  $E$ -value of  $1.0 \times 10^{-5}$  and 1 BLAST hit were used. GO terms for each sequence were obtained by converting 'GenBank Protein Accession' to 'GO ID' using the website BioDBnet<sup>6</sup>. Orthologs were divided into three datasets: one included (PSGs,  $98 > \omega > 1$ ), one included negatively selected genes (NSGs,  $0 < \omega < 1$ ), and one included strongly negatively selected genes (SNSGs,  $\omega = 0$ ). To detect which molecular functions, biological processes and cellular component were over-represented, we compared the GO terms among the three datasets using WEGO (Ye et al., 2006). GO enrichment analysis was also conducted using agriGO (Du et al., 2010).

Only two species were used in the estimation of selective pressure. Therefore, the genes under positive selection might represent adaptation to different altitudes, or other species-specific traits unrelated to altitude adaptation. In order to find out the candidates for adaptation to different altitudes, we accessed the possible functions of the positively selected genes using literature

searches and gene annotations in The *Arabidopsis* Information Resource (TAIR). The positively selected genes were used as queries to search 'TAIR10 Proteins' dataset using BLASTX with default settings<sup>7</sup>. Annotations of *Arabidopsis* genes with the highest scores were used. The positively selected genes with possible functions matching the environmental differences such as cold stimulus were identified as candidates for adaptation to different altitudes.

Extremely high ( $\geq 80\%$ ) or low GC ( $\leq 25\%$ ) content have low power to accurately estimate selective pressure (Gharib and Robinson-Rechavi, 2013). GC content for each ortholog was calculated using perl scripts to check whether the value is extremely high or low. Mean GC content for the three datasets was also calculated.

## RESULTS

### De Novo Assembly and Annotation of Unigenes

We generated c. 104 million clean reads, c. 9.3 Gb of nucleotides for *L. aberdarica* and *L. telekii* separately. The clean reads were submitted to the NCBI Sequence Reads Archive (nos. SRR3180742 and SRR3180743). The unigenes, which were assembled by using contigs, were 951 and 1004 bp on average with N50 of 1,951 and 1,997 bp, respectively, for the two species (Supplementary Table S1).

All the non-redundant unigenes were annotated using NR, Swiss-Prot, KEGG, COG, and NT. The results indicated that 79,825 unigenes of *L. aberdarica* (48%) and 84,296 unigenes of *L. telekii* (49%) have significant matches ( $E$ -value  $< 10^{-5}$ ). NR has the highest proportion of successful annotations, while COG has the lowest proportion. The three top-hits for the two species in the NR database were *V. vinifera*, *Lycopersicon esculentum*, and *Amygdalus persica* (Supplementary Figure S1).

### Selective Pressure Analyses

In total, 3,978 pairs of putative 1:1 orthologous genes between *L. aberdarica* and *L. telekii* were identified by OrthoMCL. After removing the pairs that were not consistent in the BLAST analyses, 3,182 pairs were retained. After removing the pairs with unexpected stop codons and mismatches more than 20 bp, 3049 pairs were retained. Sequences for the 3049 pairs were provided in Supplementary data.

One hundred and sixty-eight (5.5%) pairs of orthologous genes with  $98 > \omega > 1$  was recovered with the approximate analysis while 86 (2.8%) pairs was recovered with the ML analysis. For conciseness, one orthologous pair was counted as one gene hereafter. 85 (2.8%) pairs with  $98 > \omega > 1$  were shared between the two analyses. 2357 (77.3%) pairs with  $0 < \omega < 1$  were recovered in the two analyses. 480 pairs (15.7%) with  $\omega = 0$  were recovered in the two analyses.  $\omega = 0$  (caused by  $dN = 0$ ) indicates stringent negative selection. **Table 1** summarizes the results of the two analyses; **Figure 2** plots the distribution of the values of  $dN$  and  $dS$ .

<sup>6</sup><http://biodbnet.abcc.ncifcrf.gov/>

<sup>7</sup><https://www.arabidopsis.org/Blast/index.jsp>

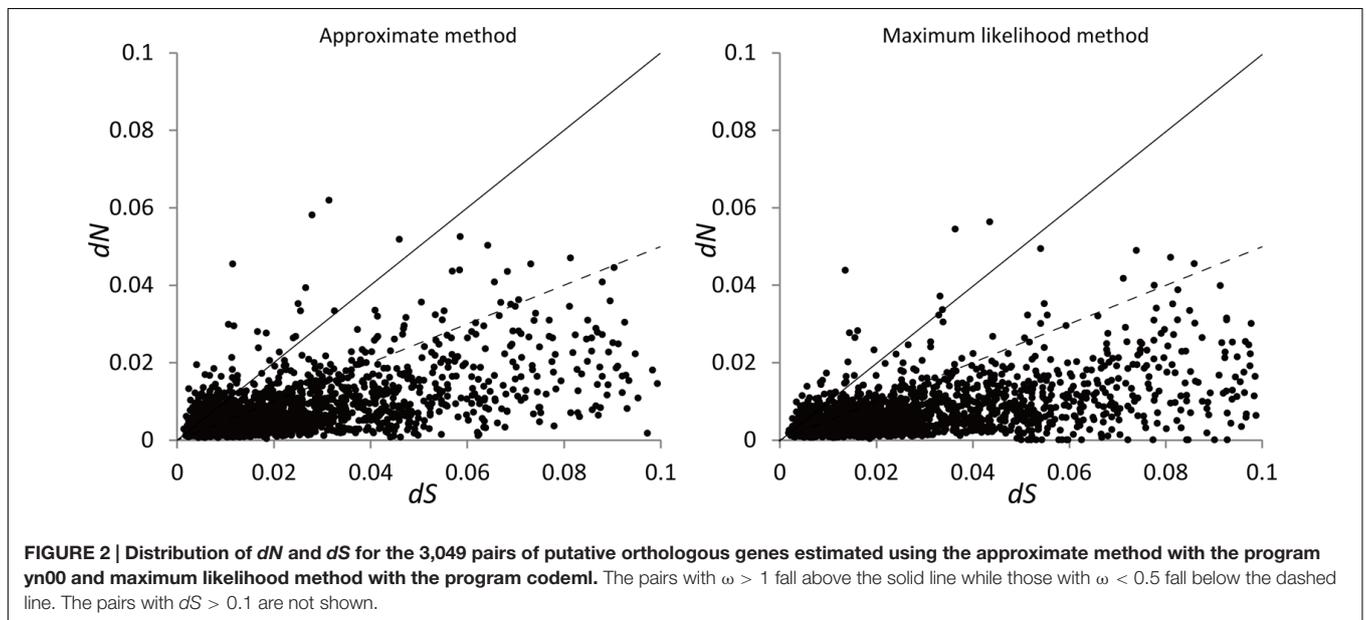
**TABLE 1 | Summary of the selective pressure analyses.**

	$1 < \omega < 98$	$0 < \omega < 1$	$dN$ or $dS = 0$
Approximate analysis	168	2358	480
ML analysis	86	2483	523
Co-existed between two analyses	85 (dataset 1)	2357 (dataset 2)	480 (dataset 4)
GC content (%)	46.3 (39.4–55.0)	45.7 (38.6–60.6)	46.2 (35.4–62.3)

Numbers of orthologous pairs recovered from approximate analysis implemented in *yn00* and maximum likelihood (ML) analysis implemented in *codeml*, numbers of pairs shared between the two analyses, and GC contents of the orthologs are shown.

## Functional Annotations

Some of the 85 positively selected genes might be related to species-specific traits rather than altitude adaptation as only two species were used in our analyses. Annotations suggested that at least 8 of the 85 positively selected genes (PSGs) are involved in functions including DNA repair, regulation of photosynthesis, response to cold, light, or desiccation stimuli (Table 2). Functions of the eight genes can be associated with the environments in which giant lobelias inhabit (Table 2). For example, *CL8493.Contig*, which encodes a putative psbP protein, is essential for regulation and stabilization of photosystem II in higher plants (Huynh et al., 2005); *Unigene78063*, which

**TABLE 2 | Candidate genes for adaptation to different altitudes.**

Seq. ID	<i>Arabidopsis</i> accession nos.	E-value (BLASTX)	Gene or protein names	Possible functions and biological process
<i>CL11124.Contig1</i>	AT5G43210.1	3e-39	Excinuclease ABC, C subunit, N-terminal	Nuclease activity; Involved in: DNA repair
<i>CL10902.Contig2</i>	AT4G25130.1	7e-90	A chloroplast-localized methionine sulfoxide reductase, a member of the MSRA family	Involved in: cellular protein modification process, cellular response to oxidative stress, protein repair, response to cytokinin, response to light stimulus
<i>CL6679.Contig</i>	AT5G62390.1	1e-83	A member of <i>Arabidopsis</i> BAG proteins	Involved in: apoptotic process, cellular response to cold, cellular response to heat, cellular response to unfolded protein, protein folding
<i>CL14161.Contig</i>	AT1G05260.1	1e-76	Encodes a cold-inducible cationic peroxidase that is Involved in: the stress response	Involved in: hyperosmotic salinity response, plant-type cell wall organization, response to cold, response to desiccation
<i>Unigene26608</i>	AT3G02820.1	6e-62	Zinc knuckle (CCHC-type) family protein	Involved in: cell cycle, replication fork protection, response to DNA damage stimulus
<i>CL8493.Contig</i>	AT1G77090.1	9e-87	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein	Involved in: photosynthesis; Located in: chloroplast
<i>Unigene78063</i>	AT5G48130.1	3e-12	Phototropic-responsive NPH3 family protein (BTB/POZ-like)	Functions in: signal transducer activity; Involved in: response to light stimulus; Located in: chloroplast
<i>Unigene29384</i>	AT1G09540.1	2e-82	Encodes putative transcription factor	Involved in: regulation of stomatal movement, regulation of transcription, DNA-templated, response to auxin, root development

The listed genes are a subset of the 85 positively selected genes based on annotations of TAIR, and a literature search. Since the two orthologs within an orthologous pair matched to same sequence, only sequence ID and E-values for sequences of *Lobelia aberdarica* are provided.

encodes putative phototropic-responsive NPH3 family protein, is located in chloroplast and involved in response to light stimulus (Motchoulski and Liscum, 1999). Four of the 85 PSGs are putative zinc finger family proteins, including *Unigene36113*, *CL11459.Contig1*, *Unigene40961*, and *CL22102.Contig1*. These PSGs are multifunctional genes, involved in the regulation of transcription, nucleic acid binding among other functions. See Supplementary Table S2 for detail information of the 85 orthologous genes.

Gene Ontology analyses using WEGO suggested that there were significant differences (Pearson Chi-Square test,  $P < 0.05$ ) among the three orthologous datasets at GO levels 3, 4, and 5. The negatively selected genes ( $0 < \omega < 1$ ) and strongly negatively selected genes ( $\omega = 0$ ) were over-represented in cytoplasmic part, hydrolase activity among others than the positively selected genes. The strongly negatively selected genes were over-represented in mitochondrial envelope, cellular component biogenesis among others than the negatively selected genes. The PSGs ( $98 > \omega > 1$ ) had a significantly higher percentage of genes with functions related to cellular response to stimulus, pyridine metabolic process, and regulation of nitrogen compound metabolic process than the strongly negatively selected genes (see Supplementary Table S3 and **Figure 3**).

Gene Ontology enrichment analysis using agriGO suggested that there was no significant enrichment (Fisher test,  $P$  value  $< 0.05$ ) at the secondary level among the three ortholog datasets. The PSGs had a higher percentage of genes with functions related to cell part, organelle part, transcription regulator activity, and binding than the other two datasets but with  $P$  value  $> 0.05$ .

GC content for ortholog ranges from 35.4 to 62.3%. No orthologs have extremely high or low GC content ( $\geq 80$  or  $\leq 25\%$ ). Average content for positively selected genes, negatively selected genes and strongly negatively selected genes are 46.3, 45.7, and 46.2%, respectively (**Table 1**).

## DISCUSSION

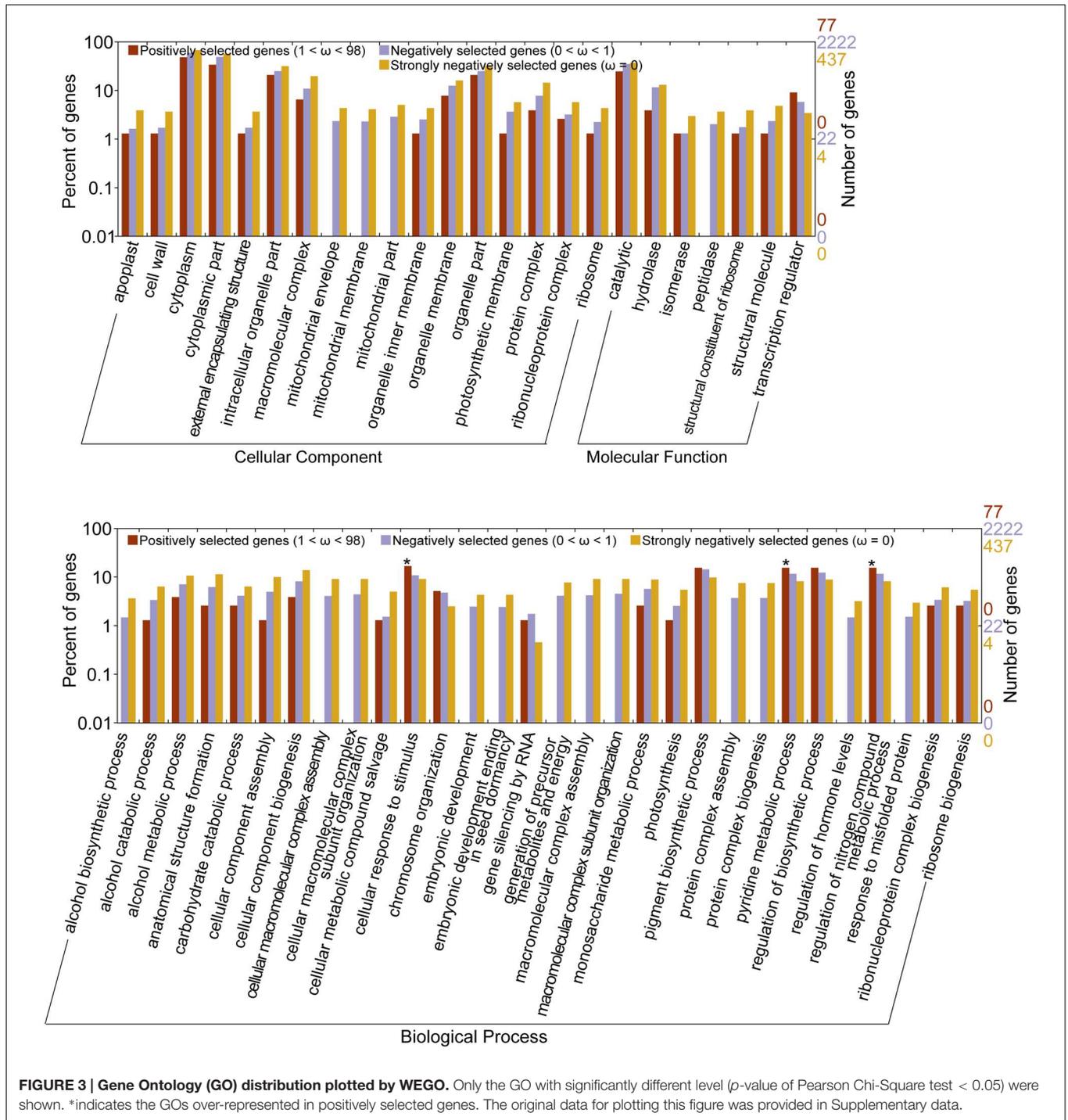
In this study, more than 100 million of RNA-seq reads were generated and assembled into c. 160,000 unique sequences for each of *L. aberdarica* and *L. telekii*. The RNA-seq data are informative for SSR marker development and population genetic studies of the giant lobelias.

Selective pressure estimation using approximate analysis and ML analysis suggested that 168 and 86 pairs of orthologous genes were under positive selection, respectively. Taken into consideration that both methods have weaknesses [such as the  $\omega$  biases when there are transition/transversion rate biases (Yang and Nielsen, 2000)], only the 85 pairs recovered in both methods were identified as PSG. *L. aberdarica* occurs in relatively mild habitats while *L. telekii* occurs in habitats with extremely strong ultraviolet radiation, fluctuating temperature and low oxygen. At least 8 of the 85 genes might be involved in adaptation to the different environments of *L. aberdarica* and *L. telekii* based on literature searches (**Table 2** and Supplementary Table S2).

*CL11124.Contig1* is a putative endonuclease subunit *slx1* gene, which is required for homologous recombination and DNA repair in eukaryotic cells (Castor et al., 2013; Gaur et al., 2015). *Unigene26608* is a putative zinc knuckle family protein, which responds to DNA damage stimulus (Srivastava et al., 2010). In consideration of the DNA damage or mutation of *L. telekii* caused by heat shocks of volcanic eruptions (Pettersson, 1961), UV and frost (Hedberg, 1969; Sinha and Häder, 2002), these genes are likely important in the genetic adaptation of this species. Similarly, transcriptome analyses of *Primula poissonii* (collected at alt. 3314 m) and *P. wilsonii* (collected at alt. 2450 m) indicated that several genes related to DNA repair and homologous recombination were under positive selection and over-represented (Zhang et al., 2013). In contrast, no negatively selected genes were indicated as putative *slx* genes or zinc knuckle family genes, thus highlighted the roles of the two genes.

Photosynthetic capacity of *L. telekii* is sensitive to temperature change (Bodner and Beck, 1987). Photosynthesis of *L. telekii* was found to be strongly reduced when temperature above 15°C while frozen leaves regained full photosynthesis immediately after thawing (Schulze et al., 1985). *CL6679.Contig* is a putative BAG family molecular chaperone regulator seven gene (*AtBAG7*) that is an essential component of the unfolded protein response during heat and cold tolerance (Williams et al., 2010). *AtBAG7* knockouts are sensitive to heat and cold stimuli (Williams et al., 2010). *CL14161.Contig* is a putative cold inducible cationic peroxidase gene, which is in response to cold or desiccation stimulus. In this study, GO enrichment analyses also indicated that the GO terms associated with response to stimulus are over-represented in the positively selected genes (**Figure 3**). The functions of these genes are consistent with the extremely cold and sometimes dry condition that *L. telekii* faces and the peculiar mechanism of frost avoidance and freezing tolerance of the species (Hedberg, 1964; Beck et al., 1982). *CL10902.Contig2* is a putative peptide-methionine sulfoxide reductase (PMSR) gene, which is involved in cellular response to oxidative stress, oxidation-reduction process, protein repair, and response to light stimulus (Gustavsson et al., 2002; Romero et al., 2004). Zhang et al. (2013) found that several orthologs related to abiotic stress (such as oxidative stress) were under positive selection, which is consistent with our results.

Several orthologs are transcription factors involved in regulation of gene expression or DNA binding. For example, *Unigene29384* encodes a putative transcription factor, which regulates stomatal movement and transcription (**Table 2** and Supplementary Table S2). Several zinc finger proteins (ZFPs) were identified in this study. ZFPs are a large family of transcription regulators in plants for modulating several stress-responsive genes. In *Arabidopsis*, functions of C<sub>2</sub>H<sub>2</sub> type ZFPs, including *ZAT6*, *ZAT10*, and *ZAT12*, have been well characterized. Previous studies have shown that modulation of ZFPs regulate plant responses to abiotic stresses including cold and drought (Davletova et al., 2005; Mittler et al., 2006; Shi et al., 2014). Since *L. aberdarica* and *L. telekii* occur in different altitudes of East African mountains, ZFPs might be involved in responses to temperature and humidity.



All the above genes might contribute to the adaptive evolution of the two species. However, it is possible that some of the identified PSGs are species-specific traits unrelated to habitat adaptation. On the other hand, some of the genes involved in habitat differences may remain obscure due to limited annotations in giant lobelias.

The main consequence of negative selection is the extinction of less-adapted variants (Loewe, 2008). The negatively

selected genes are important for maintaining key functions of *L. aberdarica* and *L. telekii*. GO classification using WEGO suggested that the negatively selected genes were constituted of genes including functions on hydrolase, macromolecular complex assembly, and generation of precursor metabolites and energy with a higher percentage than the PSGs. These functions and processes might be involved in maintaining normal growth of the two species.

The molecular mechanisms of plant adaptation to different altitudes are complex, very few studies have explored this field. In this study, we identified 85 positively selected genes between an afro-montane forest species *L. aberdarica* and an upper afro-alpine distributed species *L. telekii*. At least 8 genes related to cold stimulus, DNA repair and regulation of gene expression were positively selected, which are consistent with characters of high altitudinal environments. This study is a tentative attempt to explore the complex molecular mechanism of plant adaptation to different altitudes using non-model plants. However, it is still limited in sampling as only two species were used, and our pairwise approach cannot distinguish in which lineage positive selection occurred. Moreover, gene turnover and expression level could also reveal environmental adaptation that is not detectable using *dN/dS* analysis. Further studies with genome sequencing and more species of African giant lobelias will be carried out to explore the adaptation to different altitudes.

## AUTHOR CONTRIBUTIONS

SYZ and QFW conceived this study; SYZ, JM, GWH, and LYC carried out experimental works, data analyses and drafted the

manuscript; LYC and QFW revised the manuscript. All authors gave final approval of the version to be published.

## FUNDING

This research was supported by Grants from the National Natural Science Foundation of China (no. 31500192), the Chinese Academy of Sciences (no. XDAO5090305), and the Special Foundation for State Basic Working Program of China (no. 2013FY112300).

## ACKNOWLEDGMENTS

We thank Zhi-Duan Chen for discussion on this research; Zhu-Long Chan and Can Dai for revising this manuscript; Ya-Dong Zhou and Zhi-Xiang Zhong for assistance in plant collection.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00488>

## REFERENCES

- Agnew, A. D. Q., and Shirley-Agnew. (1994). *Upland Kenya Wild Flowers-A Flora of the Ferns and Herbaceous Flowering Plants of Upland Kenya*. Nairobi: East Africa Natural History Society.
- Beck, E., Senser, M., Scheibe, R., Steiger, H., and Pongratz, P. (1982). Frost avoidance and freezing tolerance in Afroalpine 'giant rosette' plants. *Plant Cell Environ.* 5, 215–222. doi: 10.1111/1365-3040.ep11572080
- Bodner, M., and Beck, E. (1987). Effect of supercooling and freezing on photosynthesis in freezing tolerant leaves of Afroalpine 'giant rosette' plants. *Oecologia* 72, 366–371. doi: 10.1007/BF00377565
- Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., et al. (2005). Natural selection on protein-coding genes in the human genome. *Nature* 437, 1153–1157. doi: 10.1038/nature04240
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:e421. doi: 10.1186/1471-2105-10-421
- Castor, D., Nair, N., Déclais, A. C., Lachaud, C., Toth, R., Macartney, T. J., et al. (2013). Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. *Mol. Cell* 52, 221–233. doi: 10.1016/j.molcel.2013.08.036
- Chapman, M. A., Hiscock, S. J., and Filatov, D. A. (2013). Genomic divergence during speciation driven by adaptation to altitude. *Mol. Biol. Evol.* 30, 2553–2567. doi: 10.1093/molbev/mst168
- Chen, L. Y., Renner, S. S., and Wang, Q. F. (2016). East Asian *Lobelioideae* and ancient divergence of a giant rosette *Lobelia* in Himalayan Bhutan. *Taxon*. 1. doi: 10.12705/652.6
- Chen, L. Y., Zhao, S. Y., Wang, Q. F., and Moody, M. L. (2015). Transcriptome sequencing of three *Ranunculus* species (*Ranunculaceae*) reveals candidate genes in adaptation from terrestrial to aquatic habitats. *Sci. Rep.* 5, e10098. doi: 10.1038/srep10098
- Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., et al. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218. doi: 10.1038/nature06341
- Davletova, S., Schlauch, K., Coutu, J., and Mittler, R. (2005). The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. *Plant Physiol.* 139, 847–856. doi: 10.1104/pp.105.068254
- Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, 1–10. doi: 10.1093/nar/gkq310
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Gaur, V., Wyatt, H. D., Komorowska, W., Szczepanowski, R. H., De Sanctis, D., Gorecka, K. M., et al. (2015). Structural and mechanistic analysis of the Slx1-Slx4 endonuclease. *Cell Rep.* 10, 1467–1476. doi: 10.1016/j.celrep.2015.02.019
- Gharib, W. H., and Robinson-Rechavi, M. (2013). The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol. Biol. Evol.* 30, 1675–1686. doi: 10.1093/molbev/mst062
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Gustavsson, N., Kokke, B. P., Härndahl, U., Silow, M., Bechtold, U., Poghosyan, Z., et al. (2002). A peptide methionine sulfoxide reductase highly expressed in photosynthetic tissue in *Arabidopsis thaliana* can protect the chaperone-like activity of a chloroplast-localized small heat shock protein. *Plant J.* 29, 545–553. doi: 10.1046/j.1365-313X.2002.029005545.x
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hedberg, O. (1951). Vegetation belts of the East African mountains. *Svensk. Bot. Tidskr.* 45, 140–202.
- Hedberg, O. (1957). *Afroalpine Vascular Plants: A Taxonomic Revision*. Uppsala: A. B. Lundequistska Bokhandeln.
- Hedberg, O. (1964). Features of afroalpine plant ecology. *Acta Phytogeogr. Suec.* 49, 1–144.
- Hedberg, O. (1969). Evolution and speciation in a tropical high mountain flora. *Biol. J. Linn. Soc. Lond.* 1, 135–148. doi: 10.1111/j.1095-8312.1969.tb01816.x
- Hedberg, O. (1970). Evolution of the afroalpine flora. *Biotropica* 2, 16–23. doi: 10.2307/2989783
- Huynh, L. N., Vantoi, T., Streeter, J., and Banowitz, G. (2005). Regulation of flooding tolerance of *SAG12:ipt Arabidopsis* plants by cytokinin. *J. Exp. Bot.* 56, 1397–1407. doi: 10.1093/jxb/eri141

- Iseli, C., Jongeneel, C. V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1999, 138–148.
- Knox, E. B., and Kowal, R. R. (1993). Chromosome numbers of the East African giant senecios and giant lobelias and their evolutionary significance. *Am. J. Bot.* 847–853. doi: 10.2307/2445604
- Knox, E. B., and Palmer, J. D. (1998). Chloroplast DNA evidence on the origin and radiation of the giant lobelias in eastern Africa. *Syst. Bot.* 23, 109–149. doi: 10.2307/2419583
- Lee, Y., Sultana, R., Pertea, G., Cho, J., Karamycheva, S., Tsai, J., et al. (2002). Cross-referencing eukaryotic genomes: TIGR orthologous gene alignments (TOGA). *Genome Res.* 12, 493–502. doi: 10.1101/gr.212002
- Li, L., Stoekert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Loewe, L. (2008). Negative selection. *Nat. Education* 1, 59.
- Mittler, R., Kim, Y., Song, L., Coutu, J., Coutu, A., Ciftci-Yilmaz, S., et al. (2006). Gain-and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. *FEBS Lett.* 580, 6537–6542. doi: 10.1016/j.febslet.2006.11.002
- Motchoulski, A., and Liscum, E. (1999). *Arabidopsis* NPH3: a NPH1 photoreceptor-interacting protein essential for phototropism. *Science* 286, 961–964. doi: 10.1126/science.286.5441.961
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19, 651–652. doi: 10.1093/bioinformatics/btg034
- Pettersson, B. (1961). Mutagenic effect of radiant heat shocks on phanerogamous plants. *Nature* 191, 1167–1169. doi: 10.1038/1911167a0
- Poppe, S., Dorsheimer, L., Happel, P., and Stukenbrock, E. H. (2015). Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLoS Pathog.* 11:e1005055. doi: 10.1371/journal.ppat.1005055
- Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., et al. (2012). The yak genome and adaptation to life at high altitude. *Nat. Genet.* 44, 946–949. doi: 10.1038/ng.2343
- Qu, Y., Zhao, H., Han, N., Zhou, G., Song, G., Gao, B., et al. (2013). Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat. Commun.* 4:e2071. doi: 10.1038/ncomms3071
- Romero, H. M., Berlett, B. S., Jensen, P. J., Pell, E. J., and Tien, M. (2004). Investigations into the role of the plastidial peptide methionine sulfoxide reductase in response to oxidative stress in *Arabidopsis*. *Plant Physiol.* 136, 3784–3794. doi: 10.1104/pp.104.046656
- Schulze, E. -D., Beck, E., Scheibe, R., and Ziegler, P. (1985). Carbon dioxide assimilation and stomatal response of afroalpine giant rosette plants. *Oecologia* 65, 207–213. doi: 10.1007/BF00379219
- Shi, H., Wang, X., Ye, T., Chen, F., Deng, J., Yang, P., et al. (2014). The Cysteine2/Histidine2-Type transcription factor ZINC FINGER OF ARABIDOPSIS THALIANA6 modulates biotic and abiotic stress responses by activating salicylic acid-related genes and C-REPEAT-BINDING FACTOR genes in *Arabidopsis*. *Plant Physiol.* 165, 1367–1379. doi: 10.1104/pp.114.242404
- Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., et al. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science* 329, 72–75. doi: 10.1126/science.1189406
- Sinha, R. P., and Häder, D. -P. (2002). UV-induced DNA damage and repair: a review. *Photochem. Photobiol. Sci.* 1, 225–236. doi: 10.1039/b201230h
- Srivastava, G. P., Li, P., Liu, J., and Xu, D. (2010). Identification of transcription factor's targets using tissue-specific transcriptomic data in *Arabidopsis thaliana*. *BMC Syst. Biol.* 4:S2. doi: 10.1186/1752-0509-4-S2-S2
- Thulin, M. (1984). "Lobeliaceae," in *Flora of Tropical East Africa*, ed. R. M. Polhill (Rotterdam: Balkema), 1–59.
- Williams, B., Kabbage, M., Britt, R., and Dickman, M. B. (2010). AtBAG7, an *Arabidopsis* Bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6088–6093. doi: 10.1073/pnas.0912670107
- Wu, F., Mueller, L. A., Crouzillat, D., Pétiard, V., and Tanksley, S. D. (2006). Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* 174, 1407–1420. doi: 10.1534/genetics.106.062455
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573. doi: 10.1093/oxfordjournals.molbev.a025957
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17, 32–43. doi: 10.1093/oxfordjournals.molbev.a026236
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic. Acids Res.* 34, W293–W297. doi: 10.1093/nar/gkl031
- Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329, 75–78. doi: 10.1126/science.1190371
- Zhang, L., Yan, H. F., Wu, W., Yu, H., and Ge, X. J. (2013). Comparative transcriptome analysis and marker development of two closely related Primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genomics* 14:e329. doi: 10.1186/1471-2164-14-329

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhao, Chen, Muchuku, Hu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.