CellPress
PARTNER JOURNAL

Molecular Plant
Research Article

Genome of Plant Maca (*Lepidium meyenii*) Illuminates Genomic Basis for High-Altitude Adaptation in the Central Andes

Jing Zhang[1,14], Yang Tian[2,3,14], Liang Yan[4,14], Guanghui Zhang[5,14], Xiao Wang[6,7], Yan Zeng[6,7], Jiajin Zhang[8], Xiao Ma[3], Yuntao Tan[9], Ni Long[9], Yangzi Wang[9], Yujin Ma[9], Yuqi He[10], Yu Xue[1], Shumei Hao[11], Shengchao Yang[5], Wen Wang[6], Liangsheng Zhang[12,*], Yang Dong[9,13,*], Wei Chen[3,13,*] and Jun Sheng[2,3,4,*]

[1]College of Life Science, Huazhong University of Science and Technology, Wuhan 430074, China

[2]College of Life Sciences, Jilin University, Changchun 130012, China

[3]Key Laboratory of Pu-erh Tea Science, Ministry of Education, Yunnan Agricultural University, Kunming 650201, China

[4]Pu'er Institute of Pu-erh Tea, Pu'er 665000, China

[5]Yunnan Research Center on Good Agricultural Practice for Dominant Chinese Medicinal Materials, Yunnan Agricultural University, Kunming 650201, China

[6]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

[7]University of Chinese Academy of Sciences, Beijing 100049, China

[8]School of Science and Information Engineering, Yunnan Agricultural University, Kunming 650201, China

[9]College of Life Science, Kunming University of Science and Technology, Kunming 650504, China

[10]Public Technical Service Center, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

[11]Yunnan University, Kunming 650091, China

[12]Center for Genomics and Biotechnology, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[13]Yunnan Research Institute for Local Plateau Agriculture and Industry, Kunming 650201, China

[14]These authors contributed equally to this article.

***Correspondence: Liangsheng Zhang (fafuzhang@163.com), Yang Dong (loyalyang@163.com), Wei Chen (wchenntr@gmail.com), Jun Sheng (shengj@ynau.edu.cn)**

http://dx.doi.org/10.1016/j.molp.2016.04.016

ABSTRACT

Maca (*Lepidium meyenii* Walp, 2*n* = 8*x* = 64), belonging to the Brassicaceae family, is an economic plant cultivated in the central Andes sierra in Peru (4000–4500 m). Considering that the rapid uplift of the central Andes occurred 5–10 million years ago (Ma), an evolutionary question arises regarding how plants such as maca acquire high-altitude adaptation within a short geological period. Here, we report the high-quality genome assembly of maca, in which two closely spaced maca-specific whole-genome duplications (WGDs; ~6.7 Ma) were identified. Comparative genomic analysis between maca and closely related Brassicaceae species revealed expansions of maca genes and gene families involved in abiotic stress response, hormone signaling pathway, and secondary metabolite biosynthesis via WGDs. The retention and subsequent functional divergence of many duplicated genes may account for the morphological and physiological changes (i.e., small leaf shape and self-fertility) in maca in a high-altitude environment. In addition, some duplicated maca genes were identified with functions in morphological adaptation (i.e., *LEAF CURLING RESPONSIVENESS*) and abiotic stress response (i.e., *GLYCINE-RICH RNA-BINDING PROTEINS* and *DNA-DAMAGE-REPAIR/TOLERATION 2*) under positive selection. Collectively, the maca genome provides useful information to understand the important roles of WGDs in the high-altitude adaptation of plants in the Andes.

Key words: genome sequencing, whole-genome duplication, high-altitude adaptation, Brassicaceae, *Lepidium*, leaf morphogenesis, self-incompatibility

Zhang J., Tian Y., Yan L., Zhang G., Wang X., Zeng Y., Zhang J., Ma X., Tan Y., Long N., Wang Y., Ma Y., He Y., Xue Y., Hao S., Yang S., Wang W., Zhang L., Dong Y., Chen W., and Sheng J. (2016). Genome of Plant Maca (*Lepidium meyenii*) Illuminates Genomic Basis for High-Altitude Adaptation in the Central Andes. Mol. Plant. **9**, 1066–1077.

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

INTRODUCTION

Maca (*Lepidium meyenii* Walp) is a Brassicaceae crop and herbal plant cultivated in the puna agro-ecological zones (4000–4500 m) in the central Andes (http://ecocrop.fao.org/; Figure 1A). Its potential health benefits, particularly to reproduction and fertility, have drawn great investments from pharmacological and nutritional research in recent years (Piacente et al., 2002; Shin et al., 2010). Due to its restricted cultivation area at high altitude, maca manifests robust endurance to extreme environmental stresses, including cold, strong wind, and UV-B exposure. Even though many genes and gene families involved in abiotic stress response have been identified in various plants (Hoffmann and Willi, 2008), it is still an open question that how plants such as maca adapted to the high-altitude Andean environment from an evolutionary genomics standpoint.

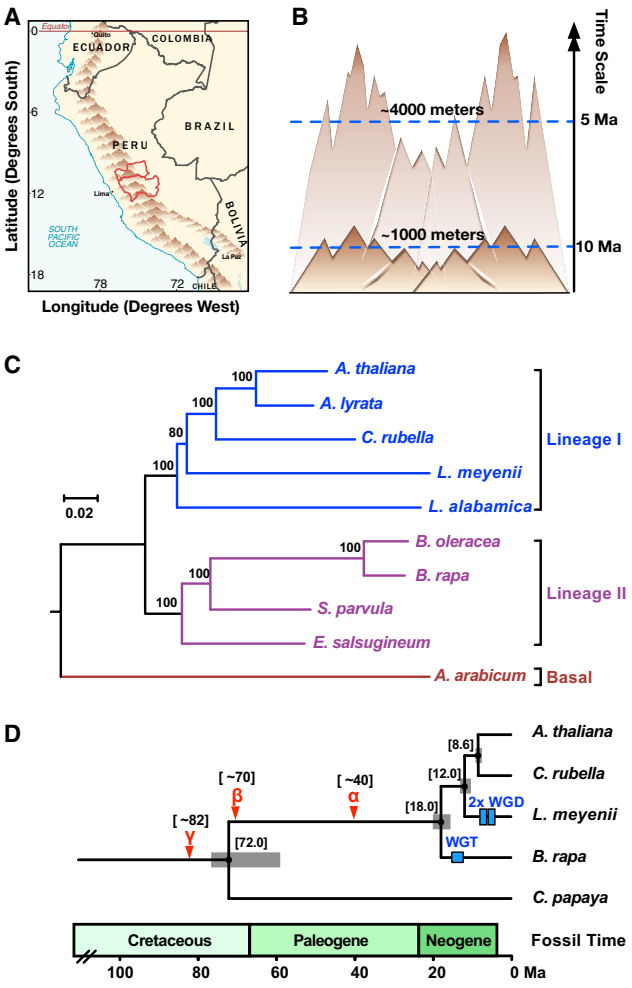


Figure 1. Phylogenetic Analysis of Representative Members in Different Lineages of Brassicaceae Family.

(A) Cultivation area of maca in the central Andes region (circled in red).

(B) Rapid elevation of the Andes Mountains between 5 and 10 million years ago.

(C) Representative members in different lineages of Brassicaceae family. Phylogenetic tree was generated from 464 single-copy orthologous genes using the maximum-likelihood method. The scale bar shows the expected number of amino acid substitutions per residue.

(D) Phylogenetic tree and divergence time of five plant species. Phylogenetic tree was generated from 503 single-copy orthologous genes using the maximum-likelihood method. Time (Ma) is listed in brackets, and corresponds to the fossil time scale. The time range for divergence is shown in gray areas. Approximate positions of the paleo-WGD events (α, β, γ) are indicated by inverted triangles. Approximate positions of the species-specific WGD events are indicated by squares.

The Andes mountain range resides alongside of the west coast of South America. The central part of the Andes, comprising alpine areas, plateaus, highland lakes, and valleys, spans ~400 km wide with an average elevation of 4000 m (Garzione et al., 2008). Regarding the formation of major high mountain ranges in the world, the process was usually steady and progressive over thousands and millions of years. However, the Andes, especially the central Andes, uniquely experienced a spurt of elevation up to 3000 m between 5 and 10 million years ago (Ma) (Figure 1B) (Garzione et al., 2008). This rapid uplift of the Andes profoundly shaped the climate and biodiversity of various ecosystems in the continent (Antonelli et al., 2009; Hoorn et al., 2010).

During rapid environmental changes, plants could rely on the dispersal of propagules to establish new colonies in desirable habitats, or undertake phenotypic plasticity to mitigate disadvantageous environmental stimuli (Jump and Penuelas, 2005; Chevin et al., 2010). However, the success of propagule dispersal and the magnitude of phenotypic plasticity both have limitations. Given the existence of genetic variability and the heritability of genetic traits in a population, genetic evolution should be instrumental for plants to survive from new local environmental conditions (Jump and Penuelas, 2005; Chevin et al., 2010).

Polyploidy, a common phenomenon arising from whole-genome duplication (WGD), acts as a critical mechanism in plants to encourage adaptation to harsh environments (Chen, 2007; Jiao et al., 2011). For example, WGD in the Brassicaceae provided ample genomic resources for functional innovation through evolution of duplicated genes and/or rewiring genetic networks (De Smet and Van de Peer, 2012; McGrath et al., 2014). The genomic analysis of *Arabidopsis thaliana* identified three major paleopolyploidy events (α, β, and γ WGDs)which occurred before ~40 Ma in the core Brassicaceae taxa (Kagale et al., 2014; Liu et al., 2014). But more importantly, it was the subsequent species-specific neo-/mesopolyploidy events that laid out the foundations for the species diversification of the Brassicaceae in the Neogene era (Kagale et al., 2014).

There are about 230 established species in the genus *Lepidium* on record widely spreading across the temperate and subtropical regions of the world (Al-Shehbaz, 1986; Warwick and Al-Shehbaz, 2006). The genus originated in the Mediterranean area as diploid species, and eventually dispersed throughout the world in the late Cenozoic period (Toledo et al., 1998). It is proposed that long-distance dispersal, rather than continental drift, propelled the rapid radiation of *Lepidium* (Toledo et al., 1998), and that allopolyploid speciation facilitated the colonization of new habitats in the Americas/Oceania (Lee et al., 2002; Dierschke et al., 2009).

Maca is one of the few *Lepidium* plants reported in both tropical and alpine areas (Al-Shehbaz, 1986). It is predominantly

self-pollinating with a disomic octoploid genome ($2n = 8x = 64$) (Quiros et al., 1996). Analyses of the floral structures and the first intron of *PISTILLATA* gene in maca and other *Lepidium* species suggested that allopolyploidization might have contributed to the polyploid genome(s), but the hybridization history in *Lepidium* is still unsolved (Lee et al., 2002). The genomic features of maca, together with its distinct habitat living at 4000 m height, make the plant an appropriate model to study both WGD events and attendant high-altitude adaptation. Here, we report the high-quality assembly and detailed analysis of the maca genome.

RESULTS

High-Quality Genome Assembly of Maca

The sequencing of the whole maca genome on the Illumina HiSeq 2500 platform yielded 1.88 billion reads in ten paired-end libraries (Supplemental Table 1). These reads represented 361.73 Gb raw data, which were about 482× coverage of the estimated maca genome (751 Mb, Supplemental Figure 1). A draft *de novo* assembly of the reads resulted in a total contig length of 728.8 Mb and a total scaffold length of 743.2 Mb, which covered 97.04% and 98.93% of the estimated maca genome, respectively (Table 1, Supplemental Figure 2, and Supplemental Table 2). The contig and scaffold N50 sizes were 81 kb and 2.4 Mb, respectively, which were among the longest in the published plant genomes based on short Illumina reads (Table 1, Supplemental Table 2). A preliminary evaluation of the assembly quality showed that 2636 scaffolds with sizes ranging from 2 kb to the maximum 8.8 Mb covered 86% of the assembled genome, and more than 97% of the clean reads from paired-end libraries could be mapped back to the assembled genome.

The metrics of the maca genome assembly came out in unexpected high quality despite maca being a disomic octoploid plant. To verify our findings, we obtained ~33.0 million RNA sequencing (RNA-seq) reads from mature leaf tissue and ~18.4 million reads from mature root tissue. Mapping of these reads back to the assembly revealed an overall mapping rate of 63.5% for the leaf and 68.8% for the root (Supplemental Table 3A). Further analysis by the core eukaryotic genes mapping approach (CEGMA) showed that 246 out of 248 ultra-conserved core eukaryotic genes were present in the maca genome assembly (Supplemental Table 3B). These benchmarks confirmed the high quality of the assembly. The heterozygosity rate of the maca genome was estimated at 0.213% (Supplemental Table 4), which was in line with the lack of apparent heterozygous peak in the *k*-mer graph (Supplemental Figure 1).

Transposable elements (TEs) constitutes 354.1 Mb or 47.65% of the maca assembly (Table 1, Supplemental Table 5). Both the total copy number and the percentage of TEs in the maca genome were the highest among all sequenced Brassicaceae species. Breakdown of the maca TE statistics showed higher proportion of retrotransposons in the genome (22.17%) than that of *A. thaliana* (11.91%) and *Brassica rapa* (17.00%). This was due to higher proportions of TE subclasses Copia (9.25%) and Gypsy (8.51%).

Gene annotation on the basis of homology and *ab initio* prediction methods identified 96 417 protein-coding genes (Table 1, Supplemental Table 6, and Supplemental Figure 3) and 18 885 non-protein-coding genes (microRNA [miRNA], tRNA, rRNA and small nuclear RNA [snRNA]; Table 1, Supplemental Table 7). Subsequent analyses of the maca leaf and root RNA-seq data

Estimated genome size (Mb)	**751**
Assembly Statistics	
De novo assembly size (Mb)	743
No. of N50 contigs	2418
N50 contig length (bp)	81 780
No. of N50 scaffolds	92
N50 scaffold length (bp)	2 420 765
No. of N90 contigs	122 974
N90 contig length (bp)	175
No. of N90 scaffolds	93 194
N90 scaffold length (Mb)	175
Transposable elements content	47.65%
Gene annotation Statistics	
Total no. of protein-coding genes	96 417
Total no. of exons	475 645
Average no. of exons per gene	4.93
Average exon size (bp)	234.92
Average intron length (bp)	222.76
Total no. of non-protein-coding genes	18 885

Table 1. Summary of the *L. meyenii* Genome Assembly Features.

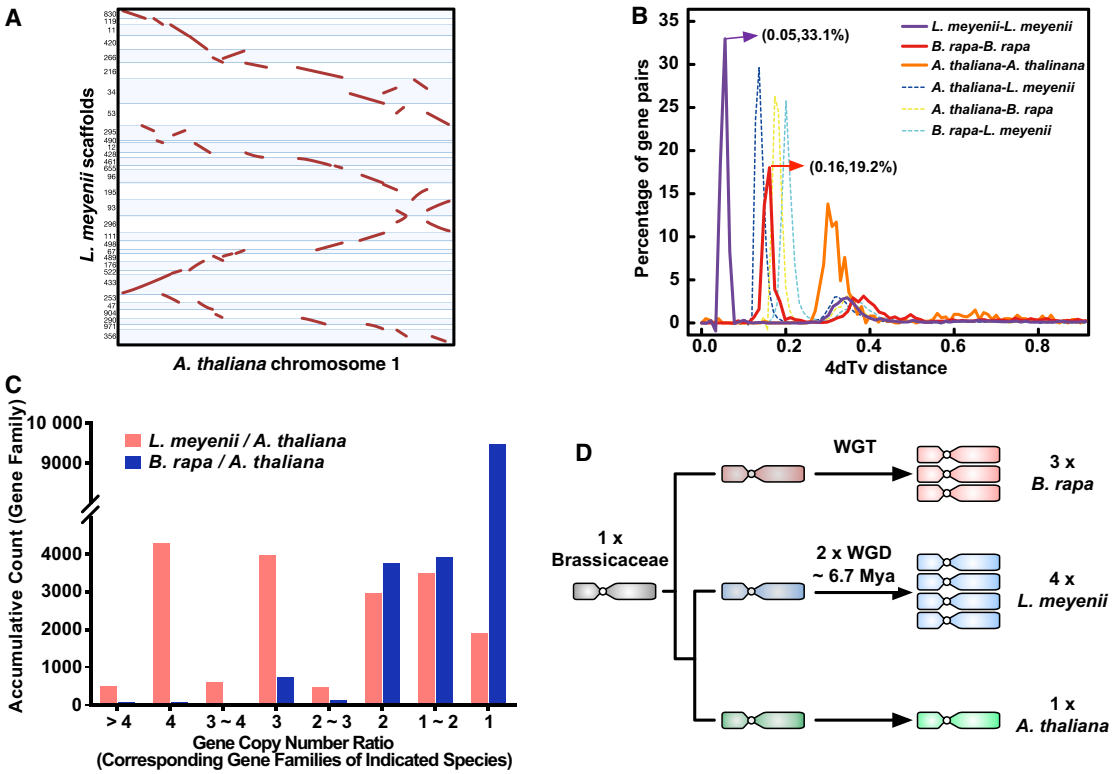


Figure 2. Whole-Genome Duplications in the *L. meyenii* Genome at around 6.7 Ma.

(A) Schematic graph of the collinear relationships between *A. thaliana* chromosome 1 and maca scaffolds. See Supplemental Figure 5 for additional collinear relationship graphs.

(B) Four-fold degenerate site transversion (4dTv) analysis estimated that two maca-specific WGDs in the maca genome happened at around 6.7 Ma.

(C) Gene copy-number ratio of the corresponding gene families between the indicated species.

(D) Schematic graph of genome evolution in *A. thaliana*, *L. meyenii*, and *B. rapa*.

showed that 61 743 (i.e., 64.04%) protein-coding genes had an FPKM value (fragments per kilobase of exon per million mapped reads) greater than 0.05.

Comparative analysis of *A. thaliana*, *Arabidopsis lyrata*, *B. rapa*, *Capsella rubella*, and maca genes identified a total of 24 269 homologous gene families, in which 13 777 gene families were shared by all five species and 3052 gene families were maca-specific (Supplemental Figure 4A). Further comparisons of these species revealed 15 900 expanded gene families in the maca genome (Supplemental Figure 4B). It is worth noting that many expanded genes and gene families were involved in abiotic stress response (i.e., WRKY, *bZIP*, *NAC*, *MYB*, and *Homeobox* transcription factor gene families), hormone signaling pathway (i.e., brassinosteroid signaling pathway), and secondary metabolite biosynthesis (i.e., glucosinolate biosynthesis).

These data collectively build up to the annotated whole-genome sequence of maca (database available at www.herbal-genome.cn).

Two Closely Spaced Maca-Specific WGDs Occurred about 6.7 Ma

Maca is phylogenetically categorized into lineage I of the widely diverse Brassicaceae (Figure 1C) (Kagale et al., 2014). It diverged from the diploid *A. thaliana* about 12.0 Ma, and *B. rapa* of lineage II about 18.0 Ma (Figure 1D).

Here, we analyzed the octoploidy of maca by aligning its assembled genome to the diploid *A. thaliana* genome with MCScanX. A 1:4 projection ratio between *A. thaliana* and maca collinear regions showed evidence of two maca-specific WGD events (Figure 2A and Supplemental Figure 5). In addition, the fourfold degenerate site transversion (4dTv) analysis of maca, *A. thaliana*, and *B. rapa* confirmed the two WGD events with a peak 4dTv distance of 0.05 in maca and the whole-genome triplication (WGT) event, with a peak 4dTv distance of 0.16 in *B. rapa* (Figure 2B). There is also a distant peak at around 0.35–0.40 that represents the ancient WGD event shared by these species. The time of the two maca-specific WGDs was estimated at around 6.7 Ma using the method described elsewhere (Simillion et al., 2002), coinciding with the rapid elevation of Andes from 1000 to 4000 m between 5 and 10 Ma (Figure 2B and 2D; see also Figure 1B).

Comparisons of homologous gene numbers in the corresponding gene families of *A. thaliana*, *B. rapa*, and maca genomes revealed a higher degree of gene retention in maca after the two maca-specific WGDs than that in *B. rapa* after WGT (Figure 2C). This result suggests that the maca-specific WGDs provided ample

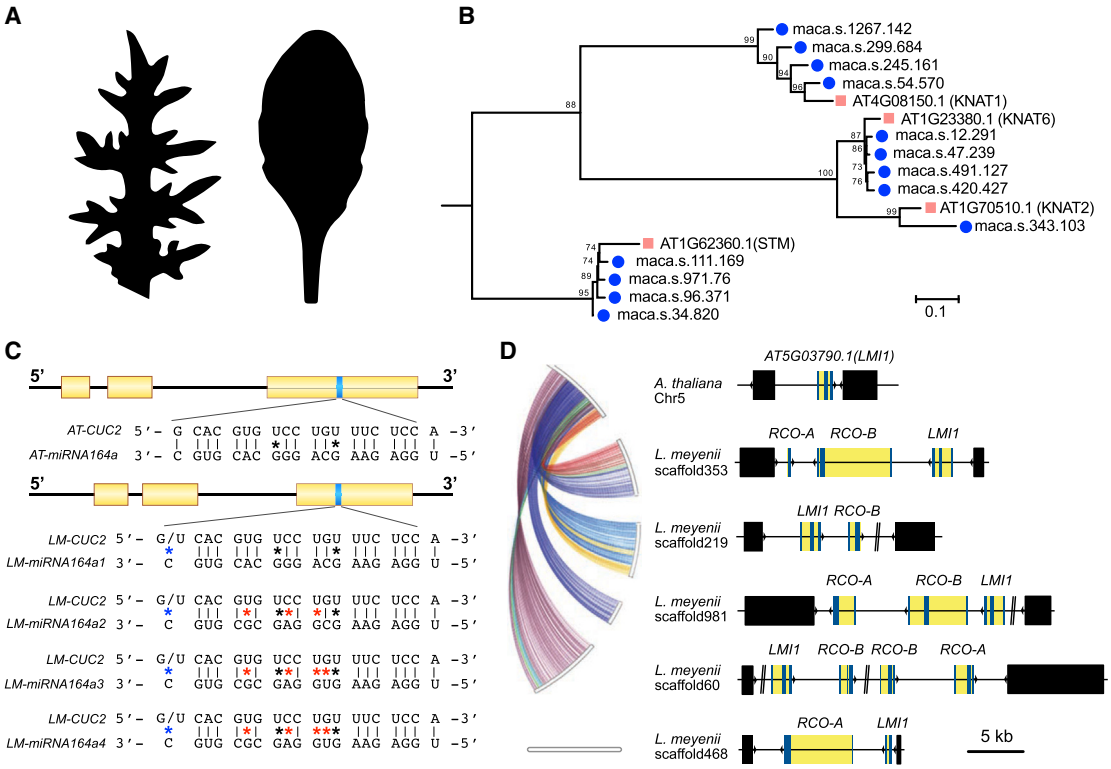


Figure 3. Evolution of Genes Involved in Leaf Morphogenesis in the Maca Genome.

(A) Leaf silhouettes of maca (left) and *A. thaliana* (right).

(B) Phylogenetic tree of class I *KNOTTED1-like homeobox* (*KNOX*) gene family from *L. meyenii* and *A. thaliana*. The scale bar shows the expected number of amino acid substitutions per residue.

(C) Maca *CUC2* genes after two WGDs could not be silenced by *miR164a* due to the presence of disruptive mismatch residual pairs between their sequences. Yellow boxes show exons of maca *CUC2* genes. The blue box shows the position of the *miR164a* binding site, and the nucleotide sequence is detailed. Black asterisks represent the two mismatch sites allowed for the binding of *miR164a* to wild-type *CUC2* transcript (*AT-CUC2*). Red asterisks represent additional mismatches in maca *miR164a* transcript that disrupt binding. Blue asterisks show a potential mismatch site that can be caused by a polymorphic U residue.

(D) Multiplication of *RCO/LMI1* gene clusters in the maca genome after two maca-specific WGDs. The *A. thaliana RCO/LMI1* gene cluster corresponds to the *RCO/LMI1* gene clusters in four maca syntenic blocks. A fifth gene cluster was identified in maca scaffold 468. Boxes with arrowheads show protein-coding genes in the region. *RCO/LMI1* genes are in yellow, with blue blocks representing exons.

genomic resources for the evolutionary adaptation to the harsh environment in the high Andes.

Evolution of Genes Involved in Leaf Morphogenesis in the Maca Genome

An adaptive strategy for plants to cope with various abiotic stresses is to decrease leaf surface area through the modification of leaf morphogenesis (i.e., maca; Figure 3A). During this intricate process, three major molecular mechanisms determine the leaf shape and complexity. On one hand, the class I *KNOTTED1-like homeobox* (*KNOX*) genes promote the leaflet initiation at the shoot apical meristem by influencing auxin-based patterning (Hake and Ori, 2002; Barkoulas et al., 2008; Blein et al., 2008). In *A. thaliana*, the lobed-leaf morphology was manifested in transgenic clones overexpressing *KNOX* genes (Lincoln et al., 1994). The comparison of maca and *A. thaliana* genomes revealed that 13 homologous *KNOX* genes in maca are resulted from the maca-specific WGDs (Figure 3B), which provided genomic evidence for its lobate leaves. On the other hand, *CUP-SHAPED COTYLEDON* (*CUC*) and *REDUCED COMPLEXITY* (*RCO*) genes both regulate the formation of serrated leaf edges, but exert actions via distinct mechanisms. In *A. thaliana*, the evolutionary divergence of *CUC* gene gave rise to three paralogs (*CUC1*, *CUC2*, and *CUC3*) as a result of neofunctionalization (Hasson et al., 2011). Despite their partially overlapping functions during leaf development, the balance between *CUC2* and its antagonist miRNA gene, *MIR164A*, primarily determines the degree of leaf-edge serration (Nikovics et al., 2006). Tipping the balance toward *CUC2* leads to large and deep serration in *A. thaliana* (Larue et al., 2009). In the maca genome, the maca-specific WGDs yielded four homologous copies each for *CUC2* and *MIR164A*. All four maca *CUC2* genes could predict protein products sharing high homology with their *A. thaliana* counterpart (Supplemental Figure 6A). The *miR164a* target sequences within maca *CUC2* mRNAs were the same except for one residual at 766 (Supplemental Figure 6B). Interestingly, at least three maca *CUC2* genes could not be silenced by *miR164a* due to the presence of disruptive mismatch residual pairs between their sequences (Figure 3C). This imbalance between *CUC2* and *MIR164A* may ultimately favor the serrated leaf edges in maca. Unaffected by the other

two mechanisms, *RCO* genes, which appear in a cluster in the genome as a result of gene duplication from the *LATE MERISTEM IDENTITY 1* (*LMI1*) type gene, sculpt developing leaflets to assist serration formation (Vlad et al., 2014). The simple leaf shape of *A. thaliana* is resulted from the loss of *RCO* genes in the cluster (Figure 3D) (Vlad et al., 2014). In the maca genome, *RCO* gene clusters were identified in the collinear regions of four scaffolds, indicating the retention of these genes after the maca-specific WGDs (Figure 3D). Another *RCO* gene cluster was found in the non-collinear region of a fifth scaffold, which may be the result of a segmental duplication as evidenced by the MCScanX analysis (Figure 3D). The abundance of *RCO* gene clusters, as well as the multiplication of *KNOX* and *CUC* genes, may have collectively provided a genomic explanation for the specialized leaf shape in maca.

Evolution of Genes Involved in Plant Cold and UV-B Adaptation Pathways

As a perennial herbaceous plant growing on tropical highlands, maca requires strong physiologic protection against frequent low nocturnal temperatures and long exposure to UV-B radiation. Despite the differences in adaptation strategies among plant species, it was proposed that: (1) most cold-adapted plants rely on a complicated cold acclimation process (Thomashow, 1999) and/or a variety of antifreeze proteins (Griffith and Yaish, 2004; Preston and Sandve, 2013) to reduce freezing injuries in the event of low temperatures; and (2) UV RESISTANCE LOCUS 8 (UVR8) protein-mediated photomorphogenic responses are essential for the establishment of UV-B tolerance in plants (Li et al., 2013). In view of the harsh growing conditions, maca is expected to retain key genes involved in these acclimation pathways after the maca-specific WGDs. Indeed, the examination of maca and *A. thaliana* genomes confirmed the multiplication and retention in the maca genome of antifreeze protein homologs (i.e., *BASIC CHITINASE* and *POLYGALACTURONASE INHIBITING PROTEIN 1*; Figure 4A), key genes in the well-characterized cold acclimation process (i.e., *CALMODULIN BINDING TRANSCRIPTIONAL ACTIVATORS*, *C-REPEAT BINDING FACTOR*, and *INDUCER OF CBF EXPRESSION*; Figure 4A) and the UVR8-mediated signaling pathway (i.e., *UVR8*, *CONSTITUTIVELY PHOTOMORPHOGENIC 1*, *ELONGATED HYPOCOTYL 5*, and *REPRESSORS OF UV-B PHOTOMORPHOGENESIS*; Figure 4A). Furthermore, the maca leaf and root RNA-seq data showed substantial expression of most of these genes (Figure 4B). These results implied the evolutionary role of the maca-specific WGDs in the development of acclimation traits in maca during Andes uplift.

Loss of Important Self-Incompatibility S-Locus Genes in the Maca Genome

The majority of flowering plant species are able to prevent inbreeding via the rejection of self-pollination, which is also known as self-incompatibility (SI). This mechanism is considered evolutionarily advantageous in that it grants offspring greater genetic diversity. Despite this argument, transition from SI to self-fertility was also a common phenomenon in many taxonomic plant groups during the course of evolution (Wright et al., 2013). Even though this transition effectively reduces the heterozygosity of a population, it can theoretically provide reproductive assurance for plants during the colonization of isolated and peripheral environments (Busch, 2005). In the Brassicaceae family, SI is determined by a single polymorphic S-locus containing two key genes: *S-LOCUS RECEPTOR KINASE* (*SRK*) and *S-LOCUS CYSTEINE-RICH PROTEIN* (*SCR*) (Sherman-Broyles et al., 2007). Predominant self-fertile species (i.e., *A. thaliana*) were proved to have inactivated *SRK* and *SCR* genes in the locus (Figure 5) (Dwyer et al., 2013). In the case of maca, the S-locus region on all four scaffolds completely lost *SRK* and *SCR* homologous genes (Figure 5), indicating its self-fertile feature. This result is in line with the low heterozygosity of the maca genome, and suggests that autogamy might play a crucial role in the colonization of new habitats of the harsh environment in the high Andes.

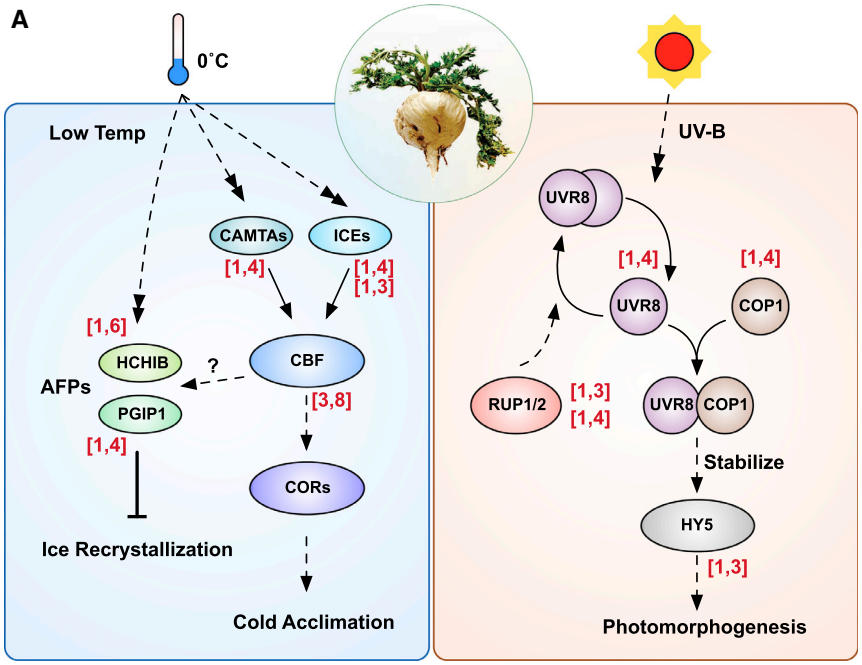
Genes under Positive Selection in Maca

Increased rate of nonsynonymous substitution (*Ka*) relative to synonymous substitution (*Ks*) within certain genes may explain the adaptive evolution of organisms at the molecular level (Qiu et al., 2012). Comparative analysis of orthologous gene pairs identified 3531 genes in maca versus *A. thaliana*, and 4280 genes in maca versus *B. rapa*, with the ratio of *Ka*/*Ks* greater than 1.0 ($p < 0.05$). A total of 683 genes were present in both lists of genes under positive selection (Supplemental Table 8). A close examination of these genes revealed candidates with putative functions related to leaf morphology, and response to cold and UV exposure. In particular, *LEAF CURLING RESPONSIVENESS* (*LCR*) regulates the development and curling phenotype of leaves (Song et al., 2012). *GLYCINE-RICH RNA-BINDING PROTEINS* (*GR-RBP*s) confer freezing tolerance to plants by regulating the opening and closing of stomata (Kim et al., 2010). The product of *DNA-DAMAGE-REPAIR/TOLERATION 2* (*DRT102*) facilitates the DNA repair from UV damage (Pang et al., 1993). A second approach based upon syntenic comparison identified 83 genes in maca versus *A. thaliana* with the ratio of *Ka*/*Ks* greater than 1.0 ($p < 0.05$, Supplemental Table 8). This gene list also contains candidates with putative functions related to cold resistance, e.g., *GLYCINE-RICH RNA-BINDING PROTEINS* (*GR-RBP*s) and *COLD SHOCK DOMAIN PROTEIN 3* (*CSP3*). Rapid evolution of these three genes might have helped maca in adapting to high-altitude environment in terms of enduring extreme living conditions.

DISCUSSION

In this study, we have reported the high-quality assembly and detailed analyses of the disomic octoploid maca genome. By comparing maca (*L. meyenii*) with the diploid *A. thaliana*, we identified two maca-specific WGD events in the maca genome, and estimated their occurrence at around 6.7 Ma during the uplift of the Andes. Subsequent analyses revealed that these two WGDs provided abundant genomic resources for reshaping maca traits and morphology in the adaptation to the harsh environments. In this regard, the whole-genome sequencing of maca not only helps accelerate genetic improvement of important ergonomical traits for Brassicaceae crops, but more importantly provides genomic resources for investigating high-altitude adaptation in plants.

A high-altitude environment poses a variety of stress factors for living organisms, including cold, strong UV-B radiation, low



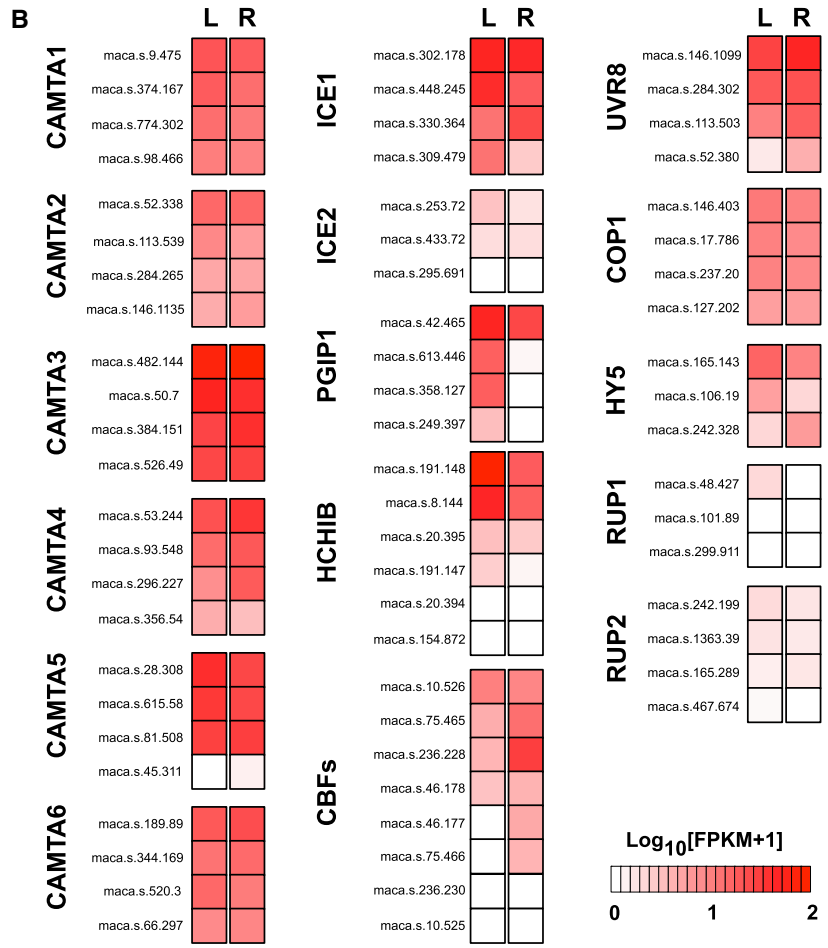


Figure 4. Evolution of Genes Involved in Plant Abiotic Stress Adaptation.

(A) Schematic pathways in response to cold and UV-B exposure in plants. The numbers of corresponding genes in *A. thaliana* and maca are shown in parentheses in the format of [*A. thaliana*, maca]. CAMTAs, CALMODULIN BINDING TRANSCRIPTIONAL ACTIVATORS; CBF, C-REPEAT BINDING FACTOR; ICE, INDUCER OF CBF EXPRESSION; CORs, COLD RESPONSIVE proteins; AFPs, antifreeze proteins; HCHIB, BASIC CHITINASE; PGIP1, POLYGALACTURONASE INHIBITING PROTEIN 1; UVR8, UV RESISTANCE LOCUS 8; COP1, CONSTITUTIVELY PHOTOMORPHOGENIC 1; HY5, ELONGATED HYPOCOTYL 5; RUP1/2, REPRESSOR OF UV-B PHOTOMORPHOGENESIS 1/2: RUP1/2 are negative regulators of the signaling pathway.

(B) Gene-expression levels of the indicated members from the cold and UV-B response pathways. The gene names and the maca gene IDs are listed on the left.

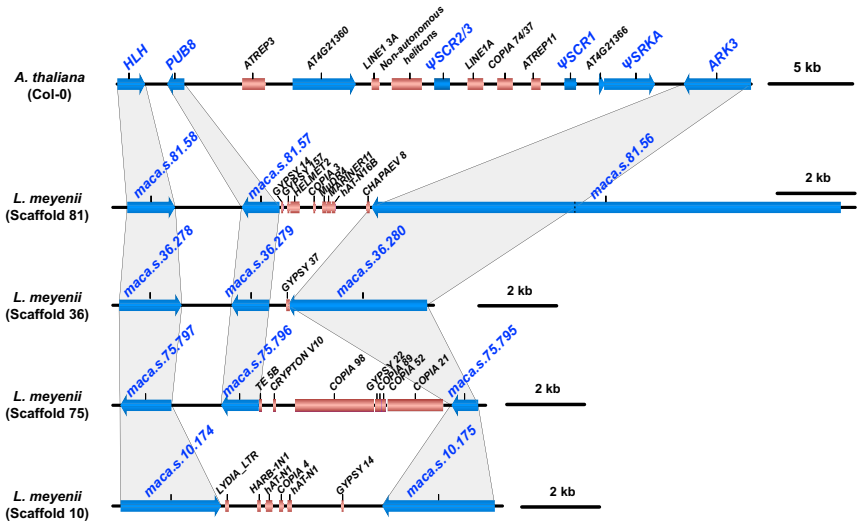


Figure 5. Zoomed-In Views of *A. thaliana* and Maca S-Locus Show the Loss of *SRK* and *SCR* Genes in Maca.

The S-locus region in *A. thaliana* (Col-0) is defined as the genomic region flanked by At4g21340 (*HLH*) and At4g21380 (*ARK3*). *A. thaliana SRK* and *SCR* genes are pseudogenes. Homologous gene pairs in *A. thaliana* and maca genomes are linked by light-gray blocks. Transposable elements are shown by red blocks. The maca gene *maca.s.81.56* on scaffold 81 is predicted to be a chimera of *ARK3* and its next adjacent gene. The dashed line marks the front part of the gene that shared homology with the *ARK3* gene in *A. thaliana*.

oxygen level, and capricious climate. For human and highland animals, hypoxia and its related complications are the most severe physiological challenges (Cheviron and Brumfield, 2012). Even though the ancient WGD events in vertebrates might have contributed to their current phenotypic complexities and evolutionary success (Freeling and Thomas, 2006), accumulating evidence showed that high-altitude adaptation in human and animals mainly arose from rapid evolution of key genes and metabolic pathways for oxygen uptake and delivery (Cheviron and Brumfield, 2012; Qiu et al., 2012; Gou et al., 2014).

In contrast, widespread WGD events during the course of evolution play a major role in the species divergence and adaptation to new habitats in land plants (De Smet and Van de Peer, 2012). For instance, the polyploid establishment in an allotment of monocots and eudicots coincided with the environmental disturbance around the Cretaceous–Paleogene extinction event, indicating that polyploidization may help mitigate the adverse effects of environmental stress in plants (Vanneste et al., 2014a, 2014b). Interestingly in the maca genome, two maca-specific WGDs occurred during the rapid uplift of the Andes. Since the genus *Lepidium* of the Brassicaceae family originates from the Mediterranean area as diploid, this polyploidization in maca provides an excellent model to study the enhancement of evolutionary fitness via WGD during environmental change.

Polyploid plants are well known for their increased tolerance to abiotic stresses. In tetraploid *A. thaliana*, potentiated drought and salinity tolerance partly result from the alterations in cell proliferation, organ size, and stoma physiology (Chao et al., 2013; del Pozo and Ramirez-Parra, 2014). Despite these observations, the genomic basis for the adaptive advantages after WGD remains unclear. One mathematic model of yeast cells showed that beneficial point mutations with strong fitness effects could accelerate the evolutionary adaptation of polyploid organisms (Selmecki et al., 2015). A second evolutionary model with yeast, however, concluded that duplication of a whole pathway was necessary for the increased fitness (van Hoek and Hogeweg, 2009). In fact, evidence supporting both hypotheses was obvious in the maca genome. On one hand, many maca genes under strong positive selection were involved in the development and abiotic stress response. On the other hand, expansion of cold and UV-B adaptation pathway, hormone signaling, and secondary metabolite biosynthesis pathway substantially strengthened the tolerance of maca to a harsh environment. As the genomic resources multiplied after WGDs, the redundant pathways and networks are subject to massive gene loss, rewiring, and reorganization for the emergence of new functions (De Smet and Van de Peer, 2012). Indeed, multiplication and evolution of genes in leaf morphogenesis and SI lead to the specialized leaf shape and mating style in maca. However, more research to observe these genes from the perspectives of pathways and networks is required.

In sum, findings from the maca genome suggested the success of polyploidization of genomes in periods of ecological upheaval, and offered insights into further investigation of adaptive advantage acquirement within a short evolutionary period in Andean plants.

METHODS

Plant Materials, Genomic DNA, and Total RNA Extraction

Yunnan Provincial Academy of Agricultural Sciences of the Alpine Economic Plant Research Institute maintained a cultivated line of maca plant in the 4200-m highland fields in southwestern China for at least 20 generations. We obtained fresh mature leaves and root from a single 5-month-old individual. Genomic DNA from 3 g of ground leaf tissues was extracted with Qiagen DNeasy Plant Mini Kits (Qiagen, MD, USA) following the manufacturer's protocol. The presence of high molecular weight DNA was confirmed by 1% agarose gel electrophoresis. About 100 μg of high-quality genomic DNA was used for library construction. Total RNA was extracted with Qiagen RNeasy Plant Mini Kits from 1 g of ground leaf tissue and 4 g of ground root tissue, respectively. The quality of the extracted RNA was verified in 1% agarose gel. For each sample, 5 μg of high-quality total RNA were used for library construction.

Library Construction and Sequencing

Purified genomic DNA was sheared into smaller fragments of random sizes by a focused ultrasonicator (Covaris, MA, USA). DNA fragments were electrophoresed in 0.8% General Purpose Agarose E-Gel (Invitrogen, CA, USA), by which fragments of desired lengths were obtained. With purified DNA fragments, we constructed the libraries with insertion sizes of 200 bp, 308 bp (three sets), 456 bp, 735 bp, 2 kb, 5 kb, 10 kb,

and 20 kb (Supplemental Table 1) using the Illumina Paired-End DNA protocol. The libraries were then sequenced on an Illumina Genome HiSeq 2500 machine using the PE-100 protocol. Total RNA-seq libraries were prepared using TruSeq RNA Library Preparation Kit v2 (Illumina, CA, USA) according to the instructions. They were also sequenced on an Illumina Genome HiSeq 2500 machine using the PE-100 protocol.

Genome Assembly

Sequencing all 10 DNA libraries generated about 1.88 billion raw reads. To ensure high-quality data for the subsequent *de novo* assembly step, we filtered out low-quality reads as specified by the following criteria: (1) if more than 5% of bases in a read were N or poly-A, (2) if more than 30 bases in a read were low quality, (3) if a read was contaminated with adaptor sequence, (4) if the size of a read was too small, and (5) if two copies of the paired-end reads had identical sequence (remove both copies). The resultant reads were then corrected by the SOAPec_v2.0.1 package (http://soap.genomics.org.cn) with default settings, yielding about 1.215 billion clean reads (Supplemental Table 1).

De novo assembly of the maca genome was performed by SOAPdenovo v2.04 (Luo et al., 2012) using default settings. We repeated the contig construction process 23 times with an array of *k*-mer values ranging from 27 to 93. The metrics of all contig constructions are summarized in Supplemental Figure 2A. The one with the highest N50 and N90 values was constructed with *k*-mer = 87, and subsequently used to assemble scaffolds. Similarly, we performed the scaffold construction process eight times with a series of *k*-mer values ranging from 47 to 73. As shown in Supplemental Figure 2B, N50 value was highest when *k*-mer was set at 57. We therefore used *k*-mer = 57 for the gapcloser step. The end result was a draft maca genome of 743 Mb, which was very close to the estimated size of 750 Mb. The maca genome size was estimated according to the 17-mer frequency distribution with the following formula: Genome size G = *K*_num/Peak_depth (Supplemental Figure 1).

RNA-Seq Data Analysis

About 33.0 million RNA-seq reads were obtained from mature leaf tissue, and about 18.4 million reads from mature root tissue. All the RNA-seq reads data were mapped against the maca genome assembly using TopHat on default settings (Trapnell et al., 2012). The FPKM value was calculated for each protein-coding gene by Cufflinks (http://cufflinks.cbcb.umd.edu) using default parameters. FPKM >0.05 was used as the cutoff value to identify expressed genes.

Repeats Annotation

Tandem repeats were identified across the genome with the help of the Tandem Repeats Finder program using default settings (Benson, 1999). TEs in the maca genome were identified by a combination of homology-based and *de novo* approaches. In terms of homology-based prediction, RepeatMasker (Tarailo-Graovac and Chen, 2009) was used to identify TEs at the DNA level against Repbase (Release 16.10; http://www.girinst.org/repbase/index.html; default settings). Additionally at the protein level, RepeatProteinMask was used to identify TEs via the RMBLAST search against TE protein database (default settings). For the *de novo* prediction, RepeatModeler (http://repeatmasker.org/) (Tarailo-Graovac and Chen, 2009) and LTR FINDER (Xu and Wang, 2007) were used to identify *de novo* evolved repeats from the assembled genome. Identification of the S-locus TEs in *A. thaliana* and maca (shown in Figure 5) was carried out by CENSOR (http://www.girinst.org/censor/) using default settings.

Protein-Coding Genes Prediction

We used both homology-based and *de novo* prediction methods to annotate protein-coding genes in the maca genome. Protein sequences of all the protein-coding genes for *Oryza sativa*, *A. lyrata*, *A. thaliana*, *Brassica oleracea*, *B. rapa*, *C. rubella*, *Panicum virgatum*, and *Thellungiella halophila* were obtained from Phytozome v9.1 (http://www.phytozome.net/), and used in the homology-based gene annotation. In brief, TBLASTN was first performed to map protein sequences of the aforementioned species to the maca genome with parameters of "*E*-value = $1e^{-5}$, −F F". For a single protein, all matching DNA sequences in the reference maca genome were concatenated by Solar after low-quality records were filtered. A protein-coding region was obtained by extending 2000 bp both upstream and downstream of the concatenated sequence. GeneWise (Birney and Durbin, 2000) was then used to predict gene structure within each protein-coding region. We used two *de novo* prediction programs, AUGUSTUS (Stanke et al., 2006) and GlimmerHMM (Majoros et al., 2004), to annotate protein-coding genes. The gene model parameters were trained from *A. thaliana*. The homology-based and *de novo* maca protein-coding gene sets were merged to form a comprehensive and non-redundant reference gene list using EVidenceModeler (Haas et al., 2008). All programs were executed under default settings unless indicated otherwise.

Non-Protein-Coding Genes Annotation

Software tRNAscan-SE (version 1.23) (Lowe and Eddy, 1997) with default parameters for eukaryote was used for tRNA annotation. Homology-based rRNA annotation was performed by mapping plant rRNAs to the maca genome using BLASTN with parameters of "*E*-value = $1e^{-5}$." The miRNA and snRNA genes were predicted by INFERNAL v0.81 (http://infernal.janelia.org) (Nawrocki et al., 2009) against the Rfam database (Release 11.0) (Gardner et al., 2011) using default parameters.

Gene Family Cluster

Protein sequences for all the protein-coding genes in *A. thaliana*, *A. lyrata*, *B. rapa*, and *C. rubella* were downloaded from Phytozome v9.1. To identify gene family clusters in these species and maca, we performed all-versus-all protein searches using BLASTP with the parameter "*E*-value = $1e^{-5}$." OrthoMCL (Version 1.4) (Li et al., 2003) was used to process high-scoring segment pairs. The MCL package in OrthoMCL was then used to define final paralogous and orthologous genes with the parameter "−abc − I = 1.5." The result was summarized in Venn diagram format using a web tool (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

Phylogenetic Tree Construction and Divergence Time Estimation

All 503 single-copy orthologous genes identified in the gene family cluster analysis from the aforementioned species were used to construct a phylogenetic tree. Multiple sequence alignments were performed for each gene using MUSCLE v.3.7 (http://www.drive5.com/muscle) with default settings (Edgar, 2004). Four-fold degenerate sites were extracted from each gene and concatenated into a "super gene" for each species. MrBayes v.3.1.2 (http://mrbayes.sourceforge.net) (Huelsenbeck and Ronquist, 2001) was used to reconstruct phylogenetic trees between species.

The MCMCTREE program within the PAML package (Yang, 2007) was used to estimate divergence time of *L. meyenii*, *C. rubella*, *A. thaliana*, *B. rapa*, *C. papaya*, and *O. sativa* (as an outgroup). The HKY85 model (model = 4) and independent rates molecular clock (clock = 2) were used for calculation. The MCMC process of MCMCTREE was performed with the samples 1 000 000 times, with a sample frequency setting of 2, after a burn-in of 200 000. The fine-tune parameters in the control file were set to "0.0058 0.0126 0.3 0.43455 0.7" to make the acceptance proportions fall into interval (0.2, 0.4).

Expansion and Contraction of Gene Families

CAFÉ v.2.1 (De Bie et al., 2006) is a tool for analyzing the evolution of gene family size based on the stochastic birth and death model. With the calculated phylogeny and the divergence time, this software was applied to identify gene families that had undergone expansion and/or contraction in the aforementioned species with the parameters

"*p*-value = 0.05, number of threads = 10, number of random = 1000, and search for lambda."

Detection of Positively Selected Genes

To detect genes under positive selection, we used the coding DNA sequence (CDS) libraries of *A. thaliana* and *B. rapa* to run BLASTN against the maca CDS library, respectively. The best hits were analyzed in *KaKs*_Calculator v.2.0 (Zhang et al., 2006) with default parameters.

In addition, a second approach based upon syntenic comparison was also performed to identify positively selected genes in maca. In brief, protein sequences were aligned against themselves using the BLASTp program. The top five alignments for each gene were kept. High-confidence collinear blocks with an E-value lower than $1E^{-10}$ and a score larger than 300 were then selected by MCScanX (Wang et al., 2012). For the paired genes inferred from the syntenic alignment, we aligned the protein sequences using CLUSTALW (Larkin et al., 2007), and used the result to guide coding-sequence alignments by PAL2NAL (Suyama et al., 2006). The *Ka* and *Ks* values were calculated using the Yang–Nielson method implemented in the yn00 program in the PAML package (Yang, 2007). A Python script was used to create a pipeline for all the calculations and is available at http://github.com/tanghaibao/biopipeline/tree/master/synonymous_calculation. All programs were run on default settings unless stated otherwise.

ACCESSION NUMBERS

The assembly and annotation of the maca genome are available at http://www.herbal-genome.cn. The sequencing reads of Illumina sequencing libraries have been deposited under NCBI Sequence Read Archive with Project ID SRP049292. All short-read data are available via Sequence Read Archive: SRR1631197, SRR1631211, SRR1631213, SRR1631216, SRR1631217, SRR1631219, SRR1631222, SRR1631223, SRR1631224, SRR1631225. The Project ID of all the RNA-seq data is SRP066645, with SRR2960160, SRR2960161 representing the transcriptome data from mature leaf and mature root tissue, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This work was supported by the Pilot Project for Establishing Social Service System through Agricultural Science and Education in Yunnan Province, Medical Plant Unit (2014NG003).

AUTHOR CONTRIBUTIONS

J.Z., Y.D., W.C., J.S., and W.W. planned and coordinated the project and wrote the manuscript. Y.T., S.H., X.M., and L.Y. collected and grew the plant material. J.Z., Y.D., and G.Z. sequenced and processed the raw data, and assembled and annotated the maca genome. X.W. and Y.Z. analyzed the genomes. J.J.Z. and Y.Q.H. set up the database. Y.T., N.L., J.Z., Y.W., and Y.M. conducted genome evolution analysis. Y.X. gave helpful suggestions on the whole project. J.Z., S.Y., and G.Z. analyzed the genes/gene families related to specific traits. L.Z. conducted whole-genome duplication-related analysis.

ACKNOWLEDGMENTS

We would also like to thank Dr. Carlos Quiros for his kind discussions of the manuscript. No conflict of interest declared.

Received: January 26, 2016
Revised: April 1, 2016
Accepted: April 26, 2016
Published: May 9, 2016

REFERENCES

Al-Shehbaz, I. (1986). Lepidieae. J. Arnold Arbor. Harv. Univ. **67**:265–311.

Antonelli, A., Nylander, J.A., Persson, C., and Sanmartin, I. (2009). Tracing the impact of the Andean uplift on neotropical plant evolution. Proc. Natl. Acad. Sci. USA **106**:9749–9754.

Barkoulas, M., Hay, A., Kougioumoutzi, E., and Tsiantis, M. (2008). A developmental framework for dissected leaf formation in the *Arabidopsis* relative *Cardamine hirsuta*. Nat. Genet. **40**:1136–1141.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27**:573–580.

Birney, E., and Durbin, R. (2000). Using GeneWise in the *Drosophila* annotation experiment. Genome Res. **10**:547–548.

Blein, T., Pulido, A., Vialette-Guiraud, A., Nikovics, K., Morin, H., Hay, A., Johansen, I.E., Tsiantis, M., and Laufs, P. (2008). A conserved molecular framework for compound leaf development. Science **322**:1835–1839.

Busch, J.W. (2005). The evolution of self-compatibility in geographically peripheral populations of *Leavenworthia alabamica* (Brassicaceae). Am. J. Bot. **92**:1503–1512.

Chao, D.Y., Dilkes, B., Luo, H., Douglas, A., Yakubova, E., Lahner, B., and Salt, D.E. (2013). Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. Science **341**:658–659.

Chen, Z.J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu. Rev. Plant Biol. **58**:377–406.

Chevin, L.M., Lande, R., and Mace, G.M. (2010). Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biol. **8**:e1000357.

Cheviron, Z.A., and Brumfield, R.T. (2012). Genomic insights into adaptation to high-altitude environments. Heredity **108**:354–361.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics **22**:1269–1271.

De Smet, R., and Van de Peer, Y. (2012). Redundancy and rewiring of genetic networks following genome-wide duplication events. Curr. Opin. Plant Biol. **15**:168–176.

del Pozo, J.C., and Ramirez-Parra, E. (2014). Deciphering the molecular bases for drought tolerance in *Arabidopsis* autotetraploids. Plant Cell Environ. **37**:2722–2737.

Dierschke, T., Mandakova, T., Lysak, M.A., and Mummenhoff, K. (2009). A bicontinental origin of polyploid Australian/New Zealand *Lepidium* species (Brassicaceae): evidence from genome *in situ* hybridization. Ann. Bot. **104**:681–688.

Dwyer, K.G., Berger, M.T., Ahmed, R., Hritzo, M.K., McCulloch, A.A., Price, M.J., Serniak, N.J., Walsh, L.T., Nasrallah, J.B., and Nasrallah, M.E. (2013). Molecular characterization and evolution of self-incompatibility genes in *Arabidopsis thaliana*: the case of the Sc haplotype. Genetics **193**:985–994.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

Freeling, M., and Thomas, B.C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. **16**:805–814.

Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., et al. (2011). Rfam: Wikipedia, clans and the "decimal" release. Nucleic Acids Res. **39**:D141–D145.

Garzione, C.N., Hoke, G.D., Libarkin, J.C., Withers, S., MacFadden, B., Eiler, J., Ghosh, P., and Mulch, A. (2008). Rise of the Andes. Science **320**:1304–1307.

Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., Yang, S., Jia, J., Kong, X., Wei, Z., et al. (2014). Whole-genome sequencing of six dog breeds

from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Res. **24**:1308–1315.

Griffith, M., and Yaish, M.W. (2004). Antifreeze proteins in overwintering plants: a tale of two activities. Trends Plant Sci. **9**:399–405.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. **9**:R7.

Hake, S., and Ori, N. (2002). Plant morphogenesis and KNOX genes. Nat. Genet. **31**:121–122.

Hasson, A., Plessis, A., Blein, T., Adroher, B., Grigg, S., Tsiantis, M., Boudaoud, A., Damerval, C., and Laufs, P. (2011). Evolution and diverse roles of the CUP-SHAPED COTYLEDON genes in *Arabidopsis* leaf development. Plant Cell **23**:54–68.

Hoffmann, A.A., and Willi, Y. (2008). Detecting genetic responses to environmental change. Nat. Rev. Genet. **9**:421–432.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartin, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., et al. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science **330**:927–931.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature **473**:97–100.

Jump, A.S., and Penuelas, J. (2005). Running to stand still: adaptation and the response of plants to rapid climate change. Ecol. Lett. **8**:1010–1020.

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., Links, M.G., et al. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell **26**:2777–2791.

Kim, J.Y., Kim, W.Y., Kwak, K.J., Oh, S.H., Han, Y.S., and Kang, H. (2010). Glycine-rich RNA-binding proteins are functionally conserved in *Arabidopsis thaliana* and Oryza sativa during cold adaptation process. J. Exp. Bot. **61**:2317–2325.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and clustal X version 2.0. Bioinformatics **23**:2947–2948.

Larue, C.T., Wen, J., and Walker, J.C. (2009). A microRNA-transcription factor module regulates lateral organ size and patterning in *Arabidopsis*. Plant J. **58**:450–463.

Lee, J.Y., Mummenhoff, K., and Bowman, J.L. (2002). Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* L. (Brassicaceae). Proc. Natl. Acad. Sci. USA **99**:16835–16840.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.

Li, J., Yang, L., Jin, D., Nezames, C.D., Terzaghi, W., and Deng, X.W. (2013). UV-B-induced photomorphogenesis in *Arabidopsis*. Protein Cell **4**:485–492.

Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K., and Hake, S. (1994). A knotted1-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. Plant Cell **6**:1859–1876.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A., Zhao, M., Ma, J., Yu, J., Huang, S., et al. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. **5**:3930.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience **1**:18.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics **20**:2878–2879.

McGrath, C.L., Gout, J.F., Johri, P., Doak, T.G., and Lynch, M. (2014). Differential retention and divergent resolution of duplicate genes following whole-genome duplication. Genome Res. **24**:1665–1675.

Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics **25**:1335–1337.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in *Arabidopsis*. Plant Cell **18**:2929–2945.

Pang, Q., Hays, J.B., Rajagopal, I., and Schaefer, T.S. (1993). Selection of Arabidopsis cDNAs that partially correct phenotypes of *Escherichia coli* DNA-damage-sensitive mutants and analysis of two plant cDNAs that appear to express UV-specific dark repair activities. Plant Mol. Biol. **22**:411–426.

Piacente, S., Carbone, V., Plaza, A., Zampelli, A., and Pizza, C. (2002). Investigation of the tuber constituents of maca (*Lepidium meyenii* Walp.). J. Agri. Food Chem. **50**:5621–5625.

Preston, J.C., and Sandve, S.R. (2013). Adaptation to seasonality and the winter freeze. Front. Plant Sci. **4**:167.

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., et al. (2012). The yak genome and adaptation to life at high altitude. Nat. Genet. **44**:946–949.

Quiros, C., Epperson, A., Hu, J., and Holle, M. (1996). Physiological studies and determination of chromosome number in Maca, *Lepidium meyenii* (Brassicaceae). Econ. Bot. **50**:216–223.

Selmecki, A.M., Maruvka, Y.E., Richmond, P.A., Guillet, M., Shoresh, N., Sorenson, A.L., De, S., Kishony, R., Michor, F., Dowell, R., et al. (2015). Polyploidy can drive rapid adaptation in yeast. Nature **519**:349–352.

Sherman-Broyles, S., Boggs, N., Farkas, A., Liu, P., Vrebalov, J., Nasrallah, M.E., and Nasrallah, J.B. (2007). S locus genes and the evolution of self-fertility in *Arabidopsis thaliana*. Plant Cell **19**:94–106.

Shin, B.C., Lee, M.S., Yang, E.J., Lim, H.S., and Ernst, E. (2010). Maca (L. meyenii) for improving sexual function: a systematic review. BMC Complemen. Altern. Med. **10**:44.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M., and Van de Peer, Y. (2002). The hidden duplication past of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **99**:13627–13632.

Song, J.B., Huang, S.Q., Dalmay, T., and Yang, Z.M. (2012). Regulation of leaf morphology by microRNA394 and its target LEAF CURLING RESPONSIVENESS. Plant Cell Physiol. **53**:1283–1294.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. **34**:W435–W439.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34**:W609–W612.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Prot. Bioinformatics **25**, 4.10:4.10.1-4.10.14.

Thomashow, M.F. (1999). PLANT COLD ACCLIMATION: freezing tolerance genes and regulatory mechanisms. Annu. Rev. Plant Physiol. **50**:571–599.

Toledo, J., Dehal, P., Jarrin, F., Hu, J., Hermann, M., Al-Shehbaz, I., and Quiros, C. (1998). Genetic variability of *Lepidium meyenii* and other Andean *Lepidium* species (Brassicaceae) assessed by molecular markers. Ann. Bot. **82**:523–530.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Prot. **7**:562–578.

van Hoek, M.J., and Hogeweg, P. (2009). Metabolic adaptation after whole genome duplication. Mol. Biol. Evol. **26**:2441–2453.

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014a). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. Genome Res. **24**:1334–1347.

Vanneste, K., Maere, S., and Van de Peer, Y. (2014b). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos. Trans. R Soc. Lond. B Biol. Sci. **369**, [pii: 20130353].

Vlad, D., Kierzkowski, D., Rast, M.I., Vuolo, F., Dello Ioio, R., Galinha, C., Gan, X., Hajheidari, M., Hay, A., Smith, R.S., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. Science **343**:780–783.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40**:e49.

Warwick, S.I., and Al-Shehbaz, I.A. (2006). Brassicaceae: chromosome number index and database on CD-Rom. Plant Syst. Evol. **259**:237–248.

Wright, S.I., Kalisz, S., and Slotte, T. (2013). Evolutionary consequences of self-fertilization in plants. Proc. Biol. Sci. **280**:20130133.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**:W265–W268.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591.

Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genom. Proteomics Bioinformatics **4**:259–263.