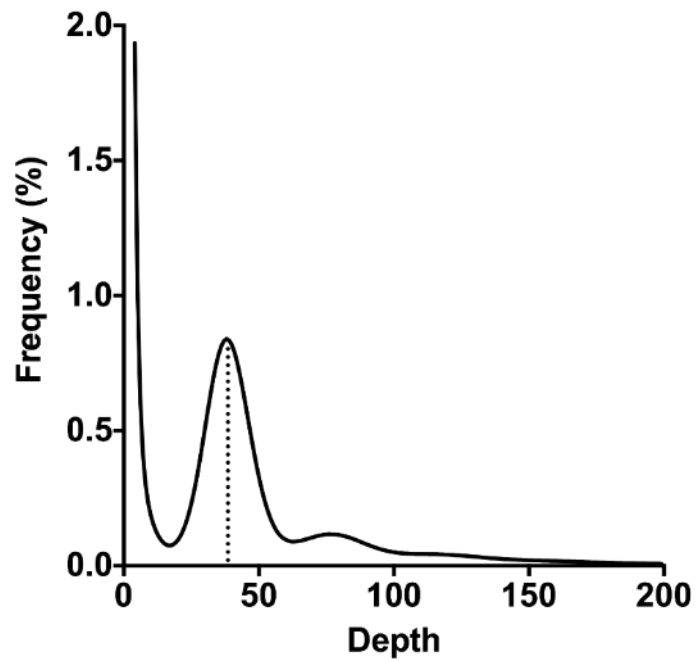


Supplemental Information

Supplemental Figures

A.



B.

<i>k</i> -mer	<i>k</i> -mer number	Peak Depth	Genome Size (bp)	Used Bases (bp)	Used Reads Num.	Coverage
17	29,528,790,000	38	777,073,421	35,434,548,000	369,109,875	45.6 ×
M*	28,539,182,510	38	751,031,118			

* M means modified *k*-mer number.

Figure S1. Related to Table 1; Estimation of the maca genome size using 17-mer method.

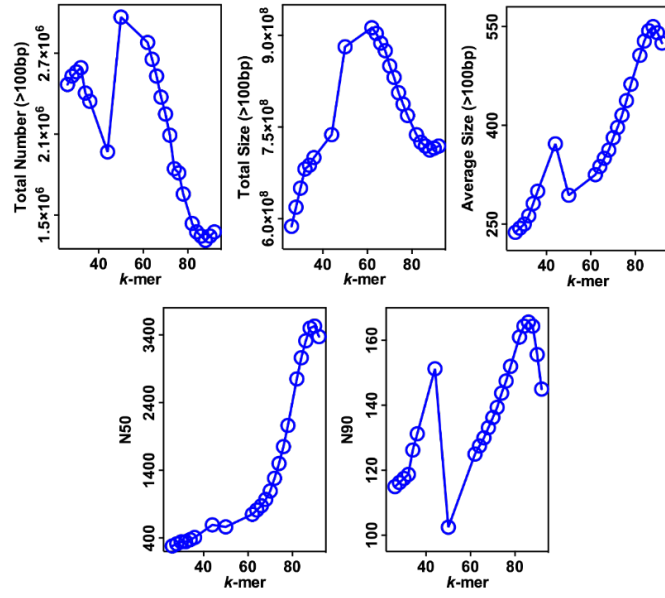
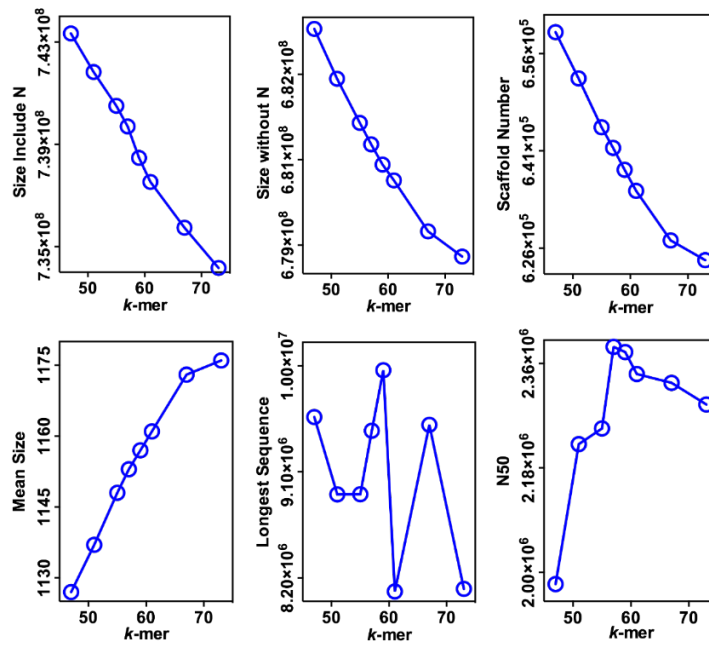
A**B**

Figure S2. Related to Table 1; Optimization of k -mer values for the construction of (A) contigs and (B) scaffolds.

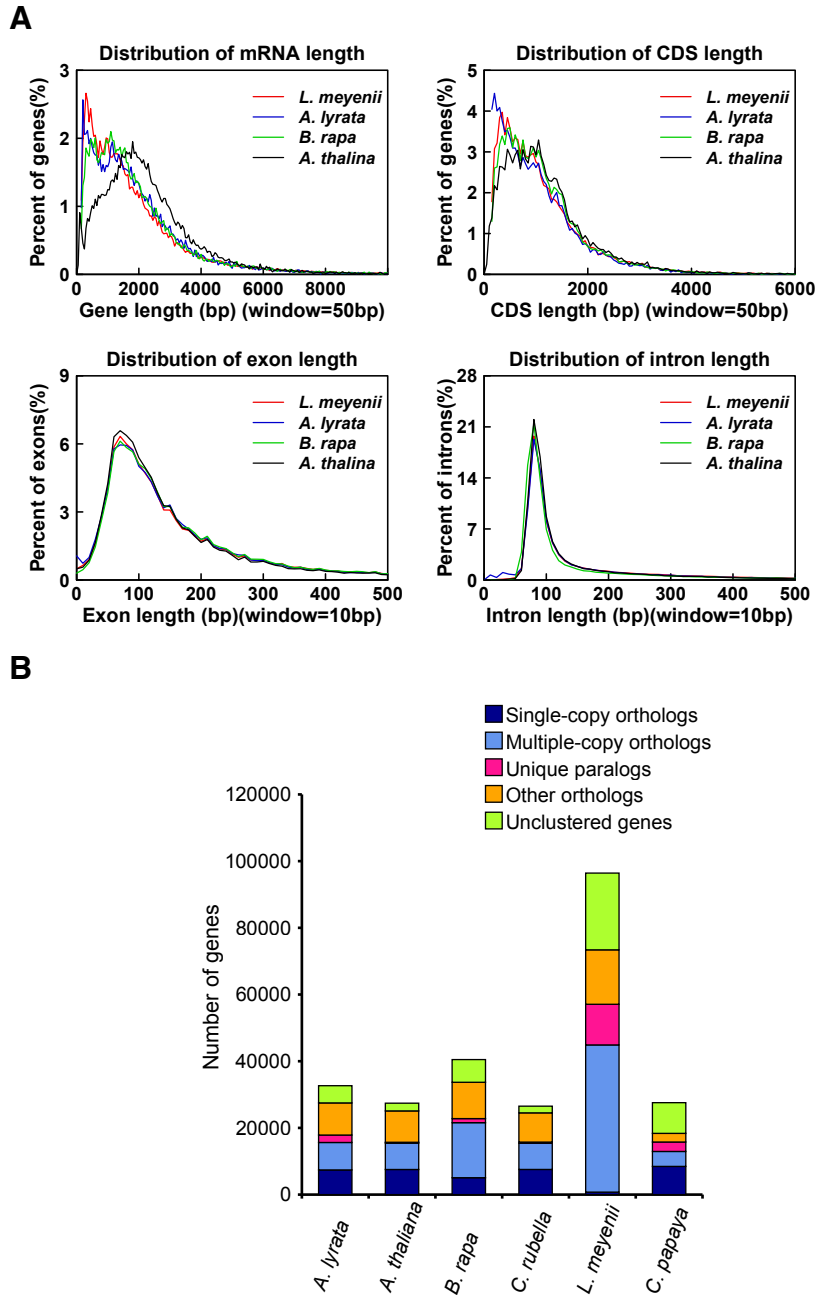
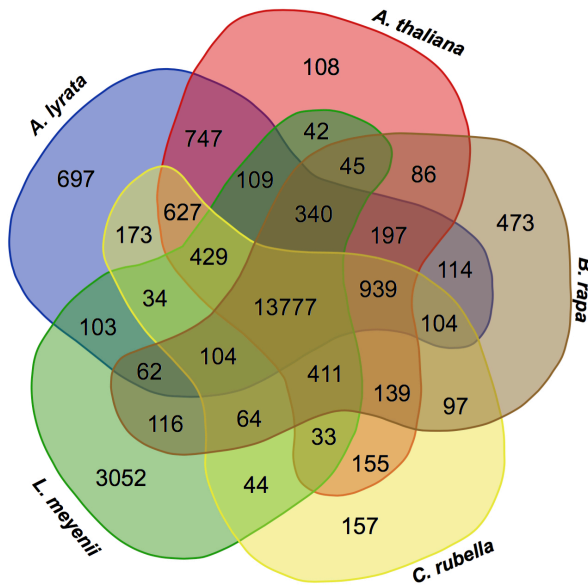


Figure S3. Related to Table 1; Parameters of annotated maca genes in comparison with those of *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Capsella rubella*, *Carica papaya* and *Brassica rapa*.

A



B

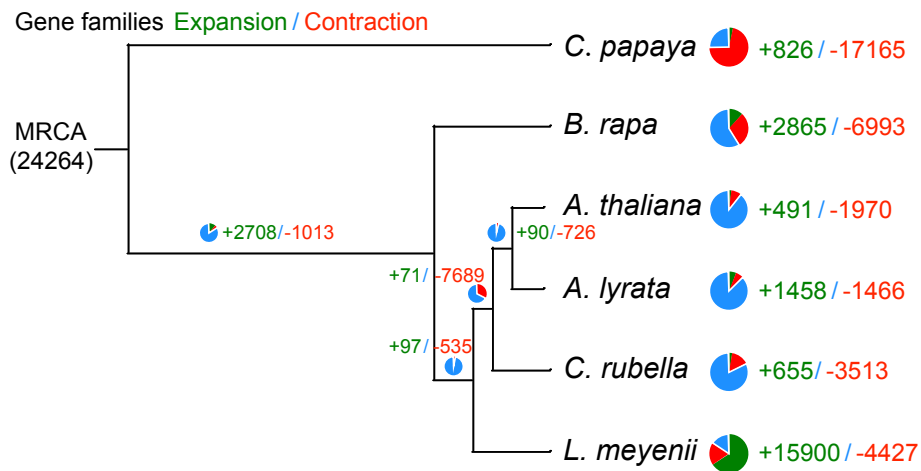


Figure S4. Related to Table 1; Gene family analyses of maca and related Brassicacease plants. (A) Gene family cluster analysis of *A. lyrata*, *A. thaliana*, *C. rubella*, maca and *B. rapa*. (B) Gene family expansion/contraction analysis of *A. lyrata*, *A. thaliana*, *C. rubella*, *C. papaya*, maca and *B. rapa*.

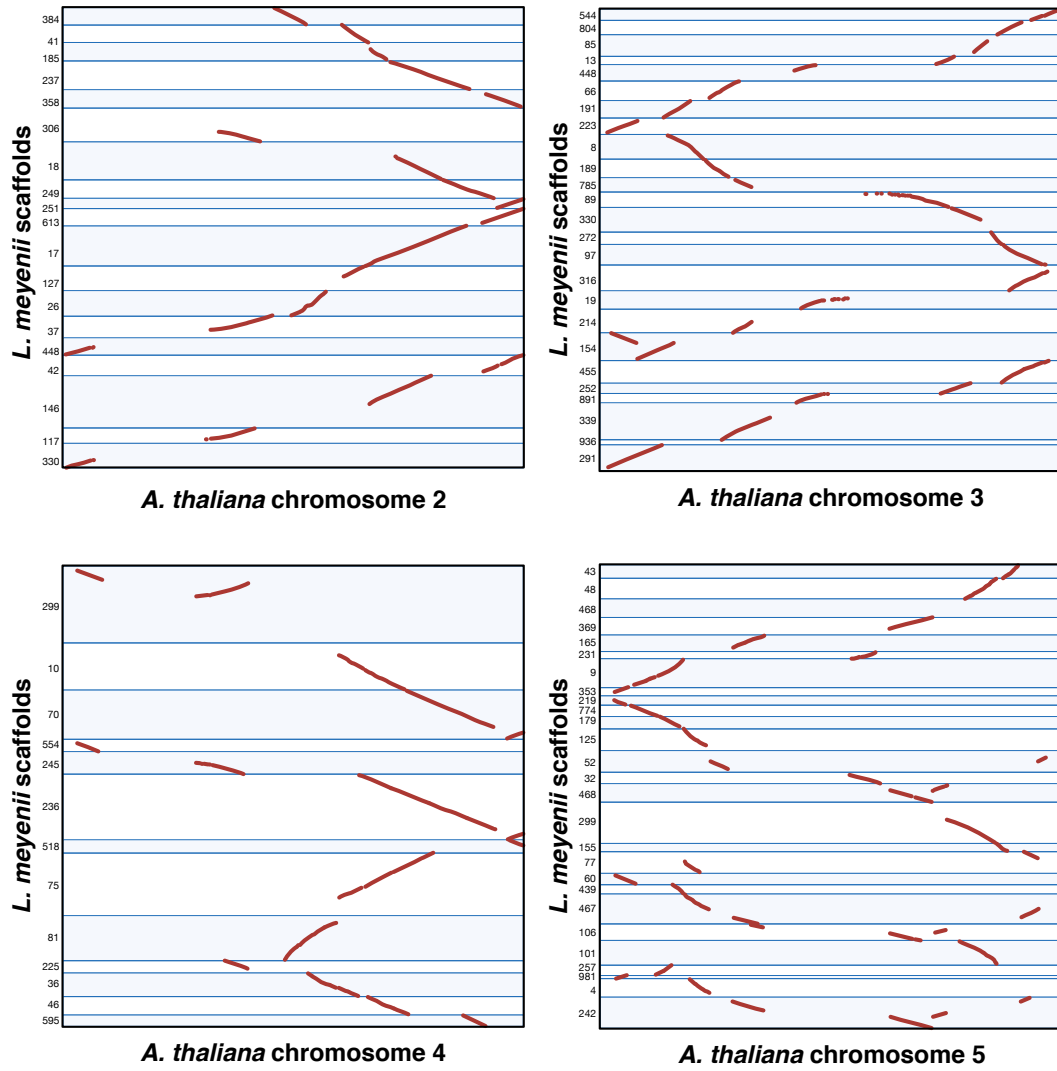
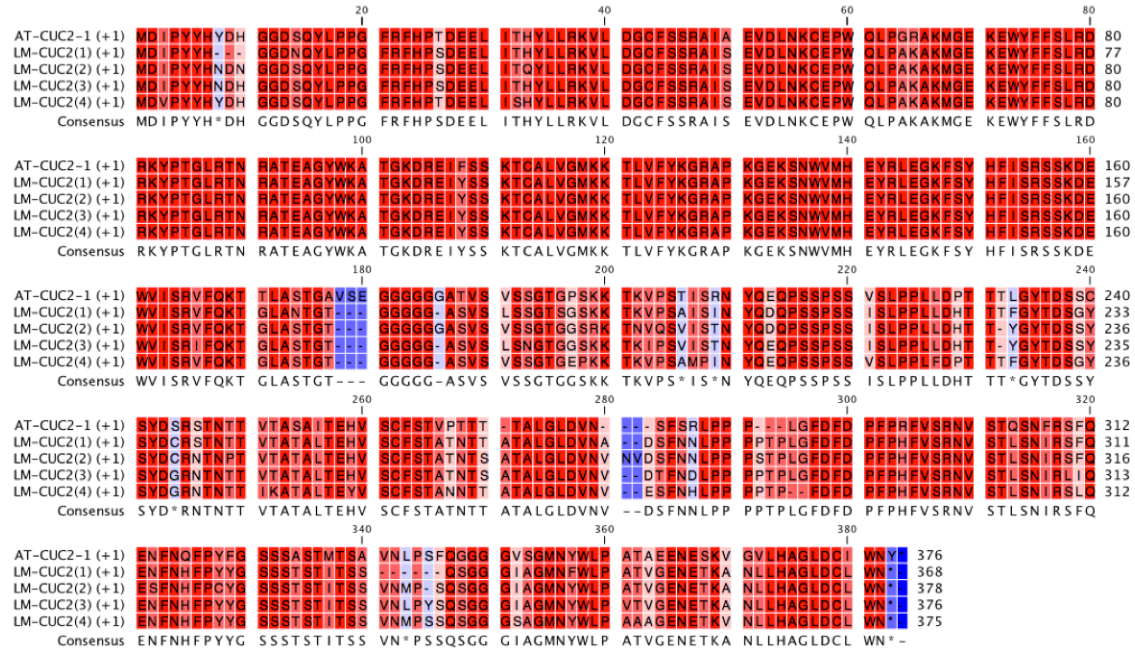


Figure S5. Related to Figure 2A; Schematic graphs of the collinear relationships between *A. thaliana* chromosomes (2, 3, 4, 5) and maca scaffolds.

A.



B.

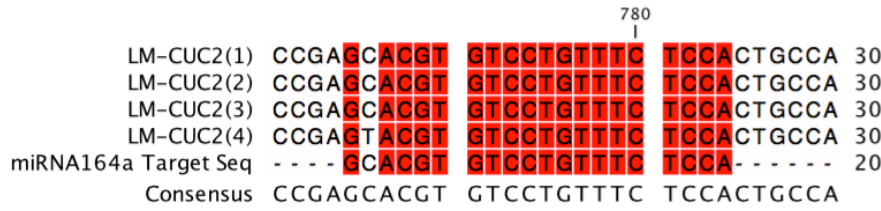


Figure S6. (A) Protein sequence alignment between *A. thaliana* and maca CUC2 proteins shows high homology. (B) Nucleotide alignment of the *CUC2* mRNAs revealed a polymorphic residual in the *miR164a* binding site.

Supplemental Tables

Table S1: Related to Table 1; Summary of sequenced paired-end libraries for the maca genome.

Insert Size (bp)	Library ID	Raw Bases (nt)	Raw Length (bp)	Raw Reads Num.	Reads Num. After Filter	Duplicate Reads Num.	Clean Reads Num.
200	Sample_MK-150_1	37,375,724,800	100_100	186,878,624	179,366,705	36,822,588	158,567,169
308	Sample_MK-428_1	40,967,253,800	100_100	204,836,269	190,291,179	6,095,531	186,853,001
308	Sample_MK-428_2	40,203,609,600	100_100	201,018,048	187,253,580	5,751,979	184,007,146
308	Sample_MK-428_3	39,801,459,200	100_100	199,007,296	187,833,671	5,878,050	184,499,837
456	Sample_MK-576_1	45,838,896,800	100_100	229,194,484	209,919,484	3,651,050	207,868,996
735	Sample_MK-855_1	27,902,681,400	100_100	139,513,407	66,474,261	52,823	66,447,142
2,000	RMKexzDMTDLAAP EI-19	33,704,395,020	90_90	187,246,639	170,713,214	142,495,047	62,165,369
5,000	RMKexzDMTDTAAPEI-20	34,521,592,500	90_90	191,786,625	171,628,116	151,580,167	43,406,319
10,000	RMKexzDMTDUAAP EI-89	30,517,727,220	90_90	169,542,929	155,949,124	133,229,402	41,110,779
20,000	RMKexzDMTDWAAPEI-90	30,909,400,920	90_90	171,718,894	151,198,023	107,411,886	80,170,650

Table S2: Related to Table 1; Summary of the final maca genome assembly after gapcloser.

	Contigs		Scaffolds	
	Size (bp)	Number	Size (bp)	Number
N90	175	122,974	175	93,194
N80	12,830	8,192	294,833	311
N70	37,026	5,007	1,035,346	186
N60	59,014	3,466	1,686,524	128
N50	81,780	2,418	2,420,765	92
Total	728,837,946	663,168	743,171,196	641,460
	(97.04%) ^a		(98.93%) ^b	

^{a,b} Percent coverage of the estimated maca genome (751 Mb).

Table S3: Related to Table 1; Evaluation of the quality of the maca genome assembly.

A. Mapping statistics of RNA-Seq reads from mature leaf and root tissues back to the maca genome assembly

	Mature Leaf Tissue	Mature Root
RNA-Seq Reads Num.	33,044,276	18,406,967
Mapped Reads in Assembly – Left	70.4%	74.0%
Mapped Reads in Assembly - Right	56.5%	63.5%
Overall Mapping Rate	63.5%	68.8%

B. Core eukaryotic genes mapping analysis of the maca genome assembly

	Protein Num. ^a	Completeness (%) ^b	Total Num. ^c	Average Num. ^d	Ortholog (%) ^e
Complete	246	99.19	1,038	4.22	97.97
Group1	66	100.00	278	4.21	100.00
Group2	56	100.00	235	4.2	96.43
Group3	61	100.00	254	4.16	100.00
Group4	63	96.92	271	4.3	95.24
Partial	247	99.60	1,083	4.38	98.38
Group1	66	100.00	288	4.36	100.00
Group2	56	100.00	248	4.43	98.21
Group3	61	100.00	267	4.38	100.00
Group4	64	98.46	280	4.38	95.31

^a Protein Num. : Number of 248 ultra-conserved core eukaryotic genes (CEGs) present in the maca genome.

^b Completeness (%) : Percentage of 248 ultra-conserved CEGs present in the maca genome.

^c Total Num. : Total number of CEGs including putative orthologs present in the maca genome.

^d Average Num. : Average number of orthologs per CEG.

^e Ortholog (%) : Percentage of detected CEGs that have more than one ortholog.

Table S4: Related to Table 1; Estimation of the heterozygosity rate of the maca genome assembly using Genome Characteristics Estimation software.

Raw_Peak	Now_Node	Low_Kmer	Now_Kmer	Cvg	Genome_Size	a[1/2]	a[1]	b[1/2]	b[1]
31	336,720,238	1,332,095,550	20,770,529,257	31.305	7.31308e+08	0.0698067	0.581339	0.0307088	0.521676

The heterozygosity rate of the maca genome can be estimated by the following formula:

$$\eta = \frac{a_{1/2}}{K(2 - a_{1/2})} \times 100\% = \frac{0.0698}{17(2 - 0.0698)} \times 100\% = 0.213\%$$

Table S5: Related to Table 1; Summary of transposable elements annotation in the maca genome assembly.

TE Classification	<i>L. meyenii</i>				<i>A. thaliana</i>		<i>B. rapa</i>	
	Total Length (bp)	Percentage in all TEs (%)	Copies	Percentage in Genome (%)	Copies	Percentage in Genome (%)	Copies	Percentage in Genome (%)
Retrotransposon	164,797,902	30.89	354,696	22.17	17,421	11.91	94,769	17.00
LTR/Copia	68,744,101	12.18	139,830	9.25	4,351	2.48	23,737	4.56
LTR/Gypsy	63,277,456	13.05	149,822	8.51	6,899	7.11	30,791	6.28
LTR/Other	1,293,234	0.77	8,799	0.17	779	0.22	3,274	0.32
LTR/Caulimovirus	3,444,506	0.47	5,419	0.46	15	-	748	0.09
LTR/ERV	2,171,360	0.83	9,567	0.29	453	0.03	1,419	0.06
LINE	25,455,402	3.30	37,931	3.43	3,963	1.91	31,926	5.53
SINE	411,843	0.29	3,328	0.06	961	0.17	2,874	0.16
DNA transposon	60,533,724	16.90	194,046	8.15	26,188	11.13	99,197	11.65
hAT	11,193,808	3.44	39,469	1.51	3,197	0.76	30,920	3.47
Helitron	3,531,798	0.93	10,636	0.48	5,249	2.49	8,604	0.85
Kolobok	47,508	0.04	506	0.01	71	-	309	0.01
PIF	2,526,799	0.64	7,325	0.34	1,034	0.33	7,568	1.04
CMC	17,600,371	4.10	47,133	2.37	4,295	1.76	16,286	1.90
MULE	19,184,456	4.98	57,139	2.58	7,264	4.12	12,513	1.72
TcMar	1,724,859	0.86	9,855	0.23	1,665	0.34	14,378	1.23
Satellite	938,949	0.16	1,834	0.13	553	0.71	1,700	0.28
DNA/other	2,086,066	1.44	16,552	0.28	2,771	0.46	4,983	0.83
Simple Repeat	1,699,110	0.31	3,597	0.23	89	0.15	1,936	0.31
Other	21,214	0.02	208	0.003	260	0.06	1,270	0.09
Unknown	128,785,533	52.19	599,304	17.33	8,251	3.06	92,777	11.59
Total	354,138,373	100.00	1,148,254	47.65	52,120	26.16	288,013	40.33

Table S6: Related to Table 1; Summary of protein-coding gene annotation in the maca genome assembly.

Methods	Gene Number	Avg. mRNA Length	Total Exon Number	Avg. Exon Length	Avg. CDS Length	Avg. Exon Number	Total Intron Number	Avg. Intron Length
<i>ab initio</i>								
AUGUSTUS	86,612	2,195.65	481,043	232.09	1,289.01	5.55	394,431	199.09
GlimmerHMM	107,142	1,735.35	442,598	245.04	1,012.25	4.13	335,456	230.95
Homology								
<i>Oryza sativa</i>	67,902	1,893.65	334,112	206.37	1,015.47	4.92	266,210	224.00
<i>Arabidopsis lyrata</i>	84,915	2,113.36	449,505	213.85	1,132.05	5.29	364,590	228.55
<i>Arabidopsis thaliana</i>	84,892	2,123.10	456,699	215.27	1,158.11	5.38	371,807	220.33
<i>Brassica oleracea</i>	84,235	2,233.71	447,304	214.55	1,139.29	5.31	363,069	253.91
<i>Brassica rapa</i>	79,672	2,375.23	436,407	215.33	1,179.48	5.48	356,735	267.05
<i>Capsella rubella</i>	83,387	2,138.97	449,397	216.01	1,164.16	5.39	366,010	222.09
<i>Panicum virgatum</i>	71,596	1,698.30	334,487	202.61	946.55	4.67	262,891	204.73
<i>Thellungiella halophila</i>	83,923	2,097.43	450,365	215.04	1,154.01	5.37	366,442	216.06
EVidenceModeler	96,417	2,035.06	475,645	234.92	1,158.89	4.93	379,228	222.76

Table S7: Related to Table 1; Summary of non-protein-coding gene annotation in the maca genome assembly.

Type	Copy Number	Avg. Length (bp)	Total Length (bp)	Pct. in Genome (%)
miRNA	346	125.16	43,304	0.00583
tRNA	2,578	74.75	192,715	0.02593
rRNA	13,698	125.34	1,716,891	0.23102
18S	6,225	161.44	1,004,970	0.13523
28S	2,791	91.69	255,910	0.03444
5.8S	1,954	121.82	238,036	0.03203
5S	2,728	79.90	217,975	0.02933
snRNA	2,263	108.41	245,339	0.03301
CD-box	1,823	101.94	185,841	0.02501
HACA-box	148	128.49	19,017	0.00256
Splicing	292	138.63	40,481	0.00545