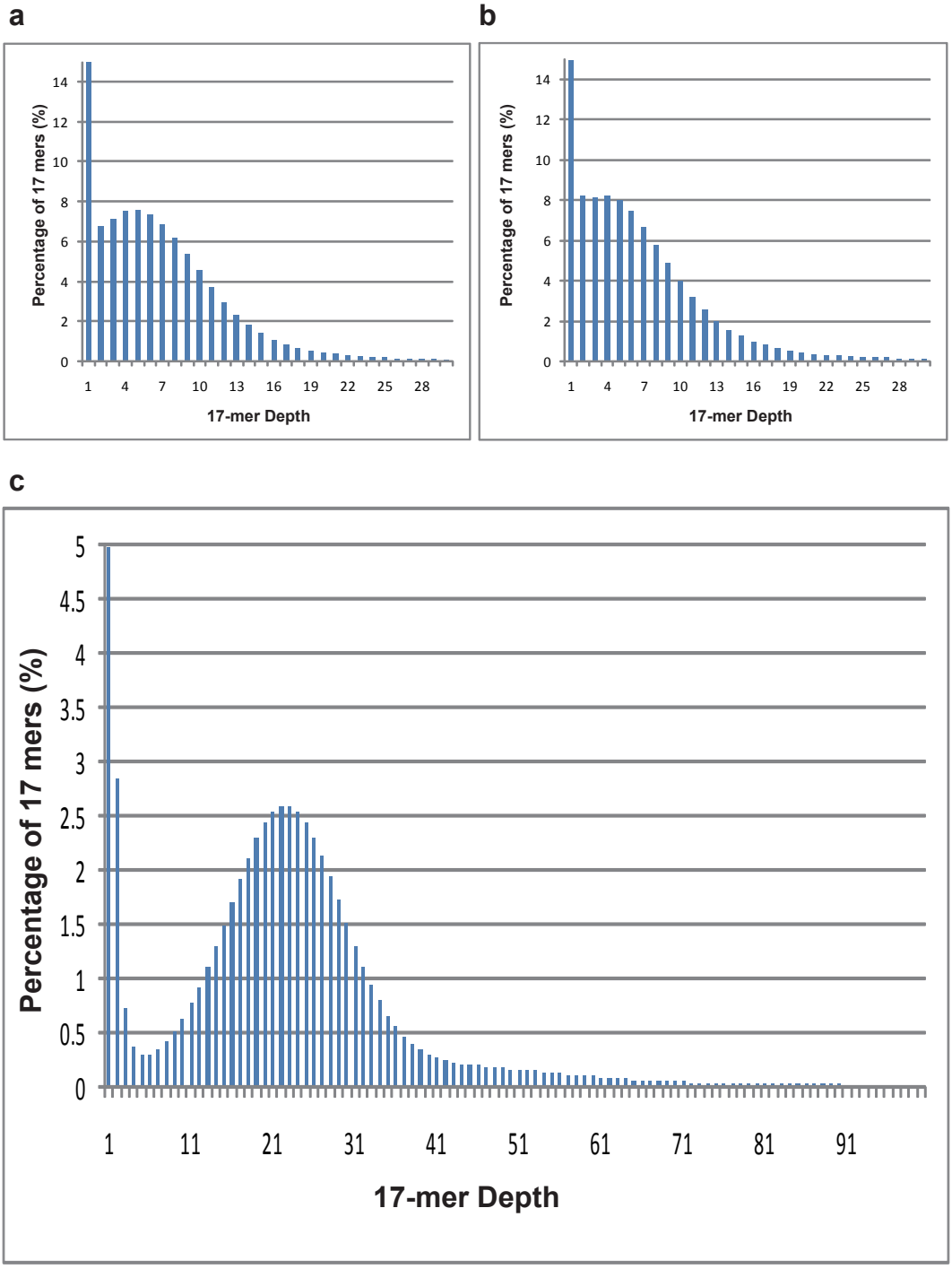
Supplementary Information

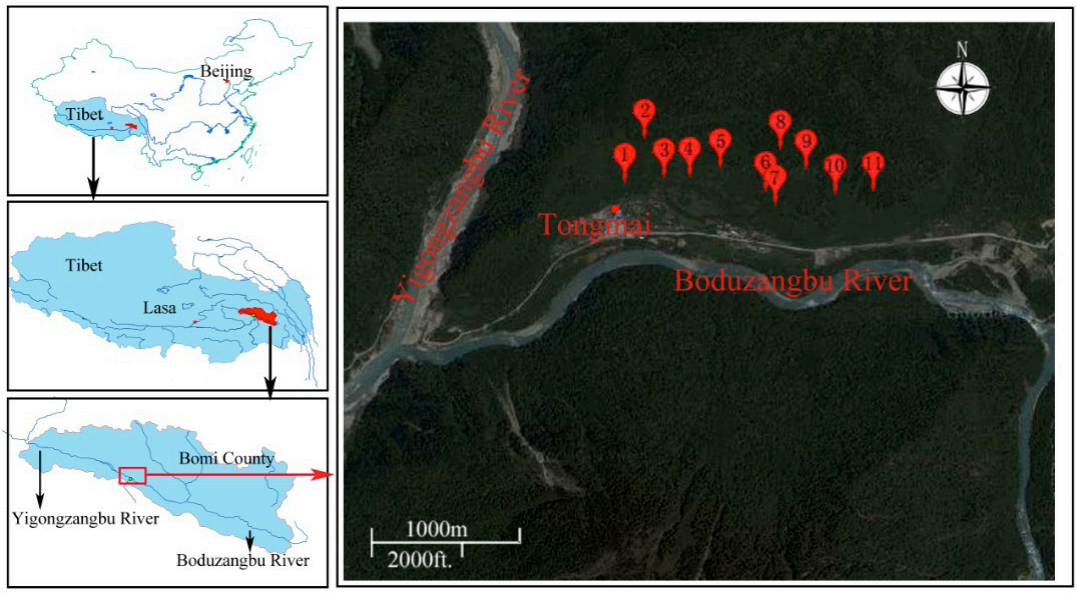
The genome of *Prunus mume*

Qixiang Zhang[1,6,*], Wenbin Chen[2,6], Lidan Sun[1,6], Fangying Zhao[3,6], Bangqing Huang[2,6], Weiru Yang[1], Ye Tao[2], Jia Wang[4], Zhiqiong Yuan[3], Guangyi Fan[2], Zhen Xing[5], Changlei Han[2], Huitang Pan[1], Xiao Zhong[2], Wenfang Shi[1], Xinming Liang[2], Dongliang Du[1], Fengming Sun[2], Zongda Xu[1], Ruijie Hao[1], Tian Lv[2], Yingmin Lv[1], Zequn Zheng[2], Ming Sun[1], Le Luo[1], Ming Cai[1], Yike Gao[1], Junyi Wang[2], Ye Yin[2], Xun Xu[2], Tangren Cheng[4,*], Jun Wang[2,*]

Inventory of Supplementary Information:

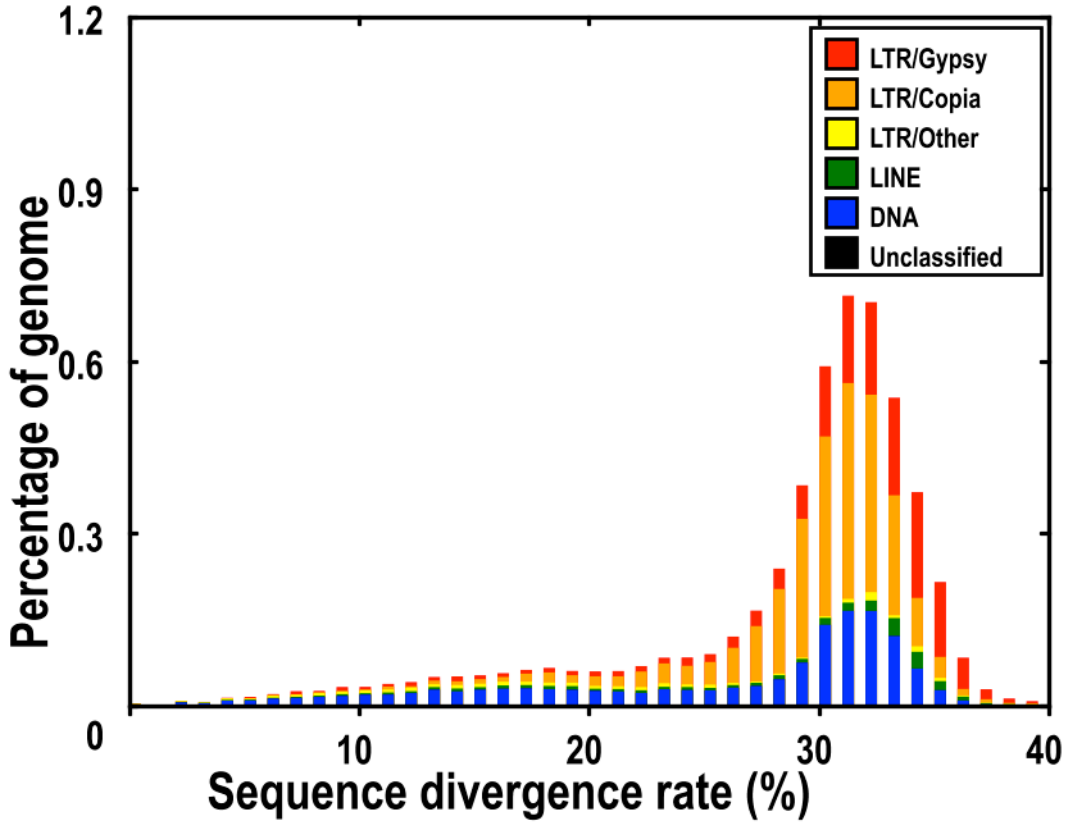


Supplementary Figure S1. K-mer analysis.
a) and b) estimating the domesticated samples, c) estimating the wild *P. mume* sample used for genome assembly. The x-axis is depth (X); the y-axis is the proportion that represents the percentage at that depth. (Without consideration of the sequence error rate, heterozygosis rate and repeat rate of the genome, the 17-mer distribution should obey the Poisson theoretical distribution. In the actual data, due to the sequence error, the low depth of 17-mer will take up a large proportion.)



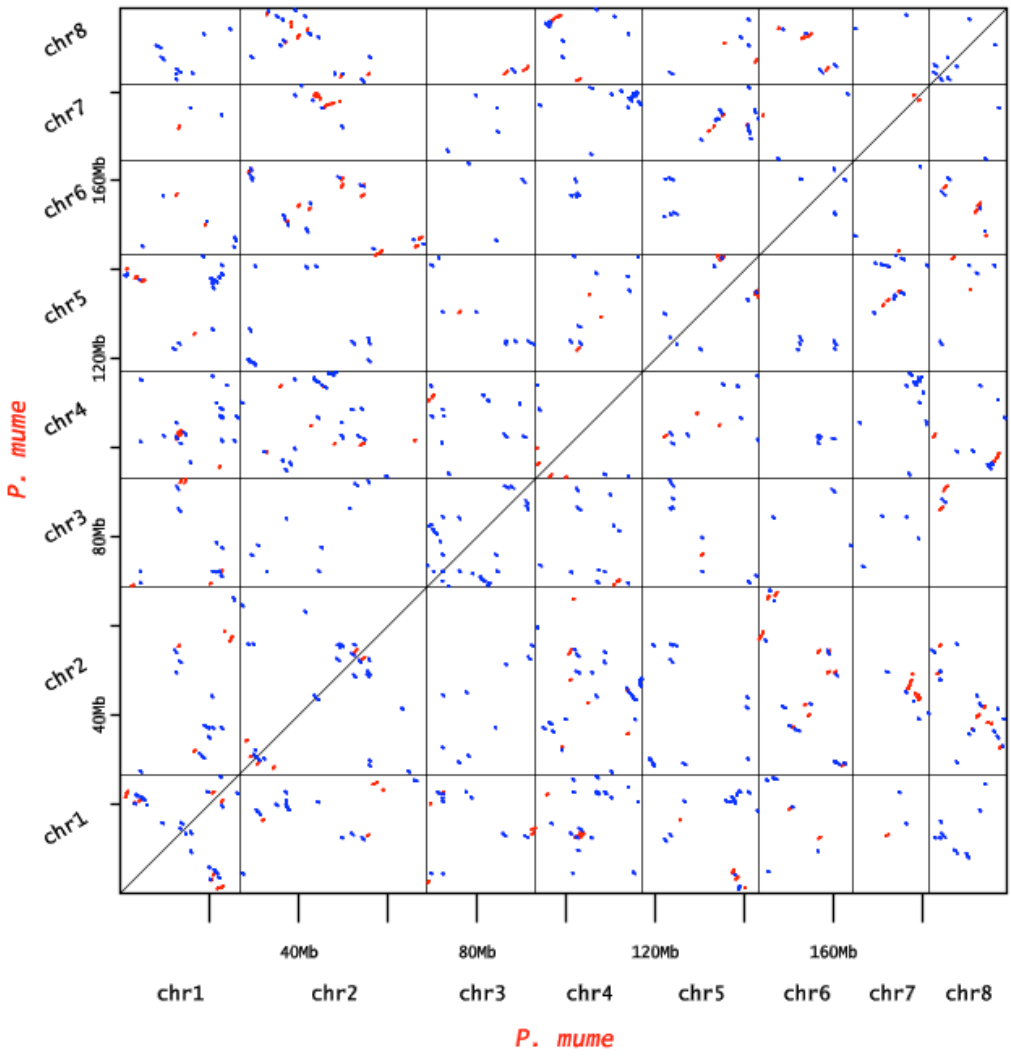
Supplementary Figure S2. The sample distribution map of GPS for *P. mume* in Tongmai town Tibet China.

Number 4 (30°06′14"N, 95°05′8"E) is the location of the sample used for *P. mume* sequencing.



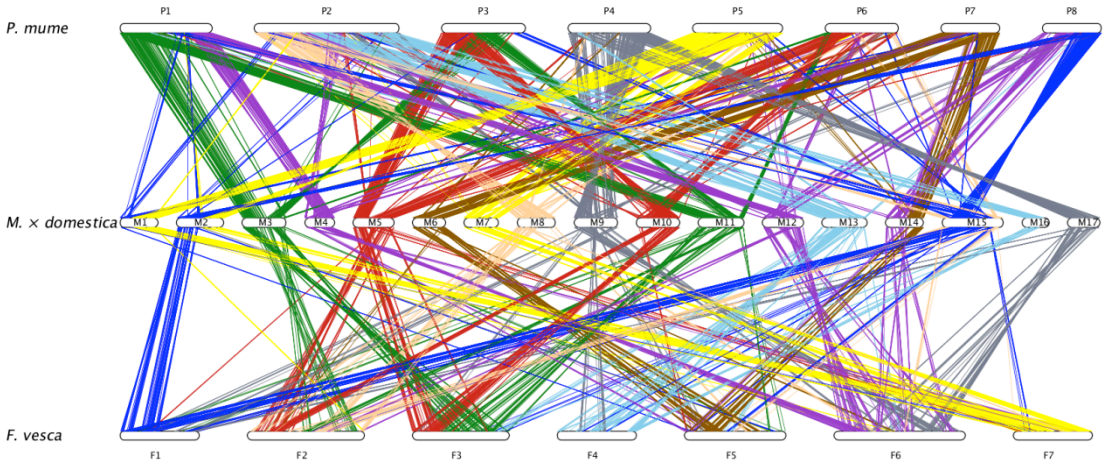
Supplementary Figure S3. Divergence rates of the transposable elements in the assembled scaffolds.

The divergence rate was calculated based on the alignment between the RepeatMasker annotated repeat copies and the consensus sequence in the repeat library.



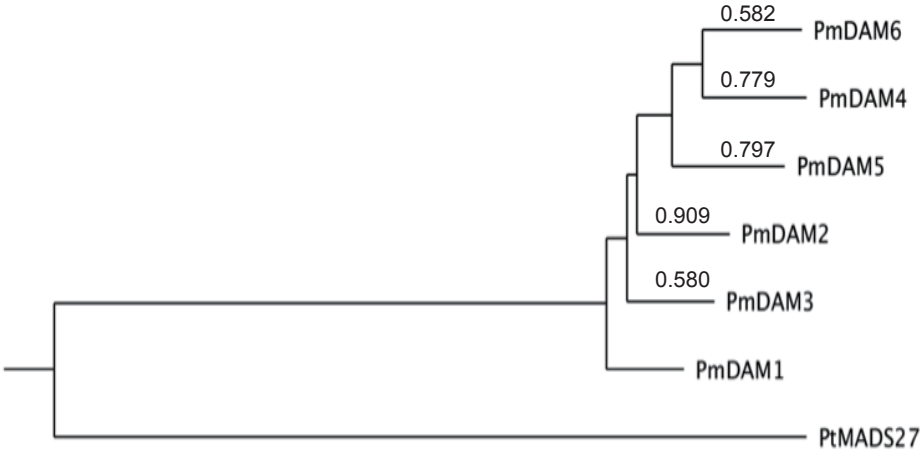
Supplementary Figure S4. Whole-genome duplication in the *P. mume* genome mapped using gene collinear order information.

Syntenic blocks are formed by red or blue dots representing best hits across any two chromosomes in the same or opposite direction, respectively.



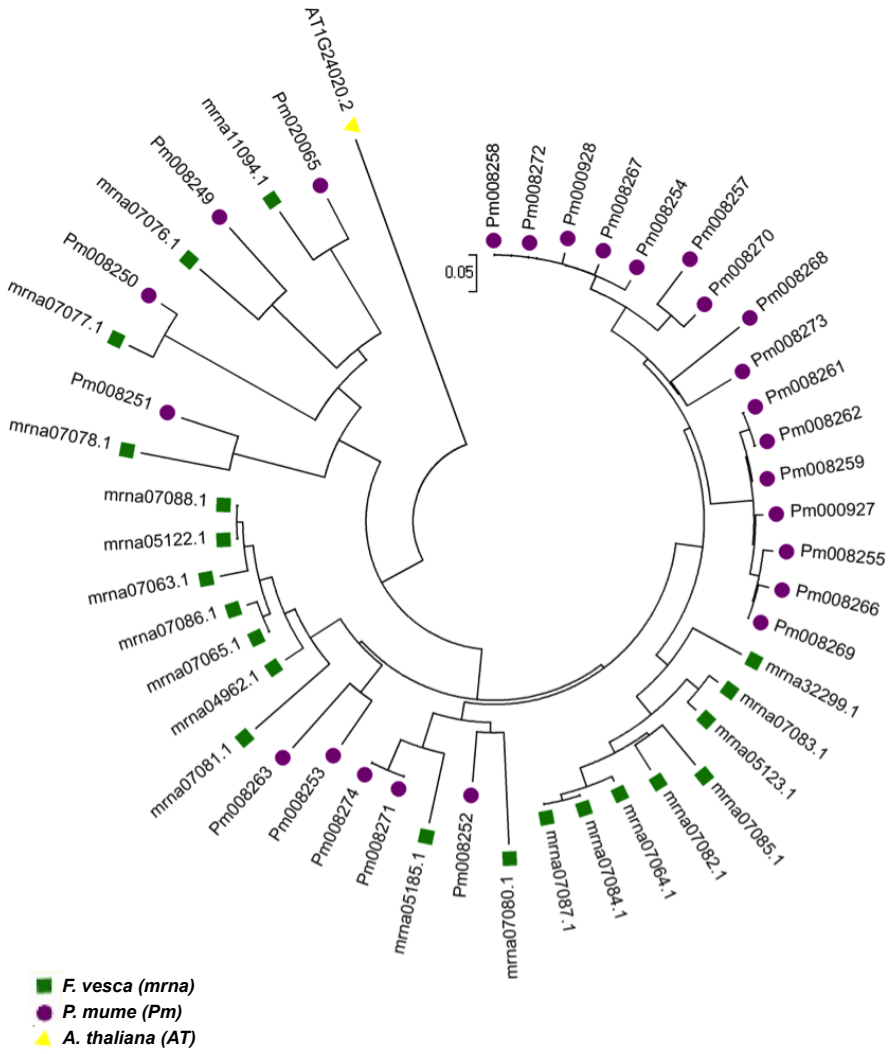
Supplementary Figure S5. The synteny between *P. mume* *F. vesca* and *M.* × *domestica*.

Schematic representation of the orthologs identified between *P. mume* (P1 to P8), *F. vesca* (F1 to F7) and *M.* × *domestica* (M1 to M17). Each line represents an orthologous gene. The nine different colors represent the blocks reflect the origin from the nine ancestral rosaceae linkage groups.



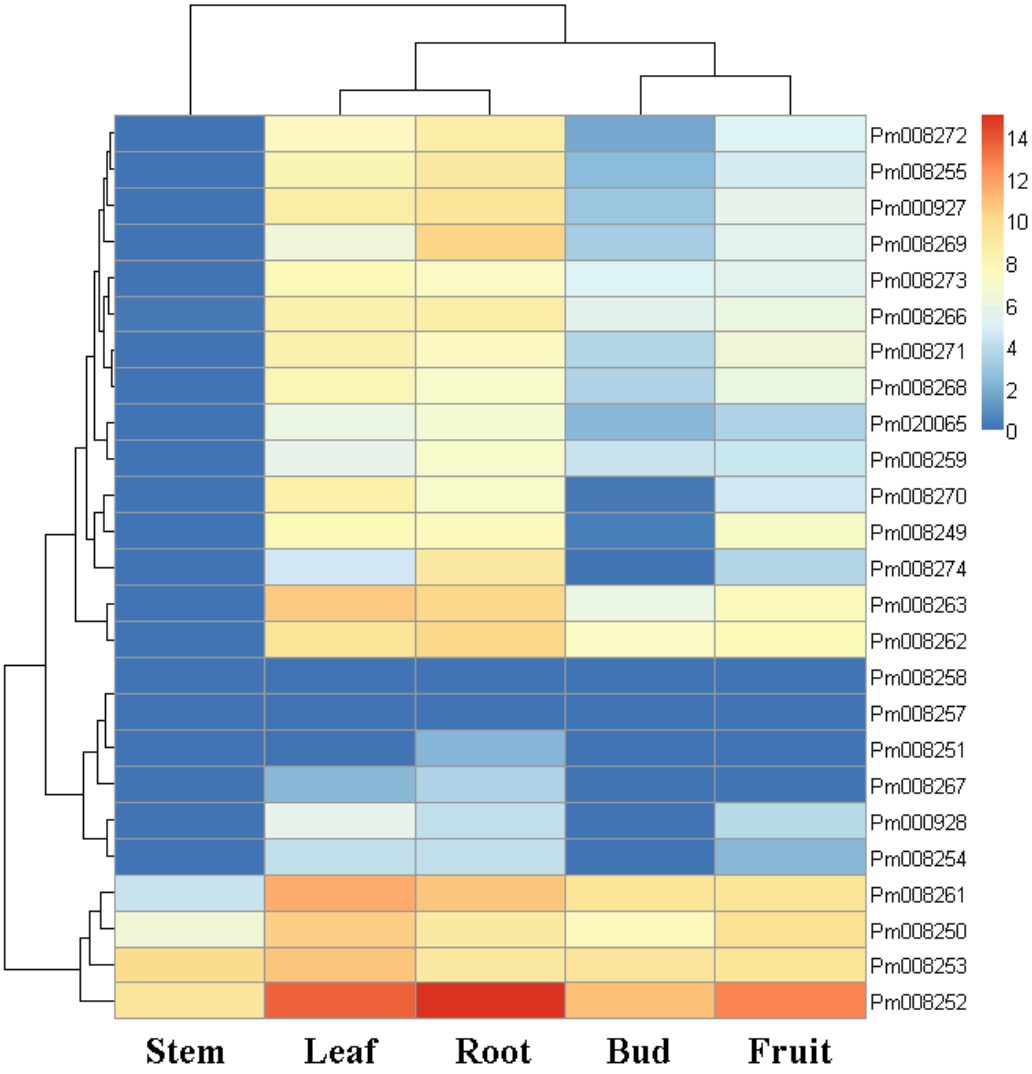
Supplementary Figure S6. Maximum likelihood rooted tree of 6 *P. mume* DAM genes.

The average ratio over all sites were shown on branches and PtMADS27 was used as outgroup.

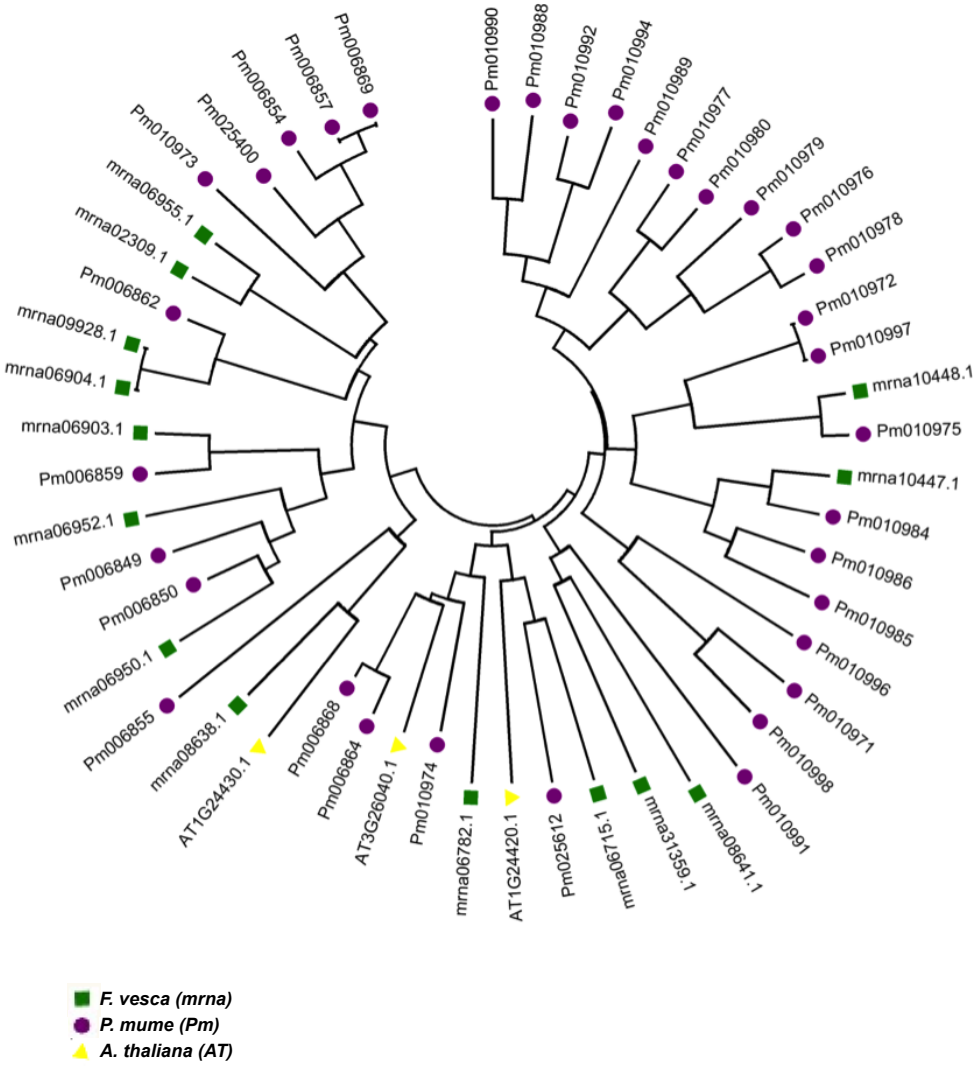


Supplementary Figure S7. Phylogenetic relationships of PR10 genes in *P. mume* (Pm) *F. vesca* (mrna) and *A. thaliana* (AT).

The phylogenetic tree was constructed using the Neighbor Joining Method with Mega 4.0 software.



Supplementary Figure S8. Relative expression of PR10 genes in tissues.
The colors represent different RPKM value.



Supplementary Figure S9. Phylogenetic relationships of BEAT genes in *P. mume* (Pm) *F. vesca* (mrna) and *A. thaliana* (AT).

The phylogenetic tree was constructed using the Neighbor Joining Method with Mega 4.0 software.

Supplementary Table S1. The sequencing data of the two domesticated *P. mume.*

NO.	Library insert size (bp)	Read length (bp)	Raw data			Filtered data		
			Total data (Gb)	Sequence depth (X)[a]	Physical depth (X)[a]	Total data (Gb)	Sequence depth (X)[a]	Physical depth (X)[a]
1	500	90PE	2.6	9.3	22.7	2.2	7.8	21.8
2	500	90PE	2.9	10.4	28.8	2.3	8.2	22.9

[a] Assumed genome size is 280 Mb.

* Physical depth= Physical coverage length/ Genome size.

Supplementary Table S2. Construction of libraries generation and filtering of sequencing data for genome assembly used.

Library insert size（bp）	Read length (bp)	Raw data			Filtered data		
		Total data (Gb)	Sequence depth (X)	Physical depth (X)	Total data (Gb)	Sequence depth (X)[a]	Physical depth (X)[a]
180	100PE	6.6	23.6	21.2	6.1	21.8	19.6
500	150PE	10.2	36.4	60.8	7.1	25.4	42.2
800	100PE	3.7	13.2	52.8	3.0	10.7	42.9
2000	45PE	2.8	10.0	222.2	2.5	8.9	198.4
5000	45PE	3.0	10.7	595.3	2.5	8.9	496.1
10000	90PE	11.4	40.7	2,261.9	4.4	15.7	873.0
20000	90PE	4.7	16.8	1,865.0	0.8	2.9	317.5
40000	50PE	8.0	28.6	11,442.9	2.0	7.1	2,871.4
Total		50.4	180.0	16,522.1	28.4	101.4	4,861.1

[a] Assumed genome size is 280 Mb.

Supplementary Table S3. Statistics the heterozygosity rate of *P. mume*.

Chromosome	Heterozygosity Site NO.	Valid site NO.	Heterozygosity rate
Chr1	7,584	24,146,087	0.0003
Chr2	9,715	38,079,607	0.0003
Chr3	7,339	21,896,908	0.0004
Chr4	6,224	21,746,526	0.0003
Chr5	8,467	23,619,679	0.0004
Chr6	5,976	19,297,980	0.0003
Chr7	5,669	15,339,843	0.0004
Chr8	3,824	15,854,464	0.0002
Total	54,798	179,981,094	0.0003

Supplementary Table S4. Chloroplast and mitochondrial sequences identification.

Chloroplast

Scaffold ID	Scaffold length	Start-End	Target ID	Target length	Start-End2
C4793941	253	1-251	NC_015206.1	155691	115485-115735
C4793941	253	1-253	NC_015996.1	159922	119315-119567
C4828461	303	177-303	NC_015996.1	159922	6633-6758
C4835979	317	151-301	NC_015206.1	155691	39193-39341
C4835979	317	151-301	NC_015996.1	159922	41145-41293
C4865545	385	253-385	NC_015206.1	155691	43729-43868
C4865545	385	170-385	NC_015996.1	159922	45613-45831
C4885025	448	266-413	NC_015206.1	155691	4610-4754
C4885025	448	1-135	NC_015996.1	159922	4242-4380
C4885565	450	59-271	NC_015206.1	155691	49609-49816
scaffold1456	7735	6784-6954	NC_000932.1	154478	4178-4348
scaffold1456	7735	1-7012	NC_015206.1	155691	4074-11480
scaffold1456	7735	1-7735	NC_015996.1	159922	3317-11725
scaffold1583	881	4-617	NC_015206.1	155691	12109-12719
scaffold1583	881	1-881	NC_015996.1	159922	12351-13233
scaffold762	5943	173-4100	NC_000932.1	154478	42581-131125
scaffold762	5943	1-5805	NC_015206.1	155691	42713-132548
scaffold762	5943	1-5943	NC_015996.1	159922	44665-136394
scaffold836	24969	1473-20991	NC_000932.1	154478	79478-154472
scaffold836	24969	1-24895	NC_015206.1	155691	79318-155678
scaffold836	24969	1-24969	NC_015996.1	159922	81615-159823

Mitochondrial

Scaffold ID	Scaffold length	Start-End	Target ID	Target length	Start-End2
C4791761	250	8-250	NC_001284.2	366924	289018-289264

C4791761	250	1-250	NC_012116.1	476890	132668-132916
C4791761	250	3-250	NC_012119.1	773279	452344-452590
C4791761	250	1-250	NC_016743.2	380861	2888-3130
C4795609	255	41-242	NC_001284.2	366924	53731-53934
C4795609	255	2-253	NC_012116.1	476890	207580-207832
C4795609	255	12-253	NC_012119.1	773279	50991-51238
C4795609	255	1-255	NC_016743.2	380861	153852-319122
C4798175	258	1-258	NC_001284.2	366924	302855-303112
C4798175	258	1-258	NC_012116.1	476890	58731-58988
C4798175	258	1-254	NC_012119.1	773279	582880-583133
C4798175	258	3-258	NC_016743.2	380861	194762-195017
C4799013	259	1-259	NC_001284.2	366924	142414-142668
C4799013	259	1-259	NC_012116.1	476890	77753-78012
C4799013	259	1-259	NC_012119.1	773279	191223-191481
C4799013	259	1-255	NC_016743.2	380861	73014-73266
C4800347	261	116-261	NC_012119.1	773279	285040-285188
C4805963	268	1-268	NC_001284.2	366924	52471-52738
C4805963	268	1-268	NC_012116.1	476890	14638-14905
C4805963	268	1-268	NC_012119.1	773279	121930-122197
C4805963	268	1-268	NC_016743.2	380861	226945-227212
C4816843	284	68-284	NC_016743.2	380861	15696-15903
C4821391	291	29-288	NC_012119.1	773279	408294-408564
C4834719	314	43-306	NC_012116.1	476890	114210-114472
C4836839	318	1-185	NC_001284.2	366924	8943-9128
C4836839	318	1-185	NC_016743.2	380861	211954-212138
C4841479	327	3-281	NC_012116.1	476890	450337-450615
C4841479	327	3-285	NC_012119.1	773279	493994-494276
C4842125	329	1-329	NC_012116.1	476890	208914-209251
C4848941	343	1-343	NC_012116.1	476890	16513-16863
C4853767	354	1-354	NC_012119.1	773279	243484-243841
C4863379	379	109-379	NC_012116.1	476890	367985-368264
C4863379	379	109-213	NC_016743.2	380861	145467-145579
C4867095	390	40-258	NC_012116.1	476890	2428-2642
C4867095	390	40-388	NC_012119.1	773279	694266-694618
C4874563	412	101-412	NC_012119.1	773279	380447-380757
C4876719	419	21-161	NC_001284.2	366924	106592-106733
C4876719	419	4-161	NC_012116.1	476890	248179-248337
C4876719	419	23-419	NC_012119.1	773279	539479-539983
C4882289	438	14-438	NC_001284.2	366924	288316-288741
C4882289	438	1-438	NC_012116.1	476890	133179-133621
C4882289	438	1-438	NC_012119.1	773279	452864-453316
C4882289	438	1-438	NC_016743.2	380861	3393-3837
C4890891	471	6-471	NC_012119.1	773279	699722-700185
C4892379	478	1-478	NC_001284.2	366924	190367-190847

C4892379	478	1-478	NC_012116.1	476890	322829-323296
C4892379	478	1-478	NC_012119.1	773279	772737-773211
C4892379	478	1-478	NC_016743.2	380861	93591-94062
C4912341	577	2-577	NC_001284.2	366924	132927-133505
C4912341	577	1-577	NC_012116.1	476890	163558-164132
C4912341	577	1-577	NC_012119.1	773279	77863-78433
C4912341	577	29-577	NC_016743.2	380861	25813-26359
C4925047	659	1-659	NC_001284.2	366924	135775-136428
C4925047	659	1-659	NC_012116.1	476890	160485-161137
C4925047	659	1-659	NC_012119.1	773279	74169-74822
C4925047	659	1-659	NC_016743.2	380861	28900-29553
C4952039	911	1-911	NC_001284.2	366924	166976-167883
C4952039	911	7-911	NC_012116.1	476890	348219-349128
C4952039	911	1-911	NC_012119.1	773279	761744-762664
C4952039	911	1-911	NC_016743.2	380861	131549-132470

Supplementary Table S5. Statistics of repeats in *P. mume* genome.

Type	Repeat Size (Mb)	% of genome
Proteinmask	17.32	7.29
Repeatmasker	12.36	5.20
Trf	10.58	4.45
Denovo	103.15	43.41
Total	106.75	44.92

Supplementary Table S6. Occurrence of transposable elements in sequenced Rosaceae genomes.

Classification	*Prunus mume*		
	Total length (Mb)	**TE coverage (%)**	**Total genome coverage (%)**
LTR/Copia	23.8	22.8	10.0
LTR/Gypsy	20.4	19.5	8.6
LTR/Other	21.8	20.8	9.2
LINE	3.1	3.0	1.3
SINE	0.9	0.9	0.4
DNA transposons	20.2	19.3	8.5
Other	1.1	1.1	0.5
Unknown	13.3	12.7	5.6
Total	104.6	100.0	44.1
Classification	***Malus* × *domestica***		

	Total length (Mb)	TE coverage (%)	Total genome coverage (%)
LTR/Copia	40.6	12.9	5.5
LTR/Gypsy	187.1	59.5	25.2
LTR/Other	3.2	1.0	0.4
LINE	48.1	15.3	6.5
SINE	-	-	-
DNA transposons	6.6	2.1	0.9
Other	-	-	-
Unknown	28.9	9.2	3.9
Total	314.5	100.0	42.4
Classification	***Fragaria vesca***		
	Total length (Mb)	**TE coverage (%)**	**Total genome coverage (%)**
LTR/Copia	10.8	22.5	5.3
LTR/Gypsy	12.9	26.8	6.4
LTR/Other	8.5	17.7	4.2
LINE	0.7	1.5	0.3
SINE	0.2	0.4	0.1
DNA transposons	12.9	26.8	6.4
Other	2.1	4.4	1.0
Unknown	-	-	-
Total	48.1	100.0	23.8

Supplementary Table S7. List of tissues and reads for whole transcriptome sequencing mapped to *P. mume* genome.

Tissue Name	**Perfect match Read NO.**	**<=5bp Mismatch Read NO.**	**Unique Match Read NO.**	**Multi-position Match Read NO.**	**Total Mapped Reads NO.**	**Total Unmapped Reads NO.**	**Total Reads NO.**	**Total BasePairs (bp)**
Bud	12,499,437	3,475,572	15,405,362	569,647	15,975,009	3,523,231	19,498,240	1,754,841,60
Fruit	17,705,866	5,789,365	22,404,254	1,090,977	23,495,231	5,338,431	28,833,662	2,595,029,58
Leaf	16,237,146	5,187,087	20,469,481	954,752	21,424,233	4,509,397	25,933,630	2,334,026,70
Root	18,373,940	6,326,385	23,528,150	1,172,175	24,700,325	6,638,567	31,338,892	2,820,500,28
Stem	12,146,280	4,286,634	15,783,948	648,966	16,432,914	3,828,296	20,261,210	1,823,508,90

Supplementary Table S8. General statistics of predicted protein-coding gene.

Gene set	**Number**	**Average length of**	**Average length of**	**#Exons per**	**Average length of**	**Average length of**

			transcribed region(bp)	CDS(bp)	gene	exon(bp)	intron(bp)
EST		4,699	2,001	562	3.1	184	701
Protein homology search	***Cucumis sativus***	24,277	2,533	1,053	4.2	253	469
	Carica papaya	27,200	2,022	913	3.7	247	411
	Fragaria vesca	29,586	2642	1043	4.0	257	521
	Arabidopsis thaliana	25,414	2412	1008	4.2	241	441
Gene finder software	**Augustus**	32,479	2,442	1,175	5.1	229	307
	Genscan	28,610	5,211	1,315	6.0	217	772
	GlimmerHMM	36,095	2,032	964	3.9	245	364
GLEAN		30,012	2,523	1,164	4.7	249	369
RNA-Seq		21,585	2,454	1,074	4.4	245	409
Combine		31,390	2,514	1,146	4.6	249	380

Supplementary Table S9. Functional annotation of predicted genes with homology or functional classification by each method.

	Database	Number	Percent
Annotated	**Swissprot**	19,696	62.8%
	InterPro	21,236	67.7%
	GO	16,822	53.6%
	KEGG	15,504	49.4%
	Trembl	25,650	81.7%
	Total	25,905*	82.5%
Unannotated		5,485	17.5%
Total		31,390	100%

***449 annotations were hits to hypothetical or uncharacterized proteins.**

Supplementary Table S10. Non-coding RNA gene fragment in the *P. mume* current assembly.

ncRNA Type	Copy #	Average length (bp)	Total length (bp)	% of genome
miRNA	**209**	**120.65**	**25,216**	**0.0106**
tRNA	**508**	**75.21**	**38,209**	**0.0012**
rRNA	**125**	**196.89**	**24,611**	**0.0103**
28S	46	348.98	16,053	0.0067
18S	17	111.29	1,892	0.0008
5.8S	11	112.55	1,238	0.0005
5S	51	106.43	5,428	0.0022
snRNA	**287**	**118.09**	**33,891**	**0.0142**
CD-box	158	98.08	15,497	0.0065
HACA-box	21	118.14	2,481	0.001

slicing	108	147.34	15,913	0.0067

Supplementary Table S11. *P. mume* genome duplication.
The table illustrates seven ancestral duplication indentified in the *P. mume* genome**.**

	Block1			Block2			Block3			Block4			Block5		
	chromosome	**start**	**end**	**chromosome**	**start**	**end**	**chromosome**	**start**	**end**	**chromosome**	**start**	**end**	**chromosome**	**start**	**end**
duplication 1	P5	16265858	18183500	P5	23396664	26040513	P7	8125911	11547528						
duplication 2	P2	6808229	12371974	P2	35608738	36935435	P4	1766505	5752689	P8	12986087	16789085			
duplication 3	P1	23177695	25259385	P2	31409881	33475726	P2	38602466	40436626	P4	292282	368157	P6	286439	3891187
duplication 4	P1	1619214	5490413	P1	20095602	22197305	P5	19956308	21313562						
duplication 5	P2	13028574	15933192	P6	5999005	11947609	P8	11039556	11238045						
duplication 6	P2	16423771	21952301	P4	19149991	23269642	P7	11799733	15655140						
duplication 7	P3	515390	3556846	P4	8365819	8496193	P4	17401251	17775106						

Supplementary Table S12. The synteny between *P. mume*, *F. vesca* and *M. × domestica*.
The number of orthologous genes per chromosome is shown in parenthesis**.**

a) Synteny between *P. mume* and *M. × domestica.*

Prunus mume **chromosome**	***Malus × domestica*** **chromosome**
P1(579)	M1(11)-M2(15)-M3(136)-M4(107)-M10(5)-M11(137)-M12(125)-M13(7)-M15(36)
P2(998)	M1(27)-M2(10)-M3(5)-M4(30)-M5(10)-M7(8)-M8(155)-M9(68)-M10(30)-M11(5)-M12(52)-M13(244)-M14(6)-M15(197)-M16(151)
P3(502)	M2(5)-M3(73)-M5(136)-M10(177)-M11(84)-M15(27)
P4(583)	M5(11)-M6(32)-M8(20)-M9(261)-M13(47)-M17(212)
P5(568)	M1(174)-M2(116)-M4(22)-M5(5)-M7(188)-M8(17)-M10(5)-M12(9)-M14(11)-M15(21)
P6(432)	M3(28)-M4(7)-M5(117)-M6(15)-M7(8)-M8(21)-M9(12)-M10(120)-M11(39)-M12(17)-M14(12)-M15(36)
P7(443)	M1(11)-M2(12)-M4(35)-M6(236)-M7(13)-M9(9)-M10(6)-M14(109)-M15(12)
P8(441)	M2(133)-M5(7)-M9(11)-M12(57)-M13(6)-M14(79)-M15(148)

b) Synteny between *F. vesca* and *M. × domestica*.

Fragaria vesca **chromosome**	***Malus × domestica*** **chromosome**
F1(211)	M2(64)-M5(5)-M9(29)-M11(9)-M15(83)-M17(21)
F2(377)	M1(6)-M3(34)-M5(53)-M6(6)-M8(72)-M10(47)-M11(24)-M12(43)-M15(92)
F3(339)	M3(81)-M5(79)-M9(10)-M10(87)-M11(63)-M13(7)-M15(7)-M16(5)
F4(169)	M9(26)-M13(88)-M15(13)-M16(42)
F5(282)	M1(5)-M2(9)-M3(5)-M4(22)-M6(106)-M8(10)-M9(10)-M10(8)-M11(9)-M13(11)-M14(64)-M15(23)
F6(384)	M1(6)-M4(53)-M6(14)-M8(18)-M9(71)-M12(96)-M13(6)-M14(40)-M15(9)-M17(71)
F7(269)	M1(112)-M2(24)-M4(10)-M5(15)-M7(92)-M8(5)-M15(11)

Supplementary Table S13. DAM gene orthologs in *P. mume.*

Gene name	Query species	ID	*P. mume* gene prediction	
			Scaffold	Genemark
PmDAM1	*Prunus persica*	gb|DQ863253.2	scaffold94	Pm004420
PmDAM2	*Prunus persica*	gb|DQ863255.1	scaffold94	Pm004419
PmDAM3	*Prunus persica*	gb|DQ863256.1	scaffold94	Pm004418
PmDAM4	*Prunus persica*	gb|DQ863250.1	scaffold94	Pm004417
PmDAM5	*Prunus persica*	gb|DQ863251.1	scaffold94	Pm004416
PmDAM6	*Prunus persica*	gb|AB437345.1	scaffold94	Pm004415

Supplementary Table S14. CBF orthologs in *Prunus mume Malus* × *domestica Fragaria vesca Populus trichocarpa Vitis vinifera Orazy sativa* and *Arabidopsis thaliana*.

Species	Number	Accession number
Prunus mume	13	Pm004870, Pm019385, Pm019386, Pm026227,Pm023766, Pm023767, Pm023768, Pm023769, Pm023770, Pm023772, Pm023775, Pm023777, Pm026221
Malus* × *domestica	10	MDP0000154764, MDP0000155057, MDP0000189347, MDP0000195376, MDP0000198054, MDP0000262710, MDP0000400129, MDP0000451365, MDP0000652413, MDP0000833641
Fragaria vesca	6	mrna13327.1, mrna13329.1, mrna30159.1, mrna30226.1, mrna32378.1, mrna32380.1
Populus trichocarpa	14	POPTR_0001s08710.1, POPTR_0001s08720.1, POPTR_0001s08740.1, POPTR_0003s12120.1, POPTR_0004s19820.1, POPTR_0006s02180.1, POPTR_0009s14990.1, POPTR_0012s13870.1, POPTR_0012s13880.1, POPTR_0013s10330.1, POPTR_0015s13830.1, POPTR_0015s13840.1, POPTR_0016s02010.1, POPTR_0019s10420.1
Orazy sativa	11	Os01t0968800-00, Os02t0558700-00, Os02t0676800-01, Os02t0677300-01, Os03t0117900-01, Os04t0572400-00, Os08t0545500-00, Os09t0522000-01, Os09t0522100-00,

		Os09t0522200-02, Os11t0242300-00
Vitis vinifera	5	GSVIVT01019860001, GSVIVT01031387001, GSVIVT01031388001, GSVIVT01033793001, GSVIVT01033795001
Arabidopsis thaliana	10	AT1G12610.1, AT1G12630.1, AT1G63030.1, AT2G35700.1, AT2G36450.1, AT4G25470.1, AT4G25480.1, AT4G25490.1, AT5G51990.1, AT5G52020.1

Supplementary Table S15. Dehydrin orthologs in *Prunus mume Malus* × *domestica Fragaria vesca Populus trichocarpa Vitis vinifera Orazy sativa* and *Arabidopsis thaliana*.

Species	Number	Accession number
Prunus mume	7	Pm000687, Pm026682, Pm026683, Pm026684, Pm020945, Pm021811, Pm006114
Malus* × *domestica	17	MDP0000126135, MDP0000129775, MDP0000178973, MDP0000196703, MDP0000265874, MDP0000269995, MDP0000360414, MDP0000529003, MDP0000595270, MDP0000595271, MDP0000629961, MDP0000689622, MDP0000698024, MDP0000770493, MDP0000862169, MDP0000868044, MDP0000868045
Fragaria vesca	7	mrna14934.1, mrna14935.1, mrna14938.1, mrna14940.1, mrna17179.1, mrna21840.1, mrna27549.1
Populus trichocarpa	8	POPTR_0002s01460.1, POPTR_0003s13850.1, POPTR_0004s16590.1, POPTR_0005s26930.1, POPTR_0009s12290.1, POPTR_0013s05870.1, POPTR_0013s05880.1, POPTR_0013s05890.1
Orazy sativa	7	Os01t0702500-01, Os02t0669100-01, Os11t0451700-00, Os11t0453900-01, Os11t0454000-01, Os11t0454200-01, Os11t0454300-01
Vitis vinifera	3	GSVIVT01018878001, GSVIVT01019440001, GSVIVT01023824001
Arabidopsis thaliana	10	AT1G20440.1, AT1G20450.1, AT1G54410.1, AT1G76180.2, AT2G21490.1, AT3G50970.1, AT3G50980.1, AT4G38410.1, AT4G39130.1, AT5G66400.1

Supplementary Table S16. Number of LRR-RLK othologous genes in each of 19 subfamilies in *Arabidopsis thaliana* (At)[30] *Pruns mume* (Pm) *Theobroma cacao* (Tc)[31] and *Populus trichocarpa* (Pt)[32].

Number of LRR-RLK for Arabidopsis, cacao and popular were obtained by

published.

Subfamily	At	Pm	Tc	Pt
LRR-I	44	15	12	19
LRR-II	14	10	10	20
LRR-III	46	33	37	63
LRR-IV	3	4	3	8
LRR-V	9	6	6	11
LRR-VI-1	5	1	5	7
LRR-VI-2	5	4	4	10
LRR-VII	8	5	6	12
LRR-VIII-1	8	6	7	15
LRR-VIII-2	12	16	19	50
LRR-IX	4	6	5	12
LRR-Xa	7	4	7	25
LRR-Xb	6	4	5	22
LRR-XI	32	69	54	54
LRR-XII	8	57	63	90
LRR-XIIIa	3	2	2	4
LRR-XIIIb	3	5	2	4
LRR-XIV	2	4	2	6
LRR-XV	2	2	4	4
Total	221	253	253	436

Supplementary Table S17. Classification of *Prunus mume* orthologous genes into one of the 19 LRR-RLK subfamilies.
One representative member of *Arabidopsis thaliana* gene[1] is cited per subfamily.

LRR-RLK subfamily	Arabidopsis representatives	*Prunus mume* accession number
LRR-I	AT3G46340	Pm002277, Pm002496, Pm006835, Pm030882, Pm014878, Pm014880, Pm014877, Pm013384, Pm014948, Pm027271, Pm005248, Pm005247, Pm030876, Pm030875, Pm020353
LRR-II	AT4G30520	Pm010269, Pm006714, Pm008622, Pm030503, Pm015448, Pm002018, Pm012811, Pm028571, Pm004091, Pm004096
LRR-III	AT3G50230	Pm020518, Pm010525, Pm001659, Pm025824, Pm006945, Pm014441, Pm021207, Pm002471, Pm005511, Pm002489, Pm005979, Pm012009, Pm024697, Pm015746, Pm022076, Pm025296,

		Pm013116, Pm008955, Pm005073, Pm027108, Pm013229, Pm019273, Pm000229, Pm003002, Pm019796, Pm024098, Pm027464, Pm013022, Pm025056, Pm026264, Pm029276, Pm018996, Pm020411
LRR-IV	AT5G51560	Pm005164, Pm017897, Pm019456, Pm023607
LRR-V	AT2G20850	Pm013128, Pm013147, Pm009700, Pm015360, Pm009945, Pm025987
LRR-VI-1	AT1G14390	Pm006551
LRR-VI-2	AT4G18640	Pm000615, Pm018968, Pm010238, Pm023844
LRR-VII	AT3G56370	Pm030375, Pm003027, Pm019069, Pm019064, Pm023650
LRR-VIII-1	AT5G37450	Pm011992, Pm011995, Pm003282, Pm002989, Pm002993, Pm000224
LRR-VIII-2	AT3G14840	Pm001341, Pm010353, Pm010355, Pm028315, Pm010187, Pm028318, Pm010184, Pm010180, Pm010179, Pm011281, Pm011285, Pm014548, Pm010657, Pm010650, Pm010658, Pm011051
LRR-IX	AT2G01820	Pm005900, Pm029428, Pm007732, Pm007731, Pm026018, Pm001495
LRR-Xa	AT1G27190	Pm005975, Pm011379, Pm013142, Pm006360
LRR-Xb	AT2G01950	Pm000305, Pm001806, Pm005928, Pm019358
LRR-XI	AT4G26540	Pm004853, Pm004852, Pm028682, Pm019206, Pm012165, Pm001768, Pm022261, Pm021906, Pm030632, Pm014378, Pm019175, Pm004847, Pm004850, Pm007681, Pm004845, Pm004842, Pm025699, Pm004844, Pm004843, Pm004830, Pm004851, Pm014573, Pm014578, Pm010946, Pm030749, Pm016991, Pm010950, Pm030977, Pm020519, Pm020526, Pm020522, Pm020527, Pm020524, Pm020528, Pm020516, Pm020529, Pm019513, Pm020500, Pm003616, Pm022449, Pm008982, Pm020415, Pm022645, Pm000370, Pm028523, Pm010866, Pm003897, Pm002307, Pm001886, Pm022247, Pm025902, Pm003639, Pm009002, Pm011209, Pm021842, Pm004312, Pm021005, Pm004310, Pm004128, Pm002311, Pm027308, Pm027305, Pm027300, Pm027299, Pm005185, Pm010102, Pm025840, Pm014536, Pm014534
LRR-XII	AT5G20480	Pm003344, Pm003343, Pm013839, Pm027761, Pm010264, Pm016619, Pm010252, Pm025421, Pm010256, Pm010251, Pm010267, Pm011739, Pm016763, Pm010244, Pm010253, Pm021185,

		Pm010415, Pm010152, Pm010248, Pm030041, Pm010285, Pm008893, Pm002944, Pm029313, Pm025787, Pm002949, Pm002952, Pm002947, Pm007467, Pm002772, Pm002785, Pm002780, Pm022102, Pm002788, Pm002778, Pm002771, Pm000729, Pm021259, Pm000909, Pm000728, Pm000911, Pm000906, Pm000727, Pm000730, Pm019477, Pm005051, Pm018903, Pm017542, Pm017541, Pm017540, Pm017543, Pm017537, Pm018902, Pm002943, Pm005356, Pm005351, Pm010287
LRR-XIIIa	AT2G35620	Pm023020, Pm024820
LRR-XIIIb	AT5G62230	Pm010851, Pm019437, Pm005587, Pm024396, Pm009223
LRR-XIV	AT2G16250	Pm027637, Pm010611, Pm021756, Pm005823
LRR-XV	AT3G02130	Pm016030, Pm030619
Unclassified		Pm008908, Pm016875, Pm028633, Pm024484, Pm016188, Pm006257, Pm026733, Pm027728, Pm010257, Pm024843

Supplementary Table S18. Numbers of orthologous genes found in *Prunus mume* (Pm) *Malus* × *domestics* (Md) *Populus trichocarpa* (Pt) *Arabidopsis thaliana* (At) *Vitis vinifera* (Vv) and *Fragaria vesca* (Fv) that encode NBS domains similar to those in plant R proteins.

(NBS data for *M.* × *domestics*, *A. thaliana*, *P. trichocarpa* and *V. vinifera* were obtained by published: Md: The genome of the domesticated apple (Malus x domestica Borkh.) Pt: Genome-wide identification of NBS resistance genes in Populus Trichocarpa At: Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis Vv: A high quality draft consensus sequence of the genome of a heterozygous grapevine variety.)

Type gene	Pm	Md	Pt	At	Vv	Fv
TIR-NBS-LRR	151	224	78	93	97	22
TIR-NBS	28	20	10	21	14	8
CC-NBS-LRR	30	181	120	51	203	37
CC-NBS	5	26	14	5	26	9
CC-TIR-NBS	0	5	-	-	-	-
CC-TIR-NBS-LRR	2	18	-	-	-	1
NBS-LRR	148	394	132	3	159	70
NBS	47	104	62	1	36	31
Total NBS with LRR	331	817	330	147	459	130
Total NBS without LRR	80	155	86	27	76	48
Total	411	972	416	174	535	178

Supplementary Table S19. Numbers of orthologous genes found in *Prunus mume* (Pm) *Malus* × *domestics* (Md) *Populus trichocarpa* (Pt) *Arabidopsis thaliana* (At) *Vitis vinifera* (Vv) *Fragaria vesca* （Fv） and *Oryza sativa* (Os) that encode pathogenesis-related proteins in plant .

Type	Pm	Md	Pt	At	Vv	Fv	Os
Pr1	17	15	18	22	11	16	34
Pr2	47	74	70	49	31	44	58
Pr3	12	15	21	14	7	9	14
Pr5	29	48	44	22	17	29	23
Pr6	2	4	4	2	2	1	1
Pr7	52	79	70	52	64	42	40
Pr8	6	21	13	1	10	7	25
Pr10	25	48	25	1	12	21	1
Total	190	304	265	163	154	169	196

Supplementary Table S20. Pathogen related genes identified in *P. mume* genome.

Gene name	Query species	ID	*P. mume* gene prediction	
			Scaffold	**Genemark**
Pr1	*Arabidopsis thaliana*	TAIR|AT2G14610.1	scaffold694	Pm021063
			scaffold694	Pm021062
			scaffold694	Pm021061
			scaffold694	Pm021060
			scaffold694	Pm021059
			scaffold694	Pm021058
			scaffold694	Pm021055
			scaffold269	Pm010664
			scaffold68	Pm002330
			scaffold144	Pm026541
			scaffold144	Pm026666
			scaffold216	Pm020984
			scaffold216	Pm020985
			scaffold2458	Pm029379
			scaffold491	Pm008682
			scaffold491	Pm008683
			scaffold130	Pm002812
Pr2	*Fragaria ananassa*	gb|AY989819.1	scaffold538	Pm007733
			scaffold236	Pm006688
			scaffold559	Pm001330
			scaffold841	Pm031195
			scaffold841	Pm031196

scaffold484	Pm027058
scaffold484	Pm027057
scaffold1057	Pm028251
scaffold1057	Pm028252
scaffold384	Pm019219
scaffold576	Pm012164
scaffold576	Pm012166
scaffold576	Pm012168
scaffold326	Pm009666
scaffold272	Pm029662
scaffold548	Pm018808
scaffold55	Pm006322
scaffold10	Pm015150
scaffold535	Pm027288
scaffold170	Pm000889
scaffold217	Pm011905
scaffold442	Pm013018
scaffold246	Pm001815
scaffold373	Pm026985
scaffold144	Pm026504
scaffold181	Pm008313
scaffold2226	Pm029335
scaffold349	Pm010418
scaffold281	Pm005809
scaffold528	Pm000021
scaffold175	Pm025784
scaffold175	Pm025785
scaffold175	Pm025782
scaffold61	Pm019057
scaffold15	Pm000196
scaffold148	Pm022936
scaffold1094	Pm028290
scaffold151	Pm021740
scaffold668	Pm027374
scaffold461	Pm009548
scaffold572	Pm013667
scaffold441	Pm016681
scaffold443	Pm009644
scaffold178	Pm025354
scaffold98	Pm004932
scaffold41	Pm027564

			scaffold182	Pm019717
Pr3	*Fragaria ananassa*	gb|AF320111.1	scaffold307	Pm008747
			scaffold580	Pm026843
			scaffold181	Pm008299
			scaffold177	Pm007648
			scaffold134	Pm007414
			scaffold1287	Pm028448
			scaffold159	Pm020963
			scaffold549	Pm012870
			scaffold198	Pm007477
			scaffold198	Pm007487
			scaffold198	Pm007488
			scaffold198	Pm007492
Pr5	*Malus × domestica*	gb|DQ318213.1	scaffold34	Pm003991
			scaffold521	Pm014435
			scaffold830	Pm023785
			scaffold144	Pm026768
			scaffold724	Pm009145
			scaffold281	Pm005777
			scaffold175	Pm025781
			scaffold175	Pm025780
			scaffold175	Pm025779
			scaffold175	Pm025777
			scaffold175	Pm025775
			scaffold175	Pm025772
			scaffold175	Pm025769
			scaffold175	Pm025739
			scaffold175	Pm025774
			scaffold61	Pm019088
			scaffold125	Pm015816
			scaffold115	Pm018833
			scaffold159	Pm020840
			scaffold159	Pm020839
			scaffold264	Pm023883
			scaffold264	Pm023884
			scaffold264	Pm023886
			scaffold264	Pm023887
			scaffold130	Pm002662
			scaffold266	Pm029535
			scaffold139	Pm006193
			scaffold139	Pm006194

			scaffold182	Pm019646
Pr6	*Arabidopsis thaliana*	gb|AY065127.1	scaffold1307	Pm028462
			scaffold880	Pm031238
Pr7	*Lycopersicon esculentum*	gb|Y17276.1	scaffold289	Pm018093
			scaffold289	Pm018094
			scaffold564	Pm018682
			scaffold500	Pm003658
			scaffold484	Pm027094
			scaffold269	Pm010634
			scaffold482	Pm026205
			scaffold326	Pm009702
			scaffold1175	Pm028353
			scaffold2134	Pm029257
			scaffold221	Pm019423
			scaffold227	Pm002421
			C5025277	Pm024247
			scaffold230	Pm000652
			scaffold230	Pm000653
			scaffold373	Pm026972
			scaffold250	Pm000069
			scaffold108	Pm021706
			scaffold897	Pm031276
			scaffold199	Pm001365
			scaffold357	Pm026384
			scaffold101	Pm020495
			scaffold61	Pm018841
			scaffold61	Pm018842
			scaffold1416	Pm016596
			scaffold157	Pm014928
			scaffold884	Pm031249
			scaffold884	Pm031251
			scaffold25	Pm023747
			scaffold25	Pm023746
			scaffold45	Pm016775
			scaffold159	Pm020848
			scaffold57	Pm013171
			scaffold57	Pm013240
			scaffold341	Pm005378
			scaffold341	Pm005377
			scaffold341	Pm005300
			scaffold341	Pm005299
			scaffold341	Pm005298
			scaffold313	Pm010336
			scaffold443	Pm009648

			scaffold473	Pm008916
			scaffold135	Pm000397
			scaffold135	Pm000396
			scaffold139	Pm006167
			scaffold139	Pm006209
			scaffold155	Pm023689
			scaffold155	Pm023690
			scaffold155	Pm023691
			scaffold155	Pm023693
			scaffold155	Pm023694
			scaffold155	Pm023702
Pr8	*Malus × domestica*	gb|DQ318214.1	scaffold303	Pm018434
			scaffold133	Pm018214
			scaffold133	Pm018201
			scaffold133	Pm018170
			scaffold182	Pm019804
			scaffold182	Pm019808
Pr10	*Prunus persica*	gb|EU117120.1	scaffold181	Pm008249
			scaffold181	Pm008250
			scaffold181	Pm008251
			scaffold181	Pm008252
			scaffold181	Pm008253
			scaffold181	Pm008254
			scaffold181	Pm008255
			scaffold181	Pm008257
			scaffold181	Pm008259
			scaffold181	Pm008261
			scaffold181	Pm008262
			scaffold181	Pm008263
			scaffold181	Pm008266
			scaffold181	Pm008267
			scaffold181	Pm008268
			scaffold181	Pm008269
			scaffold181	Pm008270
			scaffold181	Pm008271
			scaffold181	Pm008272
			scaffold181	Pm008273
			scaffold181	Pm008274
			scaffold181	Pm008258
			scaffold501	Pm000928
			scaffold501	Pm000927
			scaffold132	Pm020065

Supplementary Table S21. Numbers of orthologous genes found in *Prunus mume* (Pm) *Malus* × *domestics* (Md) *Populus trichocarpa* (Pt) *Arabidopsis thaliana* (At) *Vitis vinifera* (Vv) *Fragaria vesca*（Fv）and *Oryza sativa* (Os) that synthesis volatile molecules.

Type gene	Pm	Md	Pt	At	Vv	Fv	Os
PAL	2	6	5	4	5	2	9
ODO1	2	2	4	1	2	2	1
BPBT	13	25	27	11	12	12	29
CFAT	4	5	2	4	4	5	2
BSMT	21	32	25	23	25	15	13
CCMT	12	34	23	21	25	12	11
BEAT	34	16	17	3	4	14	--
OOMT	13	37	28	9	18	9	23
IEMT	2	44	30	15	12	14	13
EGS	9	13	18	8	18	10	7
IGS	2	10	17	8	17	10	7
POMT	6	43	32	15	12	14	11
SAMT	10	33	24	24	25	14	10
PAAS	6	4	5	2	5	6	7
ɣ-terpinene-synthas	5	21	33	9	31	27	5
β-pinene-synthase	4	20	34	10	29	27	5
germacrene	16	19	34	30	30	28	8
TPS10	8	13	13	10	7	9	8
Linalool synthase	1	2	2	1	--	1	2
CCD	6	12	16	7	7	6	5
Limonene-3-hydroxylase	68	111	108	93	50	56	97

Supplementary Table S22. The distribution of BEAT gene clusters

Chr	Source	Type	Start	End	Score	Strand	Phase	Attributes
Pm2	GLEAN	mRNA	19561090	19562397	0.541383	+	.	ID=Pm006849;
Pm 2	GLEAN	mRNA	19568381	19569688	0.801224	+	.	ID=Pm006850;
Pm 2	GLEAN	mRNA	19597787	19599115	0.999924	+	.	ID=Pm006854;
Pm 2	GLEAN	mRNA	19600651	19601989	0.737153	+	.	ID=Pm006855;
Pm 2	GLEAN	mRNA	19640968	19642296	0.999955	+	.	ID=Pm006857;
Pm 2	GLEAN	mRNA	19652034	19653344	0.999955	+	.	ID=Pm006859;
Pm 2	Cuff	mRNA	19670279	19672730	100	+	.	ID=Pm006862;
Pm 2	GLEAN	mRNA	19680726	19682066	0.978451	-	.	ID=Pm006864;
Pm 2	GLEAN	mRNA	19710735	19712075	0.978451	-	.	ID=Pm006868;
Pm 2	GLEAN	mRNA	19718018	19719346	1	-	.	ID=Pm006869;
Pm 3	GLEAN	mRNA	8086964	8088316	1	-	.	ID=Pm010971;
Pm 3	GLEAN	mRNA	8097955	8098455	0.888754	-	.	ID=Pm010972;
Pm 3	GLEAN	mRNA	8107320	8109158	0.806803	+	.	ID=Pm010973;
Pm 3	GLEAN	mRNA	8110118	8111458	1	-	.	ID=Pm010974;
Pm 3	GLEAN	mRNA	8113148	8114512	0.999903	-	.	ID=Pm010975;
Pm 3	GLEAN	mRNA	8115043	8116374	1	-	.	ID=Pm010976;
Pm 3	GLEAN	mRNA	8119387	8120736	1	-	.	ID=Pm010977;
Pm 3	GLEAN	mRNA	8121890	8123008	0.799638	-	.	ID=Pm010978;
Pm 3	GLEAN	mRNA	8129028	8130350	1	-	.	ID=Pm010979;
Pm 3	GLEAN	mRNA	8131250	8132614	1	-	.	ID=Pm010980;
Pm 3	GLEAN	mRNA	8158666	8159994	1	-	.	ID=Pm010984;
Pm 3	GLEAN	mRNA	8161540	8162847	1	-	.	ID=Pm010985;
Pm 3	GLEAN	mRNA	8169726	8171039	0.999725	-	.	ID=Pm010986;
Pm 3	GLEAN	mRNA	8192054	8193376	0.999734	-	.	ID=Pm010988;

Pm 3	GLEAN	mRNA	8204439	8205755	1	-	.	ID=Pm010989;
Pm 3	GLEAN	mRNA	8215947	8217296	0.56471	-	.	ID=Pm010990;
Pm 3	GLEAN	mRNA	8225266	8226025	0.999194	-	.	ID=Pm010991;
Pm 3	GLEAN	mRNA	8229556	8230902	0.999194	-	.	ID=Pm010992;
Pm 3	GLEAN	mRNA	8236905	8238248	1	-	.	ID=Pm010994;
Pm 3	GLEAN	mRNA	8247708	8249048	1	-	.	ID=Pm010996;
Pm 3	GLEAN	mRNA	8252941	8255332	0.67247	+	.	ID=Pm010997;
Pm 3	GLEAN	mRNA	8256871	8258223	1	+	.	ID=Pm010998;
Pm 8	GLEAN	mRNA	916180	917508	0.996351	-	.	ID=Pm025400;
Pm 8	GLEAN	mRNA	3451935	3453828	0.999999	-	.	ID=Pm025612;

Supplementary Methods

K-mer analysis

We determined the relationship between sequencing depth and the copy number of a certain K-mer (refers to a sequence with K base pairs e.g., 17-mer) and if the sequence error rate, heterozygosis rate and repeat rate of the genome were ignored, the K-mer of distribution should obey the Poisson theoretical distribution. The size of the genome was estimated using the total length of the sequence reads divided by the sequencing depth. And the peak value of the frequency curve represents the overall sequencing depth. We estimated the genome size as (N× (L-K+1)-B)/D＝G, where N is the total number of sequence reads, L is the average length of sequence reads, and K is K-mer length, defined as 17 bp. B is the total number of low-frequency (frequency <=1 in this analysis) K-mers. G is the genome size, and D is the overall depth, estimated from K-mer distribution. It must be pointed out that as the K-mer of distribution should approximate the Poisson, not all low-frequency k-mers will be errors. This might lead to an underestimate of the genome size, especially when sequencing depth is low.

The following things will also affect the genome size estimate. In the actual data, due to the sequence error, the low depth of K-mer will take up a large proportion. At the same time, for some genomes, the heterozygosis rate can cause a sub peak at the position of the half of the main peak, while a certain repeat rate can cause a repeat peak at the position of the integer multiples of the main peak. In addition, the peak position will be affected by errors that pushing the peak to the left, as errors generate more unique kmers and by presence of repeats that pushing the peak to the right. The inclusion of mitochondrial and chloroplast DNA in the sample, along with any other contamination will result in larger genome size estimation.

Estimation of heterozygosity rate

The heterozygosity rate was calculated by calling the heterozygous SNPs. All the high quality reads were mapped to the genome assembly using the software SOAP2 (http://soap.genomics.org.cn/soapaligner.html) with the cutoff less than 5 mismatches.

Then the alignment results were analyzed for SNP mining using the SOAPsnp (http://soap.genomics.org.cn/soapsnp.html). The sites that met the following criteria were searched and named as criterion effective sites: quality score of consensus genotype in the SNP mining result is greater than 20; count of all the mapped best and second best base are supported by at least 4 unique reads; sequencing depth is more than 10X; and SNPs are at least 5 bp away from each other, with an additional requirement to the criterion effective sites that the number of reads supported the best base is small than four times of the number of reads supported second best base (reads supported best base/reads supported second best base < 4) were identified as heterozygotic sites. Finally, the rate of the heterozygosity was estimated as the number of heterozygotic sites divided by the number of criterion effective sites.

Plant material and identification of RAD markers

The genetic maps that were used to develop the integrated map for anchoring the scaffolds were derived from F_1 populations, totaling 260 individuals (accession No. BJFU1210120025-0284) from the cross between 'Fenban' (accession No. BJFU1210120013) and 'Kouzi Yudie' (accession No. BJFU1210120022) from Qingdao Meiyuan. Young leaves of these *P. mume* seedlings and their parents were collected for DNA extraction. Genomic DNA was isolated from the leaves using the Plant Genomics DNA Kit (TIANGEN, Beijing, China) according to the manufacturer's recommendations.

The RAD protocols were the same as in Chutimanitsakun[15], except we used EcoRI (recognition site: 5'G^AATTC3'). Every 24 F_1 plants were pooled into one sequencing library with nucleotide multiplex identifiers (4 bp, 6 bp, and 8 bp) and each individual plant was barcoded. Approximately 830 Mb of 50-bp reads (3.1 Mb of reads data for each progeny on average) was generated on the NGS platform HiSeq2000. The SNP calling process was performed using the SOAP2+SOAPsnp pipeline[53].

Identification of repetitive elements

There are two main types of repeats in the genome, tandem repeats and interspersed repeats. We used Tandem Repeats Finder[54] (Version 4.04) and Repbase (composed of

many transposable elements, Versions 15.01) to identify interspersed repeats in the *P. mume* genome. We identified transposable elements in the genome at the DNA and protein levels. For the former, RepeatMasker (Version 3.2.7) was applied using a custom library (a combination of Repbase, a de novo transposable element library of the *P. mume* genome). For the latter, RepeatProteinMask, an updated tool in the RepeatMasker package, was used to conduct RM-BlastX searches against the transposable element protein database[55]. Identified repeats were classified into various categories.

RNAseq data generation

RNA was purified using TRIzol (Invitrogen, CA, USA) from five fresh tissues (bud, fruit, leaf, root and stem). RNA sequencing libraries were constructed using the mRNA-Seq Prep Kit (Illumina, San Diego, USA). Briefly, first strand cDNA synthesis was performed with oligo-T primer and Superscript II reverse transcriptase (Invitrogen). The second strand was synthesized with E. coli DNA Pol I (Invitrogen, CA, USA). Double stranded cDNA was purified with a Qiaquick PCR purification kit (Qiagen), and sheared with a nebulizer (Invitrogen, CA, USA) to 100-500 bp fragments. After end repair and addition of a 3'-dA overhang the cDNA was ligated to Illumina PE adapter oligo mix (Illumina), and size selected to 200 ± 20 bp fragments by gel purification. After 15 cycles of PCR amplification the 200 bp paired-end libraries were sequenced using the paired-end sequencing module (90 bp at each end) of the Illumina HiSeq 2000 platform.

Gene function annotation

Genes were aligned to the SwissProt[16] (release 2011.6) and TrEMBL (release2011.6) databases using BLASTP (1e-5) to determine the best match of the alignments. InterProScan[17] (version 4.5) motifs and domains of the genes were identified against protein databases of Pfam (release 24.0), PRINTS (release 40.0), PROSITE (release 20.52), ProDom (release 2006.1), and SMART[56] (release 6.0). Gene Ontology IDs for each gene were obtained by the corresponding InterPro entry. The genes were aligned against KEGG proteins[18] (release 58), and the matches were used to establish the KEGG pathway.

Identification of non-coding RNA genes

The tRNA genes were predicted by tRNAscan-SE[57] (Version 1.23). For rRNA identification, the rRNA template sequences (e.g., Arabidopsis thaliana and rice) were aligned against the *Prunus mume* genome using blastn to identify possible rRNAs. Other noncoding RNAs, including miRNA, snRNA, were identified using INFERNAL[58] (Version 0.81) by searching against the Rfam database (Release 9.1).

Identification of CBF, PR and BEAT genes

The CBF genes of *P. mume* were identified with *A. thaliana* CBF genes using BLASTP (E-value < 1e-10, identity > 30% and coverage > 70%). The PR genes of *P. mume* were identified with PR genes of all species using BLASTP (E-value < 1e-10, identity > 30% and coverage > 70%). The BEAT genes of *P. mume* were identified with BEAT genes (Gene Bank ID: AF043464) using BLASTP (E-value < 1e-10, identity > 30% and coverage > 70%).

Identification of LRR-LRK genes in the *P. mume* genome

LRR-LRK genes contain leucine-rich repeat motifs and like kinase domains. Based on this structural profile, we used HMMER 3.0 to search LRRs (PF00560) and kinase domains (PF00560) in *P. mume* protein sequences. We extracted the kinase domain sequences of the predicted proteins and 19 *A. thaliana* sequences of previous studies and aligned them with Clustalw[59]. Using PHYML 3.0 software, based on this alignment, we generated a phylogenetic tree by maximum likelihood method with 100 bootstrap replicates, classifying the *P. mume* sequences into 19 LRR-LRK subfamilies.

Identification of NBS domain genes

Identification of NBS genes in the *P. mume* genome were screened using HMMER 3.0 (http://hmmer.janelia.org/software) against the raw hidden Markov model (HMM), corresponding to the Pfam (http://pfam.sanger.ac.uk/) NBS (NB-ARC) family PF00931 domain (E value cutoff of 1.0). To detect TIR domains, the predicted NBS-encoding amino acid sequences were screened using the HMM model Pfam TIR PF01582 domain (E value cutoff of 1.0). To detect LRR motifs, a Pfam HMM search was used. The raw LRR_1 (PF00560), LRR_2 (PF07723), and LRR_3 (PF07725)

data were downloaded and compared against the NBS-encoding amino acids using HMMER 3.0 (E value cut off 1.0). Coiled-coil (CC) motifs were analyzed using MARCOIL[60] with a threshold probability of 90 and double-checked using prircoil2 with a P score cutoff of 0.025.

Relative expressions calculation

Firstly, we aligned the raw sequence reads to all genes using SOAPaligner or SOAP2 software included in SOAP package, and the parameters were "-m 0 -x 1000 -s 40 -l 32 -v 3 -r 2 -p 3". Secondly, we calculated the unique mapping reads of each mapped gene and used reads per kilobase per million reads sequenced (RPKM)[49] method to check the value of gene expression. RPKM = (10^(6)*C)/(NL/10^3). (Set RPKM (A) to be the expression of gene A, C to be number of reads that uniquely aligned to gene A, N to be total number of reads that uniquely aligned to all genes, and L to be the base number in the CDS of gene A.) The RPKM method is able to eliminate the influence of different gene length and sequencing discrepancy on the calculation of gene expression.

Supplementary references

53 Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966-7 (2009).

54 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573-80 (1999).

55 Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**, 462-7 (2005).

56 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).

57 Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-64 (1997).

58 Nawrocki, E.P., Kolbe, D.L. & Eddy, S.R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335-7 (2009).

59 Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8 (2007).

60 Delorenzi, M. & Speed, T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617-25 (2002).