Supplementary Figures 1-34



Supplementary Figure 1. K-mer distribution of the celery genome. (a) K-mer=17 Depth and K-mer number frequency distribution. (b) K-mer=17 Depth and K-mer type frequency distribution.



Supplementary Figure 2. The frequency and the length of the main different types of repetitive sequences in celery genome, including DNA, LINE, LTR, and SINE repeats.



Supplementary Figure 3. The density of divergence rates within and among celery, coriander, and carrot LTR repeat sequences. The 3' and 5' represent the 3'- and 5'- ends of LTR, respectively.



Supplementary Figure 4. Tandem repeat distribution in the celery genome. The green dot line indicates the tandems repeat density, and the blue dot line indicates the number of tandem repeat region. (a) the unit of the tandem repeat was =<10 bp. (b) the unit of the tandem repeat was =<50 bp.

Unit

a



Supplementary Figure 5. Tandem repeat distribution in the celery genome. Red dots indicate the tandem repeat density, and blue dots indicate the number of tandem repeat region. (a) all the unit (0~800000) of the tandem repeat. (b) the enlargement of Y axis (0~10000) of the tandem repeat.



Supplementary Figure 6. The predication of putative centromere regions of each celery chromosome according the peak and length of tandem repeat density. The blue triangle indicates the putative centromere region of each chromosome. The x axis represented the length of chromosome with 1 Mb as unit, and the y axis showed the percentages of tandem repeats along 11 chromosomes, respectively.



Evidence Support

Supplementary Figure 7. The statistics of gene set evidence supports in celery genome. De novo_0.5, EVM integrates genes supported by *Denovo* prediction; Homolog_0.5, genes supported by homologous prediction when EVM integration; RNA_0.5, genes supported by RNA-seq during EVM integration. The gene overlap is greater than 50% as a standard, and the number indicates the number of genes.



Supplementary Figure 8. (a) The statistics of gene distribution on each chromosome of celery, and other 5 species. *%5 means that we cut the 5% of the full length of the chromosome at two sides of the chromosome, and then we statistic the distribution density of gene in the two intervals. (b) The statistics of snRNA, miRNA, rRNA, and tRNA distribution on each celery chromosome. *%5 means that we cut the 5% of the full length of the chromosome at two sides of the chromosome, and then we statistic the distribution the distribution on each celery chromosome. *%5 means that we cut the 5% of the full length of the chromosome at two sides of the chromosome, and then we statistic the distribution density of each type of ncRNA in the two intervals.



Supplementary Figure 9. The chromosomal distribution of ncRNA of celery. miRNA, rRNA, snRNA, and tRNA are displayed in circles from outsides to insides, respectively.



Supplementary Figure 10. Gene family analyses. (a) The distribution of gene numbers and family sizes in celery (Ag: *A. graveolens*) and 6 other representative species (At: *Arabidopsis thaliana*; Cs: *Coriandrum sativum*; Dc: *Daucus carota*; Ls: *Lactuca sativa*; Vv: *Vitis vinifera*, Pg: *Panax ginseng*). (b) Common and lineage-specific gene families in celery and 6 species.



Supplementary Figure 11. The phylogenetic tree of grape and Apiaceae orthologous genes. (a) γ denoted by red circle, the α and ω events denoted by blue circle. A grape gene has four orthologs in the Apiaceae species genome, and named as A11, A12, A21, and A22. The species tree was produced based on our present analysis of homologous genes. (b) Homologous gene dotplot between grape and celery shows an orthologous gene ratio 1:4. (c) Homologous gene dotplot between lettuce and celery shows an orthologous gene ratio 3:4.



Supplementary Figure 12. (a) Homologous dotplot within celery genome. Median Ks value for duplicated genes on each inferred colinear block is shown, in color from red to purple as to the value of Ks



Supplementary Figure 12. (b) Homologous dotplot between celery and coriander genomes. Median Ks value for duplicated genes on each inferred colinear block is shown, in color from red to purple as to the value of Ks.



Supplementary Figure 12. (c) Homologous dotplot between celery and carrot genomes. Median Ks value for duplicated genes on each inferred colinear block is shown, in color from red to purple as to the value of Ks.



Supplementary Figure 12. (d) Homologous dotplot between celery and lettuce genomes. Median Ks value for duplicated genes on each inferred colinear block is shown, in color from red to purple as to the value of Ks.



Supplementary Figure 12. (e) Homologous dotplot between celery and grape genomes. Median Ks value for duplicated genes on each inferred colinear block is shown, in color from red to purple as to the value of Ks.



Supplementary Figure 13. Homologous alignment of Apiaceae and other relative genomes with grape as a reference. Here, we show global alignment of homologous regions in lettuce (L), carrot (D), coriander (C), celery (A) genomes and grape (V). Genomic paralogy and orthology information within and among genomes, is displayed with 48 circles, each corresponding to (sub)genomes produced by recursive polyploidizatioin. The curved lines within the innermost circle show paralogous genes among 19 grape chromosomes, in 7 colors corresponding to the 7 ancestral chromosomes before the WGT (γ). The short lines forming the innermost grape chromosome circles represent grape genes. Each of the three sets of grape paralogous regions has 3 orthologous regions in lettuce, and 4 in celery, coriander and carrot, respectively. This results in 48 circles marked according to species indicated by a capital letter.



Supplementary Figure 14. Histograms and fitted curves of Ks between colinear homologous genes. The median Ks value of each homologous block is used. (a) intragenomic; (b) intergenomic.



Supplementary Figure 15. The original and corrected synonymous nucleotide substitutions (Ks), and inferred evolutionary dates among colinear genes. (a) Ks distribution and inferred evolutionary dates within a species before evolutionary rate correction (representing the polyploidization events). (b) Ks distribution and inferred evolutionary dates within a species after evolutionary rate correction (representing the polyploidization and inferred evolutionary dates between any two species before evolutionary rate correction (representing the speciation events). (d) Ks distribution and inferred evolutionary dates between any two species after evolutionary for evolutionary dates between any two species after evolutionary for evolutionary dates between any two species after evolutionary for ev



Supplementary Figure 16. Divergence time estimated among 7 species. The tree topology was generated by MCMCtree and the CDS sequences. The red dots correspond to calibration points as mentioned in the method. Divergence estimates (Mya, million years ago) are indicated above each node and the blue nodal bars show the 95% confidence intervals (n=20,000 independent samples).



Supplementary Figure 17. (a) The retention of duplicated genes residing in 4 subgenomes of celery using the grape as reference.



Supplementary Figure 17. (b) The retention of duplicated genes residing in celery subgenomes using the carrot as reference.



Supplementary Figure 17. (c) The retention of duplicated genes residing in celery genome using the coriander as reference.



Supplementary Figure 18. Fitting a geometric distribution to gene loss frequencies in celery as to the coriander (a), carrot (b), and grape (c), respectively.



Supplementary Figure 19. The heat map of correlation coefficients was shown between any two of samples by RNA-Seq. (a) The RNA-Seq for three tissues (root, petiole, leaf) of celery. (b) The RNA-Seq for 3 different-colored sample (Green, Red, and White) of celery.



Supplementary Figure 20. Gene expression in different tissues and different celery varieties with different stem colors. The whole FPKM hierarchical clustering map, log10 (FPKM+1) values are normalized (scale number) and clustered, red indicates high expression gene, blue indicates low expression gene. The color ranges from red to blue, indicating log10 (FPKM+1) from large to small. (a) RNA-Seq analysis for three tissues (root, petiole, leaf) of celery. (b) RNA-Seq analysis for three different colors sample (green, red, and white) of celery.



Supplementary Figure 21. Venn diagrams showed the common or specific DEGs among (a) different tissues or (b) different celery varieties with different stem colors.



Supplementary Figure 22. The GO enrichment analyses of DEGs between any two tissues of celery with corrected P-value <0.05. (a) Leaf vs Root; (b) Leaf vs Petiole; (c) Root vs Petiole.



Supplementary Figure 23. The GO enrichment analyses of DEGs between any two varieties of celery with corrected *P***-value <0.05.** (a) Red vs Green; (b) White vs Green; (c) Red vs White.



Supplementary Figure 24. The KEGG enrichment analyses of DEGs between any two tissues of celery. (a) Leaf vs Root; (b) Leaf vs Petiole; (c) Root vs Petiole.



Supplementary Figure 25. The KEGG enrichment analyses of DEGs between any two varieties of celery. (a) Red vs Green; (b) White vs Green; (c) Red vs White.



Supplementary Figure 26. (a) Expression of celery genes in root obtained using RNA-seq. Homoeologous regions produced by celery were grouped in subgenomes A1-A4 as to mapped grape chromosomes. Red (1.08%-1.71%) indicates a clear higher expression, blue (85.44%-87.62%) indicates no significant difference, and green (11.31%-13.35%) indicates a clear lower expression. No significant difference between A1-A4 subgenomes.

	and the second statement of the statement of the second statement of the secon	in the base of a light before in			Vv 1
n celery petiole	and the contract of the second s				Vv 2
		NUM 1			Vv 3
	STAC STOLD THE PERSON AND IN THE STATE	ter en an and an and an	dentarik) "		Vv 4
	Manager II. Strand and Strand	an a	<u>הירי די די די</u>	,	Vv 5
		ter in the state of the second state of the se	<u>78</u>	,	Vv 6
	SC REPEARATION ADDRESS 21. 1	¹⁹ - Constant Anna ann ann an ann an ann an ann an ann an a	1 1,11 mm	,	Vv 7
	<u>an an a</u>	where a subset is a subset of the second second	n et neter fan e t de le	,	Vv 8
es		1 H Hules		,	Vv 9
alle	an en de Dans R. Steleten e Tradision			١	√v 10
n of				١	√v 11
ssio	weather the state of the state	n an	ht inn m	١	√v 12
kpre		n an	на на на 1. – П.	١	√v 13
e e)	and the second	terne and the way of the		<u>ا</u> ۱	√v 14
Gen	11. 10. 1. Al a default de la face deservant de la companya de la companya de la companya de la companya de la	e th umph ology, the		١	√v 15
U	Security State	Northe I. J		١	√v 16
	wie two ser until tot tet tet waard daar it faar it te	NY. HANY'.		١	√v 17
		W Shanding the cost of the c			√v 18
	La tereta della	an in the second se		<u>\</u>	√v 19
	o 500	1000	1500	2000	

Supplementary Figure 26. (b) Expression of celery genes in petiole obtained using RNA-seq. Homoeologous regions produced by celery were grouped in subgenomes A1-A4 as to mapped grape chromosomes. Red (1.1%-1.63%) indicates a clear higher expression, green (85.86%-87.24%) indicates no significant difference, and blue (11.14%-13.0%) indicates a clear lower expression. No significant difference between A1-A4 subgenomes.

	ander in der Statistikken statiste er ander der der der der der der der der der	Vv 1
selery leaf	nen 1942 a The Automatic Advances and Automatic Advances of the Automatic Advances of the	Vv 2
	And the second	Vv 3
	Constant and the constant and the second of the constant second second second second second second second second	Vv 4
	Ned sector for the New York and the sector watch in the Sector of the New York in the sector of the se	Vv 5
	A CONTRACTOR OF	Vv 6
	SC MININARY PROVIDED AND ADDRESSING BUILDED AND ADDRESSING BUILDED AND AND AND ADDRESS ADDRESS ADDRESS ADDRESS	Vv 7
Li	. 17. 19. 19. стали в након в Милики, възда на Малики, к. 19. стали на маке и в консекси в маке на "Даний мини	Vv 8
eles	ANNALY CONTRACTOR CONTRACTOR AND	Vv 9
fall	and white we wanted the second and the	Vv 10
ouo	STREPTION CONTRACTOR AND AND ADDRESS AND	Vv 11
essi	New Section States and Contraction (CAR) and CONTRACT	Vv 12
sxpr	A CALLER AND A CALLER	Vv 13
ene e	provide and the second s	Vv 14
Ģ	ti i di ul ultra dividi di teri dina dina i persenata ya thabana antere di 183	Vv 15
	The set of a second	Vv 16
	walk President in the Carl Server as and the Carl Server and the Carl Server and the Carl Server in the Carl Server and the Carl S	Vv 17
		Vv 18
	ta transmissione com and and a the second and a second and a second a second a second a second a second a second	Vv 19
	0 500 1000 1500 2000	

Supplementary Figure 26. (c) Expression of celery genes in leaf obtained using RNA-seq. Homoeologous regions produced by celery were grouped in subgenomes A1-A4 as to mapped grape chromosomes. Red (0.86%-1.44%) indicates a clear higher expression, green (86.0%-87.63%) indicates no significant difference, and blue (10.93%-12.81%) indicates a clear lower expression. No significant difference between A1-A4 subgenomes.

	and a second statement of the second statement of the second second statement of the second second second second	R La Martin Carl Carl Carl		Vv 1
ery	and the second			Vv 2
	Net constants to the second state of the secon			Vv 3
	State Reads for the second statements and the second	an san an a		Vv 4
	NEW CONTRACTORY OF CONTRACTORS AND A CO	11 11 11 11 11 11 11 11 11 11 11 11 11		Vv 5
ce	and a star water water and the star water and the star	e s. A AMPRIMENTAL INC.		Vv 6
hite	SC MINERAL STREAM AND ADDRESS AND ADDRESS AND ADDRESS ADDR	AN TANK AND A TANK TANK TANK		Vv 7
≥ ⊇	– 1977 in California Collegia International Additional Annalis and No. 2017, C. 1977 in the term of the second	iers konstant, av än strekskor delvise		Vv 8
eles	Where a start interaction of the constraints of the	н н		Vv 9
falle	ere das 44-antidates de 1999 (487-anti-6-anti-6-anti-			Vv 10
o uc	22007446 (AMTERIAMENTAL AND	all to the		Vv 11
ssio	marata bid telan maneri eritti atteda (2007 auf 1995) addi	t n n transformation n		Vv 12
xpre		, , link tottdb/dt∰A ing i		Vv 13
e el	The second se	lin data any ing	<u></u>	Vv 14
Ger	 C. C. L. Liberhörderfelteschier ministering in die einer die einer die einer die einer die einer die einer die die einer die ein die einer die einer die	<u>6.44</u>		Vv 15
	The set \$15.4			Vv 16
	wake theory of a settle of AMPROPERTANES. "The WARRENESS".	_		Vv 17
	ALTER CLARKER FOR THE ALTER AND THE ALTER CLARKER AND	And And Annual States and Annual States and	111	Vv 18
	the second designed with the second s			Vv 19
	0 500 1000	1500	2000	

Supplementary Figure 26. (d) Expression of celery genes in white-stem variety obtained using RNA-seq. Homoeologous regions produced by celery were grouped in subgenomes A1-A4 as to mapped grape chromosomes. Red (1.21%-1.77%) indicates a clear higher expression, green (85.62%-87.21%) indicates no significant difference, and blue (10.99%-13.17%) indicates a clear lower expression. No significant difference between A1-A4 subgenomes.

	and a state of the		Vv 1
Z	nen die der Berner alle an Berner bei		Vv 2
	Contraction of Contraction Statements of the Statement of Statement of Statements		Vv 3
	CARACTER CONTRACTOR SERVICE CONTRACTOR	namen and a statement of the statement o	Vv 4
	NAMES OF THE STATE OF A REAL STREET AND AND A COMPANY AND A STATE OF A	which are a second with the	Vv 5
cele	and the contraction of a loss of the	LA DE LA AMPROPERTIEN (SAL	Vv 6
in red o	THE REPORT OF A CONTRACT OF A	AT MANY AND ALL ALL ALL ALL ALL ALL ALL ALL ALL AL	Vv 7
	11. Substantischer Schleinen eine Kriften Leinen and Weiten schleinen eine Sterreiten sterreiten einer seiner s	tymene menetica a da frederik (1972)	Vv 8
leles	When Street Street Street All Street St	l lin n	Vv 9
ession of all	an and the statement of the statement of the statement		Vv 10
	STREET AND A DESCRIPTION OF A DESCRIPTIO		Vv 11
	ANTER OF A STREET AND A STREET AND A STREET AND A STREET AND A	t in in the second s	Vv 12
axpr	TRACTOR CONTRACTOR	en anders ander ander ander ander	Vv 13
ene e	A CARLES AND A CARL COMPANY AND A C	a da Malina da compositiva de la compositiva da la compositiva da la compositiva da la compositiva da la compos	Vv 14
Ģ	Contract and the second second second as a second	44.36, ¹⁴ 4	Vv 15
	The set of the second sec		Vv 16
	welle President und Staff Schifferen scalars 2 Ale 20 Constants (Vv 17
	TANTAN TANA ANALYA SA		WV 18
	to its detailed for the state of the second s		Vv 19
	o 500 1000	1500	2000

Supplementary Figure 26. (e) Expression of celery genes in red-stem variety obtained using RNA-seq. Homoeologous regions produced by celery were grouped in subgenomes A1-A4 as to mapped grape chromosomes. Red (0.97%-1.70%) indicates a clear higher expression, green (86.46%-87.62%) indicates no significant difference, and blue (10.69%-12.57%) indicates a clear lower expression. No significant difference between A1-A4 subgenomes.

	ander in einer statistikken kommen van in andere die einer	Herzen werden die seite Bergene		Vv	<i>'</i> 1
ery	medial . The survey of the based of	77 · · · · · · · · · · · · · · · · · ·		Vv	<i>'</i> 2
		Phile 1		Vv	<i>'</i> 3
	CARE IN COMPANY AND A COMPANY		de Parks	Vv	<i>'</i> 4
	Manage of the second states with a loss	South Marine to the second	" ** " 7 "F N	Vv	<i>'</i> 5
l ce	annen an the an and the base of the base of the second second second second second second second second second	· · · · · · · · · · · · · · · · · · ·	4 "\$ <u>8</u> 1	Vv	6
reer	SC BELLARVE MADATATION AND STUDIES &	international sector and the sector of	10 44 - 00 10 1 1 - 1	Vv	7
in gı	CONTRACTOR AND A DECEMPION OF A DECE	······································	in Antolia (1972)	Vv	′ 8
es	WWW STREET, WALLEY, WAL	F# Here		Vv	9
alle	an en 4-EMMAR MARKED FOR THE	1.1		Vv	10
n of				Vv	11
ssic	where i d the water of the statistical statistics in	adh i can i can ann i can	What were an	Vv	12
xpre		1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1	#A 101 1	Vv	13
le e)	No. 19 All States and American	- The second state of the	1. In an an arrest state of the second state of the second state of the second state of the second state of the	L Vv	14
Gen	10 - 10 - 10 - 10 an airte 10 - 10 - 10 an airte 10 - 10 - 10 - 10 - 10 - 10 - 10 - 10	nal y co urt annya, 183		Vv	15
	The second state of the second states of the second	en hande hat in the state of the		Vv	16
	wik instant, with the termination of the term	GANT, HANY' .		Vv	17
		NAME AND ADDRESS OF A DESCRIPTION OF A D		WV VV	18
	L. ROUNDED AND A THE REPORT OF A DESCRIPTION OF A DESCRIPANTA DESCRIPTION OF A DESCRIPTION OF A DESCRIPTION	NAL CONTRACTOR	<u>86</u>	Vv	19
	500	1000	1500	2000	

Supplementary Figure 26. (f) Expression of celery genes in green-stem variety obtained using RNA-seq. Homoeologous regions produced by celery were grouped in subgenomes A1-A4 as to mapped grape chromosomes. Red (0.93%-1.66%) indicates a clear higher expression, green (86.81%-88.06%) indicates no significant difference, and blue (10.28%-12.03%) indicates a clear lower expression. No significant difference between A1-A4 subgenomes.



Supplementary Figure 27. Comparative analysis of the transcription factor families of celery and 6 representative plants. (a) The distribution of 3 largest transcription factor families on 11 chromosomes in celery. Red, blue, and purple colours indicate the MYB, bHLH, and AP2/ERF gene families, respectively. (b) The number of each transcription factor in celery, grape, Arabidopsis, lettuce, ginseng, carrot, and coriander genomes.



Supplementary Figure 28. Distribution of the transcription factor families with the number larger than 100. (a) The distribution of 6 largest transcription factor families on 11 chromosomes in coriander. (b) The distribution of 5 largest transcription factor families on 9 chromosomes in carrot.



Supplementary Figure 29. (a) A phylogenetic tree of *FAR1* gene family constructed by the protein sequences of celery, coriander, carrot, lettuce, and grape. The phylogenetic tree topology was generated via MEGA 7.0. For the major nodes, neighbour-joining (NJ) bootstrap values above 40% are shown.



Supplementary Figure 29. (b) A phylogenetic tree of *GRF* gene family constructed by the protein sequences of celery, coriander, carrot, lettuce, and grape. The phylogenetic tree topology was generated via MEGA 7.0. For the major nodes, neighbour-joining (NJ) bootstrap values above 40% are shown.



Supplementary Figure 29. (c) A phylogenetic tree of *STAT* gene family constructed by the protein sequences of celery, coriander, carrot, lettuce, and grape. The phylogenetic tree topology was generated via MEGA 7.0. For the major nodes, neighbour-joining (NJ) bootstrap values above 40% are shown.



Supplementary Figure 30. (a) A FPKM hierarchical clustering map of the *FAR1* gene family. The left heatmap represents the expression values of three tissues (root, petiole, leaf) of celery. The right heatmap represents the expression values of three different colors' varieties (green, red, and white) of celery.



Supplementary Figure 30. (b) A FPKM hierarchical clustering map of the *GRF* gene family. The left heatmap represents the expression values of three tissues (root, petiole, leaf) of celery. The right heatmap represents the expression values of three different colors' varieties (green, red, and white) of celery.



Supplementary Figure 31. Numbers of NBS family genes in celery and 106 plants.



Supplementary Figure 32. Analyses of *R*-genes, including *RLK*, *RLP*, and *NBS* genes. (a) the fold change of *RLK*, *RLP*, and *NBS* genes number between celery (Agr) and grape (Vvi), Arabidopsis (Ath), lettuce (Lsa), ginseng (Pgi), carrot (Dca), and coriander (Csa). (b) The number of *RLK*, *RLP*, and *NBS* family genes in grape, Arabidopsis, lettuce, ginseng, carrot, and coriander, and celery.



Supplementary Figure 33. A phylogenetic tree of NBS gene family constructed by the protein sequences of celery, coriander, and carrot. The phylogenetic tree topology was generated via MEGA 7.0. For the major nodes, neighbour-joining (NJ) bootstrap values above 40% are shown.



Supplementary Figure 34. The interaction network between key anthocyanins and coumarins related DEGs (purple oval) and differently expressed TFs (DETs, green oval) between red variety and both green or white varieties. The orange and blue lines represent the Pearson correlation coefficient (PCC) values larger than 0.9 (positively regulated relationship) or lower than -0.9 (negatively regulated relationship), respectively. The histogram shows the number of each transcription factor category in the regulatory network.