

Supplementary Data

PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics

Michiel Van Bel^{1,2}, Tim Diels^{1,2}, Emmelien Vancaester^{1,2}, Lukasz Kreft³, Alexander Botzki³, Yves Van de Peer^{1,2,4,5}, Frederik Coppens^{1,2} and Klaas Vandepoele^{1,2,5*}

1. Ghent University, Department of Plant Biotechnology and Bioinformatics, 9052 Ghent, Belgium
2. VIB Center for Plant Systems Biology, 9052 Ghent, Belgium
3. VIB Bioinformatics Core, 9052 Ghent, Belgium
4. Genomics Research Institute, University of Pretoria, Private bag X20, Pretoria 0028, South Africa
5. Bioinformatics Institute Ghent, Ghent University, 9052 Ghent, Belgium

* To whom correspondence should be addressed. Tel: +00 32 9 3313822; Email: klaas.vandepoele@ugent.vib.be

Supplementary Methods

Supplementary Method S1: Construction species trees

The species tree for both the Dicot and Monocot instance of the PLAZA 4.0 version of the platform were initially created by using the taxonomy information provided by NCBI taxonomy (1). Trifurcations and multifurcations within the resulting trees, as well as contradictory clade indications, were resolved according to the following literature:

1. Resolving the early land plants taxonomy
 - a) Phylotranscriptomic analysis of the origin and early diversification of land plants. (2)
2. Resolving the Monocot taxonomy
 - a) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. (3)
 - b) Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. (4)
3. Resolving the Dicot taxonomy
 - a) Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set.(5)
 - b) Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. (6)
 - c) Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. (7)
 - d) Plastome phylogeny and early diversification of Brassicaceae. (8)
 - e) Phylogeny of the core Malvales: evidence from ndhF sequence data. (9)
 - f) Diversification of the phaseoloid legumes: effects of climate change, range expansion and habit shift. (10)
 - g) Mitochondrial matR sequences help to resolve deep phylogenetic relationships in rosids. (11)
 - h) Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. (12)

Supplementary Method S2: Software versions used during PLAZA build procedure

The all-versus-all protein similarity search was performed using v0.9.9 of DIAMOND (13). Clustering of protein coding genes into homologous gene families was performed using v10-201 of MCL (14), while the clustering into orthologous families was performed using v1.1.4 of OrthoFinder (15). Colinear regions were detected using v3.0.01 of i-ADHoRe (16). Multiple sequence alignments were constructed using v3.8.31 of Muscle, and the phylogenetic trees were constructed using v2.1.7 of FastTree. InterPro domains (17) were detected using InterProScan v5.24-63 (18).

Supplementary Tables

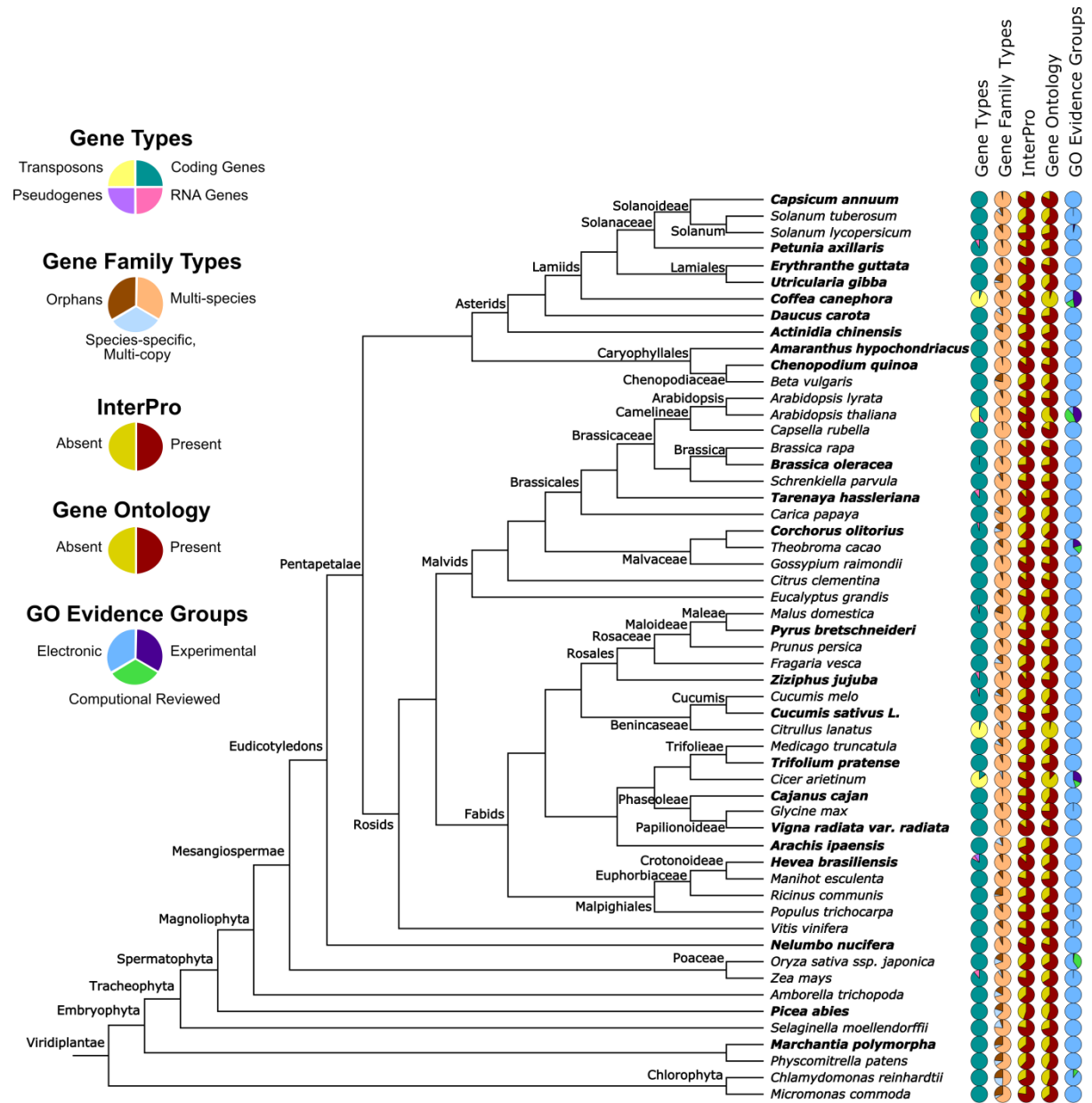
Supplementary Table S1: Data content and sources PLAZA 4.0

PLAZA Instances	Species	Common Name	Data Source	Version	Custom Gene Id Prefix	Pubmed Id
dicot	ach	Actinidia chinensis	BTI Fei lab	v1.0		24136039 (19)
dicot	ahy	Amaranthus hypochondriacus	JGI	v2.1		27898770 (20)
monocot dicot	atr	Amborella trichopoda	JGI	v1.0	ATR	24357323 (21)
monocot	aco	Ananas comosus	JGI	v3.0		26523774 (22)
dicot	aly	Arabidopsis lyrata	JGI	v2.1		26382944 (23)
monocot dicot	ath	Arabidopsis thaliana	Araport	Araport11		27862469 (24)
dicot	aip	Arachis ipaensis	JGI	v1.0		26901068 (25)
dicot	bvu	Beta Vulgaris	RefBeet	v1.2		24352233 (26)
monocot	bdi	Brachypodium distachyon	JGI	v3.1		20148030 (27)
dicot	bol	Brassica oleracea	EnsemblPlants	v2.1		24852848 (28)
dicot	bra	Brassica rapa	JGI	FPsc1.3		21873998 (29)
dicot	ccaj	Cajanus cajan	penguin.genomics.cn	v5.0		22057054 (30)
dicot	cru	Capsella rubella	JGI	v1.0		23749190 (31)
dicot	can	Capsicum annuum	peppersequence.genomics.cn	v2.0	CAN	24441736 (32)
dicot	cpa	Carica papaya	JGI	ASGPB0.4	CPA	18432245 (33)
dicot	cqu	Chenopodium quinoa	JGI	v1.0		28178233 (34)
monocot dicot	cre	Chlamydomonas reinhardtii	JGI	v5.5		17932292 (35)
dicot	car	Cicer arietinum	legumeinfo.org	v1.0		23354103 (36)
dicot	cla	Citrullus lanatus	icugi.org	v1.0		23179023 (37)
dicot	ccl	Citrus clementina	JGI	v1.0		24908277 (38)
dicot	ccan	Coffea canephora	coffee-genome.org	v1.0		25190796 (39)
dicot	col	Corchorus olitorius	GenBank	v1.0	COL	28134914 (40)
dicot	cme	Cucumis melo	melonomics.net	v3.5		22753475 (41)
dicot	csa	Cucumis sativus L.	JGI	v1.0		N/A
dicot	dca	Daucus carota	JGI	v2.0		27158781 (42)
monocot	egu	Elaeis guineensis	JGI	v1.0	EGU	23883927 (43)
dicot	egut	Erythranthe guttata	JGI	v2.0		23518688 (44)
dicot	egr	Eucalyptus grandis	JGI	v2.0		24919147 (45)
dicot	fve	Fragaria vesca	JGI	v1.1	FVE	21186353 (46)
dicot	gma	Glycine max	JGI	Wm82.a2.v1		20075913 (47)
dicot	gra	Gossypium raimondii	JGI	v2.1		22922876 (48)
dicot	hbr	Hevea brasiliensis	RefSeq	v1.0		27255837 (49)
monocot	hvu	Hordeum vulgare	EnsemblPlants	ASM32608v1	HVU	23075845 (50)
dicot	mdo	Malus domestica	gigadb.org	v2.0	MDO	20802477 (51)
dicot	mes	Manihot esculenta	JGI	v6.1		22523606 (52)
monocot dicot	mpo	Marchantia polymorpha	JGI	v3.1		28985561(53)
dicot	mtr	Medicago truncatula	JGI	Mt4.0v1		22089132 (54)
monocot dicot	mco	Micromonas commoda	JGI	v3.0	MCO	19359590 (55)
monocot	mac	Musa acuminata	banana-genome.cirad.fr/	v1.0	MAC	22801500 (56)
dicot	nnu	Nelumbo nucifera	lotus-db	v1.1		23663246 (57)
monocot	oth	Oropetium thomaeum	JGI	v1.0		26560029 (58)
monocot	obr	Oryza brachyantha	EnsemblPlants	v1.4b		23518283 (59)
monocot	osaindica	Oryza sativa ssp. Indica	www.mkbase.org	R498_v1		11935017 (60)
monocot dicot	osa	Oryza sativa ssp. Japonica	JGI	v7.0		16100779 (61)
dicot	pax	Petunia axillaris	solgenomics.net	v1.6.2		27255838 (62)
monocot	peq	Phalaenopsis equestris	genomics.org.cn	v1.0		25420146 (63)
monocot	ped	Phyllostachys edulis	www.ncgr.ac.cn	v1.0		23435089 (64)
monocot dicot	ppa	Physcomitrella patens	JGI	v3.3		18079367 (65)
monocot dicot	pab	Picea abies	ORCAE	v1.0	PAB	23698360 (66)
monocot dicot	ptr	Populus trichocarpa	JGI	v3.1		16973872 (67)
dicot	ppe	Prunus persica	JGI	v2.1		23525075 (68)
dicot	pbr	Pyrus bretschneideri	peargenome	v1.0		23149293 (69)
dicot	rco	Ricinus communis	JGI	v0.1	RCO	20729833 (70)

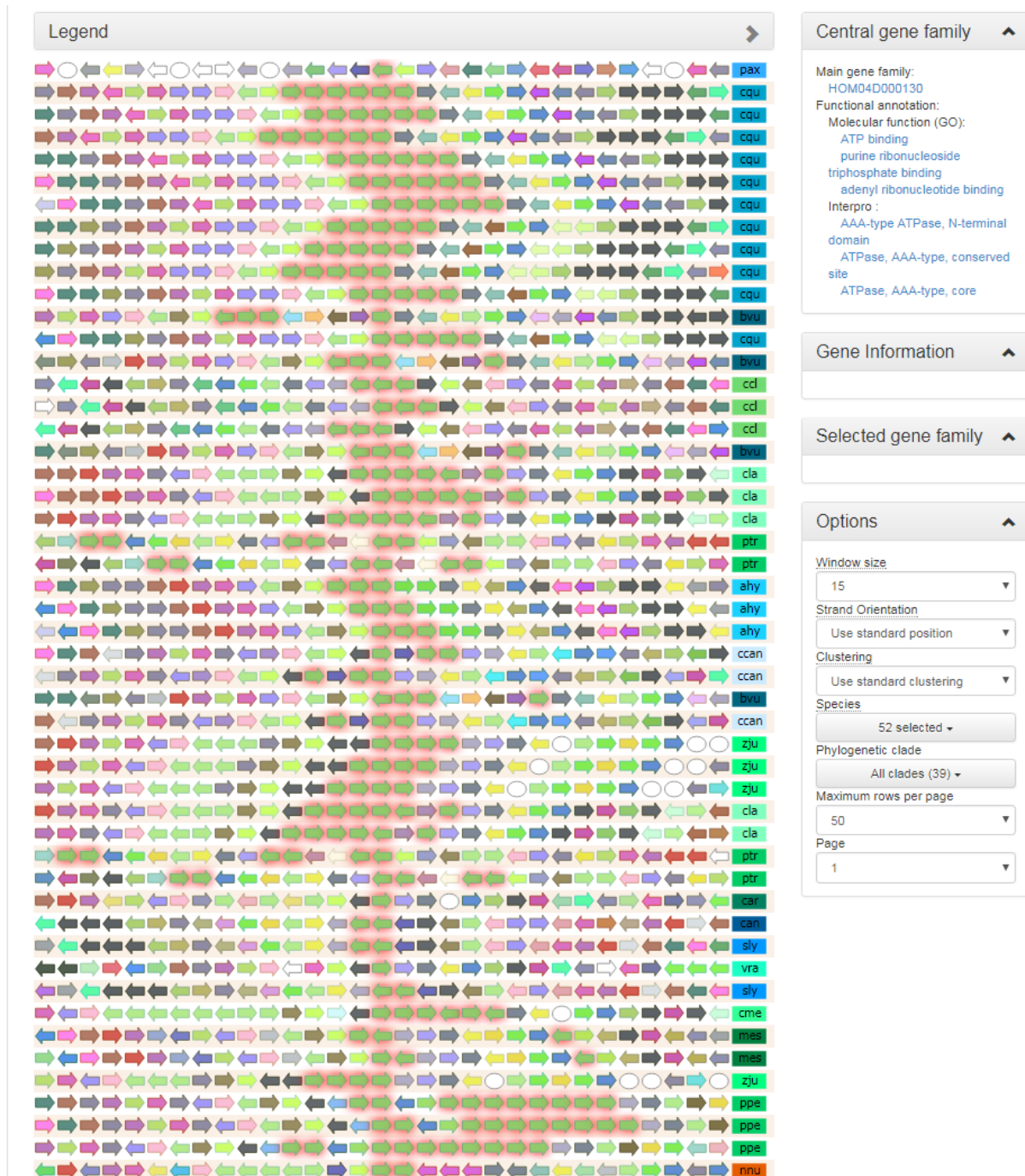
dicot	spa	Schrenkiella parvula	thellungiella.org	v2.0 (TpV84)		21822265 (71)
monocot dicot	smo	Selaginella moellendorffii	JGI	v1.0	SMO	21551031 (72)
monocot	sit	Setaria italica	JGI	v2.2		22580950 (73)
monocot dicot	sly	Solanum lycopersicum	JGI	itag2.4		22660326 (74)
dicot	stu	Solanum tuberosum	JGI	v4.03		21743474 (75)
monocot	sbi	Sorghum bicolor	JGI	v3.1.1		19189423 (76)
monocot	spo	Spirodela polyrhiza	JGI	v2.0		24548928 (77)
dicot	tha	Tarenaya hassleriana	RefSeq	v1.0	THA	23983221 (78)
dicot	tca	Theobroma cacao	RefSeq	v1.1	TCA	21186351 (79)
dicot	tpr	Trifolium pratense	JGI	v2.0	TPR	26617401 (80)
monocot	tae	Triticum aestivum	JGI	TGACv1	TAE	25035500 (81)
dicot	ugi	Utricularia gibba	genomeevolution.org	v4.0	UGI	23665961 (82)
dicot	vra	Vigna radiata var. radiata	legumeinfo.org	v6.0		25384727 (83)
monocot dicot	vvi	Vitis vinifera	JGI	Genoscope		17721507 (84)
monocot dicot	zma	Zea mays	AGP	AGPv4.36		19965430 (85)
dicot	zju	Ziziphus jujuba	RefSeq	v1.1	ZJU	25350882 (86)
monocot	zosmarina	Zostera marina	JGI	v2.2		26814964 (87)
monocot	zjn	Zoysia japonica ssp. Nagirizaki	zoysia.kazusa.or.jp	v1.1		26975196 (88)

Supplementary Figures

Supplementary Figure S1. Data content overview Dicot instance PLAZA 4.0. Species in bold are newly included species compared to previous PLAZA instances. Orphan gene families are single-copy species-specific genes. GO evidence groups are based on the following GO evidence types: Experimental (EXP, IDA, IPI, IMP, IGI, IEP), Computational Reviewed (ISS, ISO, ISA, ISM, IGC, IBA, IBD, IKR, IRD, RCA), and Electronic (TAS, NAS, IC, ND, IEA, NR).



Supplementary Figure S2. Synteny Plot showing conserved ancient tandem duplication within the Dicots.



Supplementary Figure S3. Currently available API functions within the PLAZA framework.

Schemes
HTTP

Authorize

Authentication JWT API keys and tokens More info: <https://bioinformatics.psb.ugent.be/plaza/documentation/>

POST /token_requests Request a JWT token necessary for any data retrieval API calls.

Species Species and organisms More info: <https://bioinformatics.psb.ugent.be/plaza/documentation/>

GET /species/ Get all species within the current PLAZA instance.

GET /species/{species} Get species info

Gene Gene identifiers, genes and their annotation More info: <https://bioinformatics.psb.ugent.be/plaza/documentation/>

GET /genes Get gene info of genes

GET /genes/orthologs Get integrative orthologs of genes

Gene Family Gene Families More info: <https://bioinformatics.psb.ugent.be/plaza/documentation/>

GET /gene_families Get gene families

Supplementary Figure S4. In-depth overview of the options and data types available when querying the PLAZA API for gene content.

Gene Gene identifiers, genes and their annotation More info: <https://bioinformatics.psb.ugent.be/plaza/documentation/>

GET /genes Get gene info of genes

Get gene info of a set of genes, optionally including the InterPro domains, GO terms and/or gene families of each gene

Parameters Try it out

Name	Description
gene * <small>required</small>	PLAZA gene identifiers (e.g. <code>AT5G50870</code>) separated by <code>,</code> (max number of genes is limited to 100)
array [string] <i>(query)</i>	
include	Which of the optional gene attributes to include in the result. Possible values are any combination of <code>cds</code> , <code>interpro_domains</code> , <code>go_terms</code> , <code>gene_families</code> and <code>attributes</code> separated by <code>,</code>
array [string] <i>(query)</i>	

Responses Response content type `application/json`

Code	Description
200	<p>Successful query. Returns gene info for the given gene ids and a list of which input genes are unknown.</p> <p>Example Value Model</p> <pre>{ "result": { "genes": [{ "id": "AT5G50870", "species": "ath", "main_transcript": "AT5G50870.1", "bio_type": "coding", "translation_table": 1, "cds": "ATGATAGATTTTCAGTCGAATCCAGAAAGACTCCGGGCTGCCCCAAATCCGATAACTCTGTTCTGCTTAG", "interpro_domains": [{ "interpro": "IPR002921", "description": "Fungal lipase-like domain", "start": "368", "stop": "498" }], "go_terms": [{ "go": "GO:0006629", "evidence": "IEA", "description": "lipid metabolic process" }], "gene_families": [{ "id": "H0M04M000024", "genes_by_species": { "species": "ath", "genes": ["AT5G50870"] } }], "attributes": [{ "name": "symbol", "values": ["UBC27"] }] }], "unknown_genes": ["AT5G50870"] } }</pre>
400	<p>Invalid parameter input.</p> <p>Example Value Model</p> <pre>{ "error": "Too many input genes provided" }</pre>
401	<p>JWT API key is missing, invalid or expired.</p> <p>Example Value Model</p> <pre>{ "error": "Invalid username/password combination" }</pre>

References

1. Coordinators, N.R. (2017) Database Resources of the National Center for Biotechnology Information. *Nucleic acids research*, **45**, D12-D17.
2. Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A. *et al.* (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E4859-4868.
3. Barrett, C.F., Baker, W.J., Comer, J.R., Conran, J.G., Lahmeyer, S.C., Leebens-Mack, J.H., Li, J., Lim, G.S., Mayfield-Jones, D.R., Perez, L. *et al.* (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *The New phytologist*, **209**, 855-870.
4. Mathews, S., Tsai, R.C. and Kellogg, E.A. (2000) Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. *American journal of botany*, **87**, 96-107.
5. Stull, G.W., Duno de Stefano, R., Soltis, D.E. and Soltis, P.S. (2015) Resolving basal lamiid phylogeny and the circumscription of Icacinaeae with a plastome-scale data set. *American journal of botany*, **102**, 1794-1813.
6. Bremer, K., Friis, E.M. and Bremer, B. (2004) Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. *Systematic biology*, **53**, 496-505.
7. Zeng, L., Zhang, N., Zhang, Q., Endress, P.K., Huang, J. and Ma, H. (2017) Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *The New phytologist*, **214**, 1338-1354.
8. Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X., Zhang, D., Ma, T., Hu, Q., Al-Shehbaz, I.A. *et al.* (2017) Plastome phylogeny and early diversification of Brassicaceae. *BMC genomics*, **18**, 176.
9. Alverson, W.S., Whitlock, B.A., Nyffeler, R., Bayer, C. and Baum, D.A. (1999) Phylogeny of the core Malvales: evidence from ndhF sequence data. *American journal of botany*, **86**, 1474-1486.
10. Li, H., Wang, W., Lin, L., Zhu, X., Li, J., Zhu, X. and Chen, Z. (2013) Diversification of the phaseoloid legumes: effects of climate change, range expansion and habit shift. *Frontiers in plant science*, **4**, 386.
11. Zhu, X.Y., Chase, M.W., Qiu, Y.L., Kong, H.Z., Dilcher, D.L., Li, J.H. and Chen, Z.D. (2007) Mitochondrial matR sequences help to resolve deep phylogenetic relationships in rosids. *BMC evolutionary biology*, **7**, 217.
12. Bruneau, A., Doyle, J.J., Herendeen, P., Hughes, C., Kenicer, G., Lewis, G., Mackinder, B., Pennington, R.T., Sanderson, M.J., Wojciechowski, M.F. *et al.* (2013) Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon*, **62**, 217-248.
13. Buchfink, B., Xie, C. and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature methods*, **12**, 59-60.
14. Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*, **30**, 1575-1584.
15. Emms, D.M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology*, **16**, 157.
16. Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y. and Vandepoele, K. (2012) i-ADHoRe 3.0--fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic acids research*, **40**, e11.

17. Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y., Dosztanyi, Z., El-Gebali, S., Fraser, M. *et al.* (2017) InterPro in 2017-beyond protein family and domain annotations. *Nucleic acids research*, **45**, D190-D199.
18. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236-1240.
19. Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., Zhang, L., Niu, X., Zhang, X., Meng, M. *et al.* (2013) Draft genome of the kiwifruit *Actinidia chinensis*. *Nature communications*, **4**, 2640.
20. Clouse, J.W., Adhikary, D., Page, J.T., Ramaraj, T., Deyholos, M.K., Udall, J.A., Fairbanks, D.J., Jellen, E.N. and Maughan, P.J. (2016) The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly. *The plant genome*, **9**.
21. Amborella Genome, P. (2013) The Amborella genome and the evolution of flowering plants. *Science*, **342**, 1241089.
22. Ming, R., VanBuren, R., Wai, C.M., Tang, H., Schatz, M.C., Bowers, J.E., Lyons, E., Wang, M.L., Chen, J., Biggers, E. *et al.* (2015) The pineapple genome and the evolution of CAM photosynthesis. *Nature genetics*, **47**, 1435-1442.
23. Rawat, V., Abdelsamad, A., Pietzenek, B., Seymour, D.K., Koenig, D., Weigel, D., Pecinka, A. and Schneeberger, K. (2015) Improving the Annotation of *Arabidopsis lyrata* Using RNA-Seq Data. *PloS one*, **10**, e0137391.
24. Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D. (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant journal : for cell and molecular biology*, **89**, 789-804.
25. Bertoli, D.J., Cannon, S.B., Froenicke, L., Huang, G., Farmer, A.D., Cannon, E.K., Liu, X., Gao, D., Clevenger, J., Dash, S. *et al.* (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nature genetics*, **48**, 438-446.
26. Dohm, J.C., Minoche, A.E., Holtgrawe, D., Capella-Gutierrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sorensen, T.R., Stracke, R., Reinhardt, R. *et al.* (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*, **505**, 546-549.
27. International Brachypodium, I. (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763-768.
28. 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. *Nature communications*, **5**, 3930.
29. Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F. *et al.* (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature genetics*, **43**, 1035-1039.
30. Varshney, R.K., Chen, W., Li, Y., Bharti, A.K., Saxena, R.K., Schlueter, J.A., Donoghue, M.T., Azam, S., Fan, G., Whaley, A.M. *et al.* (2011) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature biotechnology*, **30**, 83-89.
31. Slotte, T., Hazzouri, K.M., Agren, J.A., Koenig, D., Maumus, F., Guo, Y.L., Steige, K., Platts, A.E., Escobar, J.S., Newman, L.K. *et al.* (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature genetics*, **45**, 831-835.
32. Kim, S., Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A., Seo, E., Choi, J., Cheong, K., Kim, K.T. *et al.* (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature genetics*, **46**, 270-278.
33. Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V., Lewis, K.L. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*, **452**, 991-996.
34. Jarvis, D.E., Ho, Y.S., Lightfoot, D.J., Schmockel, S.M., Li, B., Borm, T.J., Ohyanagi, H., Mineta, K., Michell, C.T., Saber, N. *et al.* (2017) The genome of *Chenopodium quinoa*. *Nature*, **542**, 307-312.

35. Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L. *et al.* (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*, **318**, 245-250.
36. Varshney, R.K., Song, C., Saxena, R.K., Azam, S., Yu, S., Sharpe, A.G., Cannon, S., Baek, J., Rosen, B.D., Tar'an, B. *et al.* (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature biotechnology*, **31**, 240-246.
37. Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W.J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., Wang, Z. *et al.* (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature genetics*, **45**, 51-58.
38. Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J. *et al.* (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature biotechnology*, **32**, 656-662.
39. Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G. *et al.* (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*, **345**, 1181-1184.
40. Islam, M.S., Saito, J.A., Emdad, E.M., Ahmed, B., Islam, M.M., Halim, A., Hossen, Q.M., Hossain, M.Z., Ahmed, R., Hossain, M.S. *et al.* (2017) Comparative genomics of two jute species and insight into fibre biogenesis. *Nature plants*, **3**, 16223.
41. Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., Gonzalez, V.M., Henaff, E., Camara, F., Cozzuto, L., Lowy, E. *et al.* (2012) The genome of melon (*Cucumis melo* L.). *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 11872-11877.
42. Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro, P. *et al.* (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nature genetics*, **48**, 657-666.
43. Singh, R., Ong-Abdullah, M., Low, E.T., Manaf, M.A., Rosli, R., Nookiah, R., Ooi, L.C., Ooi, S.E., Chan, K.L., Halim, M.A. *et al.* (2013) Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, **500**, 335-339.
44. Yang, R., Jarvis, D.E., Chen, H., Beilstein, M.A., Grimwood, J., Jenkins, J., Shu, S., Prochnik, S., Xin, M., Ma, C. *et al.* (2013) The Reference Genome of the Halophytic Plant *Eutrema salsugineum*. *Frontiers in plant science*, **4**, 46.
45. Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D. *et al.* (2014) The genome of *Eucalyptus grandis*. *Nature*, **510**, 356-362.
46. Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P. *et al.* (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nature genetics*, **43**, 109-116.
47. Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J. *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178-183.
48. Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S. *et al.* (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nature genetics*, **44**, 1098-1103.
49. Tang, C., Yang, M., Fang, Y., Luo, Y., Gao, S., Xiao, X., An, Z., Zhou, B., Zhang, B., Tan, X. *et al.* (2016) The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature plants*, **2**, 16073.
50. International Barley Genome Sequencing, C., Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K. *et al.* (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711-716.

51. Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D. *et al.* (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nature genetics*, **42**, 833-839.
52. Prochnik, S., Marri, P.R., Desany, B., Rabinowicz, P.D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., Harkins, T. *et al.* (2012) The Cassava Genome: Current Progress, Future Directions. *Tropical plant biology*, **5**, 88-94.
53. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F. *et al.* (2017) Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell*, **171**, 287-304 e215.
54. Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Bedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H. *et al.* (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature*, **480**, 520-524.
55. Worden, A.Z., Lee, J.H., Mock, T., Rouze, P., Simmons, M.P., Aerts, A.L., Allen, A.E., Cuvelier, M.L., Derelle, E., Everett, M.V. *et al.* (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, **324**, 268-272.
56. D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M. *et al.* (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, **488**, 213-217.
57. Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L.T., Zhang, Q., Kim, M.J., Schatz, M.C., Campbell, M. *et al.* (2013) Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome biology*, **14**, R41.
58. VanBuren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E. *et al.* (2015) Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature*, **527**, 508-511.
59. Jacquemin, J., Bhatia, D., Singh, K. and Wing, R.A. (2013) The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Current opinion in plant biology*, **16**, 147-156.
60. Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**, 79-92.
61. International Rice Genome Sequencing, P. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793-800.
62. Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S., Bliet, M., Boersma, M.R., Borghi, L., Bruggmann, R. *et al.* (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nature plants*, **2**, 16074.
63. Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.C., Liu, K.W., Chen, L.J., He, Y., Xu, Q., Bian, C. *et al.* (2015) The genome sequence of the orchid *Phalaenopsis equestris*. *Nature genetics*, **47**, 65-72.
64. Peng, Z., Lu, Y., Li, L., Zhao, Q., Feng, Q., Gao, Z., Lu, H., Hu, T., Yao, N., Liu, K. *et al.* (2013) The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nature genetics*, **45**, 456-461, 461e451-452.
65. Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y. *et al.* (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*, **319**, 64-69.
66. Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A. *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*, **497**, 579-584.
67. Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596-1604.
68. International Peach Genome, I., Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J. *et al.* (2013) The high-quality

- draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature genetics*, **45**, 487-494.
69. Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M.A., Tao, S., Korban, S.S., Wang, H. *et al.* (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome research*, **23**, 396-408.
 70. Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J., Chen, G. *et al.* (2010) Draft genome sequence of the oilseed species *Ricinus communis*. *Nature biotechnology*, **28**, 951-956.
 71. Dassanayake, M., Oh, D.H., Haas, J.S., Hernandez, A., Hong, H., Ali, S., Yun, D.J., Bressan, R.A., Zhu, J.K., Bohnert, H.J. *et al.* (2011) The genome of the extremophile crucifer *Thellungiella parvula*. *Nature genetics*, **43**, 913-918.
 72. Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A. *et al.* (2011) The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science*, **332**, 960-963.
 73. Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z., Wang, W. *et al.* (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nature biotechnology*, **30**, 549-554.
 74. Tomato Genome, C. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635-641.
 75. Potato Genome Sequencing, C., Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R. *et al.* (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189-195.
 76. Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberler, G., Hellsten, U., Mitros, T., Poliakov, A. *et al.* (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature*, **457**, 551-556.
 77. Wang, W., Haberler, G., Gundlach, H., Glasser, C., Nussbaumer, T., Luo, M.C., Lomsadze, A., Borodovsky, M., Kerstetter, R.A., Shanklin, J. *et al.* (2014) The Spirodela polyrhiza genome reveals insights into its neotenus reduction fast growth and aquatic lifestyle. *Nature communications*, **5**, 3311.
 78. Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C. *et al.* (2013) The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of crucifers. *The Plant cell*, **25**, 2813-2830.
 79. Argout, X., Salse, J., Aury, J.M., Gaultier, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N. *et al.* (2011) The genome of *Theobroma cacao*. *Nature genetics*, **43**, 101-108.
 80. De Vega, J.J., Ayling, S., Hegarty, M., Kudrna, D., Goicoechea, J.L., Ergon, A., Rognli, O.A., Jones, C., Swain, M., Geurts, R. *et al.* (2015) Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. *Scientific reports*, **5**, 17394.
 81. International Wheat Genome Sequencing, C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.
 82. Ibarra-Laclette, E., Lyons, E., Hernandez-Guzman, G., Perez-Torres, C.A., Carretero-Paulet, L., Chang, T.H., Lan, T., Welch, A.J., Juarez, M.J., Simpson, J. *et al.* (2013) Architecture and evolution of a minute plant genome. *Nature*, **498**, 94-98.
 83. Kang, Y.J., Kim, S.K., Kim, M.Y., Lestari, P., Kim, K.H., Ha, B.K., Jun, T.H., Hwang, W.J., Lee, T., Lee, J. *et al.* (2014) Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nature communications*, **5**, 5443.
 84. Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463-467.
 85. Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112-1115.

86. Liu, M.J., Zhao, J., Cai, Q.L., Liu, G.C., Wang, J.R., Zhao, Z.H., Liu, P., Dai, L., Yan, G., Wang, W.J. *et al.* (2014) The complex jujube genome provides insights into fruit tree biology. *Nature communications*, **5**, 5315.
87. Olsen, J.L., Rouze, P., Verhelst, B., Lin, Y.C., Bayer, T., Collen, J., Dattolo, E., De Paoli, E., Dittami, S., Maumus, F. *et al.* (2016) The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature*, **530**, 331-335.
88. Tanaka, H., Hirakawa, H., Kosugi, S., Nakayama, S., Ono, A., Watanabe, A., Hashiguchi, M., Gondo, T., Ishigaki, G., Muguerza, M. *et al.* (2016) Sequencing and comparative analyses of the genomes of zoysiagrasses. *DNA research : an international journal for rapid publication of reports on genes and genomes*, **23**, 171-180.