*D1442–D1447* *Nucleic Acids Research, 2022, Vol. 50, Database issue* *Published online 1 November 2021*
*https://doi.org/10.1093/nar/gkab1014*

# GreeNC 2.0: a comprehensive database of plant long non-coding RNAs

**Marco Di Marsico[1], Andreu Paytuvi Gallart[2], Walter Sanseverino[2] and Riccardo Aiese Cigliano[2,*]**

[1]Dipartimento di Scienze Agrarie, Alimentari e Ambientali, Università degli Studi di Perugia, Borgo XX Giugno 74, 06121 Perugia, Italy and [2]Sequentia Biotech SL, Carrer de Pamplona 88, 08018, Barcelona, Spain

Received September 14, 2021; Revised October 08, 2021; Editorial Decision October 11, 2021; Accepted October 12, 2021

## ABSTRACT

**The Green Non-Coding Database (GreeNC) is one of the reference databases for the study of plant long non-coding RNAs (lncRNAs). Here we present our most recent update where 16 species have been updated, while 78 species have been added, resulting in the annotation of more than 495 000 lncRNAs. Moreover, sequence clustering was applied providing information about sequence conservation and gene families. The current version of the database is available at: http://greenc.sequentiabiotech.com/wiki2/Main_Page.**

## INTRODUCTION

Long non-coding RNAs (lncRNAs) used to be considered as transcriptional noise in the past decades, but lately, this class of molecules has gained increasing attention in epigenetic research and it is now recognized to have an important role in mediating the transmission and the expression of genetic information (1). LncRNAs are RNA molecules longer than 200 nucleotides with no protein-coding ability (2), despite this, they are involved in fundamental biological processes, and their activities are complex and diverse. In fact, lncRNAs could help in the regulation of protein modification but also chromatin remodelling, RNA metabolism, transcription, DNA methylation, and many other processes (1). Due to their activity in a wide number of pathways, lncRNA are very well studied in human and clinical applications (3). In plants, many lncRNAs have been characterized in model organisms as *Arabidopsis thaliana*, *Zea mays*, and *Solanum lycopersicum*. For instance, *Arabidopsis* lncRNA APOLO (*AUXIN-REGULATED PROMOTER LOOP*) regulates the expression of un-related distant auxin-responsive genes during the lateral root development by modulating local chromatin conformation (4). *Zea mays* PILNCR1 is involved in the plant adaptation to phosphate deficiency (5). LncRNA1459, detected in *Solanum lycopersicum*, has been shown to be involved in the fruit ripening process (6).

In order to help the scientific community to study plant lncRNA sequences and functions, we developed and published the GreeNC database in 2015 (7). Since then, the database has been accessed >250 000 times and it has become a reference for the plant scientific community working on lncRNAs. In the last years tens of new plant species have been sequenced and for many of the species shown in GreeNC new or updated reference genomes have been published, for this reason we present GreeNC 2.0 a new update where lncRNAs from new 78 species were added and 16 species were updated. In addition, we performed an extensive sequence clustering in order to detect orthologous groups of lncRNAs both between species and within species. With this additional information researchers will be able to detect whether candidate lncRNAs belong to gene families and if they are conserved across species.

## MATERIALS AND METHODS

### Genome and annotations

FASTA sequences of transcripts were downloaded from Phytozome v13 and Plants Ensembl version 51. The assembly version of each species is reported in Supplementary Table S1. Only un-restricted genomic data were used (8–85). For *Oryza* and for *Brassica rapa*, transcripts were downloaded from Plant Ensembl 51 and Phytozome v13, respectively.

### Identification of lncRNAs

As in the previous version of GreeNC (7), two bash scripts were used to identify lncRNAs among the downloaded transcript sequences (Figure 1). With the first script coding potential is calculated maintaining only transcripts with a minimum length of 200 nucleotides and an ORF shorter than 120 amino acids by using Ugene (38.1). Sequences were blasted (v2.9.0) against SwissProt (2021/04). CPC

*To whom correspondence should be addressed. Tel: +34 633042534; Email: raiesecigliano@sequentiabiotech.com

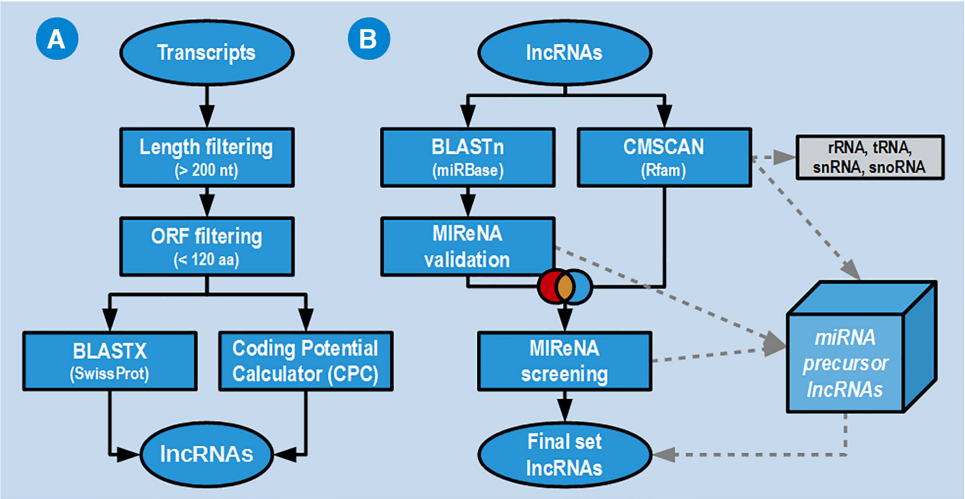


**Figure 1.** Overview of the in-house developed computational pipeline for lncRNA annotation, which consists of script 1 (**A**) and script 2 (**B**).

(0.9-r2) was used to assess the protein-coding potential of transcripts. To discriminate other non-coding transcripts from lncRNAs, and to identify possible miRNA precursors, a second script was used. Transcripts were analyzed by cmscan (Infernal 1.1rc4) against the RFAM database (release 14.6). BLASTn (2.6.0) was used against a database of mature plant miRNA sequences from miRBase (release 22.1). The final list of lncRNAs was divided into high- and low confidence. Transcripts without hits in BLASTX described as non-coding by CPC and not considered as miRNA precursors, were classified as high-confidence lncRNAs. Those without hits in BLASTX but considered coding by CPC, those with BLASTX hits considered noncoding by CPC, and those considered miRNA precursors, were marked as low-confidence lncRNAs. To exclude putative transposons, RepeatMasker was used, in order to identify transcripts containing predicted repetitive regions. These transcripts are also classified as low-confidence. RepeatMasker (4.1.0) was executed with a custom library obtained by RepBase (86) with the following parameters *-no_is, -gff, -nolow*.

### Relational database

Data was imported into a MySQL-based relational database stored on an Ubuntu server (Ubuntu 18.04.4 LTS). This database was then integrated into a MediaWiki by mapping relational data fields against predefined templates via Semantic MediaWiki. Transcript sequences in a FASTA file were formatted using makeblastdb. Sequence retrieval is based on blastdbcmd. An Express Node.js API web service was created to expose both sequence retrieval and BLAST searches via client JavaScript from the MediaWiki interface.

### OrthoFinder

To evaluate sequence similarity and cluster lncRNAs in orthogrups, an OrthoFinder (87) analysis was executed with the following parameters *-d, -f, -S diamond_ultra_sens*. As input files, lncRNA sequences from all the species were used.

## RESULTS

The previous version of GreeNC (7) included 43 species, resulting in a total of 120 000 annotated lncRNAs. After this update, GreeNC 2.0 includes information on >495 000 transcripts from 94 species between plants and algae (Figure 2). More than 327 000 transcripts were annotated as high confidence lncRNA. With this update, the highest percentages of lncRNAs were annotated in *Triticum dicoccoides* (7.7%), and *Aegilops tauschii* (6.9%) and *Hordeum vulgare* (4.8%), while the lowest in *Juglans regia* (0.13%), *Chara braunii* (0.12%) and *Cyanidioschyzon merolae* (0.02%).

Even if it is known that lncRNAs do not show high conservation at nucleotide level (88), we decided to perform a sequence clustering based on the Orthofinder algorithm in order to provide information about highly conserved lncRNAs. About 39% of the 542 656 identified transcripts were assigned to orthogroups. In total, 65 191 orthogroups were identified however, as expected, no orthogroups were present in all the species. Despite this, shared orthogroups were identified between species of the same genus, suggesting the presence of genus-specific lineages of lncRNAs (i.e. *Triticum, Arabidopsis thaliana, Oryza, Gossypium, Brassica*). Moreover, the presence of species-specific orthogroups highlights that long non-coding transcripts may be organized in gene families.

A total of 24 743 orthogroups were identified as species-specific, with a mean of 242 orthogroups per species. The highest number of species-specific orthogroups was recorded in *Triticum dicoccoides* (3487 orthogroups), while the lowest was detected in *Cyanidioschyzon merolae* (2). A total of 81 446 transcripts (15% of the total) were classified as species-specific, with a mean of 798 transcripts per species. Also in this case, the highest and lowest values were recorded in *Triticum dicoccoides* (17 234 transcripts) and *Cyanidioschyzon merolae* (4), respectively.

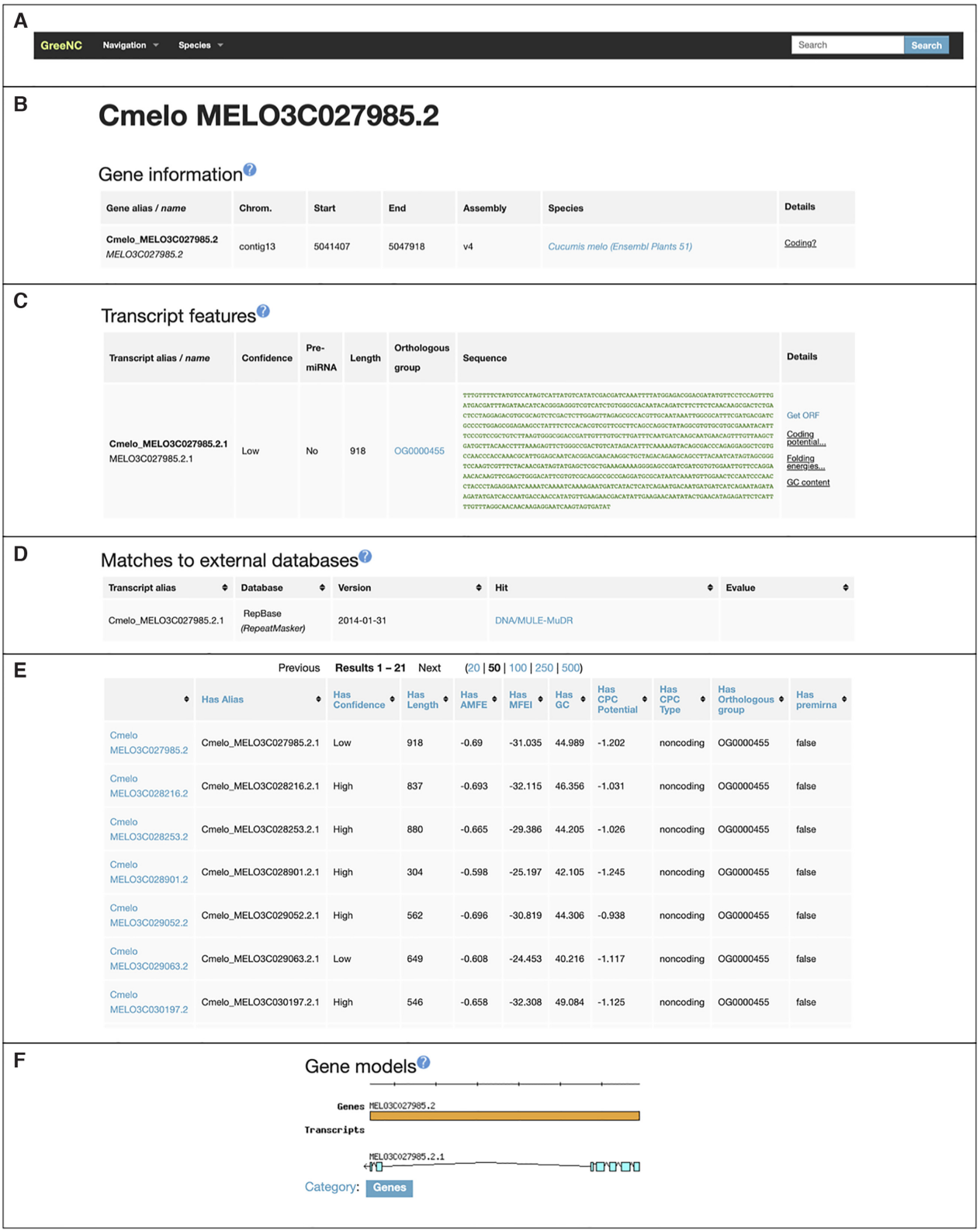


**Figure 2.** A snapshot of a *Cucumis melo* entry from the GreeNC database. (**A**) Header, to navigate through the website and access to the tools and the pages of the species; (**B**) table of gene information reporting genomic coordinates, genome version, the source of the genome assembly and if the gene encodes at least one coding transcript; (**C**) table of transcript features reporting the kind of lncRNA (low-/high-confidence), if it is a precursor of miRNAs, length, orthologous group, sequence and links to get the Open Reading Frame (ORF), the Coding Potential, the folding energy and the GC content; (**D**) an optional table that provides links to other databases, when applicable, and giving information about the version of the database and the e-value of the match; (**E**) table of transcripts belonging to the same orthogroup reporting the kind of lncRNA, length, folding energies (AMFE, MFEI), GC content (**F**) a schematic representation of the gene and transcript models.

## DATA AVAILABILITY

The GreeNC database is a MySQL relational database and it is freely accessible at: http://greenc.sequentiabiotech.com/wiki2/Main_Page.

The pipeline for lncRNA prediction is available at: https://github.com/sequentiabiotech/GreeNC.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Słodowska-Curie Grant Agreement No 101007438.

## FUNDING

European Union's Horizon 2020 Research and Innovation Program under the Marie Słodowska-Curie [101007438]. Funding for open access charge: Self funding.
*Conflict of interest statement.* None declared.

## REFERENCES

1. Wu,L., Liu,S., Qi,H., Cai,H. and Xu,M. (2020) Research progress on plant long non-coding RNA. *Plants*, **9**, 408.
2. Mercer,T.R., Dinger,M.E. and Mattick,J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.
3. Li,G., Deng,L., Huang,N. and Sun,F. (2021). The biological roles of lncRNAs and future prospects in clinical application. *Diseases*, **9**, 8.
4. Ariel,F., Lucero,L., Christ,A., Mammarella,M.F., Jegu,T., Veluchamy,A., Mariappan,K., Latrasse,D., Blein,T., Liu,C. *et al.* (2020) R-loop mediated trans action of the APOLO long noncoding RNA. *Mol. Cell*, **77**, 1055–1065.
5. Du,Q., Wang,K., Zou,C., Xu,C. and Li,W.-X. (2018) The PILNCR1-miR399 regulatory module is important for low phosphate tolerance in maize. *Plant Physiol.*, **177**, 1743–1753.
6. Li,R., Fu,D., Zhu,B., Luo,Y. and Zhu,H. (2018). CRISPR/Cas9-mediated mutagenesis of lncRNA1459 alters tomato fruit ripening. *Plant J.*, **94**, 513–524.
7. Paytuví Gallart,A., Hermoso Pulido,A., Anzar Martínez de Lagrán,I., Sanseverino,W. and Aiese Cigliano,R. (2016) GREENC: a Wiki-based database of plant lncRNAs. *Nucleic Acids Res.*, **44**, D1161–D1166.
8. Peng,Z., Bredeson,J.V., Wu,G.A., Shu,S., Rawat,N., Du,D., Parajuli,S., Yu,Q., You,Q., Rokhsar,D.S. *et al.* (2020) A chromosome-scale reference genome of trifoliate orange (*Poncirus trifoliata*) provides insights into disease resistance, cold tolerance and genome evolution in Citrus. *Plant J.*, **104**, 1215–1232.
9. Hane,J.K., Ming,Y., Kamphuis,L.G., Nelson,M.N., Garg,G., Atkins,C.A., Bayer,P.E., Bravo,A., Bringans,S., Cannon,S. *et al.* (2017) A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. *Plant Biotechnol. J.*, **15**, 318–330.
10. McGrath,J.M., Funk,A., Galewski,P., Ou,S., Townsend,B., Davenport,K., Daligault,H., Johnson,S., Lee,J., Hastie,A. *et al.* (2020) A contiguous de novo genome assembly of sugar beet EL10 (*Beta vulgaris* L.). bioRxiv doi: https://doi.org/10.1101/2020.09.15.298315, 08 October 2020, preprint: not peer reviewed.
11. Mamidi,S., Healey,A., Huang,P., Grimwood,J., Jenkins,J., Barry,K., Sreedasyam,A., Shu,S., Lovell,J.T., Feldman,M. *et al.* (2020) A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci. *Nat. Biotechnol.*, **38**, 1203–1210.
12. Carballo,J., Santos,B. a. C.M., Zappacosta,D., Garbus,I., Selva,J.P., Gallo,C.A., Díaz,A., Albertini,E., Caccamo,M. and Echenique,V. (2019) A high-quality genome of *Eragrostis curvula* grass provides insights into *Poaceae* evolution and supports new strategies to enhance forage quality. *Sci. Rep.*, **9**, 10250.
13. Pilkington,S.M., Crowhurst,R., Hilario,E., Nardozza,S., Fraser,L., Peng,Y., Gunaseelan,K., Simpson,R., Tahir,J., Deroles,S.C. *et al.* (2018) A manually annotated *Actinidia chinensis* var. chinensis (kiwifruit) genome highlights the challenges associated with draft genomes and gene prediction in plants. *BMC Genomics*, **19**, 257.
14. Cormier,F., Lawac,F., Maledon,E., Gravillon,M.-C., Nudol,E., Mournet,P., Vignes,H., Chaïr,H. and Arnau,G. (2019) A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theor. Appl. Genet.*, **132**, 1733–1744.
15. Zhang,J., Zhang,X., Tang,H., Zhang,Q., Hua,X., Ma,X., Zhu,F., Jones,T., Zhu,X., Bowers,J. *et al.* (2018) Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.*, **50**, 1565–1573.
16. Castanera,R., Ruggieri,V., Pujol,M., Garcia-Mas,J. and Casacuberta,J.M. (2019) An improved melon reference genome with single-molecule sequencing uncovers a recent burst of transposable elements with potential impact on genes. *Front. Plant Sci.*, **10**, 1815.
17. Zhou,R., Macaya-Sanz,D., Rodgers-Melnick,E., Carlson,C.H., Gouker,F.E., Evans,L.M., Schmutz,J., Jenkins,J.W., Yan,J., Tuskan,G.A. *et al.* (2018) Characterization of a large sex determination region in *Salix purpurea* L. (*Salicaceae*). *Mol. Genet. Genomics*, **293**, 1437–1452.
18. Roth,M.S., Cokus,S.J., Gallaher,S.D., Walter,A., Lopez,D., Erickson,E., Endelman,B., Westcott,D., Larabell,C.A., Merchant,S.S. *et al.* (2017) Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E4296–E4305.
19. Islam,M.S., Saito,J.A., Emdad,E.M., Ahmed,B., Islam,M.M., Halim,A., Hossen,Q.M.M., Hossain,M.Z., Ahmed,R., Hossain,M.S. *et al.* (2017) Comparative genomics of two jute species and insight into fibre biogenesis. *Nat. Plants*, **3**, 16223.
20. Valliyodan,B., Cannon,S.B., Bayer,P.E., Shu,S., Brown,A.V., Ren,L., Jenkins,J., Chung,C.Y.-L., Chan,T.-F., Daum,C.G. *et al.* (2019) Construction and comparison of three reference-quality genome assemblies for soybean. *Plant J.*, **100**, 1066–1082.
21. Beier,S., Himmelbach,A., Colmsee,C., Zhang,X.-Q., Barrero,R.A., Zhang,Q., Li,L., Bayer,M., Bolser,D., Taudien,S. *et al.* (2017) Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. *Sci. Data*, **4**, 170044.
22. Hufford,M.B., Seetharam,A.S., Woodhouse,M.R., Chougule,K.M., Ou,S., Liu,J., Ricci,W.A., Guo,T., Olson,A., Qiu,Y. *et al.* (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*, **373**, 655–662.
23. Ling,H.-Q., Zhao,S., Liu,D., Wang,J., Sun,H., Zhang,C., Fan,H., Li,D., Dong,L., Tao,Y. *et al.* (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*, **496**, 87–90.
24. Kang,Y.J., Satyawan,D., Shim,S., Lee,T., Lee,J., Hwang,W.J., Kim,S.K., Lestari,P., Laosatit,K., Kim,K.H. *et al.* (2015) Draft genome sequence of adzuki bean, *Vigna angularis*. *Sci. Rep.*, **5**, 8069.
25. 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. *Nat. Biotechnol.*, **31**, 240–246.
26. Maccaferri,M., Harris,N.S., Twardziok,S.O., Pasam,R.K., Gundlach,H., Spannagl,M., Ormanbekova,D., Lux,T., Prade,V.M., Milner,S.G. *et al.* (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.*, **51**, 885–895.
27. Janouškovec,J., Liu,S.-L., Martone,P.T., Carré,W., Leblanc,C., Collén,J. and Keeling,P.J. (2013) Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS One*, **8**, e59001.
28. Li,Z.-M., Zheng,X.-M. and Ge,S. (2011) Genetic diversity and domestication history of African rice (*Oryza glaberrima*) as inferred from multiple gene sequences. *TAG Theor. Appl. Genet. Theor. Angew. Genet.*, **123**, 21–31.
29. Cannarozzi,G., Plaza-Wüthrich,S., Esfeld,K., Larti,S., Wilson,Y.S., Girma,D., de Castro,E., Chanyalew,S., Blösch,R., Farinelli,L. *et al.* (2014) Genome and transcriptome sequencing identifies breeding

targets in the orphan crop tef (*Eragrostis tef*). *BMC Genomics*, **15**, 581.
30. Briskine,R.V., Paape,T., Shimizu-Inatsugi,R., Nishiyama,T., Akama,S., Sese,J. and Shimizu,K.K. (2017) Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Mol. Ecol. Resour.*, **17**, 1025–1036.
31. Reyes-Chin-Wo,S., Wang,Z., Yang,X., Kozik,A., Arikit,S., Song,C., Xia,L., Froenicke,L., Lavelle,D.O., Truco,M.-J. *et al.* (2017) Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.*, **8**, 14953.
32. Willing,E.-M., Rawat,V., Mandáková,T., Maumus,F., James,G.V., Nordström,K.J.V., Becker,C., Warthmann,N., Chica,C., Szarzynska,B. *et al.* (2015) Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nat. Plants*, **1**, 14023.
33. Unver,T., Wu,Z., Sterck,L., Turktas,M., Lohaus,R., Li,Z., Yang,M., He,L., Deng,T., Escalante,F.J. *et al.* (2017) Genome of wild olive and the evolution of oil biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E9413–E9422.
34. Xu,X., Pan,S., Cheng,S., Zhang,B., Mu,D., Ni,P., Zhang,G., Yang,S., Li,R., Wang,J. *et al.* (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189–195.
35. 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. *Nat. Commun.*, **5**, 5443.
36. 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. *Nat. Genet.*, **46**, 270–278.
37. Luo,M.-C., Gu,Y.Q., Puiu,D., Wang,H., Twardziok,S.O., Deal,K.R., Huo,N., Zhu,T., Wang,L. and Wang,Y. (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*, **551**, 498–502.
38. Matsuzaki,M., Misumi,O., Shin-I,T., Maruyama,S., Takahara,M., Miyagishima,S.-Y., Mori,T., Nishida,K., Yagisawa,F., Nishida,K. *et al.* (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, **428**, 653–657.
39. Wang,L., Yu,S., Tong,C., Zhao,Y., Liu,Y., Song,C., Zhang,Y., Zhang,X., Wang,Y., Hua,W. *et al.* (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.*, **15**, R39.
40. Chen,Z.J., Sreedasyam,A., Ando,A., Song,Q., De Santiago,L.M., Hulse-Kemp,A.M., Ding,M., Ye,W., Kirkbride,R.C., Jenkins,J. *et al.* (2020) Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.*, **52**, 525–533.
41. Lovell,J.T., MacQueen,A.H., Mamidi,S., Bonnette,J., Jenkins,J., Napier,J.D., Sreedasyam,A., Healey,A., Session,A., Shu,S. *et al.* (2021) Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature*, **590**, 438–444.
42. Gordon,S.P., Contreras-Moreira,B., Levy,J.J., Djamei,A., Czedik-Eysenberg,A., Tartaglio,V.S., Session,A., Martin,J., Cartwright,A., Katz,A. *et al.* (2020) Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat. Commun.*, **11**, 3670.
43. Marrano,A., Britton,M., Zaini,P.A., Zimin,A.V., Workman,R.E., Puiu,D., Bianco,L., Pierro,E.A.D., Allen,B.J., Chakraborty,S. *et al.* (2020) High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome. *GigaScience*, **9**, giaa050.
44. Hufnagel,B., Marques,A., Soriano,A., Marquès,L., Divol,F., Doumas,P., Sallet,E., Mancinotti,D., Carrere,S., Marande,W. *et al.* (2020) High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nat. Commun.*, **11**, 492.
45. 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.
46. Brawley,S.H., Blouin,N.A., Ficko-Blean,E., Wheeler,G.L., Lohr,M., Goodson,H.V., Jenkins,J.W., Blaby-Haas,C.E., Helliwell,K.E., Chan,C.X. *et al.* (2017) Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (*Bangiophyceae*, *Rhodophyta*). *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E6361–E6370.
47. Chalhoub,B., Denoeud,F., Liu,S., Parkin,I.A.P., Tang,H., Wang,X., Chiquet,J., Belcram,H., Tong,C., Samans,B. *et al.* (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**, 950–953.
48. De Vega,J.J., Ayling,S., Hegarty,M., Kudrna,D., Goicoechea,J.L., Ergon,Å., 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. *Sci. Rep.*, **5**, 17394.
49. Paterson,A.H., Wendel,J.F., Gundlach,H., Guo,H., Jenkins,J., Jin,D., Llewellyn,D., Showmaker,K.C., Shu,S., Udall,J. *et al.* (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*, **492**, 423–427.
50. Guo,S., Zhao,S., Sun,H., Wang,X., Wu,S., Lin,T., Ren,Y., Gao,L., Deng,Y., Zhang,J. *et al.* (2019) Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.*, **51**, 1616–1623.
51. THE INTERNATIONAL WHEAT GENOME SEQUENCING CONSORTIUM (IWGSC), Appels,R., Eversole,K., Stein,N., Feuillet,C., Keller,B., Rogers,J., Pozniak,C.J., Choulet,F., Distelfeld,A. *et al.* (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, **361**, eaar7191.
52. Lightfoot,D.J., Jarvis,D.E., Ramaraj,T., Lee,R., Jellen,E.N. and Maughan,P.J. (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.*, **15**, 74.
53. 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.
54. Chaw,S.-M., Liu,Y.-C., Wu,Y.-W., Wang,H.-Y., Lin,C.-Y.I., Wu,C.-S., Ke,H.-M., Chang,L.-Y., Hsu,C.-Y., Yang,H.-T. *et al.* (2019) Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. *Nat. Plants*, **5**, 63–73.
55. Filiault,D.L., Ballerini,E.S., Mandáková,T., Aköz,G., Derieg,N.J., Schmutz,J., Jenkins,J., Grimwood,J., Shu,S., Hayes,R.D. *et al.* (2018) The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *eLife*, **7**, e36426.
56. Hu,T.T., Pattyn,P., Bakker,E.G., Cao,J., Cheng,J.-F., Clark,R.M., Fahlgren,N., Fawcett,J.A., Grimwood,J., Gundlach,H. *et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.*, **43**, 476–481.
57. 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.
58. Liu,S., Liu,Y., Yang,X., Tong,C., Edwards,D., Parkin,I.A.P., Zhao,M., Ma,J., Yu,J., Huang,S. *et al.* (2014) *The Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.*, **5**, 3930.
59. Slotte,T., Hazzouri,K.M., Ågren,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. *Nat. Genet.*, **45**, 831–835.
60. Nishiyama,T., Sakayama,H., De Vries,J., Buschmann,H., Saint-Marcoux,D., Ullrich,K.K., Haas,F.B., Vanderstraeten,L., Becker,D. and Lang,D. (2018) The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell*, **174**, 448–464.
61. 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.
62. Kagale,S., Koh,C., Nixon,J., Bollina,V., Clarke,W.E., Tuteja,R., Spillane,C., Robinson,S.J., Links,M.G., Clarke,C. *et al.* (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.*, **5**, 3706.
63. Jarvis,D.E., Ho,Y.S., Lightfoot,D.J., Schmöckel,S.M., Li,B., Borm,T.J.A., Ohyanagi,H., Mineta,K., Michell,C.T., Saber,N. *et al.* (2017) The genome of *Chenopodium quinoa*. *Nature*, **542**, 307–312.
64. Lonardi,S., Muñoz-Amatriaín,M., Liang,Q., Shu,S., Wanamaker,S.I., Lo,S., Tanskanen,J., Schulman,A.H., Zhu,T.,

Luo,M.-C. *et al.* (2019) The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.*, **98**, 767–782.
65. Dohm,J.C., Minoche,A.E., Holtgräwe,D., Capella-Gutiérrez,S., Zakrzewski,F., Tafer,H., Rupp,O., Sörensen,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.
66. Shirasawa,K., Isuzugawa,K., Ikenaga,M., Saito,Y., Yamamoto,T., Hirakawa,H. and Isobe,S. (2017) The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Res.*, **24**, 499–508.
67. Scaglione,D., Reyes-Chin-Wo,S., Acquadro,A., Froenicke,L., Portis,E., Beitel,C., Tirone,M., Mauro,R., Lo Monaco,A., Mauromicale,G. *et al.* (2016) The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. *Sci. Rep.*, **6**, 19427.
68. Lovell,J.T., Jenkins,J., Lowry,D.B., Mamidi,S., Sreedasyam,A., Weng,X., Barry,K., Bonnette,J., Campitelli,B., Daum,C. *et al.* (2018) The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. *Nat. Commun.*, **9**, 5213.
69. 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.
70. Yang,X., Hu,R., Yin,H., Jenkins,J., Shu,S., Tang,H., Liu,D., Weighill,D.A., Cheol Yim,W., Ha,J. *et al.* (2017) The *Kalanchoë* genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nat. Commun.*, **8**, 1899.
71. Young,N.D., Debellé,F., Oldroyd,G.E.D., Geurts,R., Cannon,S.B., Udvardi,M.K., Benedito,V.A., Mayer,K.F.X., Gouzy,J., Schoof,H. *et al.* (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature*, **480**, 520–524.
72. Guo,L., Winzer,T., Yang,X., Li,Y., Ning,Z., He,Z., Teodor,R., Lu,Y., Bowser,T.A., Graham,I.A. *et al.* (2018) The opium poppy genome and morphinan production. *Science*, **362**, 343–347.
73. McCormick,R.F., Truong,S.K., Sreedasyam,A., Jenkins,J., Shu,S., Sims,D., Kennedy,M., Amirebrahimi,M., Weers,B.D., McKinley,B. *et al.* (2018) The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.*, **93**, 338–354.
74. Badouin,H., Gouzy,J., Grassa,C.J., Murat,F., Staton,S.E., Cottret,L., Lelandais-Brière,C., Owens,G.L., Carrère,S., Mayjonade,B. *et al.* (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, **546**, 148–152.
75. Ouyang,S., Zhu,W., Hamilton,J., Lin,H., Campbell,M., Childs,K., Thibaud-Nissen,F., Malek,R.L., Lee,Y., Zheng,L. *et al.* (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.*, **35**, D883–D887.
76. Zhang,L., Chen,F., Zhang,X., Li,Z., Zhao,Y., Lohaus,R., Chang,X., Dong,W., Ho,S.Y.W., Liu,X. *et al.* (2020) The water lily genome and the early evolution of flowering plants. *Nature*, **577**, 79–84.
77. Alioto,T., Alexiou,K.G., Bardil,A., Barteri,F., Castanera,R., Cruz,F., Dhingra,A., Duval,H., Fernández i Martí,Á. and Frias,L. (2020) Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *Plant J.*, **101**, 455–472.
78. Zeng,L., Tu,X.-L., Dai,H., Han,F.-M., Lu,B.-S., Wang,M.-S., Nanaei,H.A., Tajabadipour,A., Mansouri,M. and Li,X.-L. (2019) Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biol.*, **20**, 79.
79. Chen,J., Huang,Q., Gao,D., Wang,J., Lang,Y., Liu,T., Li,B., Bai,Z., Luis Goicoechea,J., Liang,C. *et al.* (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying Oryza genome evolution. *Nat. Commun.*, **4**, 1595.
80. Avni,R., Nave,M., Barad,O., Baruch,K., Twardziok,S.O., Gundlach,H., Hale,I., Mascher,M., Spannagl,M., Wiebe,K. *et al.* (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, **357**, 93–97.
81. Xu,S., Brockmöller,T., Navarro-Quezada,A., Kuhl,H., Gase,K., Ling,Z., Zhou,W., Kreitzer,C., Stanke,M., Tang,H. *et al.* (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 6133–6138.
82. Lee,C.-R., Wang,B., Mojica,J.P., Mandáková,T., Prasad,K.V.S.K., Goicoechea,J.L., Perera,N., Hellsten,U., Hundley,H.N., Johnson,J. *et al.* (2017) Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat. Ecol. Evol.*, **1**, 119.
83. Sato,S., Tabata,S., Hirakawa,H., Asamizu,E., Shirasawa,K., Isobe,S., Kaneko,T., Nakamura,Y., Shibata,D., Aoki,K. *et al.* (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.
84. International Cassava Genetic Map Consortium(ICGMC) (2015) High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3 Genes Genomes Genet.*, **5**, 133–144.
85. Huang,S., Li,R., Zhang,Z., Li,L.I., Gu,X., Fan,W., Lucas,W.J., Wang,X., Xie,B. and Ni,P. (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.*, **41**, 1275–1281.
86. Bao,W., Kojima,K.K. and Kohany,O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA*, **6**, 11.
87. Emms,D.M. and Kelly,S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, **20**, 238.
88. Wang,J., Zhang,J., Zheng,H., Li,J., Liu,D., Li,H., Samudrala,R., Yu,J. and Wong,G.K.-S. (2004) Neutral evolution of 'non-coding' complementary DNAs. *Nature*, **431**, 1–2.