Plant Communications
Correspondence

CellPress
Partner Journal

CGRP: A versatile platform for exploring Cucurbitales genomics

Dear Editor,

Cucurbitales, one of the most diverse angiosperm orders that originated in the Cretaceous, contains at least 2680 species representing 128 genera across eight families (Society, 2016; Li et al., 2019). This order includes numerous economically important crops and plays a crucial role in human survival and ecological sustainability, thereby attracting widespread scientific interest (Guo et al., 2020; Ma et al., 2022). The continuous accumulation of genome assemblies and associated databases has provided strong support for studying the evolution, functional genomics, and breeding of Cucurbitales species (Yu et al., 2023; Schwacke et al., 2025). However, paleopolyploidization (whole-genome duplication [WGD]), which represents a key driver of genetic innovation, has profoundly shaped the complex genomic architecture of Cucurbitales species through chromosome reorganization and gene family expansion (Wang et al., 2022). These processes pose great challenges for tracing evolutionary trajectories (ETs) and mining key trait-associated genes in Cucurbitales species. Here, we established the Cucurbitales Genome Research Platform (CGRP; https://cucurbitales.cgrpoee.top/), a comprehensive resource designed to decipher genome evolution and function, particularly by reconciling WGD-driven genome reorganization with functional innovation across Cucurbitales (Figure 1). CGRP integrates large-scale data resources and systematic analytical tools through 87 interactive interfaces across eight functional portals to serve genomic exploration and molecular breeding in Cucurbitales (Supplemental Note 1; Supplemental Figure 1).

CGRP contains the most extensive genome assembly and annotation resources for Cucurbitales (Supplemental Note 2). It integrates 213 genomes in a unified format, including all published genomes of 34 Cucurbitales species as well as our newly generated *Hemsleya chinensis* genome, which represents an early-diverging lineage within Cucurbitaceae (Figure 1A). Currently, detailed gene annotations are available for 29 Cucurbitales species genomes (29-CGs), and their high-quality versions were utilized in this study to generate comparative genomics resources (Supplemental Figure 2). We integrated JBrowse (v2.14) (Diesh et al., 2023) for browsing genome details and developed a toolbox for users to easily access genes in species-specific chromosomal regions (Supplemental Figure 3). For the 29-CGs and grape (*Vitis vinifera*), we annotated 3 761 602 biological function terms and metabolic pathways for 720 404 proteins and identified 18,367,214 transposable elements, 86 894 regulatory proteins, 927 302 duplicated genes, and 606 N^6-methyladenosine regulators, using six standardized pipelines (Supplemental Methods; Supplemental Figures 4–9). To facilitate access to these data, we developed six user-friendly, web-accessible interfaces and established a CGRP-specific identifier full association system, enabling users to efficiently utilize these annotations for innovative studies and molecular breeding (Supplemental Figure 10).

Central genomic synteny (GS) interfaces enable scientists to systematically study WGD-associated genomic evolution, functional innovation, and trait-related genes in Cucurbitales (Supplemental Note 3; Supplemental Figure 11A–11C). We identified 12 083 671 gene pairs across 1 208 371 syntenic blocks based on 900 pairwise genome comparisons. These data were integrated into the Syn-Blocks interface and SynView tool to empower the inference of WGD- and speciation-associated syntenic blocks (Figure 1B and Supplemental Figure 12). We generated 2700 GS dot plots and incorporated them into the Syn-Dotplots interface featuring the DotView tool (Figure 1B), which enables detailed analyses of WGD events, chromosomal rearrangements, and genomic fractionation in Cucurbitales (Supplemental Figures 13–15). The Deco-Alignment interface provides a large-scale multi-genome alignment matrix with 225 × 23 647 gene identifiers related to seven WGDs across 29-CGs and grape (Figure 1C and Supplemental Figure 16). Integrated with DecoBrowse, this interface enables users to dissect WGD-driven evolution of agronomic trait genes and efficiently identify trait-associated candidate genes in Cucurbitales to accelerate crop molecular breeding. The applicability of Deco-Alignment in inferring evolutionary history and identifying candidates for the flower sex regulator (CmWIP1) in Cucurbitales is demonstrated (Supplemental Figures 17 and 18). Moreover, the WGDs-Genes interface contains 270 489 paralogs that were generated by seven WGD events in 29-CGs, allowing users to quickly identify trait-associated genes derived from specific WGD events (Supplemental Figure 19). In this study, we found that 53% of heat shock factors in Cucurbitaceae originated from WGDs, suggesting that WGDs made substantial genetic contributions to the radiative evolution of Cucurbitales during the Paleocene–Eocene extreme heat event (Guo et al., 2020).

Ancestral genomes are critical reference resources for deciphering plant ETs and functional innovations (FornMura et al., 2017). Although we previously reconstructed ancestral Cucurbitales karyotypes (Wang et al., 2022), their applicability has remained limited. Here, for the first time, we present the Anc-Genomes and Anc-Aligning interfaces, providing the most comprehensive collection of ancestral Cucurbitales genomes (ACGs) to date (Supplemental Note 4; Figure 1D). Anc-Genomes integrates ACGs representing nine evolutionary nodes and visualizes detailed ETs of paleochromosomes in both static and animated modes (Supplemental Figure 20). Anc-Aligning provides access options and comparative visualization functions for ACGs, allowing users to easily infer the ETs of newly sequenced genomes (Supplemental Figure 21). Together, these data resources and tools enable scientists to investigate the evolutionary history of

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

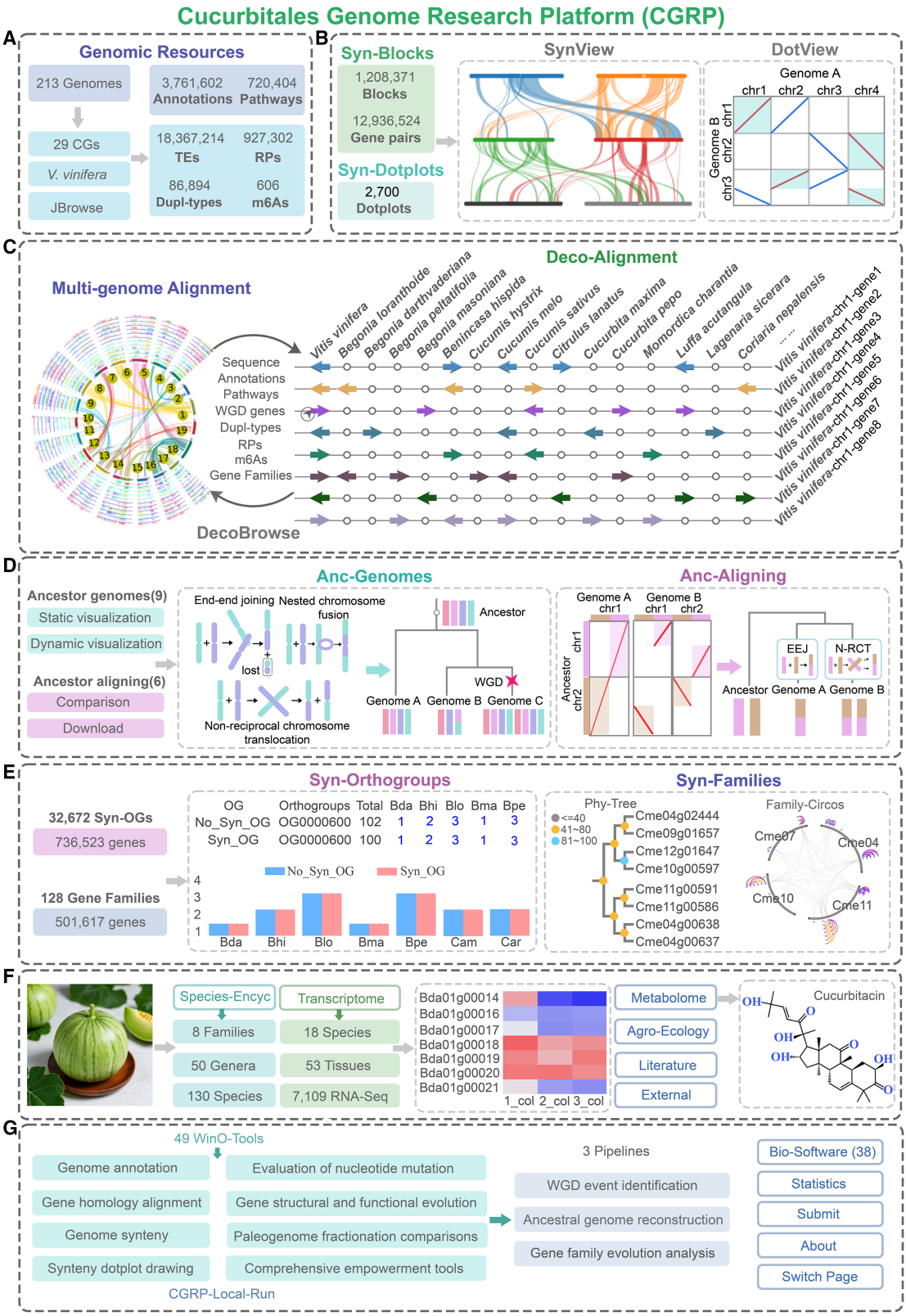


Figure 1. Overview of data resources and functional modules in CGRP.

(A) Genomic resources in CGRP.

(B) The Syn-Blocks interface includes GS resources and the SynView visualization tool for Cucurbitales genomes. The Syn-Dotplots interface includes GS dot plots and the DotView visualization tool.

(legend continued on next page)

ACGs and identify trait-associated paleogenome reshuffling. Our previous study suggested that paleogenome reshuffling shaped the diversity of cucurbitacin biosynthesis in Cucurbitales (Wang et al., 2022).

The Specialized Syn-Orthogroups and Syn-Families interfaces for gene families enable users to trace the genetic basis of agronomic traits across Cucurbitales (Supplemental Note 5; Figure 1E). Syn-Orthogroups offers 32 672 synteny-based orthogroups across Cucurbitales and outgroup species and provides user-friendly data acquisition and visualization capabilities. It supports the identification of trait-related genes in Cucurbitales and the investigation of associated genomic evolutionary events. Using Syn-Orthogroups, we identified synteny-based orthogroups and candidate genes involved in the biosynthesis of bitter compounds and elucidated the associated gene duplication events(Supplemental Figures 22 and 23). Furthermore, we identified 501 617 trait-associated genes from 128 families involved in multiple aspects of growth, development, and environmental adaptability of Cucurbitales plants. The Syn-Families interface is integrated with the Phy-Tree and Family-Circos tools, enabling users to explore the evolutionary origins of specific traits and facilitating targeted crop improvement using these family resources, as demonstrated by analysis of the cytochrome P450 family (Supplemental Figures 24 and 25).

We developed six interactive interfaces to comprehensively integrate diverse data resources for functional genomics research and breeding applications in Cucurbitales (Figure 1F and Supplemental Figure 26). Species-Encyc curates phylogenetic classifications, karyotypes, genome sizes, and living habits of 130 species across 50 genera from eight Cucurbitales families and makes these data freely accessible. The Transcriptome interface contains 7109 transcriptomes from 53 tissues across 18 Cucurbitales species, including 33 newly generated transcriptomes from *Hemsleya chinensis*. It houses genome-wide gene expression data, access, and visualization modules, enabling users to study the expression and function of genes in Cucurbitales crops (Supplemental Figure 27). The Metabolome interface integrates the largest-scale metabolome resources for Cucurbitales, empowering scientists to decipher metabolic pathways and key genes involved in metabolite biosynthesis (Supplemental Figure 28). The Agro-Ecology, Literature, and External interfaces enable users to easily access ecological and plant-related knowledge, literature, and external database resources for Cucurbitales. More detailed resources, functions, and application cases of these six interfaces are provided in Supplemental Note 6.

A series of portals empowers CGRP with enhanced mining capabilities, continuous updates, and user-friendly features (Supplemental Note 7 and Figure 1G). In WinO-Tools, we built 49 web-based bioinformatics tools, 15 of which were newly developed by us. These tools feature user-friendly interfaces, instructions, and demonstrations, enabling users to easily study genome annotation, evolution, and function. In addition to running independently, these tools can be customized into commonly used pipelines for exploring WGDs, ancestral genomes, and gene family evolution. In particular, the WGD-associated genome analysis pipeline proposed in this study is considered a gold standard for deciphering complex genomes (Caspermeyer, 2018; Wang et al., 2018). To overcome network and computational resource limitations, we developed the CGRP-Local-Run toolkit, which encapsulates all tools, allowing users to run them locally. Details of this toolkit are available on CGRP (https://cucurbitales.cgrpoee.top/tools/docker). We also collected downloads and usage guides for 38 bioinformatics software packages in Bio-Software, providing convenient alternatives for genome analysis. Moreover, the Statistics portal includes a series of interactive charts that facilitate users' understanding of the outline and biological significance of data resources in CGRP. The Submit and About portals feature community modules to enhance collaborative updates and improve the overall user experience of CGRP. Finally, the Switch Page provides a user-friendly tool for the comprehensive exploration of genomic, transcriptomic, and metabolomic data resources in CGRP (Supplemental Note 7).

In summary, CGRP combines rich and distinctive data resources, powerful analytical tools, and user-friendly interactive functions to continuously deliver high-value services for innovative research that support breeding practices in Cucurbitales.

DATA AND CODE AVAILABILITY

The newly assembled genome sequences have been deposited in the NCBI database under accession number PRJNA1392842. The source code for platform construction is available on GitHub (https://github.com/songtengfei1126/CGRP/tree/master). All other datasets are available at https://cucurbitales.cgrpoee.top/, which is open for research use without user registration.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32470676 and 32170236 to Jinpeng Wang), the Hebei Natural Science Foundation (Grant No. C2020209064 to Jinpeng Wang), and the Central Guidance on Local Science and Technology Development Fund of Hebei Province (Grant No. 246Z2508G to Lixia Gou).

ACKNOWLEDGMENTS

No conflict of interest is declared.

AUTHOR CONTRIBUTIONS

Jinpeng Wang conceived and initiated the project. Jinpeng Wang, T.S., S. W., Z.L., and Z.W. supervised and managed the project. Genomic analyses were led by T.S., X.G., X.W., J.Y., X.L., F.L., and L.G. Data collection

(C) The Deco-Alignments interface stores multi-genome alignments and is equipped with DecoBrowse to explore the origin and evolution of agronomic trait genes.

(D) The Anc-Genomes interface enables static and dynamic visualization of ACGs and ETs. The Anc-Aligning interface enables alignment and visualization analyses of new genomes against inferred ancestral genomes, thereby facilitating exploration of their ETs.

(E) The Syn-Orthogroups and Gene-Families interfaces contain large-scale gene family resources, incorporated with user-friendly data acquisition options and visualization tools to facilitate exploration of the origins of important agronomic traits.

(F) Diverse data resource interfaces supporting functional genomics and molecular breeding research in Cucurbitales plants.

(G) Functional modules that empower CGRP with advanced data mining, continuous updates, and user-friendly features.

and genome annotations were performed by T.S., Y.C., Jiaqi Wang, and Y.L. T.S., Z.Y., and T.L. contributed to programming and database construction. The manuscript was organized, written, and revised by Jinpeng Wang, T.S., S.W., Z.L., and Z.W. All authors read and approved the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: March 6, 2025
Revised: November 26, 2025
Accepted: December 26, 2025
Published: December 27, 2025

Tengfei Song[1,8], Zijian Yu[1,8], Xintong Gao[1,8], Xinhong Wei[1,8], Tianyu Lei[1,2,3], Junjie Yan[1], Xinyu Li[1], Feng Long[1], Yufan Cao[1], Xuan Wang[1], Jiaqi Wang[1], Lixia Gou[1], Yuxian Li[1], Zhenyi Wang[1,], Zhikun Li[4,*], Shengdan Wu[5,*] and Jinpeng Wang[1,6,7,*]*

[1]Department of Bioinformatics, School of Life Sciences, North China University of Science and Technology, Tangshan, Hebei 063000, China
[2]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]State Key Laboratory of North China Crop Improvement and Regulation, North China Key Laboratory for Crop Germplasm Resources of the Ministry of Education, Co-Innovation Center for Cotton Industry of Hebei Province, Hebei Agricultural University, Baoding, China
[5]State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, College of Ecology, Lanzhou University, Lanzhou 730000, China
[6]Key Laboratory of Hebei for Integrated Utilization of Saline Alkali Land in Medical Engineering, North China University of Science and Technology, Tangshan, Hebei 063000, China
[7]Key Laboratory of Quality Control for Salt Alkali Resistant Traditional Chinese Medicine of Hebei Administration of Traditional Chinese Medicine, North China University of Science and Technology, Tangshan, Hebei 063000, China
[8]These authors contributed equally to this article.
***Correspondence: Zhenyi Wang (zhenyiwang0301@163.com), Zhikun Li (lzk790319@163.com), Shengdan Wu (wusd@lzu.edu.cn), Jinpeng Wang (wangjinpeng@ncst.edu.cn)**
https://doi.org/10.1016/j.xplc.2025.101688

REFERENCES

Caspermeyer, J. (2018). Exploring an Ancient Event in Pumpkin, Gourd, and Melon Evolution. Mol. Biol. Evol. **35**:256–257. https://doi.org/10.1093/molbev/msx291.

Diesh, C., Stevens, G.J., Xie, P., De Jesus Martinez, T., Hershberg, E. A., Leung, A., Guo, E., Dider, S., Zhang, J., Bridge, C., et al. (2023). JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol. **24**:74. https://doi.org/10.1186/s13059-023-02914-z.

Murat, F., Armero, A., Pont, C., Klopp, C., and Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. Nat. Genet. **49**:490–496.

Guo, J., Xu, W., Hu, Y., Huang, J., Zhao, Y., Zhang, L., Huang, C.H., and Ma, H.J.M.P. (2020). Phylotranscriptomics in Cucurbitaceae Reveal Multiple Whole-Genome Duplications and Key Morphological and Molecular Innovations. Mol Plant. **13**:1117–1133.

Li, H.T., Yi, T.S., Gao, L.M., Ma, P.F., Zhang, T., Yang, J.B., Gitzendanner, M.A., Fritsch, P.W., Cai, J., Luo, Y., et al. (2019). Origin of angiosperms and the puzzle of the Jurassic gap. Nat. Plants **5**:461–470.

Ma, L., Wang, Q., Zheng, Y., Guo, J., Yuan, S., Fu, A., Bai, C., Zhao, X., Zheng, S., Wen, C., et al. (2022). Cucurbitaceae genome evolution, gene function, and molecular breeding. Horticult. Res. **9**:uhab057. https://doi.org/10.1093/hr/uhab057.

Schwacke, R., Bolger, M.E., and Usadel, B. (2025). PubPlant – a continuously updated online resource for sequenced and published plant genomes. Preprint at bioRxiv. https://doi.org/10.1101/2025.03.12.642823.

The Angiosperm Phylogeny Group, Chase, M.W., Christenhusz, M.J. M., Fay, M.F., Byng, J.W., Judd, W.S., Soltis, D.E., Mabberley, D.J., Sennikov, A.N., Soltis, P.S., and Stevens, P.F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical J. Linnean Soc. **181**.

Wang, J., Yuan, M., Feng, Y., Zhang, Y., Bao, S., Hao, Y., Ding, Y., Gao, X., Yu, Z., Xu, Q., et al. (2022). A common whole-genome paleotetraploidization in Cucurbitales. Plant Physiol. **190**:2430–2448. https://doi.org/10.1093/plphys/kiac410.

Wang, J., Sun, P., Li, Y., Liu, Y., Yang, N., Yu, J., Ma, X., Sun, S., Xia, R., Liu, X., et al. (2018). An Overlooked Paleotetraploidization in Cucurbitaceae. Mol. Biol. Evol. **35**:16–26. https://doi.org/10.1093/molbev/msx242.

Yu, J., Wu, S., Sun, H., Wang, X., Tang, X., Guo, S., Zhang, Z., Huang, S., Xu, Y., Weng, Y., et al. (2023). CuGenDBv2: an updated database for cucurbit genomics. Nucleic Acids Res. **51**:D1457–D1464. https://doi.org/10.1093/nar/gkac921.

Plant Communications, Volume 7

Supplemental information

CGRP: A versatile platform for exploring Cucurbitales genomics

Tengfei Song, Zijian Yu, Xintong Gao, Xinhong Wei, Tianyu Lei, Junjie Yan, Xinyu Li, Feng Long, Yufan Cao, Xuan Wang, Jiaqi Wang, Lixia Gou, Yuxian Li, Zhenyi Wang, Zhikun Li, Shengdan Wu, and Jinpeng Wang

Supplemental information

CGRP: A versatile platform for exploring the genomics of Cucurbitales

Tengfei Song, Zijian Yu, Xintong Gao, Xinhong Wei, Tianyu Lei, Junjie Yan, Xinyu Li, Feng Long, Yufan Cao, Xuan Wang, Jiaqi Wang, Lixia Gou, Yuxian Li, Zhenyi Wang, Zhikun Li, Shengdan Wu, Jinpeng Wang

Supplemental information

CGRP: A versatile platform for exploring the genomics of Cucurbitales

Tengfei Song[1, 8], Zijian Yu[1, 8], Xintong Gao[1, 8], Xinhong Wei[1, 8], Tianyu Lei[1, 2, 3], Junjie Yan[1], Xinyu Li[1], Feng Long[1], Yufan Cao[1], Xuan Wang[1], Jiaqi Wang[1], Lixia Gou[1], Yuxian Li[1], Zhenyi Wang[1, *], Zhikun Li[4, *], Shengdan Wu[5, *], Jinpeng Wang[1, 6, 7, *]

[1] Department of Bioinformatics, School of Life Sciences, North China University of Science and Technology, Tangshan, Hebei 063000, China

[2] State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Science, Beijing 100093, China

[3] University of Chinese Academy of Sciences, Beijing 100049, China

[4] State Key Laboratory of North China Crop Improvement and Regulation, North China Key Laboratory for Crop Germplasm Resources of Education Ministry, Co-Innovation Center for Cotton Industry of Hebei Province, Hebei Agricultural University, Baoding, China

[5] State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, College of Ecology, Lanzhou University, Lanzhou 730000, China

[6] Key Laboratory of Hebei for Integrated Utilization of Saline alkali Land in Medical Engineering, North China University of Science and Technology, Tangshan, Hebei 063000, China

[7] Key Laboratory of Quality Control for Salt alkali Resistant Traditional Chinese Medicine of Hebei Administration of Traditional Chinese Medicine, North China University of Science and Technology, Tangshan, Hebei 063000, China

[8] These authors contributed equally to this article.

[*]**Correspondence:** Jinpeng Wang (wangjinpeng@ncst.edu.cn), Shengdan Wu (wusd@lzu.edu.cn), Zhikun Li (lzk790319@163.com), Zhenyi Wang (zhenyiwang0301@163.com)

This PDF file includes:

Supplemental notes S1-S7

Supplemental Materials and Methods

Supplemental Figures 1-30

Supplemental Tables 1-8

Supplemental References

Supplemental notes…….…...…..5

S1. Overview of CGRP……..5

S2. Enrich genomic resources for free access……6

S3. Web services for genomic synteny (GS)……..……..10

S4. Web services for Ancestral Cucurbitales genomes…..…..……15

S5. Web services for orthogroups and gene families……...……17

S6. Diverse data resources for the studies of functional genomics research and breeding in Cucurbitales comprehensively……..20

S7. Empowering CGRP with new mining, continuous updates, and user-friendly…….....23

Supplemental Materials and Methods…….29

Supplemental Figures 1-30……...…….34

Supplemental Tables 1-8……...…….....60

Supplemental References……...82

52 **Supplemental Information**

53 **Supplemental Notes**

54 **Supplemental note S1. Overview of CGRP**

55 Within CGRP, we compiled 8 functional portals, including the **Resource**, **Browse**, **Gene**
56 **Families**, **WinO-Tools**, **Bio-Software**, **Statistics**, **Submit**, and **About (Supplemental Figure**
57 **1)**. In **Resource**, we integrated all published genomes of the Cucurbitales, as well as many
58 transcriptomes and metabolomes. In this portal, we also collected encyclopedias, ecological
59 resources, literatures, and links of external databases for species in Cucurbitales. The **Browse**
60 portal mainly includes various genome annotations and distinctive genomic synteny (GS)
61 resources. Genomic annotations, including the potential biological functions and metabolic
62 pathways of proteins, as well as identified transposable elements (TEs), regulatory proteins
63 (RPs), gene duplication types (Dupl-types), and N^6-methyladenosine regulators (m^6As).
64 Specifically, we built a series of interfaces in Browse for exploring the GS associated with
65 evolutionary events and the ancestral Cucurbitales genomes. In **Gene Families**, we constructed
66 synteny-based orthogroups across the genomes of Cucurbitales and outgroups, and identified
67 128 gene families that play important roles in the growth and development of Cucurbitales
68 plants. For these portals, we have configured user-friendly interactive interfaces that could
69 better serve the researches of Cucurbitales evolution and functional studies.

70 The other five portals aim to fully empower the platform to cope with the continuous
71 emergence of new genomes and provide user-friendly and continuous updates for the resources
72 and tools in CGRP. The **WinO-Tools** has compiled 49 web-based operating tools, which were
73 involving the 3 pipelines for the identification of polyploidization, ancestral genome

reconstruction, and gene family evolution. In **Bio-Software**, we collected 38 bioinformatics software and provided detailed installation and operation commands. The **Statistics** portal provides statistical charts of various resources in the platform, allowing for an overview of data and quick search for the underlying database. The **Submit** portal is designed to support researchers in submitting various genomic data of the Cucurbitales, which facilitates collaborative updates of CGRP. The **About** portal lists the usage instructions for CGRP and provides ways to contact us. In addition, we designed front-end web applications for CGRP, which can access and acquire back-end available resources and call bioinformatics tools. Conveniently, the front-end interface is equipped with *Switch Page* that provides a simple search function, allowing users to submit gene names or protein sequences and easily search for all relevant information in CGRP with just one click.

Supplemental note S2. Enrich genomic resources for free access

The acquisition of genomic resources is crucial for knowledge mining in genomics. Herein, we collected all published genomes of Cucurbitales and the genomes of representative species from the main angiosperm clades. Especially, we provided the genome of *Hemsleya chinensis*, a medicinal plant from the Cucurbitaceae that we recently deciphered. For these Cucurbitales genomes, we annotated their proteins with detailed biological functions (BFs) and identified the TEs, RPs, Dupl-types and m^6As using the unified pipelines, respectively. These genomics resources have been integrated into the Resource and Browse portals, providing users with flexible interactive operation options to easily access relevant data.

In Genome interface, we integrated the genomes of Cucurbitales and outgroups in a unified format and provided browsing and downloading functions. For Cucurbitales, we collected 185 genomes of 34 species, including the species with multiple versions of genomes **(Supplemental Figure 2 and Supplemental Table 1)**. Within the Cucurbitales, the genomes of 29 species (29-CGs) can get the gene coding and protein sequences (CDS and PEP), and the locations on the chromosomes (GFF). The 29-CGs contain our newly deciphered genome of *Hemsleya chinensis* from the Cucurbitaceae. We also collected the genomes of 28 representative species from major angiosperm clades to explore the genetic innovation of Cucurbitales plants within a large-scale. For these genomes, we provided the download options and configured the JBrowse (v2.14.0) (Diesh et al., 2023) for browsing the details. In addition, we developed the interactive options for users to access the PEPs and CDS from specific chromosome regions of species, as well as the genomic locations of genes **(Supplemental Figure 3)**.

The annotations of BFs of proteins is helpful in exploring the molecular basis of the origin of agronomic traits. Selecting the latest high-quality genome version, we annotated the BFs of proteins in 29-CGs and *V. vinifera*. By comparing the PEPs with 16 databases, we obtained 3,761,602 BF annotation entries **(Supplemental Figure 4)**. For each entry, we listed the gene name, genomic location, Gene Ontology (GO) number from AmiGO (Carbon et al., 2009), InterPro number from InterPro (Paysan-Lafosse et al., 2023), and BF descriptions. Moreover, comparing the PEPs with Kyoto Encyclopedia of Genes and Genomes (KEGG) database using GhostKoala (Kanehisa et al., 2016), we inferred the signaling pathways of 720,404 proteins in 29-CGs and *V. vinifera* **(Supplemental Figure 5 and Supplemental Table 2)**.

Accurately identifying and classifying the functional elements of the genome has important scientific value for studying species adaptive evolution. Using unified pipelines, we identified the TEs, RPs, Dupl-types, and m^6As in the genomes of 29-CGs. TEs can often lead to the generation of new mutations in DNA sequences, reshape gene regulatory networks, and even increase genome size by altering the structures of genome (Naville et al., 2019; Nishihara, 2020; Usha et al., 2022). We identified a total of 18,367,314 TEs, including 6,168,937 retrotransposons, 1,867,367 DNA transposons, and 1,947,228 other types of transposons. RPs play an important role in plant growth and development, biotic and abiotic stress, and are classified into transcription factors (TFs), transcription regulators (TRs) and protein kinases (PKs) (Manna et al., 2021; Weidemuller et al., 2021; Zheng et al., 2016b). We identified a total of 86,894 RPs. These RPs can be subdivided into 48,876 TFs from 69 families, 10,736 TRs from 23 families, and 27,282 PKs from 130 families. Moreover, we found that the RPs families exhibited rapid expansion in some species that experienced high-frequency WGDs, while other some species did not show an expansion pattern **(Supplemental Figures 6-8)**. For example, the genera *Cucurbita* and *Begonia* have undergone two WGDs events since the origin of the Cucurbitales, but their TFs exhibited extremely different expansion patterns. This seems to be mainly attributed to the RPs loss of preference after the WGDs, which facilitated the adaptation of differences in plants of Cucurbitales.

Furthermore, we identified the Dupl-types and m^6As in the genomes of 29-CGs and *V. vinifera*. First, we conducted genome-wide search for Dupl-types in these genomes and identified a total of 927,302 duplicated genes. Among these duplicates, 475,437 originated from dispersed duplications (DSDs), 23,716 from proximal duplications (PDs), 148,300 from transposed

duplications (TRDs), 75,411 from tandem duplications (TDs), and 204,438 from whole genome duplications [WGDs]/segmental duplications (WSDs) (**Supplemental Table 3)**. In these five types of duplicates, we found that the most duplicates were associated with DDs, implying that the TEs play an important role in shaping genomic structural diversity **(Supplemental Figure 9)**. Subsequently, we detected a large number of WSD and TD produced duplicates in plants of Cucurbitales, indicating that WGD and TD events greatly contribute to the improvement of genomic plasticity. In addition, we identified a total of 606 m^6A genes, including 139 writers, 205 readers, and 262 erasers. The writers were further classified into 92 MTA70, 44 WTAP, and 3 VIRILIZER (**Supplemental Table 4)**.

Finally, to support further exploration of functional elements of genomes, we developed six interactive interfaces in Browse portal: Fun-Annotations, KEGG-Annotations, Trans-elements, Regu-proteins, Dupl-types, and N^6-methyladenosine **(Supplemental Figure 10)**. The annotated entries are stored separately in the Fun-Annotations and KEGG-Annotations interfaces, which were implanted the flexible interactive options that allow users to access target annotations through the gene names, sequences, or other entries. In Trans-elements, we provide options for obtaining TEs from both genome-wide and family scales. Users can freely access RPs and their sequence, pfam domains, and genomic annotations through the Regu-proteins interface. More helpfully, each TR entry provides links to the pfam and IPR databases for exploring its biological functions. In Dupl-types, the identified Dupl-types are on standby for scientists to investigate the origin of the target genes. Finally, the N^6-methyladenosine interface is equipped with user-friendly web service options for m^6As.

Specifically, we standardized the naming of the gene datasets (referred as CGRP-specific identifiers) to facilitate cross-functional module searches for comprehensive annotation landscape of target genes. The CGRP-specific identifier contains gene from which species, chromosome, and gene order. We established a CGRP-specific identifier full-association system for the data information of each functional module. Users can easily access the comprehensive genomic and BF annotations of target genes through the CGRP-specific identifiers in any functional module.

Supplemental note S3. Web services for genomic synteny (GS)

GS provides important clues for revealing the evolution of complex genome structures and gene functions. By comparing the genomes of 29-CGs and *V. vinifera*, we generated a vast pairwise GS resource, multiple genome alignments across species, and paralogous genes generated from WGD events. Through GS comparisons, we reconstructed the ancestral Cucurbitales genomes and inferred the evolutionary trajectories of karyotypes in extant species. For these resources, we established 6 interactive interfaces with user-friendly functional modules in the Browse portal, providing high-energy web services for studying the evolution of genome structures and functions in Cucurbitales.

In Syn-Blocks interface, we offered synteny blocks (SBs) for pairwise genomes and compiled operational options for identifying the SBs associated with evolutionary events **(Supplemental Figure 11A)**. We detected 1,208,371 SBs from 900 pairwise genome comparisons, containing 12,936,524 synteny gene pairs (SGPs). For each SGP, we calculated the non-synonymous and synonymous substitution rates (*Ka* and *Ks*) of nucleotides, and calculated the median and

average *Ks* values of SGPs from the SBs. Using the preset operation options, researchers can freely access the target SBs and infer the evolutionary events related to their origins **(Supplemental Figure 12 A-C)**. Given the application scenarios: (1) It is convenient to select the WGD and species divergence generated SBs according to our previous studies (Wang et al., 2018; Wang et al., 2022), and then calculated the *Ks* peaks associated with evolutionary events through the 'Draw *Ks* peak' option. For example, the *Ks* peaks corresponding to the two WGDs (ECH and CCT) events in *C. melo* (Cme) were ~1.32 and ~1.05, respectively, while the *Ks* peak for the divergence between *C. melo* and *C. sativus* (Csa) was ~0.14 **(Supplemental Figure 12 A-C)**. (2) Then, according to the median *Ks* of SGPs in SBs approximately equal to the *Ks* peak of evolutionary events, the first SB between chromosomes Cme02) and Cme10 of *C. sativus* was identified as paralogous genomic region (PGR) generated from ECH event, and the second SB between Cme01 and Cme11 identified as CCT produced PGR **(Supplemental Figure 12A)**. Similarly, the third SB between Cme11 and Csa02 contains 66 SGPs with a median *Ks* of ~0.1 that identified as the orthologous region (OR) generated from the *C. melo* and *C. sativus* divergence **(Supplemental Figure 12C)**. Finally, we built the **SynView** web service port by integrating the Synvisio (Bandi et al., 2022), allowing researchers to examine the SBs at various genomic scales **(Supplemental Figure 12D)**.

In Syn-Dotplots interface, we provided the genomic synteny dotplots (GSDs) and visualization functions for users to infer the genome evolution of Cucurbitales **(Supplemental Figure 11B)**. Comparing the genomes of 29-CGs and *V. vinifera*, we mapped 2,700 GSDs in B-dotplot, BB-dotplot, and BBK-dotplot. The B-dotplot visualizes the Blast matches to display SBs, presenting a comprehensive range of anchoring gene pairs (AGPs) but with significant

background noise in syntenic regions (SRs). The BB-dotplot could be used to visualize SBs and Blast matches, with clear SRs but lack discrete AGPs. The BBK-dotplot integrates the *Ks* medians of SBs into BB-dotplot. These dotplots have important guiding significance for revealing the polyploidization and chromosome evolution of Cucurbitales, as demonstrated in previous studies (Wang *et al.*, 2018; Wang *et al.*, 2022). For example, using the *V. vinifera* and *C. melo* dotplots, we detected that their orthologous ratio was 1:2, indicating that the *C. melo* experienced another round of WGD after it divergence with *V. vinifera* **(Supplemental Figures 13 and 14)**. In addition, we developed **DotView** for the local visualization of dotplot to investigate the syntenic relationships between specific chromosomes and genes. For example, we detected that the ORs of Cme12 were located in chromosomes 6, 10, and 17 of *V. vinifera*, and many orthologs were missing from these ORs, suggesting that the large-scale genome reorganization of *V. vinifera* and *C. melo* relative to the ancestral eudicot genome **(Supplemental Figure 15)**.

In Deco-Alignment interface, we offered the multi-genome alignment (MGA) across Cucurbitales species, and developed the **DecoBrowse** for studying the evolution and function of important trait genes **(Supplemental Figure 11C)**. Deconvolution of the synteny regions associated with 7 WGD events (ECH, CCT, BCT, CST, SSD, CSD1, and CSD2), we established an MGA across 28 species from the Cucurbitaceae, Begoniaceae, and Coriariaceae in Cucurbitales with *V. vinifera* genome as reference **(Supplemental Figure 16)**. This MGA is a massive matrix consisting of 225×23,647 gene identifiers (IDs), with each gene linked to its genomic and BF annotations. The compilation of flexible interaction options and powerful **DecoBrowse** tool allows researchers to easily infer how and when homologs were generated

and diverged and whether there were gene deletions after certain evolutionary events, thereby contributing to the elucidation of evolutionary and functional innovation trajectories of genes, gene families, regulatory pathways, and important economic and agronomic traits. With important agricultural applicability value, the candidate genes of key agronomic traits from specific Cucurbitales species can be quickly screened through the Deco-Alignment interface to improve the efficiency of crop molecular breeding.

Applicability of Deco-alignment interface in evolution and mining of important trait genes. The control of flower sex is of great significance for agricultural breeding and production, and the Cucurbitaceae has long been recognized as an excellent model system for studying flower sex differentiation (Zhang et al., 2022). A recent study revealed that the CmWIP1 is a key regulatory factor in determining flower sex in *C. melo* (Zhang *et al.*, 2022). Here, we demonstrated the evolution of CmWIP1 gene and the screening of candidate genes for flower sex determination in Cucurbitales species using the Deco-Alignment interface and its toolbox. Aligning the CmWIP1 protein with MGA, we determined that the CGRP identifier of CmWIP1 in *C. melo* is *Cme05g00149* and the paralogous gene is *Cme05g00149*, which was generated from the CCT event **(Supplemental Figure 17)**. Browsing the local MGA, we detected that two orthologous genes *Cmo02g01279* and *Cmo2g00081* of CmWIP1 in *C. moschata* were generated from the CST event, but species-specific deletion occurred in *B. hispida*. This result suggests that the regulatory networks of flower sex in *C. moschata* and *B. hispida* have undergone their own reshaping after the divergence of species. Moreover, the IDs of the orthologs and paralogs of CmWIP1 from any Cucurbitaceae species in MGA were linked to their GFFs, PEPs, GO items, KEGG pathways, and the corresponding RPs and GTDs

(Supplemental Figures 17 and 18). In addition, by integrating the gene homology relationships and detailed annotations provided by Deco-Alignment, it is easy to screen key candidate genes that determine flower sex in Cucurbitales crops. The duplicates of CmWIP1 in specific Cucurbitales species can be selected as candidate genes for flower sex determination, with priority given to its orthologs, paralogs, and other homologs. For example, the two orthologs (*Cmo02g01279* and *Cmo2g00081*) of CmWIP1 are the most prioritized, followed by the two CCT-paralogs (*Cmo06g00072* and *Cmo14g00447*), and other duplicates are the last candidates. Similarly, users can accurately and efficiently decipher the WGDs-driven evolutionary history of other agronomic trait genes in Cucurbitales and screen candidates for specific crops.

In WGDs-Genes interface, we offered the WGDs generated paralogous gene pairs (PGPs) in the genomes of 28-CGs and *V. vinifera*. To support scientists in querying WGD produced genes, we identified a total of 270,489 PGPs generated from 7 WGDs (**Supplemental Table 5)**. Using the prepared operation options, the specific WGD produced PGPs can be quickly queried through PEPs or IDs of the target genes **(Supplemental Figure 19)**. The Cucurbitaceae originated in the late Cretaceous and successfully evolved through radiation after experiencing the Paleocene-Eocene extreme heat event, becoming one of the largest surviving plant clades and mainly distributed in tropical and subtropical areas (Guo et al., 2020). During the evolutionary history of Cucurbitaceae plants, the recursive WGDs have played an important role in enhancing their plasticity in response to heat stress. Here, we identified 462 heat shock factors (HSFs) across 23 Cucurbitaceae genomes, of which 53.9% (249) were associated with WGDs (**Supplemental Table 6)**. Especially, among the 27 HSFs identified in *C. moschata*,

51.8% (14) were CCT genes, and 14.8% (4) were CST genes, suggesting that the WGDs provide crucial genomic dynamics for the strong heat stress plasticity of *C. moschata*.

Supplemental note S4. Web services for ancestral Cucurbitales genomes

In Anc-Genomes interface, we reconstructed the ancestral Cucurbitales genomes (ACGs) and evolutionary trajectories of species karyotypes, and provided static and dynamic visualization tools **(Supplemental Figure 20)**. In previous studies, we inferred the ancestral Cucurbitales karyotypes (ACKs) of multiple evolutionary nodes (ENs) and elucidated the large-scale chromosomal rearrangement events that occurred during the diversification of extant species (Wang *et al.*, 2022; Xie et al., 2019). Here, we further introduced the *C. nepalensis* of Coriariaceae and reconstructed the ACKs covered the Cucurbitaceae, Begoniaceae, and Coriariaceae in Cucurbitales and inferred the protogenes in ACKs. Now, we have reconstructed the ACGs of 9 ENs in Cucurbitales, containing the genome of the most recent common ancestor (MRCA-G), ancestral Coriariaceae genome I and II (ACoG-I and ACoG-II), common ancestral Cucurbitaceae and Begoniaceae genome (CACB-G), ancestral Begoniaceae genome I and II (ABG-I and ABG-II), ancestral Cucurbitaceae genome I, II, and III (ACuG-I, ACuG-II, and ACuG-III) **(Supplemental Figure 20B)**. Employing these ACGs as references, we inferred the evolutionary histories of Cucurbitales karyotypes. For example, we found that the 17 chromosomes of MRCA-G underwent 20 fusions to form the ACoG-II karyotype after the CSD1 event, and then evolved into the *C. nepalensis* karyotype after the CSD2 event. Distinctively, we developed static and dynamic visualization interfaces for browsing paleogenome reshaping imprints in contemporary species.

In Anc-Aligning interface, we offered ACG resources and aligning functions, which can facilitate the exploration of the evolution of new genomes **(Supplemental Figure 21)**. ACG resources include the chromosomes and gene content for 6 Ens of Cucurbitales (MRCA-G, ACoG-I, CACB-G, ABG-I, ACuG-I, and ACuG-III). Comparing the genomes with the ACGs, users can infer the evolutionary trajectories of the genome structures. For example, comparing the *B. hispida* genome with ACoG-III, we identified 5 chromosomal rearrangements, including 3 end-to-end chromosomal fusions and 2 translocations **(Supplemental Figure 21B)**.

Apart from deciphering the evolutionary history of Cucurbitales genomes, these ancestral genome resources have important practical value in elucidating the genetic basis of trait origins. Our previous study found that the species-specific paleochromosome reshuffling contributed to the diversification of cucurbitacin biosynthesis genes (CB) in Cucurbitales (Wang *et al.*, 2022). Conservative ancestral genomic regions provided excellent genetic conditions for the expansion of CBGs, while chromosomal rearrangements (CRs) are accompanied by fractionation of CBGs in Cucurbitales species. Furthermore, the CRs may be a key genomic variation that leads to the emergence of species-specific innovative traits. As previously reported, a paleochromosome fusion resulted in a new STORR gene, establishing the morphine biosynthesis pathway in poppy (Guo et al., 2018). Here, we provided the first largest ACGs and detailed ETs, fully serving the mining of paleochromosome reshuffling associated with Cucurbitales crop traits. It provides new data resources and analysis tools for optimizing the genetic breeding, stress resistance, and environmental adaptability of Cucurbitales crops through chromosome engineering.

Supplemental note S5. Web services for orthogroups and gene families

The identification and phylogenetic analysis of orthogroups and gene families is one of the most effective ways to reveal genomic events related to the diversity of plant clades and important traits (Guo *et al.*, 2020; Jiao et al., 2011; One Thousand Plant Transcriptomes, 2019)., Here, we identified large-scale orthogroups (OGs) and gene families (GFs) with important biological functions across the genomes of Cucurbitales and outgroup. It is particularly important to integrate GS into OGs and GFs for more accurate inference of genomic evolutionary events (Jiao et al., 2014; Wu et al., 2020). Therefore, we constructed OGs and GFs containing GS relationships across Cucurbitales species, which are referred to as Syn-OGs and Syn-GFs, respectively. For these resources, we built user-friendly interactive interfaces Syn-Orthogroups and Syn-Families in the Gene Families portal **(Supplemental Figures 22 and 23)**.

In Syn-Orthogroups interface, we offered the OGs and Syn-OGs and compiled the user-friendly toolbox, enabling users to efficiently explore the genomic events of important traits in Cucurbitales **(Supplemental Figure 22)**. We identified a total of 32,676 OGs contained 736,523 genes, and established the Syn-OGs by integrating the GS relationships across the genomes of 29-CGs and *V. vinifera*. Users can search for the OGs and Syn-OGs according to the gene IDs or PEPs from target species. For each OG, we provided the homologous genes included in the OG, statistical bar charts, associated SBs, phylogenetic tree, and associated genomic and BF annotations. These comprehensive OG resources are beneficial for users to quickly and accurately trace the genetic background of important traits across species, which

has important practical significance for accelerating Cucurbitales crop germplasm improvement.

Given an application scenario. In previous study, scientists discovered that the main switch gene Bt (*Melo3C005611*) for the biosynthesis of cucurbitacin, a bitter substance in *C. melo* fruit (Zhou et al., 2016). Here, comparing the protein sequence of Bt gene with OG database, we found BT gene from OG0000600 **(Supplemental Figure 22)**. This OG contains 102 genes from 29-CGs and *V. vinifera*, of which 98% (100) were supported by orthologous or paralogous GS relationships. According to the orthologous relationship, we found that the Bt gene has an orthologous gene in *V. vinifera*, indicating that the Bt originated at least from the common ancestor of core-eudicots. Along with various scales of gene duplication and loss during the diversification of Cucurbitales, resulting in highly divergent copy numbers of Bt genes in extant species. In the Syn-OG tree, we found that 6 Bt genes were generated from tandem duplication and distributed on very different branches of the tree in *C. melo*, suggesting that these gene duplications before the divergence of Cucurbitales **(Supplemental Figure 23)**. We detected 2 orthologs of Bt gene generated from the genus Cucurbita specific WGD (CST) event in *C. argyrosperma* and *C. pepo*, respectively. It is worth noting that it is difficult to identify the two pairs of duplicates in Cucurbita generated from WGD if only relying on OG without synteny support. Moreover, the genes contained in OG0000600 can be selected as candidate genes for bitter biosynthesis in Cucurbitales plants, and their priority order can be determined according to their genetic relationship with Bt gene.

In Syn-Families interface, we identified the genes of important trait-associated families in Cucurbitales plants, and developed acquisition and visualization analysis modules. In 29-CGs and *V. vinifera*, we identified a total of 501,617 genes from 128 trait-associated families, which were involved the growth and development, biosynthesis of secondary metabolites, biotic and abiotic stress, cold stress, and other adaptive traits **(Supplemental Figure 24A and Supplemental Table 7)**. For each trait-associated gene, we linked its corresponding genomic BF annotations, and GS blocks through the CGRP-specific identifier full-association system. We developed an acquisition module that allows users to search for all information about trait-associate families through any of CGRP-specific identifiers, proteins, and trait descriptions. Specifically, we developed **Phy-Tree** and **Family-Circos** tools for Syn-Families, which enable researchers to study the phylogenetic relationships and expansion patterns of trait-associated family genes. In summary, Syn-Families provides new data resources and analysis tools for mining the genetic basis of specific agronomic traits in Cucurbitales, which is of great significance for targeted improvement of crop breeding.

An application case of the Cytochrome P450 family. In Syn-Families, we searched 110 P450 genes from *C. melo* and constructed the phylogenetic tree using the Phy-Tree **(Supplemental Figure 24B)**. From the P450 tree, we found that the P450 genes were unevenly distributed on 12 chromosomes of *C. melo*, among which the Cme10 and Cme04 have the highest number of genes, with 20 and 17 genes, respectively, and the genes from these two chromosomes generated from CCT event. Using the Family-Circos, we visualized the P450 genes of *C. melo*, and found 4 clusters distributed on the Cme04, Cme07, Cme10, and Cme11 **(Supplemental Figure 25)**. Calculating the *Ks* of gene pairs within each gene cluster, we found that the average

373 *Ks* values of 4 clusters were ~1.41, ~1.36, ~1.38, and ~1.78, respectively, which were likely
374 generated from the common ancestor of Cucurbitaceae. These results also revealed that the
375 WGD and tandem duplications have important genetic contributions to the secondary
376 metabolites of P450 biosynthesis in *C. melo*.

377 **Supplemental note S6. Diverse data resources for the studies of functional genomics**
378 **research and breeding in Cucurbitales comprehensively**

379 Effective management of diverse data resources has important practical value for the
380 development of functional genomics and breeding of Cucurbitales crops. In CGRP, we
381 integrated the species encyclopedias, multi-omics data, ecological and breeding knowledge,
382 research literature, and external databases for Cucurbitales. These resources are stored in six
383 user-friendly web interfaces of the Resource portal.

384 In Species-Encyc interface, we collected encyclopedia information for 130 species from 50
385 genera in 8 families of Cucurbitales **(Supplemental Figure 26A)**. We integrated the Latin
386 names, common names, phylogenetic classifications, karyotypes, genome sizes, and living
387 habits for Cucurbitales species. This interface provides support for users to quickly understand
388 the fundamental background of genome evolution and functional research.

389 In Transcriptome interface, we integrated 7,109 transcriptomes of 53 tissues from 18
390 Cucurbitales plants **(Supplemental Figure 26B and Supplemental Table 8)**. Specifically, we
391 provided 33 transcriptomes of four tissues (root, stem, tuber, and leave) from our newly
392 sequenced *H. chinensis*. These transcriptomes can be downloaded in batches by species and
393 tissue types for gene expression analysis. In addition, we conducted genome-wide gene

expression quantification for 103 transcriptomes of 14 important Cucurbitales crops. Users can search for the gene expression levels (FPKM and TPM values) in specific tissues of Cucurbitales species and use visualization tool in CGRP to present gene expression heat maps. It efficiently supports expression and functional research of genes in Cucurbitaceae crops.

An applicable case of Transcriptome interface. We searched the expression values of four paralogs in *C. maxima* which were generated from CST event **(Supplemental Figure 27)**. Among the three tissues (leaf, stem, and root) of *C. maxima*, we detected three paralogs exhibited divergent expression patterns, suggesting that these CST duplicates experienced neo-functionalization or sub-functionalization. Interestingly, we found that the CST duplicates *Cma01g00047* and *Cma14g01241* have consistent expression patterns, implying that they were played gene dose-response role in *C. maxima*. By searching for BF annotations of this pair duplicates using CGRP-specific identifier, it was found that both of them are related to plant growth hormone regulation. These results showed that the paralogs (*Cma01g00047* and *Cma14g01241*) may serve as candidate targets for artificial improvement in the growth regulation of Cucurbitales crops, and its growth plasticity may be enhanced through dose adjustment.

In Metabolome interface, we established a database containing 5,751 metabolome entries for 85 plants in Cucurbitales, including 119 important metabolites detected by us from the 4 tissues of *H. chinensis* using broad target detection **(Supplemental Figure 26C)**. This interactive interface provides bioinformatics support for researchers to explore important secondary metabolite biosynthesis pathways and associated enzyme genes in Cucurbitales. For example,

by entering Cucurbitacin IIA and selecting *H. chinensis* in the access options of the platform, it returned the detailed annotations, which are expected to include the chemical formula, compounds, numbering of metabolites in Pubchem database (CID), image of compound structure, primary classification of substances (Class I), secondary classification of substances (Class II), CAS registry number (CAS), unique identifier (cpd_ID), and KEGG pathway (kegg_map) **(Supplemental Figure 28)**. Here, no regulatory pathway involving Cucurbitacin IIA was found, indicating that the biosynthesis of this secondary metabolite remains unclear. However, other information about this metabolite provides valuable clues for inferring possible pathways through analog associations. Furthermore, we detected that the regulatory pathway number of cucurbitacin D is ko01110. The candidate genes for cucurbitacin D can be searched through ko01110.

Moreover, we established the Agro-Ecology interface, providing common pests and disease control strategies, economic values, and cultivation techniques for Cucurbitales crops to assist in their agricultural planting **(Supplemental Figure 26D)**. We established the Literature interface with convenient search options, integrating the 19,631 research papers related to Cucurbitales plants published in the past 20 years **(Supplemental Figure 26E)**. Finally, the platform contains 53 external database links related to Cucurbitales in the External Databases interface, allowing users to easily access additional resources across platforms to expand their innovative research and breeding decision **(Supplemental Figure 26F)**.

Supplemental note S7. Empowering CGRP with new mining, continuous updates, and user-friendly

Although CGRP has integrated large-scale genomic resources that have been published and generated by us, new data are always coming. Empowering the platform with new mining capabilities, we built 49 web-based bioinformatics tools (WBBT) in the platform **(Supplemental Figure 29)**. Each tool was configured with an intuitive user interface in WinO-Tools portal and provided easy-to-use application examples.

To support the identification of polyploidization events (API), reconstruction of ancestral genomes (AKR), and exploration of gene family evolution (GFE), the CGRP has built 34 tools related to the three comparative genomics analysis pipelines (CGAPs) proposed by us recently (Yu et al., 2025). These tools include: 1) the Blast (v2.10.1+) (Hu and Kurgan, 2019), Diamond (v0.9.25) (Buchfink et al., 2014), and SequenceServer (v2.0.0) (Priyam et al., 2019) for gene homology alignment (GHA), 2) the MCScanX (Wang et al., 2012), Collinearscan (Wang et al., 2006), and Synvisio (Bandi *et al.*, 2022)) for the detection and visualization of genome synteny, 3) the ParaAT (v2.0) (Zhang et al., 2012), *KaKs*_Calculator (v2.0) (Wang et al., 2010), and our NG-*KaKs*-Cal and Syn_Ks_Dis (Yu *et al.*, 2025) for the calculation of nucleotide variation rate, 4) our B_Dotter, BB_Dotter, and BBK_Dotter (Yu *et al.*, 2025) for genome synteny dotplot drawing, 5) our Paleo-gene identifier, Paleo-gene Rl, Paleo-gene Rll, and P-index-calculator (Yu *et al.*, 2025) for the comparisons of paleogenome fractionation (PFC), 6) the Codonw (v1.4.4) (Shil et al., 2006), CpGfinder (Sujuan et al., 2008), MEME (Bailey et al., 2015), GSDS (v2.0) (Hu et al., 2015), Hmmer (v3.3) (Potter et al., 2018), Pfam (v34.0) (Mistry et al., 2021), FastTree (v2.2) (Price et al., 2009), Phyml (v3.3) (Guindon et al., 2010), Seq alignment and IQ-Tree2 (v2.1.1) (Nguyen et al., 2015) for the analysis of gene structural and functional evolution (GSFE), and 7) the ShinyCircos (Yu et al., 2018) , and our

458 SS-extractor, SAF-converter, BDl-finder, BDl-deduplicator, BDl-extractor, and BDl-combiner
459 (Yu *et al.*, 2025) are used to assist in visualizing and processing data for the three CGAPs (refer
460 to as comprehensive empowerment tools: CETs).

461 Furthermore, we built Interproscan (v5.62) (Quevillon et al., 2005), iTAK (Zheng et al., 2016a),
462 and Dupgene finder (Qiao et al., 2019) for the BF annotation of genes, as well as the
463 identification of RPs and Dupl-types. In the GHA section, we added the Blast match and Seq-
464 fetch script enable large-scale comparison of gene homology. In GS section, we developed the
465 Syn-blocks-stat script to count the overall numbers and genomic locations of SBs and SGPs
466 generated by pairwise GS comparisons. In PFC section, we developed the Syn-deco and Multi-
467 syn-phytree tools to construct MGA and phylogenetic trees for Syn-OGs. In GSFE section, we
468 developed the Heatmap, Family-circos, and Gmapping tools for visualizing gene expressions,
469 gene family expansion patterns, and genomic locations of genes. Finally, we developed the
470 GC-calculator, Seqlenth-calculator, Genstr-counter, and Gam-Calculator in CET section, which
471 are convenient for calculating the sequence feature of genes and genomes.

472 Compared with existing bioinformatics platforms, the WinO-Tools in CGRP have distinct
473 innovations. Firstly, it has built 49 comparative genomics tools, of which 15 are our newly
474 developed scripts. Secondly, these WBBTs are user-friendly, especially for botanists and
475 breeding experts without bioinformatics background. Thirdly, apart from independent
476 operation, these tools can be flexibly customized as universal comparative genomics pipelines
477 (API, AKR, and GFE), as demonstrated in our previous studies (Kong et al., 2023b; Wang et
478 al., 2024; Wang *et al.*, 2018; Wang et al., 2017; Wang et al., 2020; Wang et al., 2019). Especially,

the WGD-associated genome analysis pipeline is considered as the gold standard for deciphering complex genomes (Caspermeyer, 2018; Wang *et al.*, 2018), and their execution steps are detailed in **Supplementary Materials and Methods** and CGRP. Fourthly, these WBBTs and pipelines can be used for the analysis of genomes from CGRP and new genomes of Cucurbitales. Fifthly, to avoid network and computing resources limitations, we developed CGRP-Local-Run toolkit that encapsulates all WBBTs, allowing users to run them locally.

To further enhance the practicality and efficiency of CGRP, we implemented the Bio-Software and Statistics portals **(Supplemental Figure 29)**. In Bio-Software portal, we collected download resources for 38 bioinformatics software and provided detailed installation and operation command lines, as well as common errors during operation. To help users better understand the performance of software and interpret the output results, we provided the references for each software. We also offered the quick links between these software and tools in WinO-Tools portal to facilitate users in choosing and utilizing WBBTs. In Statistics portal, we offered a series of visual statistical charts to assist users in summarizing the overview and biological significance of data resources in the platform. These charts involve the phylogenetic relationships of species in Cucurbitales whose genome have been sequenced, functional elements of genomes, biological annotations of genes, genome synteny, and gene families related to important agronomic traits. The charts also linked to the underlying data sources, allowing for quick querying of target data.

With continuous updates and user-friendly throughout the CGRP, we incorporated the Submit and About portals into the platform **(Supplemental Figure 29)**. The Submit portal encourages

users to submit genomic resources of Cucurbitales plants to CGRP. Data submission should be subject to user agreements and publishing policies to ensure the secure handling of data. Currently, we have submitted our *H. chinensis* genome as a project case. The About portal provides the manual and architectural features of CGRP and instructions of genomic analysis pipelines. We also provided our contact information in this portal for timely communication of various issues related to CGRP in the future.

Switch Page interactive interface for comprehensive exploration of multi-omics resources. To integrate genomic, transcriptomic, and metabolomic resources to study the annotation, evolution, and function of genes, we developed a Switch Page comprehensive search module for CGRP **(Supplemental Figure 30)**. Users can input the CGRP-specific identifiers or protein sequences of genes in Switch Page to explore all associated three types of data resources.

Here, we present a case study of the protein *Csa3G903540*, which catalyzes the conversion of Cucurbitadfenol to 19-hydroxy-cuol in cucumber (Shang et al., 2014). In genomic resources, it is easy to search for the CGRP-specific identifier (*Csa03g04634*), sequence information, and biological functional annotations of this gene **(Supplemental Figure 30A-B)**. All types of duplicates are presented, such as its tandem duplicated gene *Csa03g04635* **(Supplemental Figure 30C)**. According to the gene-associated analyses of syntenic block, *Ks* distribution, and syntenic dotplot, we determined that *Csa03g04634* generated from the Cucurbitales common WGD event (CCT) **(Supplemental Figure 30D)**. In Deco-Alignment, we detected the evolutionary events (WGD and speciation) associated homologs of *Csa03g04634* across Cucurbitales species **(Supplemental Figure 30E)**. For example, the orthologs of *Csa03g04634*

are *Bda11g00301*and *Bda13g00455* in *Begonia darthvaderiana*, which were generated from Begoniaceae common WGD event (BCT). These two genes could be considered as candidate enzyme genes involved in catalyzing the biosynthesis of 19-hydroxy-cuol in *Begonia darthvaderiana*.

In Anc-Genome, we found that *Csa03g04634* originated from a relatively conserved region of chromosome 12 in ancestral genome AcuG-I, which may have laid excellent genetic conditions for the duplication of this gene **(Supplemental Figure 30F)**. In addition, we detected that the orthogroup of *Csa03g04634* is OG0001892, and it belongs to the Cytochrome P450 family. The Syn-Orthogroups and Syn-Families provide the phylogenetic relationship and expansion pattern of this gene in Cucurbitales **(Supplemental Figures 30G-H)**.

Furthermore, the Switch Page also provides data information for *Csa03g04634* at the transcriptome and metabolome levels. The expression patterns of *Csa03g04634* and its CCT duplicated gene (*Csa03g00192*) in different tissues can be easily obtained from Transcriptome module **(Supplemental Figure 30I)**. These paralogs exhibits extremely different expression patterns, suggesting that they have undergone functional differentiation after CCT. We found that *Csa03g04634* has almost identical expression patterns with its orthologs *Cma08g00978* and *Lsi09g01869* in *Cucurbita maxima* and *Lagenaria siceraria*, suggesting that the three genes have similar biological functions. In Metabolome module, we found that the metabolic pathway associated with *Csa03g04634* is ko00996, as well as the associated secondary metabolites and their detailed annotations **(Supplemental Figure 30J)**.

In summary, Switch Page provides a user-friendly tool for integrating multi-omics data resources to explore the evolution and important trait gene functions of Cucurbitales.

Supplemental Materials and Methods

Collection of genomic data

We collected the Latin names of Cucurbitales and associated publications from plaBiPD (Gui et al., 2023), along with the genomes, GFF, and the sequences of CDS and PEP. A total of 185 Cucurbitales genomes were collected from NCBI (https://www.ncbi.nlm.nih.gov/), CuGenDBv2 (Yu et al., 2023), and CNCB (https://www.cncb.ac.cn/) **(Supplemental Table 1)**. Details of these datasets can be searched in the Resource portal of CGRP. For species with multiple genome versions, prioritize selecting the high-quality, latest, or widely used version.

Images and descriptions of plants were mainly obtained from Wikipedia (http://wikipedia.org) and Plants of the World Online (https://powo.science.kew.org/). The literature were extracted from PubMed (https://pubmed.ncbi.nlm.nih.gov/). Cucurbitales transcriptomes were obtained from the SRA database (https://www.ncbi.nlm.nih.gov/sra/). Metabolomes were sourced from the NPASS database (https://bidd.group/NPASS/index.php).

BF annotation of proteins and identification of genomic functional elements

BF annotation of proteins. The InterProScan (v5.57) (Jones et al., 2014) was used to annotate the BFs of proteins by searching the InterPro database (Blum et al., 2021). The GhostKoala (Kanehisa *et al.*, 2016) was employed to annotate the KEGG pathways of proteins.

Identification of TEs. A comprehensive approach combining evidence-based searches and de novo predictions was employed to identify the TEs of Cucurbitales. For evidence-based searches, the RepeatMasker (v4.1.4) (Tarailo‐Graovac and Chen, 2009) was used with default

parameters. For de novo predictions, the RepeatModeler (v2.0.3) (Flynn et al., 2020) was applied to generate a consensus library, and the DeepTE (-sp P) was used to reclassify unknown TEs, improving classification accuracy. Then, the results from both methods were integrated using RepeatMasker (v4.1.4) (Tarailo‐Graovac and Chen, 2009).

Identification of RPs GTDs, and m^6As. The RPs were predicted using iTAK (Zheng *et al.*, 2016a) with default settings. To infer the GTDs, the DupGen_Finder (Qiao *et al.*, 2019) was used with default parameters. The m6A regulators includes three functional groups: writers, readers, and erasers (Yue et al., 2019), each associated with different Pfam families. According to the Pfam IDs PF05063, PF17098, PF18408, and PF15912, four families (MTA70, WTAP, HAKAI, and VIRILIZER) were identified among writers. Using PF04146 and PF13532, the IYT521-B family was identified among readers and erasers.

These methods for genomic BF annotation and element identification have been organized into six commonly used pipelines, and their detailed operational instructions are listed in the About portal of CGRP (https://cucurbitales.cgrpoee.top/documentation).

Inference of genomic synteny

Detection of SBs and calculation of Ka and Ks. The anchored gene pairs within and between genomes were identified using BLASTP (v2.10.1+) (Hu and Kurgan, 2019) with stringent criteria (E-value < 1e-5 and score > 100). The BLASTP matches were then used as input for ColinearScan (Wang *et al.*, 2006) to detect SBs, with the "maximum gap" set to 50 intervening genes. The *Ka* and *Ks* values of SPGs were calculated using the NG_KaKs_Cal (Yu *et al.*, 2025). *Ks* peaks associated with evolutionary events were analyzed using our previously

developed method (Kong et al., 2023a; Wang *et al.*, 2022).

Drawing of GSDs. The GSDs were constructed to visualize anchor gene pairs from BLASTP and SBs within and between genomes. The three types of GSDs were produced using the tools B_Dotter, BB_Dotter, and BBK_Dotter (Yu *et al.*, 2025). For the best matches from BLASTP results were represented by red dots, while the others were shown with blue dots. Each SBs in GSDs contains at least 10 SGPs and integrates BLASTP matching information.

Construction of the MGAs. By integrating GS comparisons and GSD analyses, we identified orthologs and paralogs across the compared genomes. The MGAs were constructed using our previous hierarchical comparison algorithm (Wang *et al.*, 2018). Assuming that the ploidy of the common paleopolyploid for the reference genome and k Cucurbitales genomes is N, the MGAs were built according to the scheme:

$$The\ columns\ of\ MGAs\ = N \times \left(\sum_{i}^{k} \left(i \times MulWGD \right) \right)$$

where the MulWGD is the multiplier of WGD after the *i-th* species is added to MGA after it divergence with the outgroups. For example, applying this formula, we calculated the columns of the MGA with *V. vinifera* as the reference to be:

$$The\ V.\ vinifera\ MGAs\ = 1 \times \left(\sum_{i}^{17} (i \times 2) + \sum_{i}^{10} (i \times 4) + \sum_{i}^{1} (i \times 1) \right) = 225$$

Reconstruction of the ACGs. The construction of ACGs mainly involves two steps: inference of ACKs and ancestral gene content. Regarding the inference of AKs, the detailed pipelines can be found in previous studies (Yu *et al.*, 2025). Based on the constructed MGA, we inferred

the ancestral genes in ACGs. In MGA, if a *V. vinifera* gene has orthologs in Cucurbitales genome, it is considered an ancestral gene in the reconstructed ACGs. Based on this criterion, the 23,647 genes were inferred in the MRCA-G. Using the same method, we constructed the other MGAs, and inferred the ancestral gene content in the genomes of ACoG-I, CACB-G, ABG-I, ACuG-I, and ACuG-III.

Then, the animations of ancestral karyotypes and their evolution were created with Adobe animation (Teplá and Klímová, 2015).

Identification of Syn-OGs and of GFs across Cucurbitales

Identification of the Syn-OGs. Orthologous and paralogous genes across the Cucurbitales were identified using OrthoFinder (v2.5.1) (Emms and Kelly, 2019) with protein similarity assessed by BLASTP (v2.10.1+) (E-value < 1e-5, score > 100) (Hu and Kurgan, 2019). Clustering was performed using the Markov clustering algorithm (MCL) with an inflation value of 4.5. The phylogenetic trees of OGs was then constructed using the neighbor-joining (NJ) algorithm implemented in FastTree (v2.2) (Price *et al.*, 2009). The Syn-OGs were constructed by searching for the GS database in platform.

Identification of GFs. The query sequences for 128 GFs were downloaded from the TAIR database (https://www.arabidopsis.org/). Gene families in each genome were identified using BLASTP (v2.10.1+) (Hu and Kurgan, 2019) and Pfam (v34.0) (Mistry *et al.*, 2021), with the thresholds of E-value < 1e-3 and score > 100.

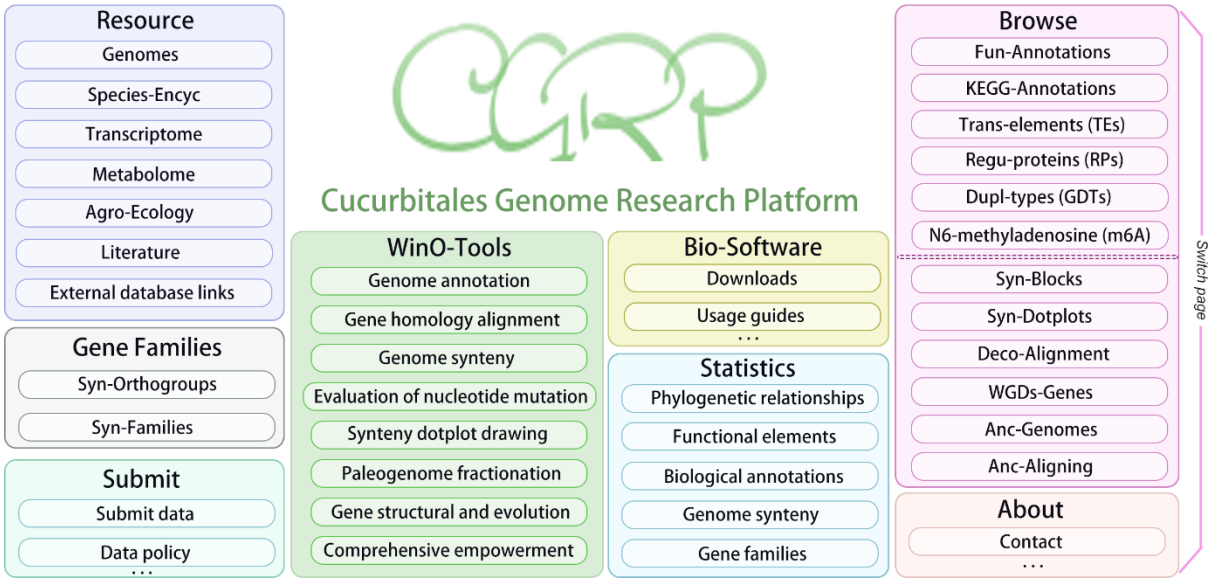
Transcriptome analysis

To quantify gene expression levels, we used Trimmomatic (v0.38) (Bolger et al., 2014) to ensure transcriptome quality. RNA-seq reads were mapped to the reference genome using Hisat2 (v2.1.0) (Kim et al., 2019) and samtools (v1.12) (Li et al., 2009). Finally, the quantification of gene expressions was performed using the StringTie (v2.1.5) (Pertea et al., 2016) with the parameters -e and -A.

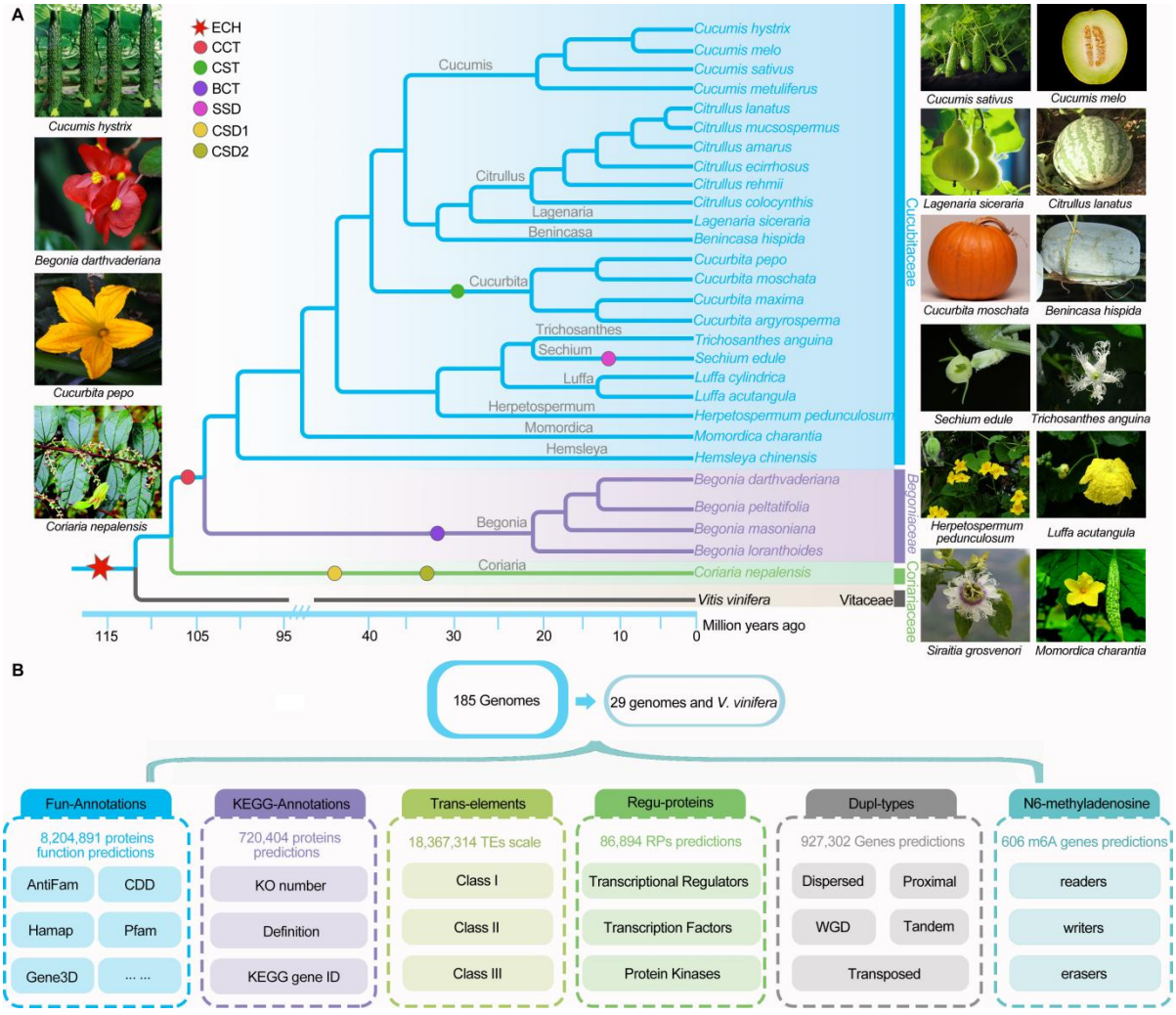
Platform construction

As an open web server, CGRP efficiently manages data stored in MySQL using scripts and software. The frontend is developed with technologies including the HTML (https://www.w3.org), CSS (https://www.w3.org), JavaScript (https://developer.mozilla.org/), Echarts (https://echarts.apache.org/), and Bootstrap (https://getbootstrap.com/). The backend is built on the Python-based Django v4.2 framework (https://www.djangoproject.com). The application layer supports users to upload data, with both the frontend and backend processing in parallel to provide real-time feedback on software execution, enabling the efficient mining and integration of large-scale biological data.

Supplemental Figures 1-30

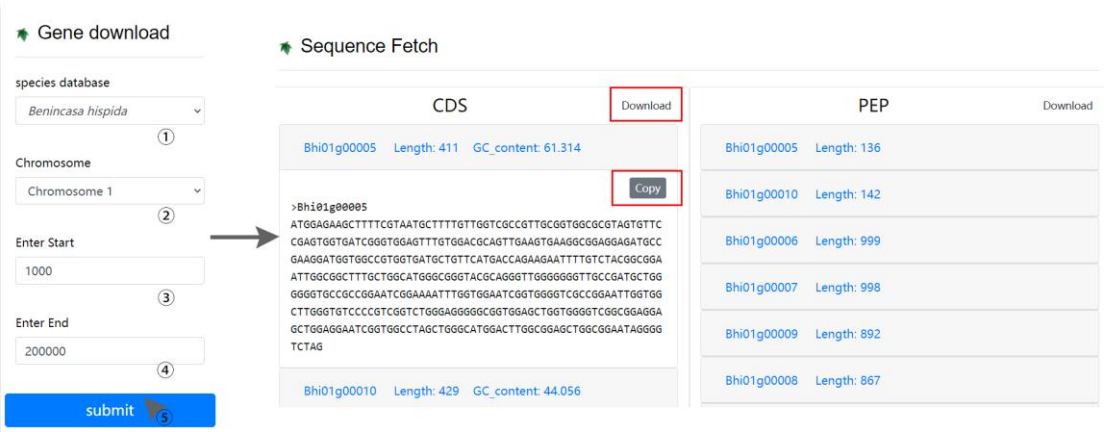


Supplemental Figure 1. The portals of CGRP and their contents.

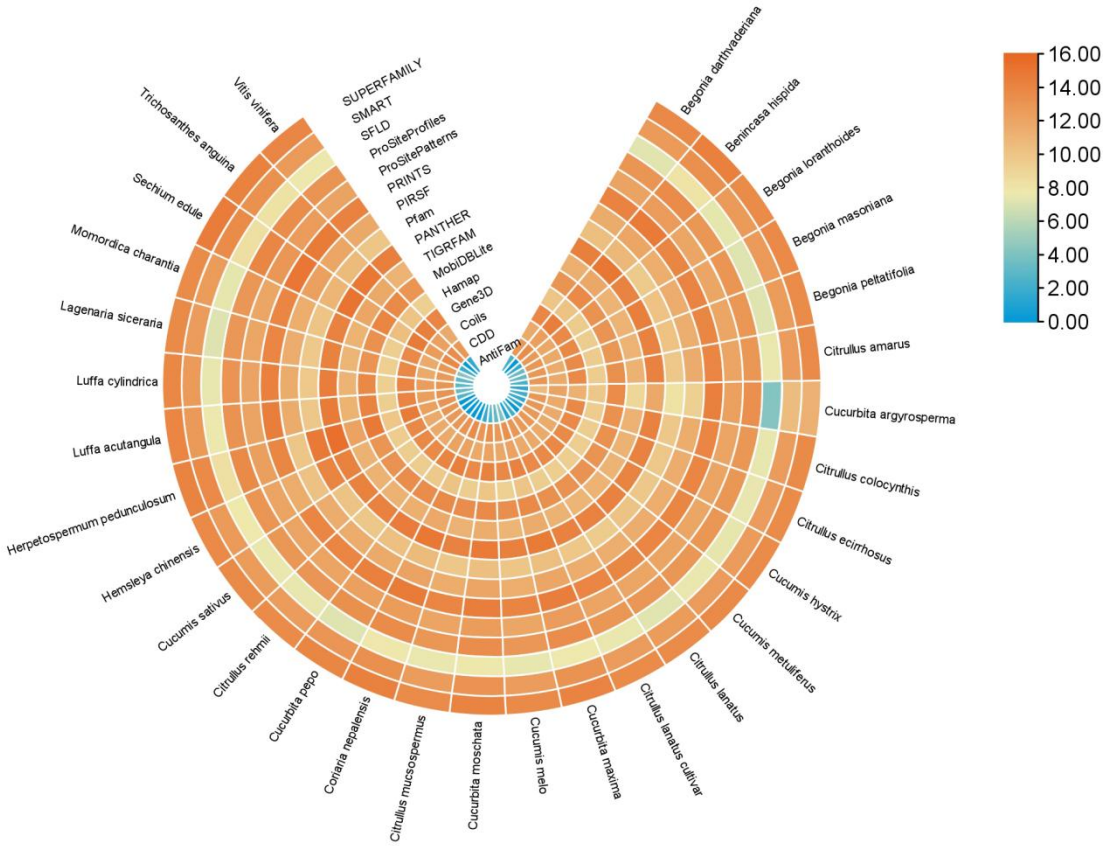


Supplemental Figure 2. Phylogenetic relationships and genomic resources of sequenced Cucurbitales

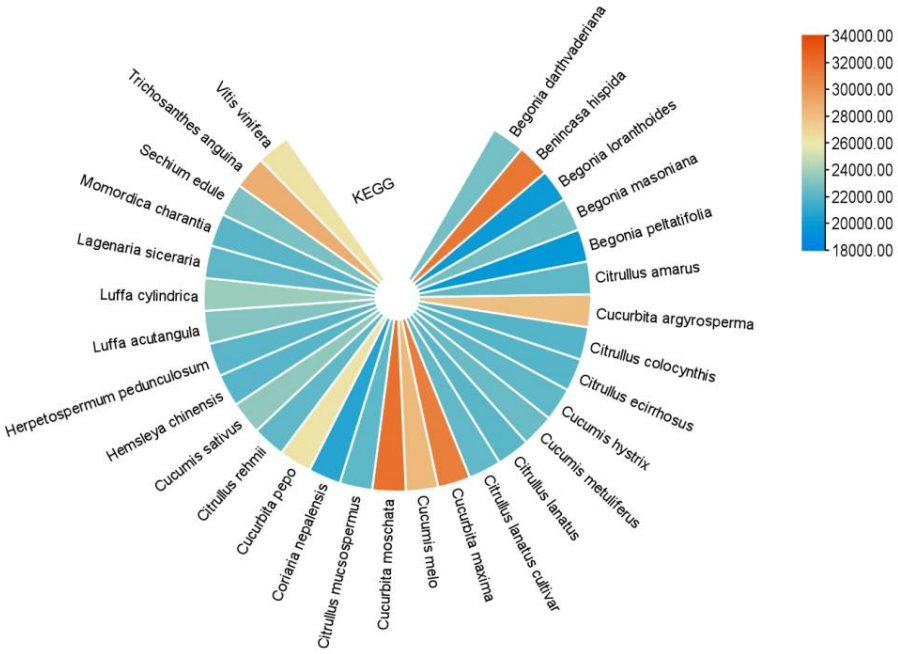
and outgroup species. (A) Phylogenetic tree of 29 Cucurbitales plants and *V. vinifera*, as well as the images of representative species from the main branches of Cucurbitales. The polyploidization events were marked on evolutionary nodes, including the core eudicot common hexaploidization (ECH), Cucurbitales common tetraploidization (CCT), Cucurbita-specific tetraploidization (CST), Begoniaceae common tetraploidization (BCT), Sechium-specific duplication (SSD), and Coriariaceae specific duplication 1 and 2 (CSD1 and CSD2). **(B)** Genomic resources in CGRP.



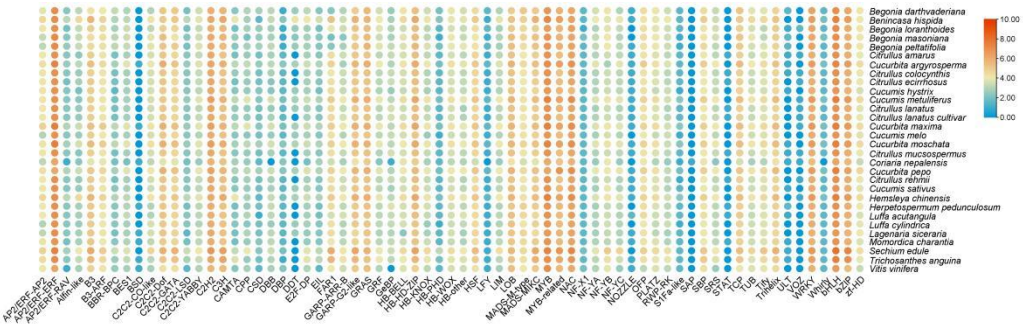
Supplemental Figure 3. Gene download interface and operation examples. Users can select genes from specific chromosome region of target species for batch download.



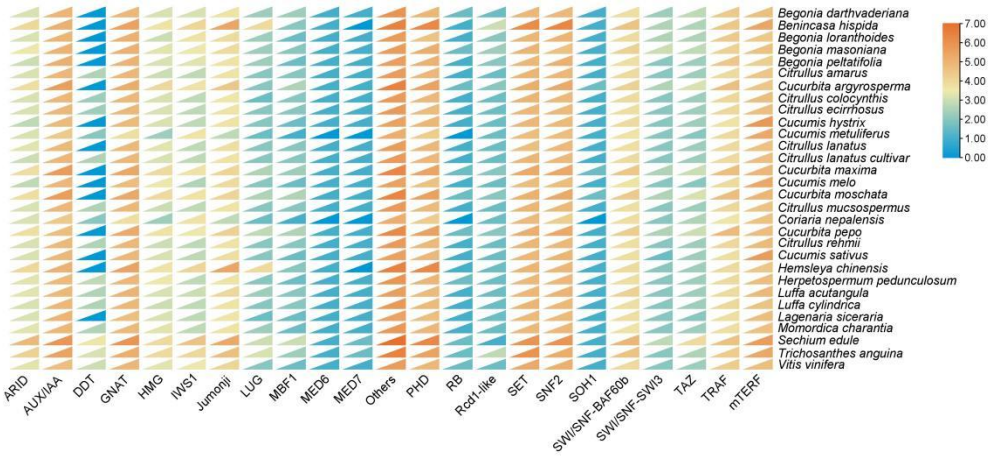
Supplemental Figure 4. The statistical heatmap of biological function annotations of proteins in the genomes of 29-CGs and *V. vinifera*. Each square represents the number of gene annotations in each genome. These gene annotations have undergone log10 consensus conversion.



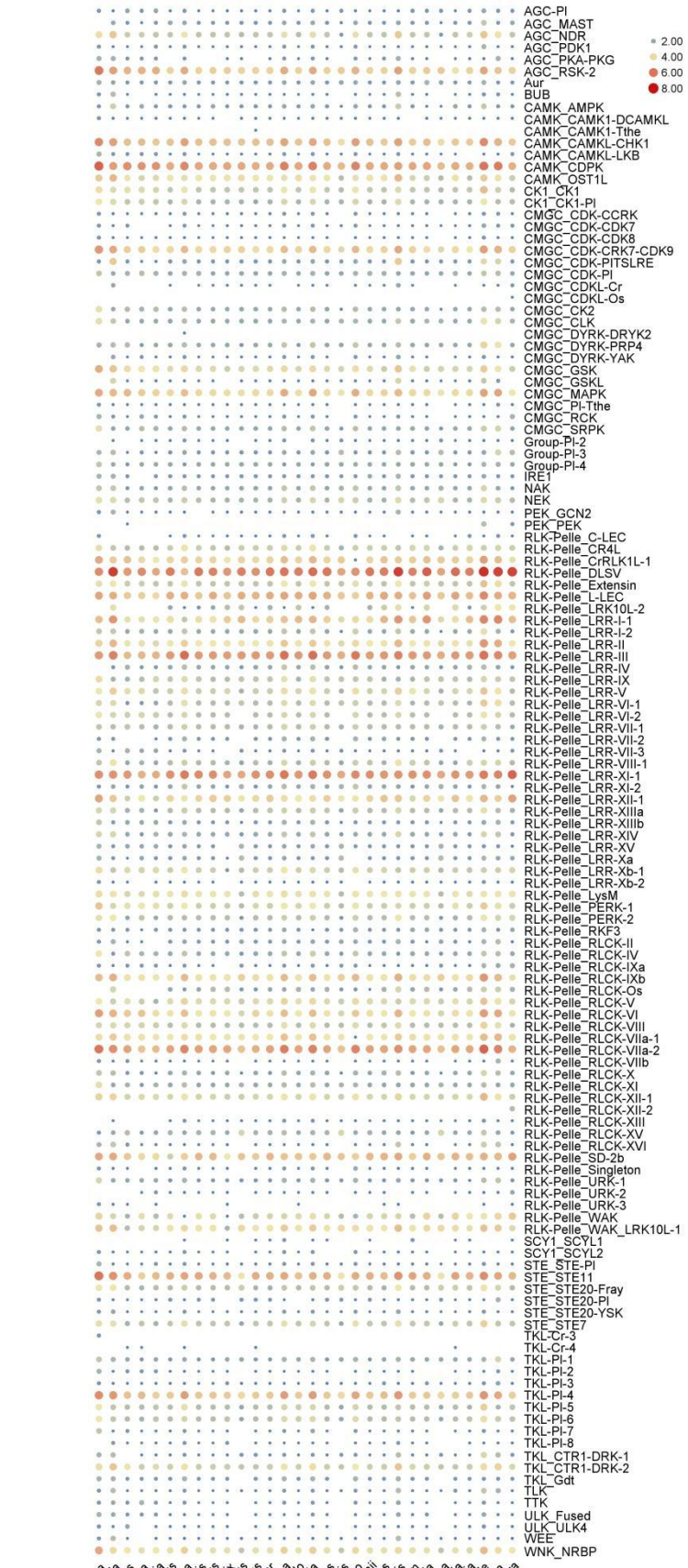
Supplemental Figure 5. The statistical heatmap of KEGG pathways in the genomes of 29-CGs and *V. vinifera*. Each square represents the number of KEGG pathways in each genome. These numbers of KEGG pathways have undergone log10 consensus conversion.



Supplemental Figure 6. The statistical heatmap of transcription factors (TFs) in the genomes of 29-CGs and *V. vinifera*. Each dot represents the number of RPs identified in each genome. The counts of RPs in each genome have undergone log10 consensus conversion.

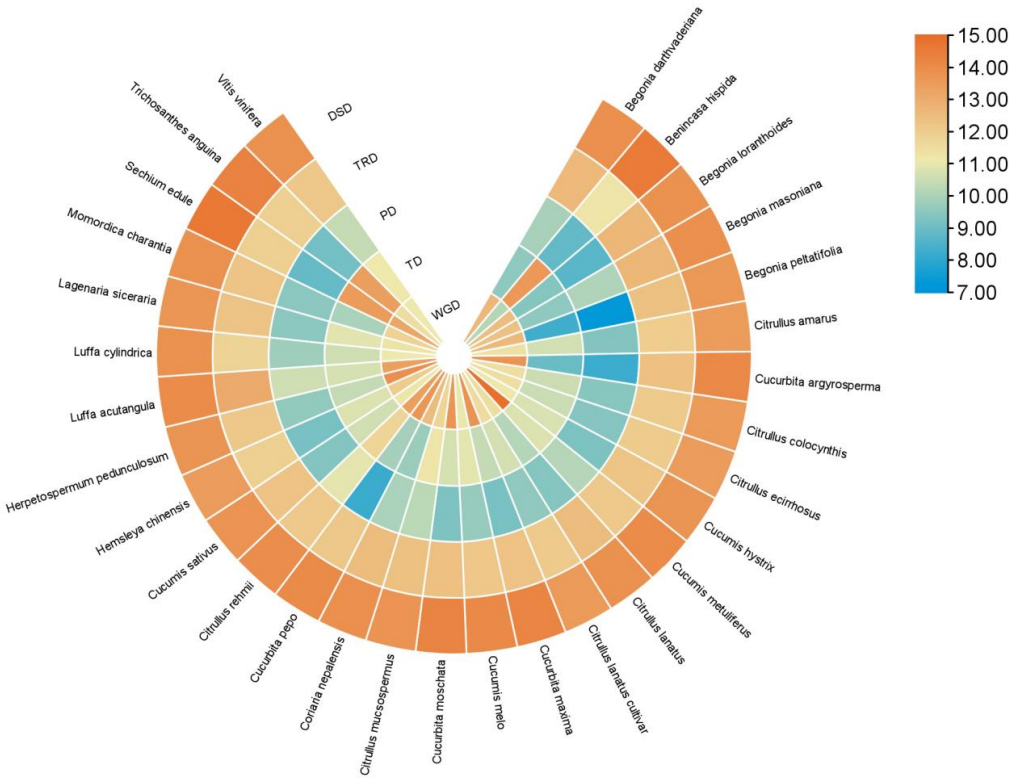


Supplemental Figure 7. The statistical heatmap of transcription regulators (TRs) in the genomes of 29-CGs and *V. vinifera*. Each triangle represents the number of identified TRs in each genome. These counts of TRs have undergone log10 consensus conversion.

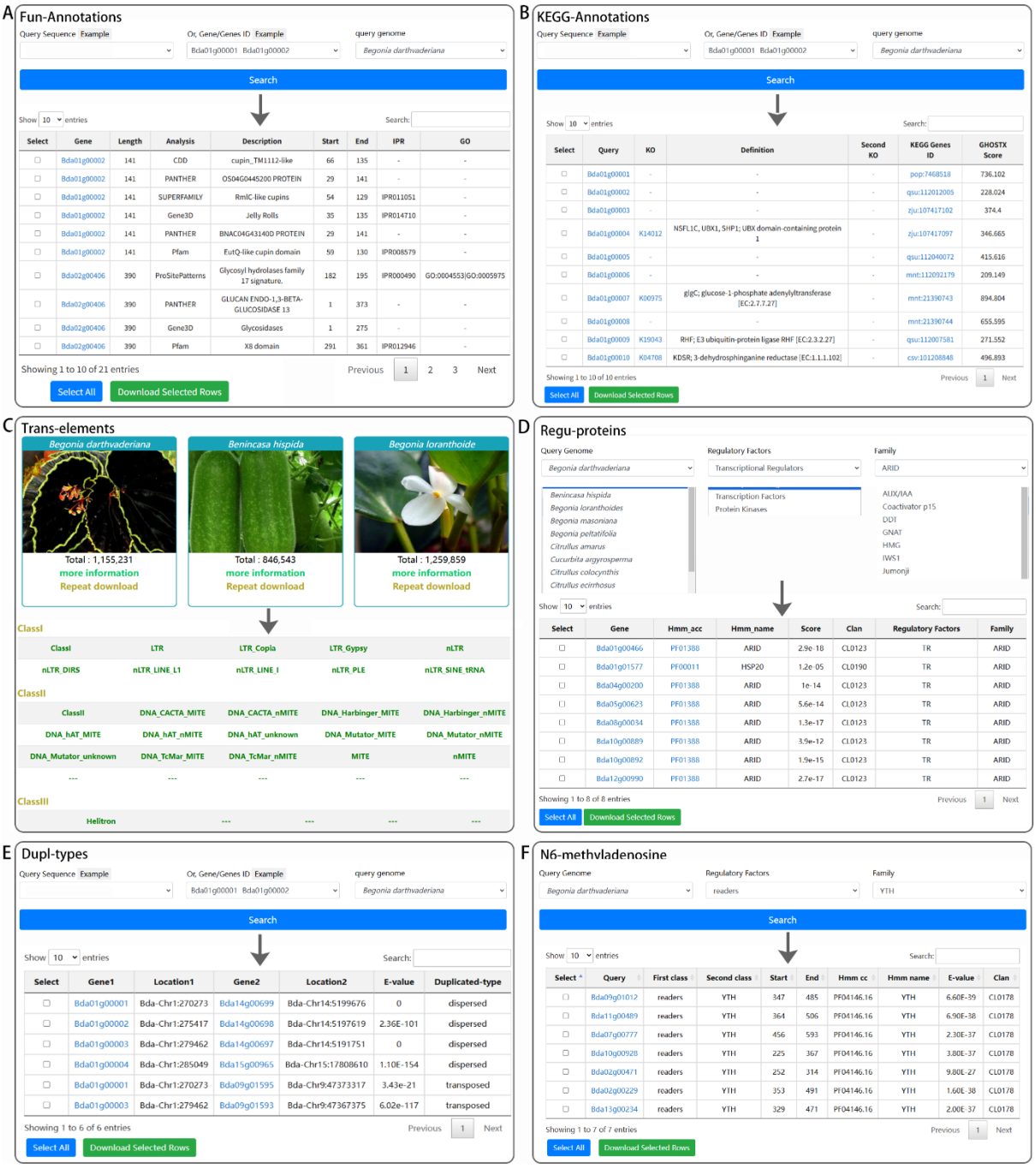

AGC-PI
AGC_MAST
AGC_NDR
AGC_PDK1
AGC_PKA-PKG
AGC_RSK-2
Aur
BUB
CAMK_AMPK
CAMK_CAMK1-DCAMKL
CAMK_CAMK1-Tthe
CAMK_CAMKL-CHK1
CAMK_CAMKL-LKB
CAMK_CDPK
CAMK_OST1L
CK1_CK1
CK1_CK1-PI
CMGC_CDK-CCRK
CMGC_CDK-CDK7
CMGC_CDK-CDK8
CMGC_CDK-CRK7-CDK9
CMGC_CDK-PITSLRE
CMGC_CDK-PI
CMGC_CDKL-Cr
CMGC_CDKL-Os
CMGC_CK2
CMGC_CLK
CMGC_DYRK-DRYK2
CMGC_DYRK-PRP4
CMGC_DYRK-YAK
CMGC_GSK
CMGC_GSKL
CMGC_MAPK
CMGC_PI-Tthe
CMGC_RCK
CMGC_SRPK
Group-PI-2
Group-PI-3
Group-PI-4
IRE1
NAK
NEK
PEK_GCN2
PEK_PEK
RLK-Pelle_C-LEC
RLK-Pelle_CR4L
RLK-Pelle_CrRLK1L-1
RLK-Pelle_DLSV
RLK-Pelle_Extensin
RLK-Pelle_L-LEC
RLK-Pelle_LRK10L-2
RLK-Pelle_LRR-I-1
RLK-Pelle_LRR-I-2
RLK-Pelle_LRR-II
RLK-Pelle_LRR-III
RLK-Pelle_LRR-IV
RLK-Pelle_LRR-IX
RLK-Pelle_LRR-V
RLK-Pelle_LRR-VI-1
RLK-Pelle_LRR-VI-2
RLK-Pelle_LRR-VII-1
RLK-Pelle_LRR-VII-2
RLK-Pelle_LRR-VII-3
RLK-Pelle_LRR-VIII-1
RLK-Pelle_LRR-XI-1
RLK-Pelle_LRR-XI-2
RLK-Pelle_LRR-XII-1
RLK-Pelle_LRR-XIIIa
RLK-Pelle_LRR-XIIIb
RLK-Pelle_LRR-XIV
RLK-Pelle_LRR-XV
RLK-Pelle_LRR-Xa
RLK-Pelle_LRR-Xb-1
RLK-Pelle_LRR-Xb-2
RLK-Pelle_LysM
RLK-Pelle_PERK-1
RLK-Pelle_PERK-2
RLK-Pelle_RKF3
RLK-Pelle_RLCK-II
RLK-Pelle_RLCK-IV
RLK-Pelle_RLCK-IXa
RLK-Pelle_RLCK-IXb
RLK-Pelle_RLCK-Os
RLK-Pelle_RLCK-V
RLK-Pelle_RLCK-VI
RLK-Pelle_RLCK-VIII
RLK-Pelle_RLCK-VIIa-1
RLK-Pelle_RLCK-VIIa-2
RLK-Pelle_RLCK-VIIb
RLK-Pelle_RLCK-X
RLK-Pelle_RLCK-XI
RLK-Pelle_RLCK-XII-1
RLK-Pelle_RLCK-XII-2
RLK-Pelle_RLCK-XIII
RLK-Pelle_RLCK-XV
RLK-Pelle_RLCK-XVI
RLK-Pelle_SD-2b
RLK-Pelle_Singleton
RLK-Pelle_URK-1
RLK-Pelle_URK-2
RLK-Pelle_URK-3
RLK-Pelle_WAK
RLK-Pelle_WAK_LRK10L-1
SCY1_SCY1
SCY1_SCYL2
STE_STE-PI
STE_STE11
STE_STE20-Fray
STE_STE20-PI
STE_STE20-YSK
STE_STE7
TKL-Cr-3
TKL-Cr-4
TKL-PI-1
TKL-PI-2
TKL-PI-3
TKL-PI-4
TKL-PI-5
TKL-PI-6
TKL-PI-7
TKL-PI-8
TKL_CTR1-DRK-1
TKL_CTR1-DRK-2
TKL_Gdt
TLK
TTK
ULK_Fused
ULK_ULK4
WEE
WNK_NRBP
2.00
4.00
6.00
8.00
Begonia darthvaderiana
Benincasa hispida
Begonia loranthoides
Begonia masoniana
Begonia peltatifolia
Cucurbita argyrosperma
Citrullus amarus
Citrullus colocynthis
Citrullus ecirrhosus
Cucumis melo
Cucumis myriocarpus
Citrullus lanatus
Cucumis sativus
Cucurbita maxima
Cucurbita moschata
Cucumis melo
Citrullus mucospermus
Coniaria repandensis
Cucurbita pepo
Cucumis hystrix
Cucumis sativus
Herpetospermum pedunculosum
Hensleya chinensis
Luffa acutangula
Luffa cylindrica
Lagenaria siceraria
Momordica charantia
Sechium edule
Trichosanthes anguina
Vitis vinifera

671

Supplemental Figure 8. The statistical heatmap of protein kinases (PKs) in the genomes of 29-CGs and *V. vinifera*. Each triangle represents the number of identified TRs in each genome. These counts of TRs have undergone log10 consensus conversion.

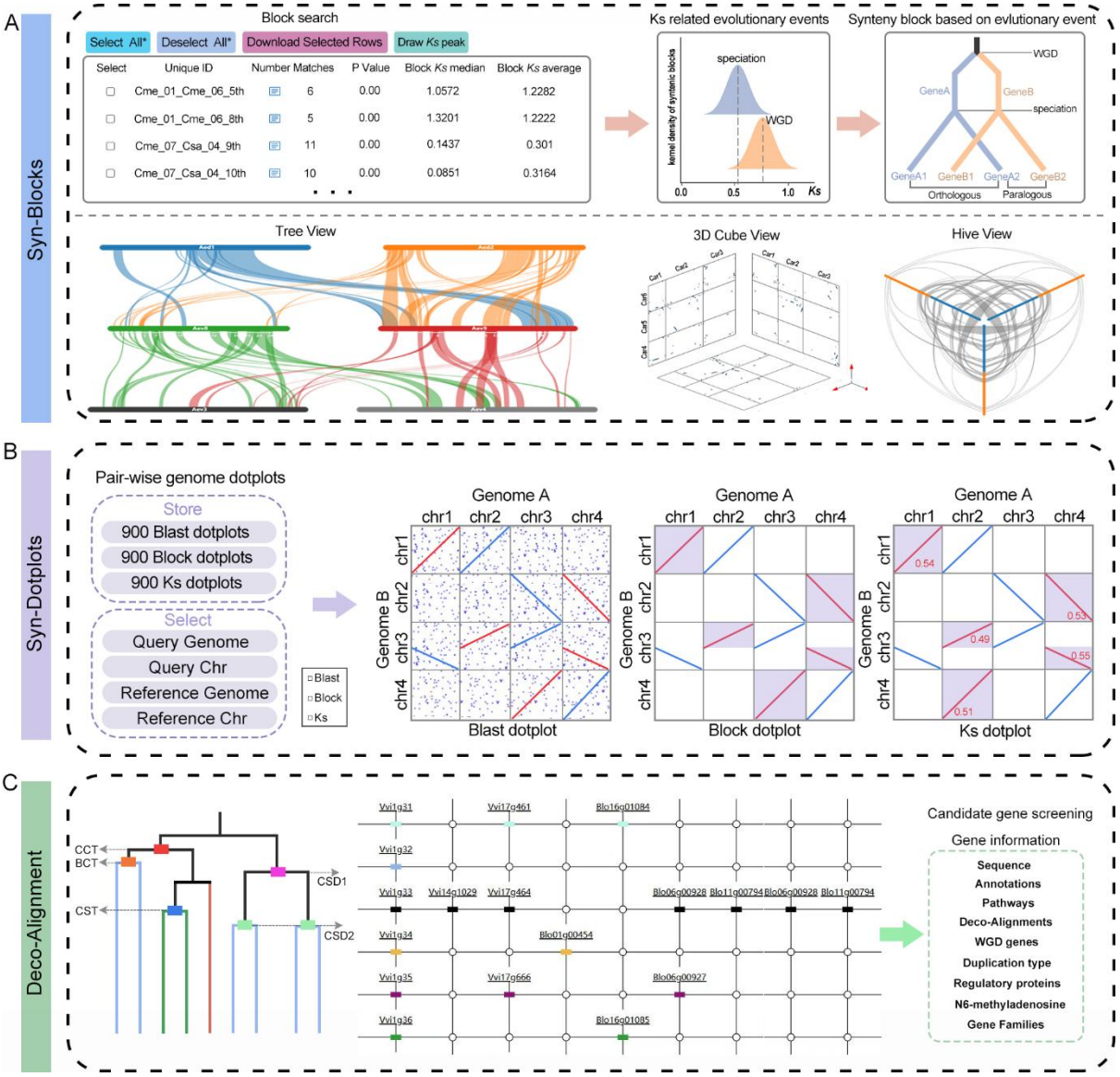


Supplemental Figure 9. The statistical heatmap of gene duplications in the genomes of 29-CGs and *V. vinifera*. Each square represents the number of gene duplications in each genome. These gene annotations have undergone log10 consensus conversion.



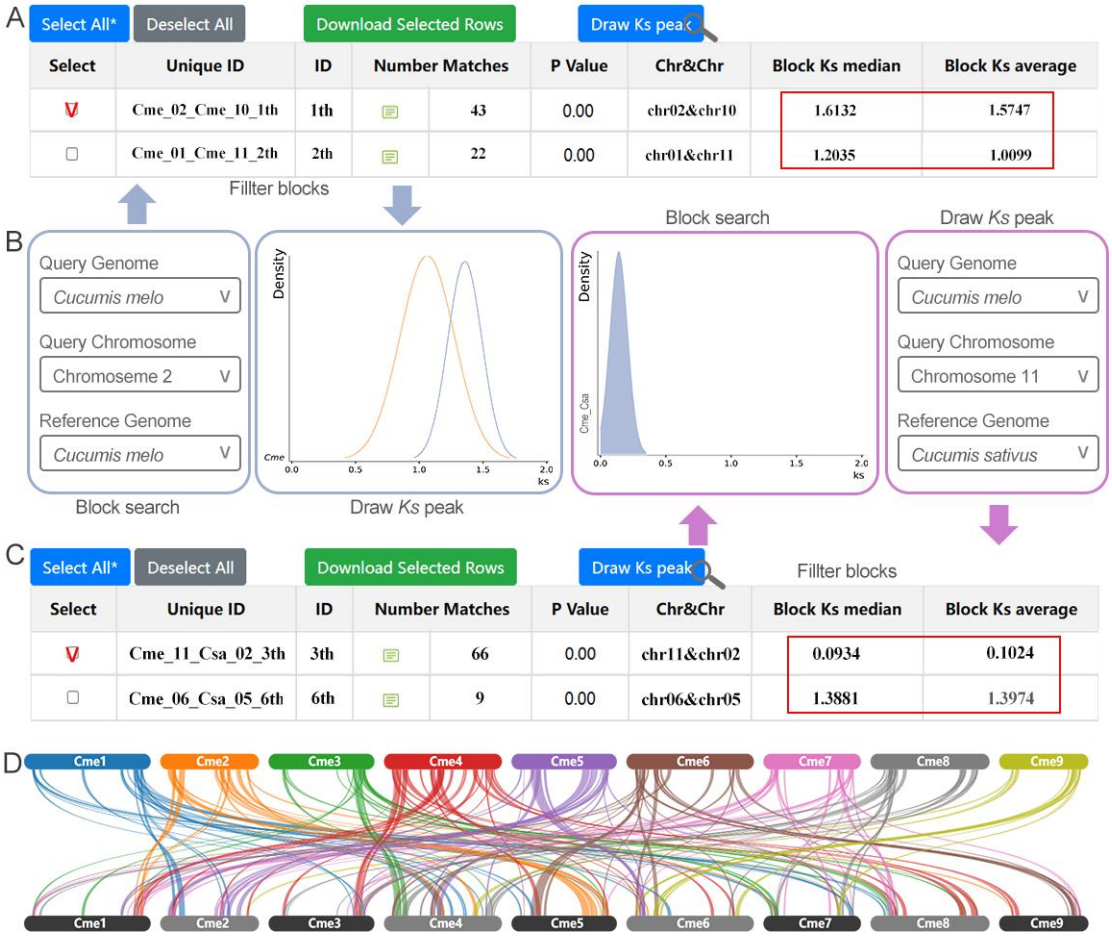
679

680 **Supplemental Figure 10. Six interactive interfaces for genome annotations and functional element**
681 **resources in Browse portal. (A)** Fun-Annotations interface and operation examples. **(B)** KEGG-
682 Annotations interface and operation examples. **(C)** Trans-elements interface and operation examples. **(D)**
683 Regu-proteins interface and operation examples. **(E)** Dupl-types interface and operation examples. **(F)** N6-
684 methyladenosine interface and operation examples.

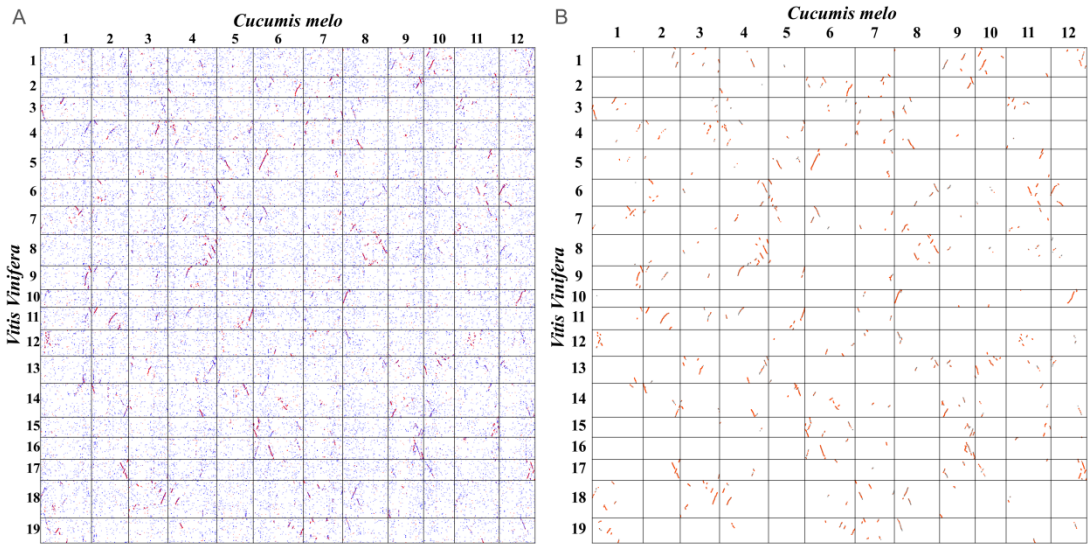


685

686 **Supplemental Figure 11. High-value interfaces for genome synteny and ancestral genomes. (A)** The
687 Syn-Blocks interface includes synteny blocks and visualization for the genomes of Cucurbitales plants. **(B)**
688 The Syn-Dotplots interface includes pairwise genome synteny dotplots and visualization analysis functions.
689 **(C)** The Deco-Alignments interface stores multiple genome alignment and was equipped with powerful
690 browsing capabilities to explore the origin and evolution of target genes.

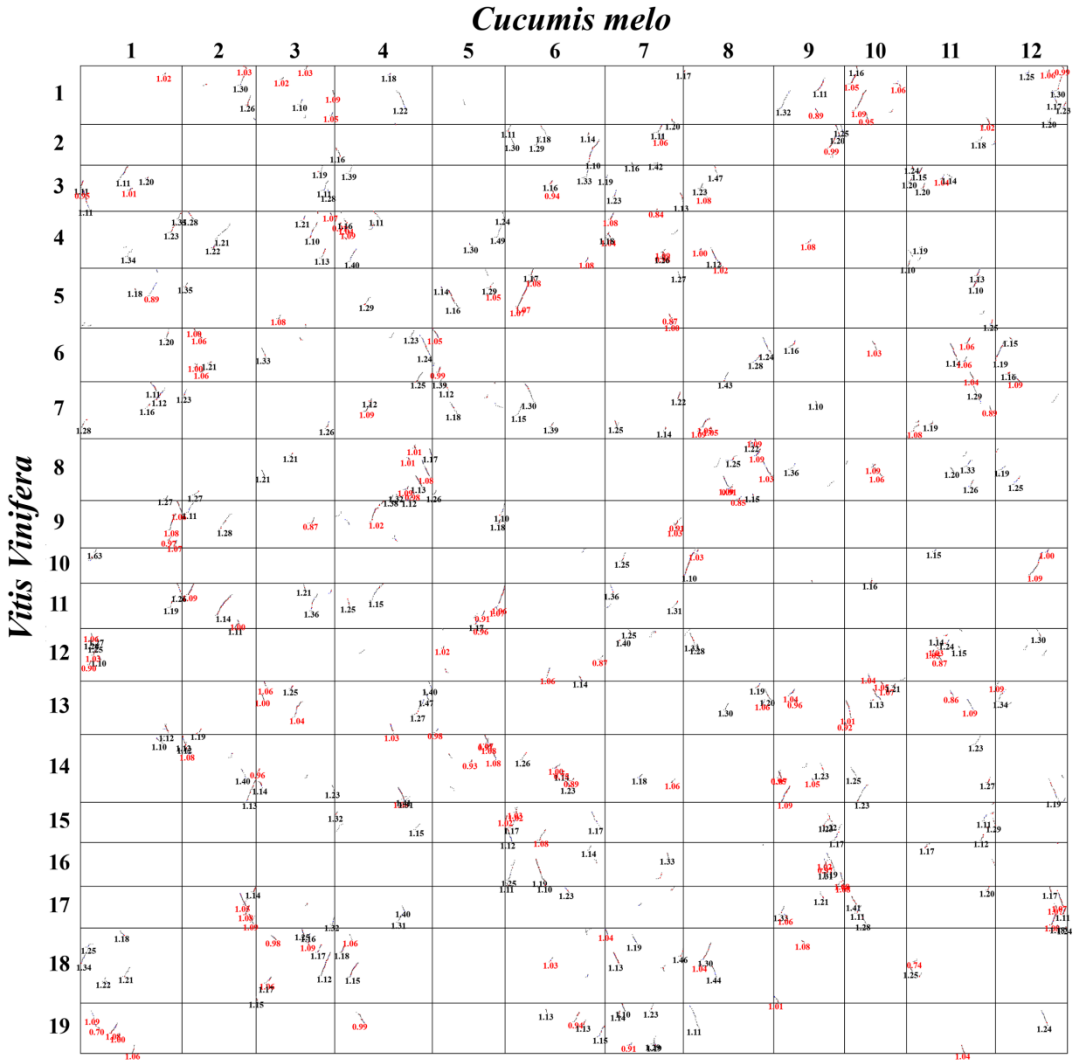


Supplemental Figure 12. Synteny comparisons across the genomes of *Cucumis melo* and *Cucumis sativus* in Syn-Blocks interface. (A-C) Workflow diagrams for the 'Block search', 'Filter blocks', and 'Draw *Ks* peaks' options. In Syn-Blocks interface, we selected the SBs within the *Cucumis melo* to draw *Ks* peaks. From the *Ks* distribution, we inferred that the *Cucumis melo* experienced two WGD events, with the ECH event has *Ks* peak at ~1.32 and the CCT event has *Ks* peak at ~1.05. **(D)** For syn-blocks, we built the SynView tool, allowing researchers to freely visualize the specific SBs in their preferred format.



698

699 **Supplemental Figure 13. Genomic comparisons between the *V. vinifera* and *C. melo* genomes. (A)** The
700 dotplot of BLASTP homologous gene pairs between the genomes of *V. vinifera* and *C. melo* showed an
701 orthology ratio of 1:2. **(B)** The synteny dotplot between the genomes of *V. vinifera* and *C. melo* showed an
702 orthology ratio of 1:2. Genomic syntenic blocks (≥10 gene pairs) inferred from ColinearScan were shown in
703 dotplot according to the genomic locations of *V. vinifera* and *C. melo*. Best-hit genes are represented by red
704 dots, other genes by gray dots.



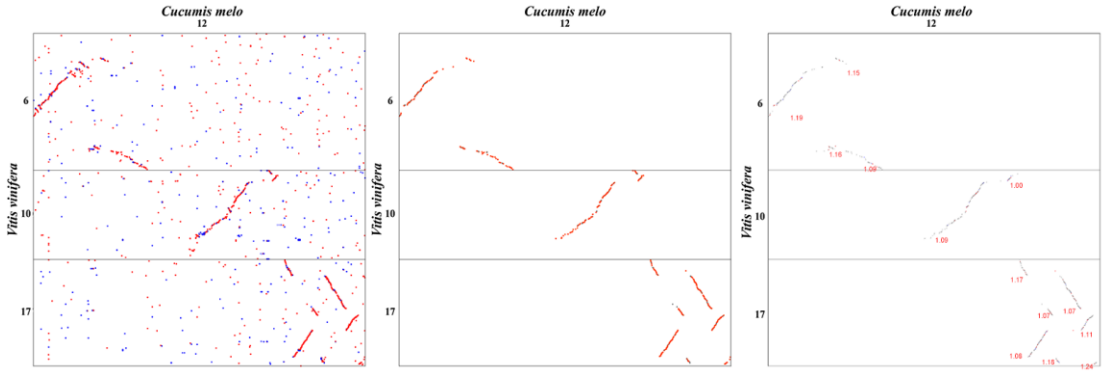
705

706 **Supplemental Figure 14. Intergenomic comparison analyses of the *V. vinifera* and *C. melo* genomes.**

707 The best and other matched homologous gene pairs were shown by red and gray dots, respectively. Median

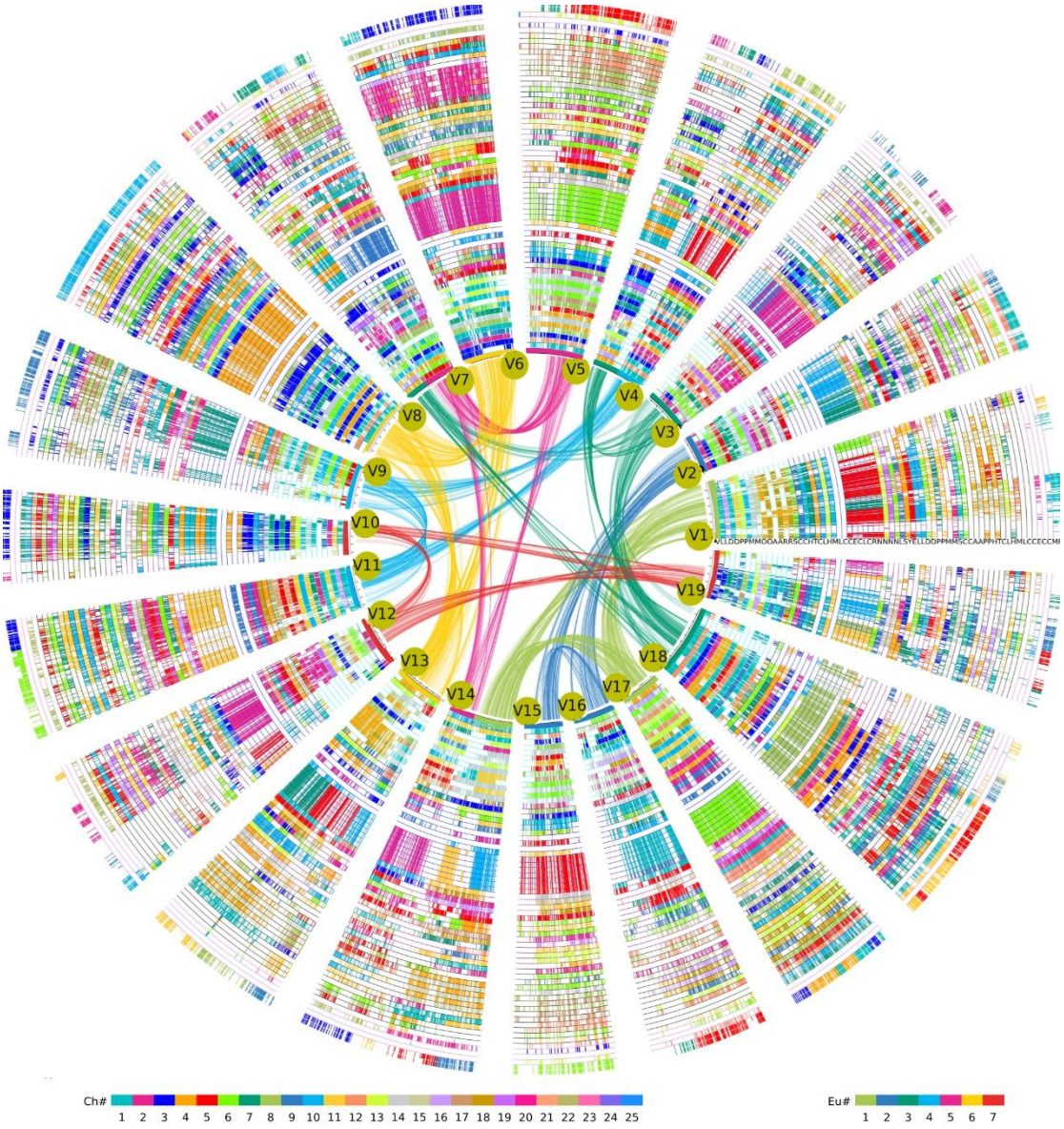
708 *Ks* values of each inferred syntenic blocks were exhibited near their corresponding regions. The *Ks* values

709 <1.05 were in red, and others are in black.



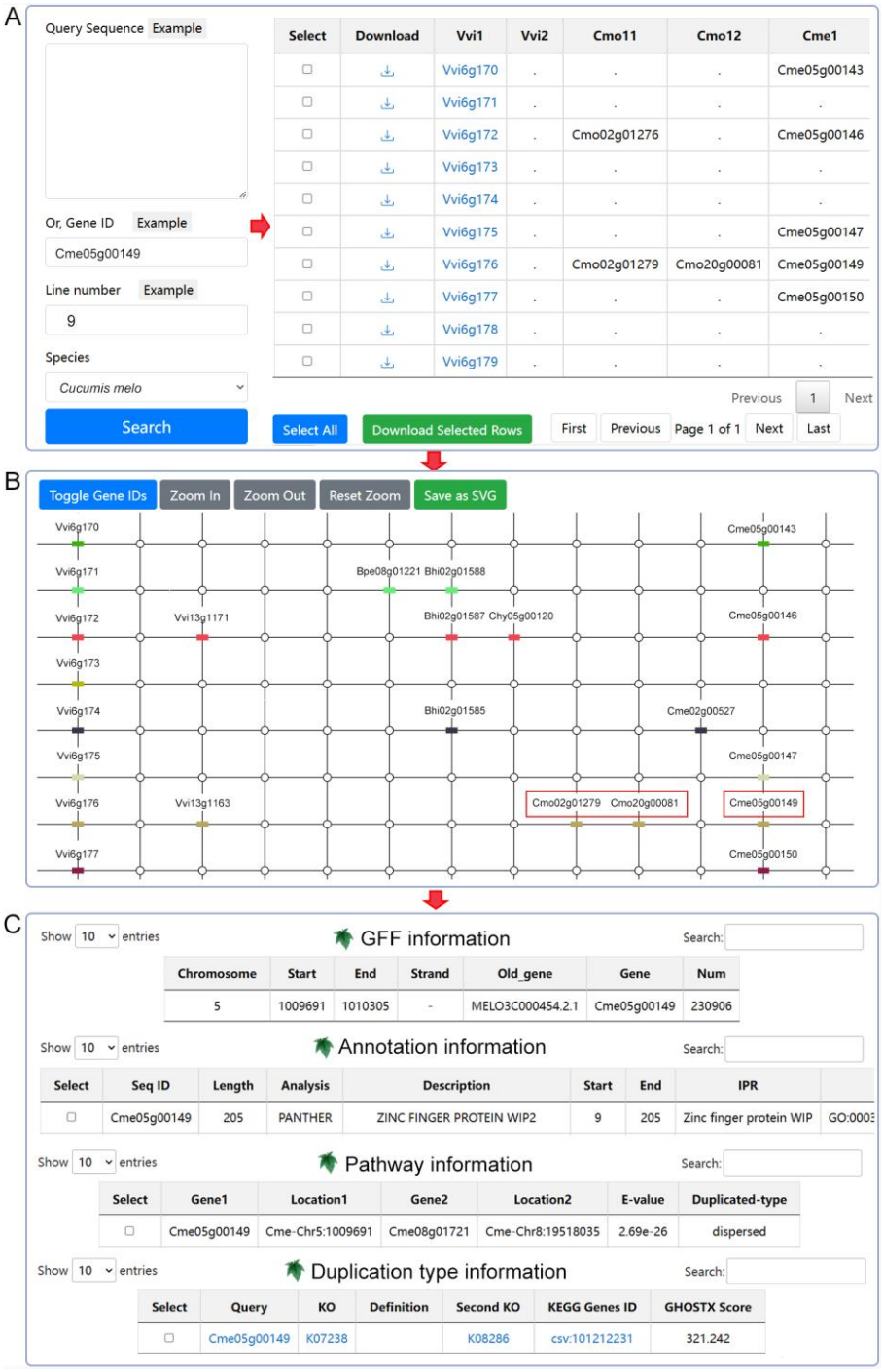
710

Supplemental Figure 15. Genomic synteny dotplots generated from the Pair-wise dotplots interface. **(A)** The local BALSTP dotplot comparisons between *V. vinifera* and *C. melo* shows that chromosome 12 of *C. melo* was fused by the chromosome 6,10 and 12 of *V. vinifera*. **(B)** The local block dotplot shows the syntenic gene pairs. **(C)** The local *Ks* dotplot shows the *Ks* values of blocks around 1.07, suggesting that these blocks produced by the divergence of *V. vinifera* and *C. melo*.



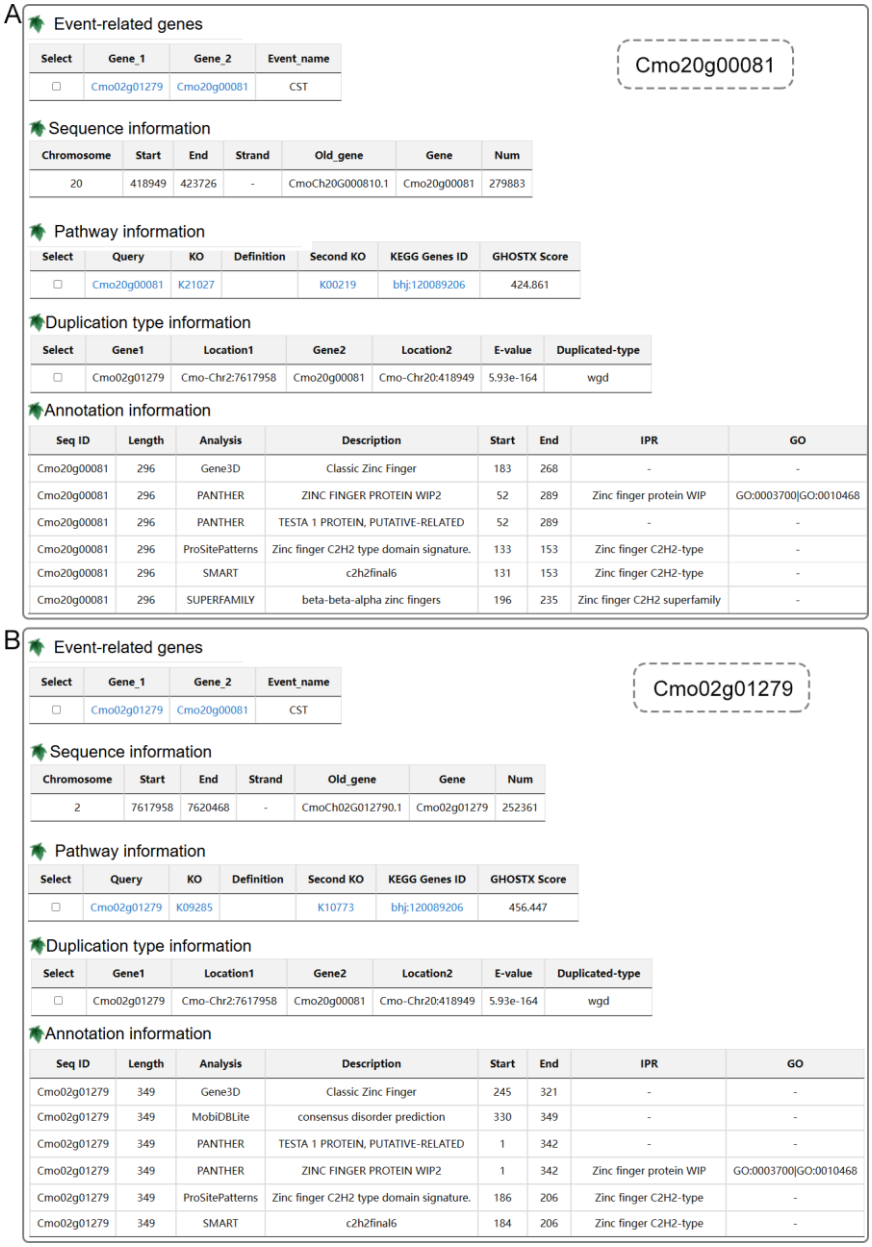
Supplemental Figure 16. Multiple genome alignment across Cucurbitales species with *V. vinifera* as reference. The *V. vinifera* chromosomes form the innermost circle, and their paralogous genes generated from core eudicot hexaploidization were linked by curves, which were colored according to the 7 ancestral eudicot chromosomes inferred previously (Wang *et al.*, 2018; Wang *et al.*, 2022). This alignment displayed in 75 circles, and each circle represents one subgenome related to the polyploidizations as shown in Figure

2. Short lines between two circles shows syntenic genes, corresponding to the colors of genes from which chromosome. The colors of chromosomes were shown at the bottom.

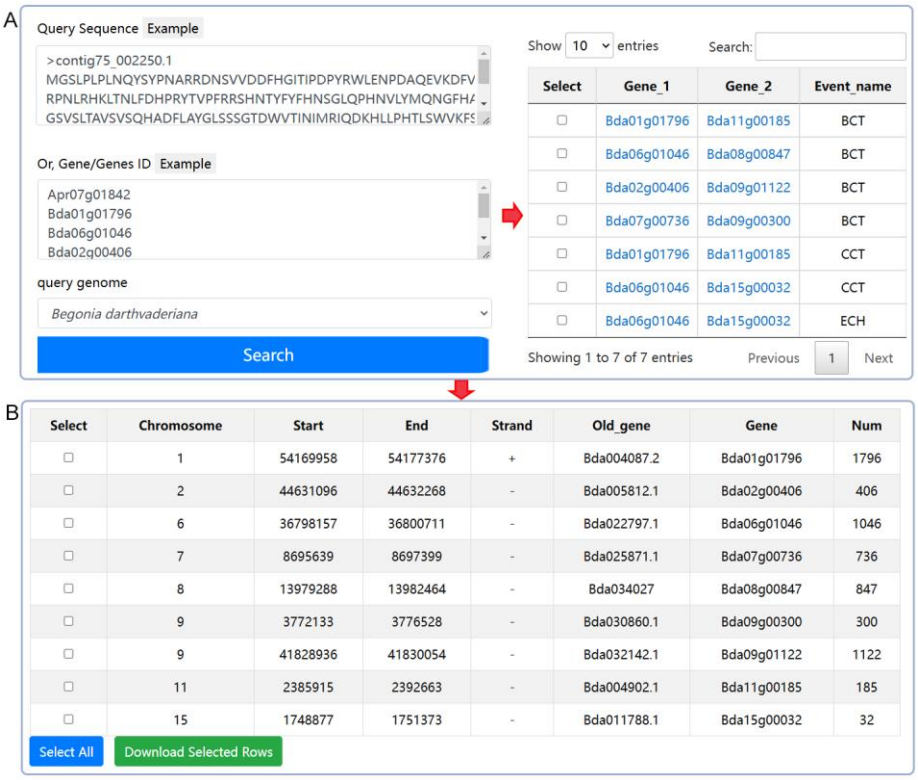


Supplemental Figure 17. Searching for genes in the Deco-Alignments interface. The CmWIP1 gene was identified in *Cucumis melo* genome. (A) In the input page of the "Deco-Alignments" interface, the BLAST alignment shows that the CmWIP1 gene in *C. melo* genome is *Cme05g00149*. The multiple genome alignment shows that the orthologous genes in *Cucurbita moschata* genome is *Cmo02g01279* and

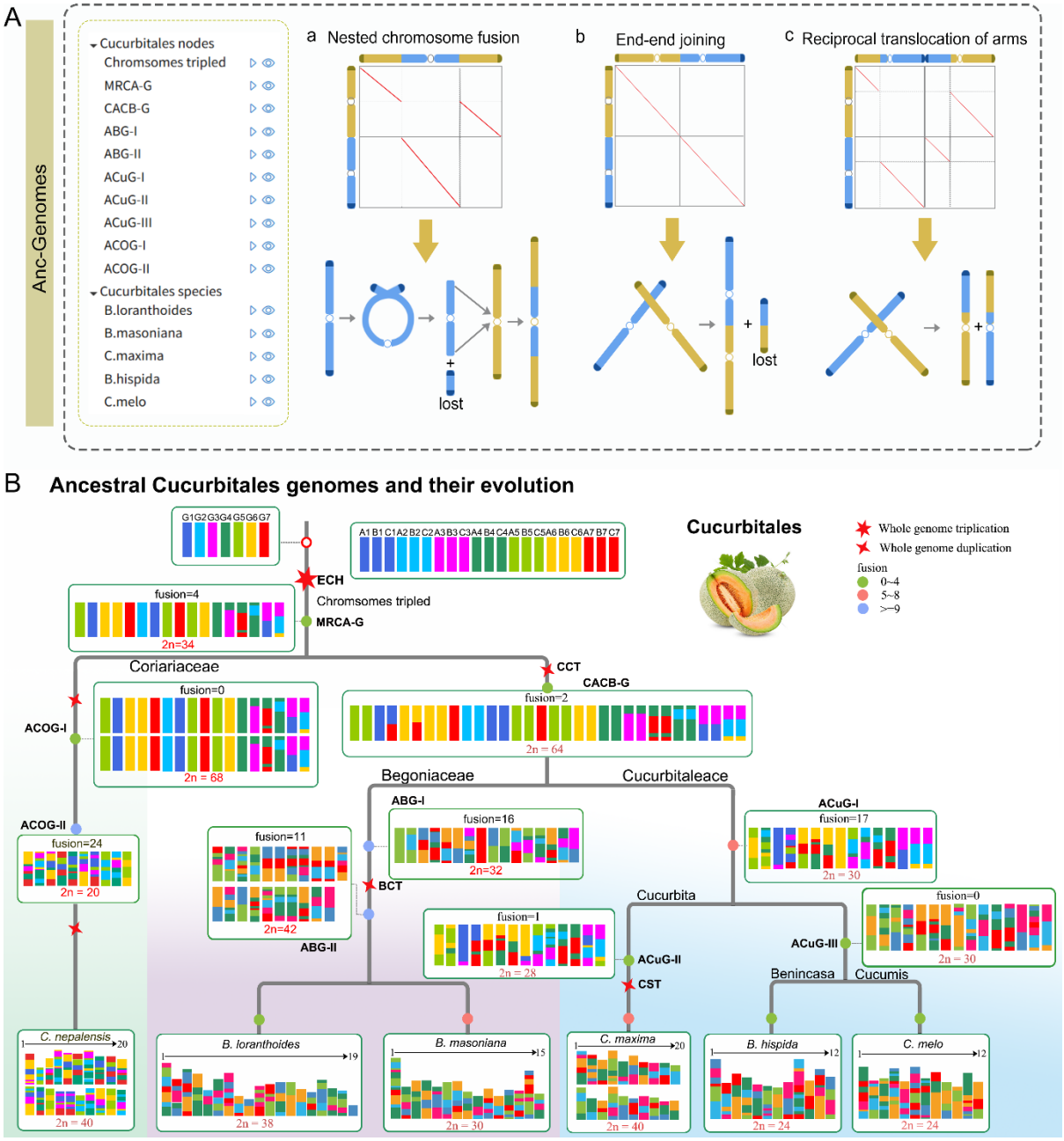
Cmo20g00081. **(B)** The DecoBrowse visualization interface. **(C)** Various annotations corresponding to the gene in our platform.



Supplemental Figure 18. The *Cmo02g01279* and *Cmo20g00081* gene annotations searched from our platform. (A) The *Cmo20g00081* gene annotation searched from our platform. **(B)** The *Cmo02g01279* gene annotation searched from our platform.



Supplemental Figure 19. Searching for polyploidization generated paralogous genes in the WGDs-Genes interface. (A) Users can directly query the database with gene ID, or enter protein sequence to compare with the genome of target species to search the gene whether generated from a specific polyploidization event. **(B)** The annotations related to target gene searched from our platform.



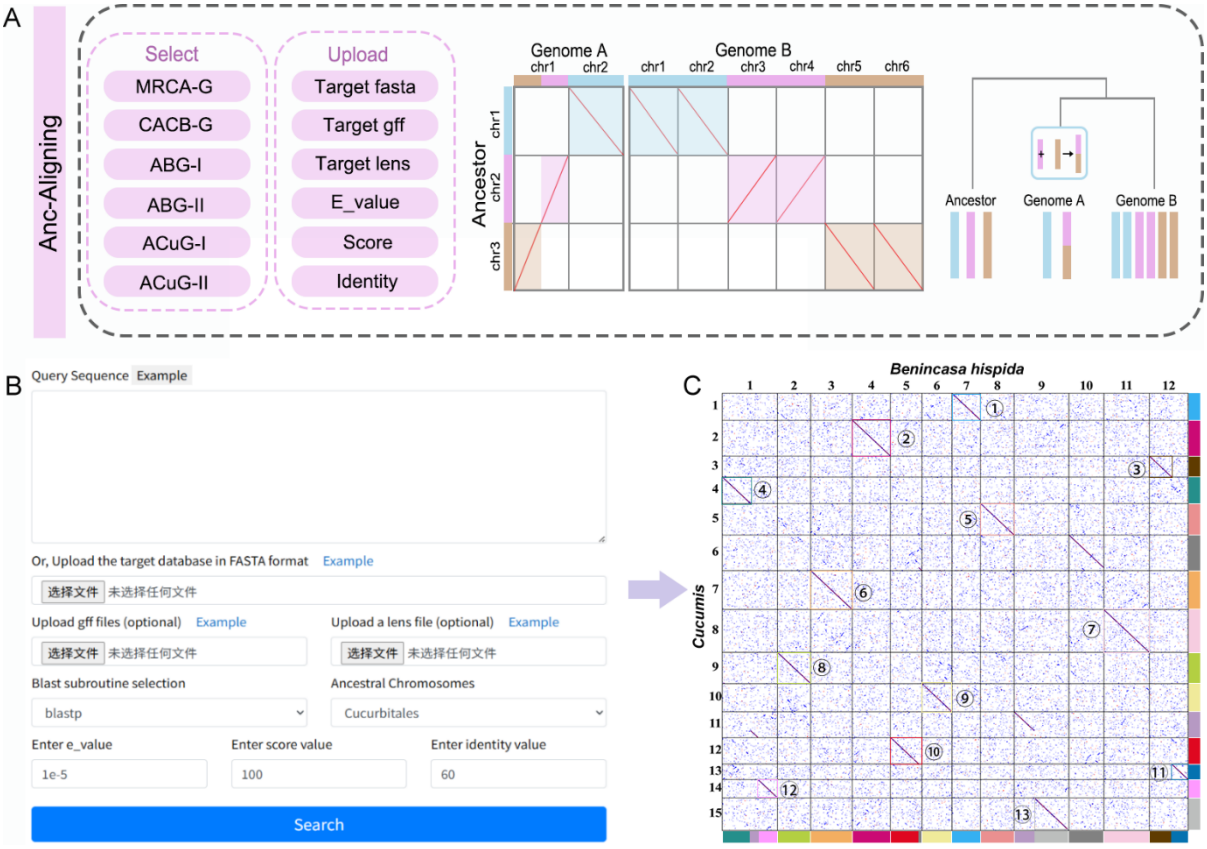
741

742 **Supplemental Figure 20. The Anc-Genomes interface. (A)** The Anc-Genomes interface enables static and

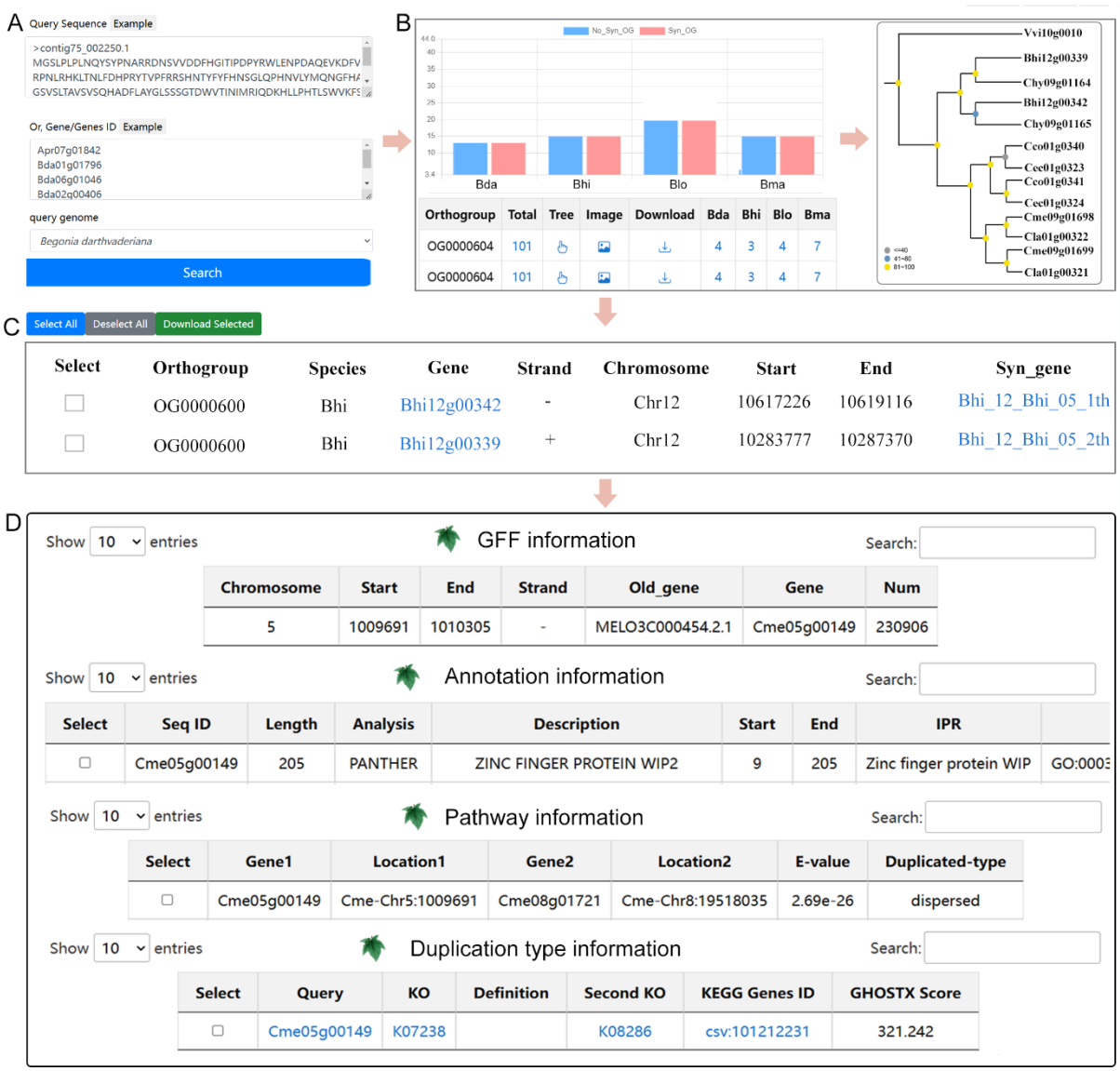
743 dynamic visualization of the ancestral Cucurbitales genomes and evolutionary trajectories. **(B)** The ancestral

744 genomes and evolutionary trajectories of Cucurbitales plants.

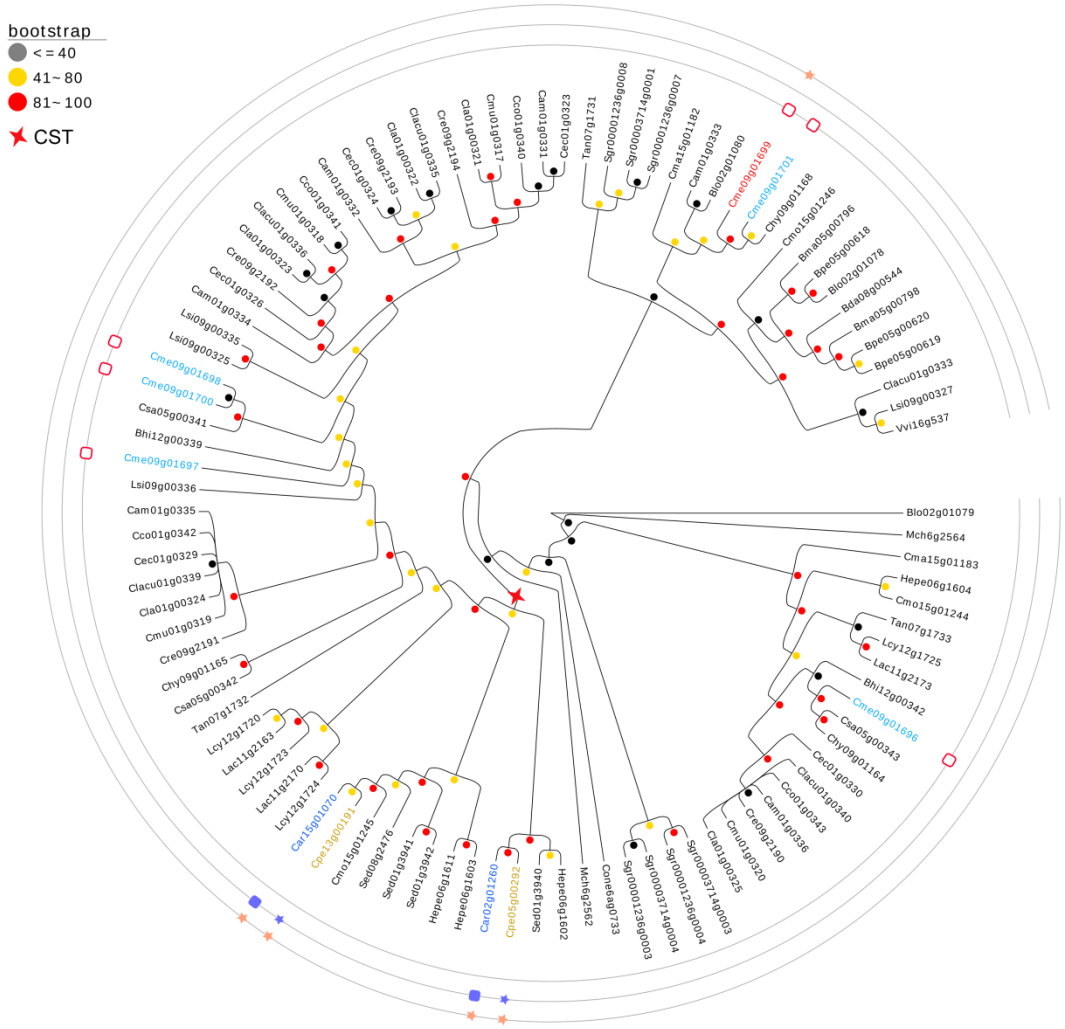
745



Supplemental Figure 21. The Anc-Aligning interface and its usage case. (A) The Anc-Aligning interface empowers alignment and visualization analysis for investigating the evolution of structures in new genomes. **(B)** The Anc-Aligning interface. **(C)** Inter-genomic GSDs between of the genomes of ancestral *Cucumis* and *B. hispida*. The 15 inferred ancestral *Cucumis* chromosomes were displayed in different colors. The complete ancestral chromosomes in *B.hispida* genome were highlighted with colored boxes corresponding to the ancestral *Cucumis* chromosomes. The numbers indicate the retained complete ancestral chromosomes.

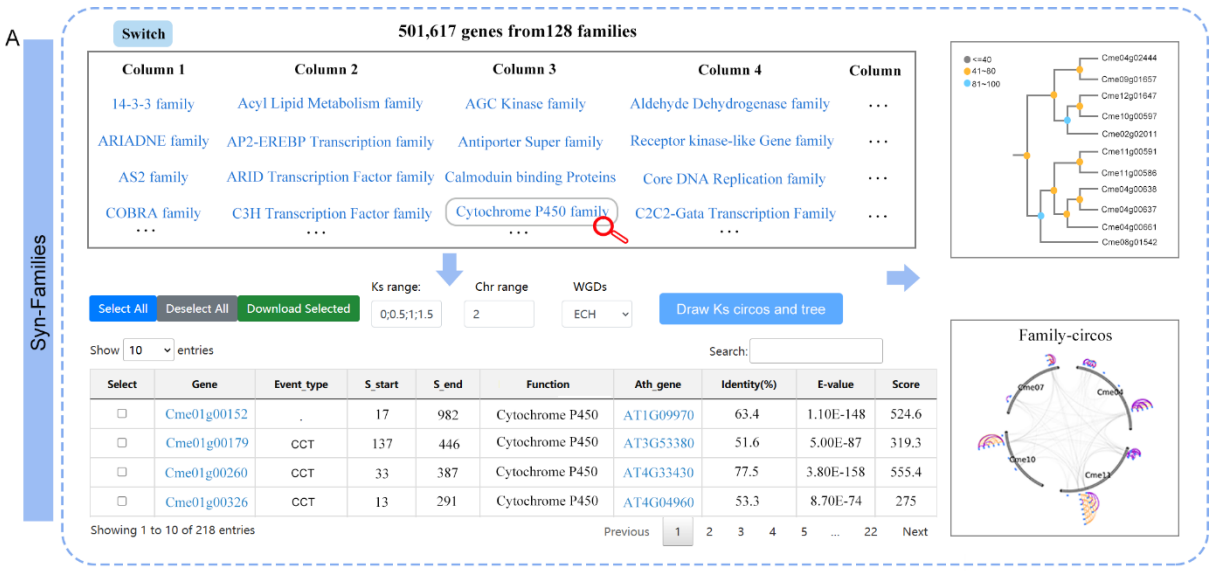


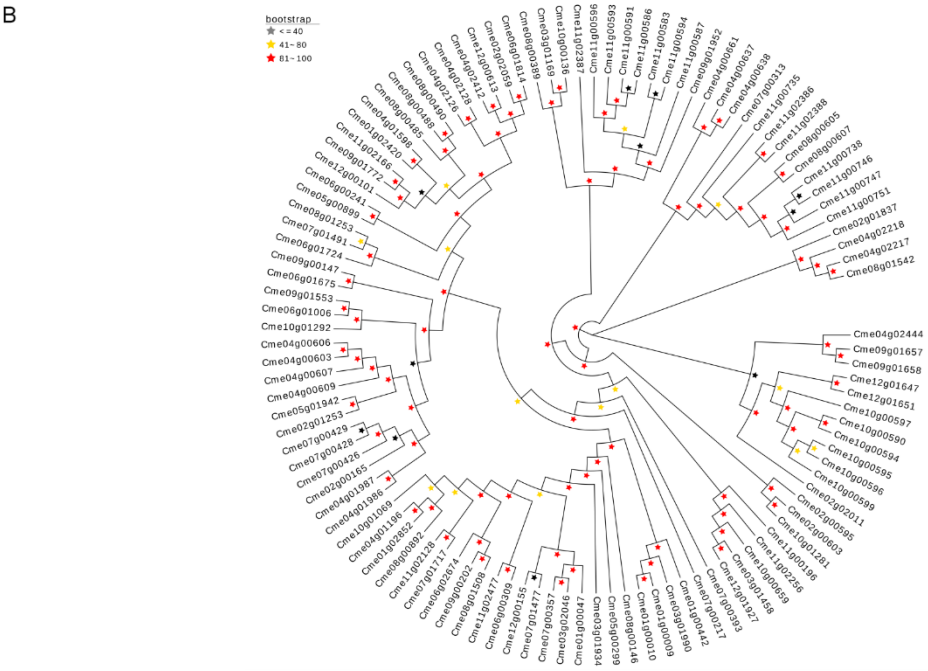
Supplemental Figure 22. The genomic resources and functional modules in the Syn-Orthogroups interface and illustrating an example of orthogroups (OGs) in the Syn-Orthogroups interface. (A) Users can either directly query the database using gene IDs or input protein sequences for comparison with the target genome. **(B)** A bar chart depicting the number of genes across species in the orthogroups, along with the counts of OGs and Synteny OGs. **(C)** By clicking the gene counts, users can search the detailed synteny information for the genes. **(D)** The related annotations related to the OGs.



760

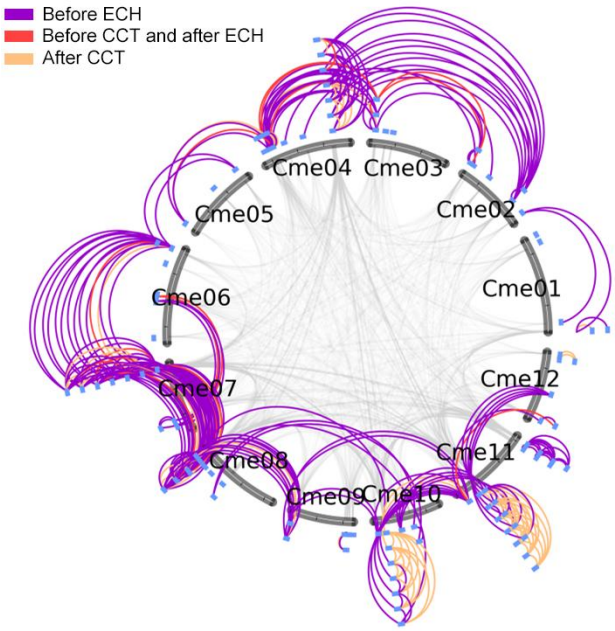
761 **Supplemental Figure 23. Phylogenetic tree of the genes in OG0000600 across the 28-CGs and *V.***
762 ***vinifera* genomes.** Genes from different species are marked with different colors, with the *Bt* gene labelled
763 red. The dots represent in three layers indicate the genes with tandem duplication, paralogous, and
764 orthologous relationships (inner to outer ring, respectively).





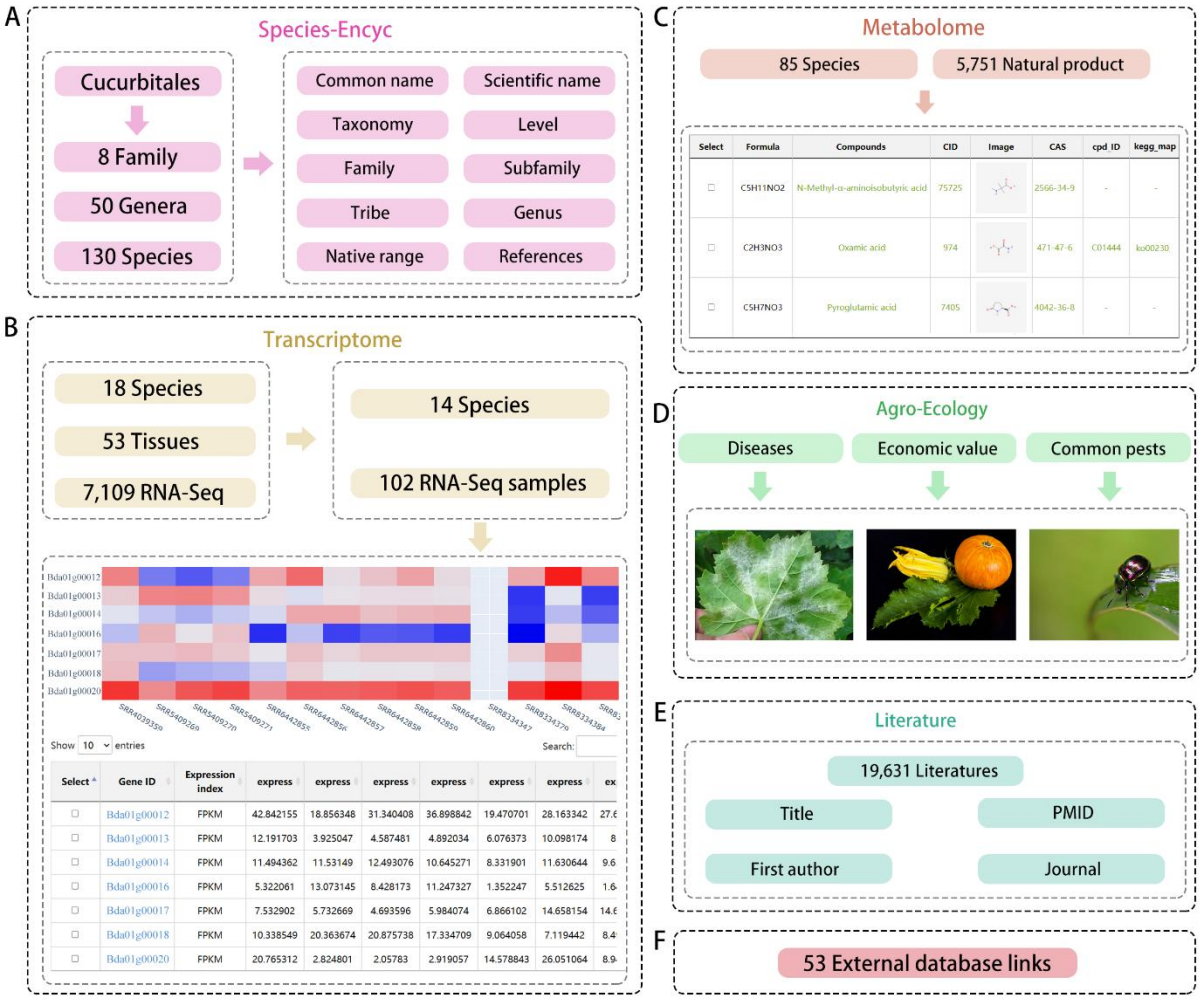
765

766 **Supplemental Figure 24. The genomic resources and functional modules in the Gene-families interface**
767 **and case**. **(A)** The Gene-families interface. This interface contains genes from 128 important trait gene
768 families across the species in Cucurbitales, along with their genomic and biological functional annotations,
769 and synteny information. Both interfaces were equipped with user-friendly access options and visualization
770 tools. **(B)** Phylogenetic tree of the genes from Cytochrome P450 gene family in *C. melo* genome. Blue dots
771 indicate the bootstrap values, with the size of the dots representing the magnitude of the bootstrap values.
772 Larger dots correspond to the higher bootstrap values.

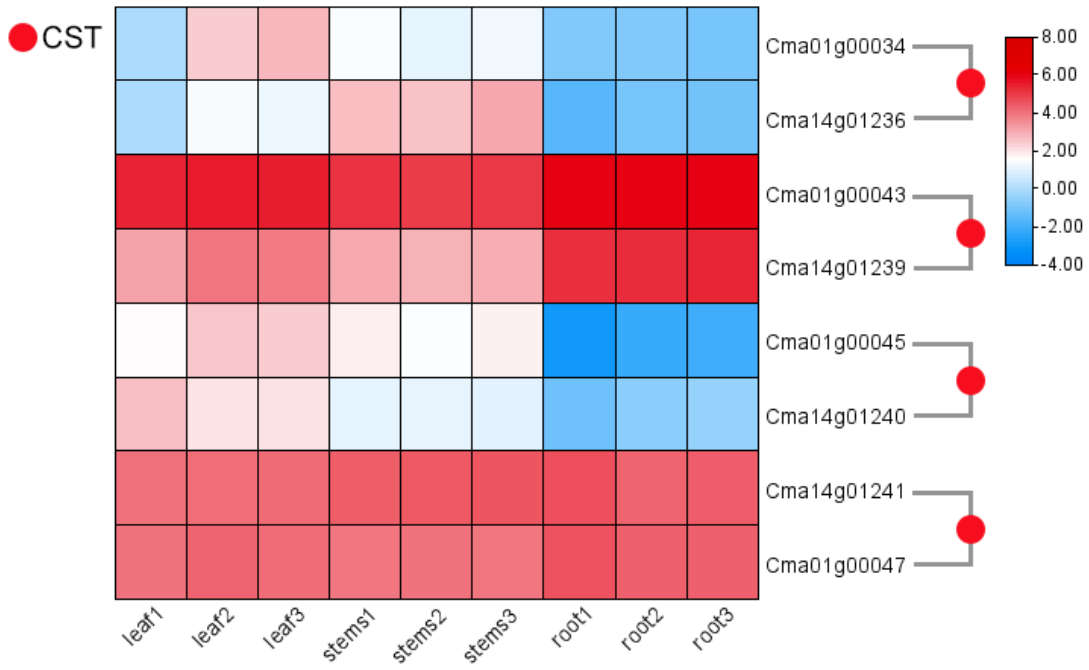


773

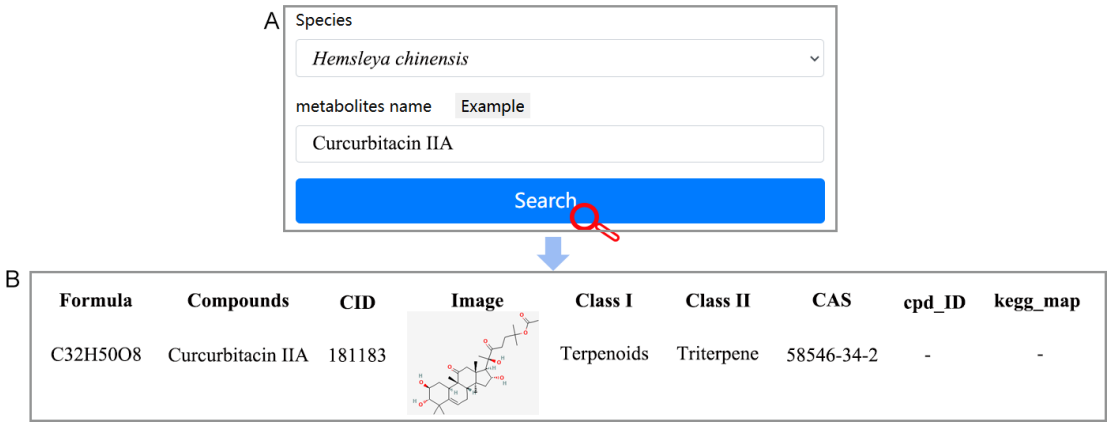
774 **Supplemental Figure 25. Expansion patterns of Cytochrome P450 gene family in *C. melo* genome.**
775 Cytochrome P450 genes are displayed in blue patches. Curved lines (gray) link the gene pairs like generated
776 from the CCT event. Curved lines (yellow, red, and purple) link gene pairs generated after the CCT (yellow),
777 before the CCT and after the ECH (red), and before the ECH (purple), respectively.



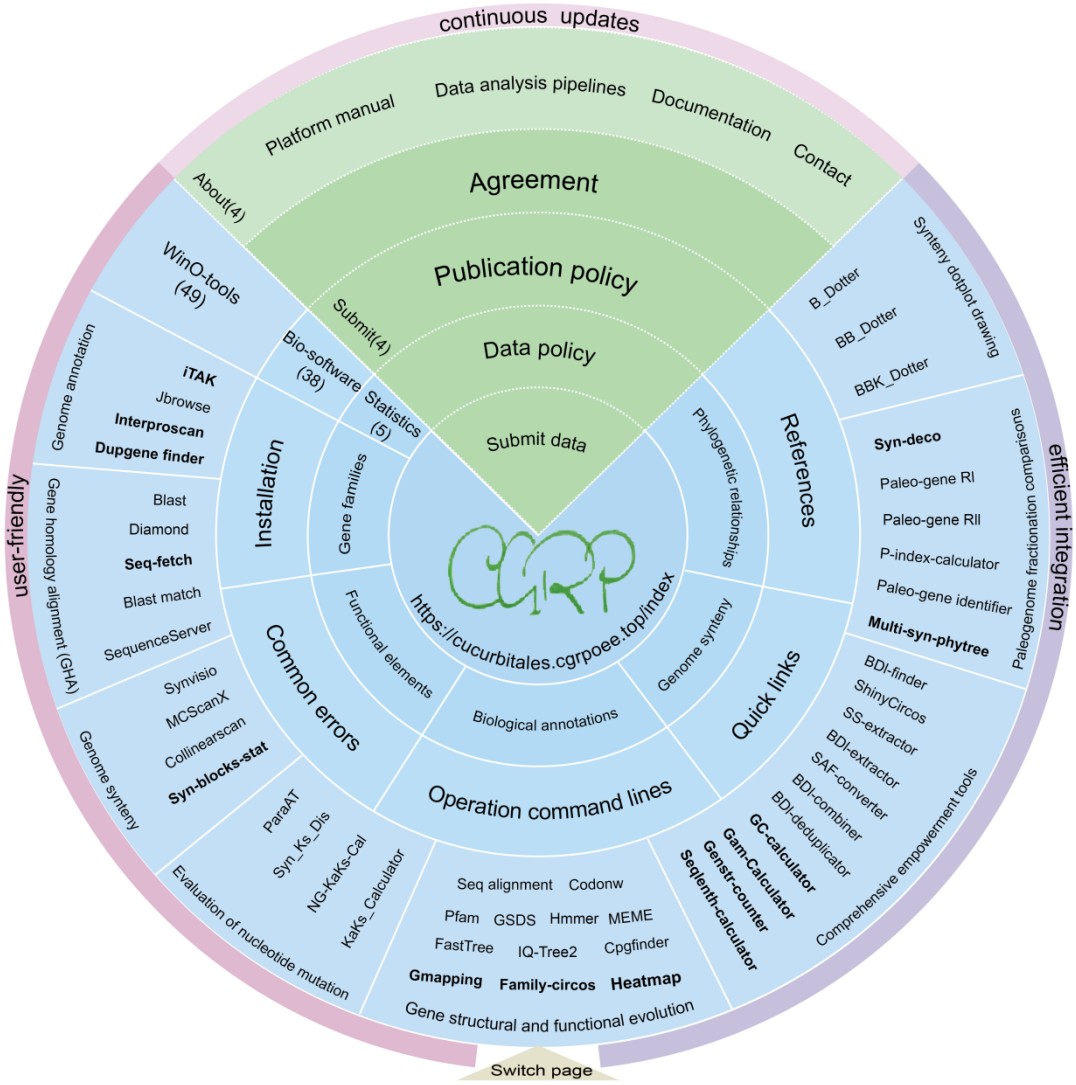
Supplemental Figure 26. The interfaces related to comprehensive resources that are beneficial for genomics research of Cucurbitales. (A) Species encyclopedia interface of Cucurbitales. **(B)** Transcriptome interface and its visualization analysis module. **(C)** Metabolome interface and its search function. **(D)** Agriculture and ecological resources interface. **(E)** Interface for searching research papers related to Cucurbitales. **(F)** Cross platform database query interface.



Supplemental Figure 27. The heatmap showing the expression levels of four paralogous gene pairs generated from CST events in *C. maxima* genome. Each square represents the value of gene expression after log2 conversion.

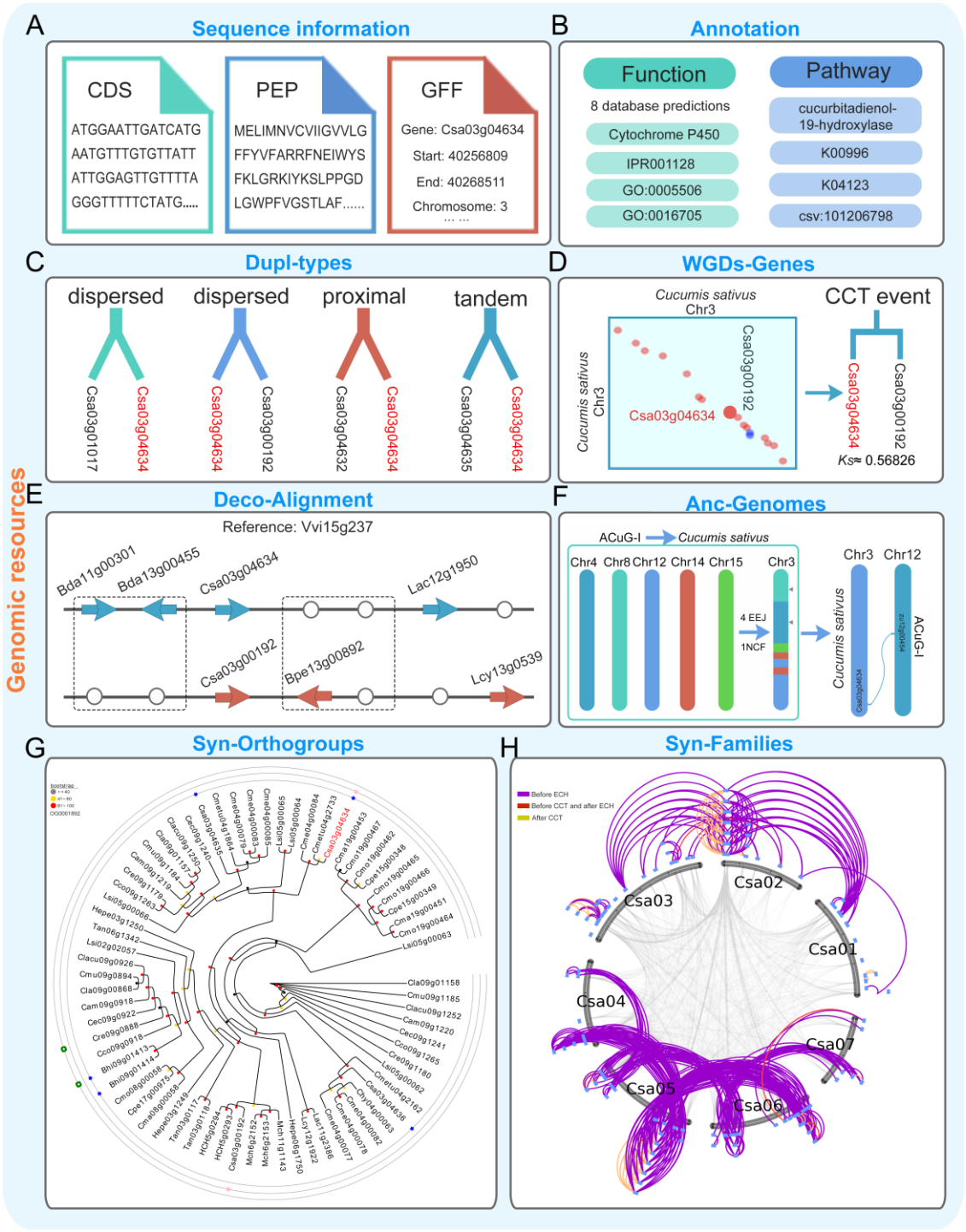


Supplemental Figure 28. The Cucurbitacin IIA generated from the *H. chinensis* were searched in the Metabolome interface. (A) Users can directly select target species and metabolites to search the detailed information of metabolites in our platform. **(B)** The searched results include the Formula, Compounds, the number of metabolites (CID) in Pubchem database, structural images of compounds, primary classification of substances (Class I), secondary classification of substances (Class II), CAS registry number (CAS), unique identifier (cpd_ID) of the compounds, and KEGG pathways (kegg_map).



796

797 **Supplemental Figure 29. The portals that empower the CGRP with new mining, continuous updating,**
798 **and use-friendly features.** The software involved in the WinO-Tools portal and its functional modules are
799 highlighted in bold in our newly developed script (outermost ring). The number and functionality of the
800 software relate to biological software, statistical portals and their corresponding interfaces (intermediate
801 rings). Outside the sectors are the contents of the About and Submit portals. This makes CGRP accurate,
802 efficient, user-friendly and constantly updated.


Genomic resources
A
Sequence information
CDS
ATGGAATTGATCATG
AATGTTTGTGTTATT
ATTGGAGTTGTTTTA
GGGTTTTTCTATG.....
PEP
MELIMNVCVIIGVVLG
FFYVFARRFNEIWYS
FKLGRKIYKSLPPGD
LGWPFVGSTLAF.....
GFF
Gene: Csa03g04634
Start: 40256809
End: 40268511
Chromosome: 3
B
Annotation
Function
8 database predictions
Cytochrome P450
IPR001128
GO:0005506
GO:0016705
Pathway
cucurbitadienol-19-hydroxylase
K00996
K04123
csv:101206798
C
Dupl-types
dispersed
dispersed
proximal
tandem
Csa03g01017
Csa03g04634
Csa03g04634
Csa03g00192
Csa03g04632
Csa03g04634
Csa03g04635
Csa03g04634
D
WGDs-Genes
Cucumis sativus Chr3
Csa03g00192
Csa03g04634
Cucumis sativus Chr3
CCT event
Csa03g04634
Csa03g00192
Ks≈ 0.56826
E
Deco-Alignment
Reference: Vvi15g237
Bda11g00301
Bda13g00455
Csa03g04634
Lac12g1950
Csa03g00192
Bpe13g00892
Lcy13g0539
F
Anc-Genomes
ACuG-I
Cucumis sativus
Chr4 Chr8 Chr12 Chr14 Chr15 Chr3
4 EEJ
1NCF
Cucumis sativus
Chr3 Chr12
ACuG-I
G
Syn-Orthogroups
H
Syn-Families
Before ECH
Before CCT and after ECH
After CCT
Csa02
Csa03
Csa01
Csa04
Csa07
Csa05
Csa06
I
Transcriptome
Csa03g00192
Csa03g04634
Cma08g00978
Lsi09g01869
2
1
0
-1
-2
root root root root leaf leaf leaf fruit fruit fruit
J
Metabolome
Acetyl-CoA
IPP/DMAPP
Squalene
HcSE1/HcSE2
SE
2,3-Oxidosqualene
Cucurbitadienol (Cuol)
Csa03g04634
(ko00996
C30H50O2)
19-hydroxy-cuol
OSC
HcOSC
CsBi
ClBi
CmBi
Cucumis sativus

Supplemental Figure 30. The Switch Page interactive interface integrates multi-omics data resources to explore the enzyme gene (*Csa3G903540*) that catalyzes Cucurbitadfenol in cucumber. By entering the proteins of *Csa3G903540* in Switch Page, we obtained the CGRP-specific identifier (*Csa03g04634*), sequence information, and biological functional annotations of this gene **(A-B)**. **(C)** Various types of duplicates of *Csa03g04634*. **(D)** The genomic synteny data resources demonstrates that the *Csa03g04634* generated from Cucurbitales common WGD event (CCT). **(E)** Multiple gene synteny alignment across Cucurbitales species for *Csa03g04634*. **(F)** The chromosomal localizations of *Csa03g04634* in cucumber and ancestral (AcuG-I) genomes, as well as the involved chromosomal rearrangements. **(G)** Phylogenetic position of the *Csa03g04634* in OG0001892 across Cucurbitales genomes. **(H)** *Csa03g04634* belongs to the Cytochrome P450 gene family and its expansion patterns in cucumber genome. **(I)** The expression patterns of *Csa03g04634* and its homologs in different tissues of corresponding Cucurbitales species. **(J)** Secondary metabolites and biosynthetic pathways associated with *Csa03g04634* in cucumber.

Supplemental Tables 1-8

Supplemental Table 1. Overview of genomic data and its original sources.

Num	Species name	Assembly Level	Assembly accession	Journal	Reference
1	*Citrullus lanatus*	Chromosome	GCA_039639715.1		
2	*Citrullus colocynthis*	Chromosome	GCA_039645085.1		
3	*Citrullus amarus*	Chromosome	GCA_039644775.1		
4	*Citrullus rehmii*	Chromosome	GCA_039644925.1		
5	*Citrullus lanatus subsp. mucosospermus*	Chromosome	GCA_039644915.1		
6	*Citrullus naudinianus*	Chromosome	GCA_039645035.1		
7	*Citrullus ecirrhosus*	Chromosome	GCA_039644935.1		
8	*Coriaria nepalensis*	Chromosome	GCA_027186245.1		
9	*Cucumis sativus*	Chromosome	GCF_000004075.3	Gigascience	(Li, et al. 2019)
10	*Cucumis melo*	Chromosome	GCA_025177605.1	Nature Communications	(Vaughn, et al. 2022)
11	*Cucurbita pepo subsp. pepo*	Chromosome	GCF_002806865.2	Plant Biotechnology Journal	(Montero-Pau, et al. 2017)
12	*Gynostemma pentaphyllum*	Chromosome	GCA_028015215.1	Horticulture Research	(Zhang, et al. 2023)
13	*Luffa aegyptiaca*	Chromosome	GCA_017139565.1	Horticulture Research	(Wu, et al. 2020)
14	*Lagenaria siceraria*	Chromosome	GCA_002890555.2	Plants Basel	(Mastrochirico, et al. 2023)
15	*Benincasa hispida*	Chromosome	GCA_009727055.1	Nature Communications	(Xie, et al. 2019)
16	*Sicyos edulis*	Chromosome	GCA_036507275.1		
17	*Luffa acutangula*	Chromosome	GCA_036850905.1		
18	*Cucurbita ficifolia*	Chromosome	GCA_033971965.1		
19	*Cucumis hystrix*	Chromosome	GCA_025688845.1		
20	*Trichosanthes cucumerina*	Chromosome	GCA_036507285.1		
21	*Herpetospermum pedunculosum*	Chromosome	GCA_028017165.1	Genome Biology and Evolution	(Yang, et al. 2023)

22	*Cucurbita argyrosperma subsp. sororia*	Chromosome	GCA_018691285.1	Horticulture Research	(Barrera-Redondo, et al. 2021)
23	*Begonia masoniana*	Chromosome	GCA_022432975.1	New Phytologist	(Li, et al. 2022)
24	*Begonia loranthoides*	Chromosome	GCA_022433065.1	New Phytologist	(Li, et al. 2022)
25	*Begonia peltatifolia*	Chromosome	GCA_022433055.1	New Phytologist	(Li, et al. 2022)
26	*Momordica charantia*	Scaffold	GCA_001995035.1		
27	*Momordica charantia*	Scaffold	GCF_001995035.1		
28	*Cucurbita moschata*	Scaffold	GCA_002738365.1	Molecular Plant	(Sun, et al. 2017)
29	*Cucurbita moschata*	Scaffold	GCF_002738365.1	Molecular Plant	(Sun, et al. 2017)
30	*Cucurbita maxima*	Scaffold	GCA_002738345.1	Molecular Plant	(Sun, et al. 2017)
31	*Cucurbita maxima*	Scaffold	GCF_002738345.1	Molecular Plant	(Sun, et al. 2017)
32	*Trichosanthes truncata*	Scaffold	GCA_033996785.1		
33	*Datisca glomerata*	Scaffold	GCA_003255025.1	Science	(Griesmann, et al. 2018)
34	*Telfairia occidentalis*	Scaffold	GCA_026257165.2	Biodivers Genomes	(Pirro, et al. 2022)
35	*Begonia fuchsioides*	Scaffold	GCA_003255005.1	Science	(Griesmann, et al. 2018)
36	*Cucumis cinereus*	Scaffold	GCA_037974625.1		
37	*Sechiopsis laciniata*	Scaffold	GCA_037833695.1		
38	*Cyclanthera cordifolia*	Scaffold	GCA_037832605.1		
39	*Sicyos baderoa*	Scaffold	GCA_037832685.1		
40	*Cucumis variabilis*	Scaffold	GCA_037974545.1		
41	*Gerrardanthus grandiflorus*	Scaffold	GCA_040966115.1		
42	*Gurania insolita*	Scaffold	GCA_040966235.1		
43	*Frantzia villosa*	Scaffold	GCA_040966175.1		
44	*Gurania lobata*	Scaffold	GCA_040966215.1		
45	*Citrullus lanatus*	Chromosome	GCA_039619345.1		
46	*Citrullus lanatus*	Chromosome	GCA_039645055.1		
47	*Citrullus lanatus*	Chromosome	GCA_039637665.1		
48	*Citrullus lanatus*	Chromosome	GCA_039637515.1		

49	*Citrullus lanatus*	Chromosome	GCA_039638265.1		
50	*Citrullus lanatus*	Chromosome	GCA_039638695.1		
51	*Citrullus lanatus*	Chromosome	GCA_039639425.1		
52	*Citrullus lanatus*	Chromosome	GCA_039638095.1		
53	*Citrullus lanatus*	Chromosome	GCA_039637945.1		
54	*Citrullus lanatus*	Chromosome	GCA_039639725.1		
55	*Citrullus lanatus*	Chromosome	GCA_039637835.1		
56	*Citrullus colocynthis*	Chromosome	GCA_039644965.1		
57	*Citrullus colocynthis*	Chromosome	GCA_039644945.1		
58	*Citrullus colocynthis*	Chromosome	GCA_039645205.1		
59	*Citrullus amarus*	Chromosome	GCA_039644905.1		
60	*Citrullus amarus*	Chromosome	GCA_039644995.1		
61	*Citrullus amarus*	Chromosome	GCA_039644975.1		
62	*Citrullus amarus*	Chromosome	GCA_039645015.1		
63	*Citrullus lanatus subsp. mucosospermus*	Chromosome	GCA_039644955.1		
64	*Coriaria nepalensis*	Chromosome	GCA_027190085.1		
65	*Cucumis sativus*	Chromosome	GCA_016163735.1	Nature Communications	(Li, et al. 2022)
66	*Cucumis sativus*	Chromosome	GCA_016161785.1	Nature Communications	(Li, et al. 2022)
67	*Cucumis sativus*	Chromosome	GCA_016161805.1	Nature Communications	(Li, et al. 2022)
68	*Cucumis sativus*	Chromosome	GCA_016161715.1	Nature Communications	(Li, et al. 2022)
69	*Cucumis sativus*	Chromosome	GCA_016161855.1	Nature Communications	(Li, et al. 2022)
70	*Cucumis sativus*	Chromosome	GCA_016161875.1	Nature Communications	(Li, et al. 2022)
71	*Cucumis sativus*	Chromosome	GCA_016161885.1	Nature Communications	(Li, et al. 2022)
72	*Cucumis sativus*	Chromosome	GCA_016161765.1	Nature Communications	(Li, et al. 2022)
73	*Cucumis sativus*	Chromosome	GCA_016163705.1	Nature Communications	(Li, et al. 2022)
74	*Cucumis sativus*	Chromosome	GCA_016163745.1	Nature Communications	(Li, et al. 2022)
75	*Cucumis sativus*	Chromosome	GCA_016161775.1	Nature Communications	(Li, et al. 2022)
76	*Cucumis melo*	Chromosome	GCA_026261895.1	Plant Journal	(Oren, et al. 2022)

77	*Cucumis melo*	Chromosome	GCA_040436765.1	Plant Journal	(Oren, et al. 2022)
78	*Cucumis melo*	Chromosome	GCA_026262045.1	Plant Journal	(Oren, et al. 2022)
79	*Cucumis melo*	Chromosome	GCA_026262125.1	Plant Journal	(Oren, et al. 2022)
80	*Cucumis melo*	Chromosome	GCA_026262055.1	Plant Journal	(Oren, et al. 2022)
81	*Cucumis melo*	Chromosome	GCA_026261985.1	Plant Journal	(Oren, et al. 2022)
82	*Cucumis melo*	Chromosome	GCA_026262315.1	Plant Journal	(Oren, et al. 2022)
83	*Cucumis melo*	Chromosome	GCA_026262265.1	Plant Journal	(Oren, et al. 2022)
84	*Cucumis melo*	Chromosome	GCA_026262095.1	Plant Journal	(Oren, et al. 2022)
85	*Cucumis melo*	Chromosome	GCA_020920065.2	Plant Journal	(Oren, et al. 2022)
86	*Cucumis melo*	Chromosome	GCA_025177625.1	Plant Journal	(Oren, et al. 2022)
87	*Cucumis melo*	Chromosome	GCA_026262135.1	Plant Journal	(Oren, et al. 2022)
88	*Cucumis melo*	Chromosome	GCA_026261935.1	Plant Journal	(Oren, et al. 2022)
89	*Cucumis melo*	Chromosome	GCA_026261875.1	Plant Journal	(Oren, et al. 2022)
90	*Cucumis melo*	Chromosome	GCA_011762645.1	Plant Journal	(Oren, et al. 2022)
91	*Cucumis melo*	Chromosome	GCA_026262325.1	Plant Journal	(Oren, et al. 2022)
92	*Cucumis melo*	Chromosome	GCA_026262305.1	Plant Journal	(Oren, et al. 2022)
93	*Cucumis melo*	Chromosome	GCA_026261825.1	Plant Journal	(Oren, et al. 2022)
94	*Cucumis melo*	Chromosome	GCA_026261845.1	Plant Journal	(Oren, et al. 2022)
95	*Cucumis melo*	Chromosome	GCA_026261915.1	Plant Journal	(Oren, et al. 2022)
96	*Cucumis melo*	Chromosome	GCA_026262175.1	Plant Journal	(Oren, et al. 2022)
97	*Cucumis melo*	Chromosome	GCA_026262215.1	Plant Journal	(Oren, et al. 2022)
98	*Cucumis melo*	Chromosome	GCA_026262015.1	Plant Journal	(Oren, et al. 2022)
99	*Cucumis melo*	Chromosome	GCA_026262255.1	Plant Journal	(Oren, et al. 2022)
100	*Cucumis melo*	Chromosome	GCA_026262345.1	Plant Journal	(Oren, et al. 2022)
101	*Cucumis melo*	Chromosome	GCA_026262025.1	Plant Journal	(Oren, et al. 2022)
102	*Cucumis melo*	Chromosome	GCA_020920055.1	Plant Journal	(Oren, et al. 2022)
103	*Cucumis melo*	Chromosome	GCA_026262165.1	Plant Journal	(Oren, et al. 2022)
104	*Cucumis melo subsp. agrestis*	Chromosome	GCA_030578075.1		

105	*Cucumis melo subsp. agrestis*	Chromosome	GCA_014525375.1	iScience	(Yang, et al. 2020)
106	*Cucumis melo var. inodorus*	Chromosome	GCA_009760825.1	iScience	(Zhang, et al. 2020)
107	*Citrullus lanatus*	Chromosome	GCA_029034555.1	Frontiers in Plant Science	(Nie, et al. 2023)
108	*Citrullus lanatus*	Chromosome	GCA_029034585.1	Frontiers in Plant Science	(Nie, et al. 2023)
109	*Citrullus lanatus*	Chromosome	GCA_000238415.2	Frontiers in Plant Science	(Nie, et al. 2023)
110	*Citrullus lanatus*	Chromosome	GCA_004801215.2	Frontiers in Plant Science	(Nie, et al. 2023)
111	*Citrullus lanatus subsp. cordophanus*	Chromosome	GCA_018142915.1	PNAS	(Renner, et al. 2021)
112	*Cucurbita pepo subsp. pepo*	Chromosome	GCA_002806865.2	Plant Biotechnology Journal	(Montero-Pau, et al. 2017)
113	*Gynostemma pentaphyllum*	Chromosome	GCA_020536105.1	DNA Research	(Huang, et al. 2021)
114	*Cucumis melo*	Scaffold	GCA_902497455.1	Frontiers in Plant Science	(Castanera, et al. 2020)
115	*Cucumis sativus*	Scaffold	GCA_000224045.1	Public Library of Science ONE	(Wóycicki, et al. 2011)
116	*Benincasa hispida*	Chromosome	GCA_027475165.1		
117	*Citrullus colocynthis*	Chromosome	GCA_963978565.1		
118	*Citrullus amarus*	Chromosome	GCA_033088355.1		
119	*Citrullus amarus*	Chromosome	GCA_033088365.1		
120	*Luffa acutangula*	Chromosome	GCA_012295215.1	Scientific Reports	(Saensuk, et al. 2022)
121	*Citrullus amarus*	Chromosome	GCA_030954535.1		
122	*Citrullus lanatus subsp. mucosospermus*	Chromosome	GCA_030954485.1		
123	*Cucurbita argyrosperma subsp. argyrosperma*	Chromosome	GCA_004115005.2	Molecular Plant	(Barrera-Redondo, et al. 2019)
124	*Momordica charantia*	Scaffold	GCA_013281855.1	PNAS	(Matsumura, et al. 2020)
125	*Lagenaria siceraria*	Scaffold	GCA_003268545.1	Plant Journal	(Wu, et al. 2017)
126	*Cucumis sativus*	Contig	GCA_001483825.3		
127	*Cucumis melo*	Scaffold	GCA_902497455.1	Frontiers in Plant Science	(Castanera, et al. 2020)
128	*Cucumis melo var. makuwa*	Scaffold	GCA_005549215.1		
129	*Cucumis melo var. makuwa*	Scaffold	GCA_005549225.1	Scientific Data	(Shin, et al. 2019)

130	*Momordica charantia*	Scaffold	GCA_900491585.1		
131	*Luffa aegyptiaca*	Scaffold	GCA_016170045.1		
132	*Luffa aegyptiaca*	Contig	GCA_012295205.1		
133	*Benincasa hispida*	Scaffold	GCA_947786645.1		
134	*Benincasa hispida*	Scaffold	GCA_016987355.1		
135	*Benincasa hispida*	Scaffold	GCA_012071895.1		
136	*Lagenaria siceraria*	Scaffold	GCA_000466325.1	Plant Journal	(Xu, et al. 2013)
137	*Citrullus colocynthis*	Scaffold	GCA_031780015.1		
138	*Cucumis melo*	Chromosome	GCA_036687305.1		
139	*Begonia darthvaderiana*	Chromosome	GCA_022432945.1	New Phytologist	(Li, et al. 2022)
140	*Cucumis sativus*	Scaffold	GCA_000224045.1	Public Library of Science ONE	(Wóycicki, et al. 2011)
141	*Cucumis sativus*	Scaffold	GCF_000224045.1	Public Library of Science ONE	(Wóycicki, et al. 2011)
142	*Cucumis melo*	Scaffold	GCA_000313045.1	PNAS	(Garcia-Mas, et al. 2012)
143	*Cucumis melo*	Scaffold	GCF_000313045.1	PNAS	(Garcia-Mas, et al. 2012)
144	*Cucumis sativus* L. var. *sativus*	Chromosome		GigaScience	(Li, et al. 2019)
145	*Cucumis sativus L. var. sativus*	Chromosome			
146	*Cucumis sativus L. var. sativus*	Chromosome		Molecular Genetics and Genomics	(Osipowski, et al. 2020)
147	*Cucumis sativus var. hardwickii*	Chromosome		Nature Genetics	(Qi, et al. 2013)
148	*Cucumis hystrix*	Chromosome		Horticulture Research	(Qin, et al. 2021)
149	*Citrullus lanatus subsp. vulgaris*	Chromosome		Nature Genetics	(Guo, et al. 2019)
150	*Citrullus lanatus subsp. Vulgaris*	Chromosome		Plant Biotechnology Journal	(Wu, et al. 2019)
151	*Citrullus lanatus subsp. Cordophanus*	Chromosome		Proc Natl Acad Sci USA	(Renner, et al. 2021)
152	*Citrullus mucosospermus*	Chromosome		Plant Biotechnology Journal	(Wu, et al. 2023)
153	*Citrullus amarus*	Chromosome		Plant Biotechnology Journal	(Wu, et al. 2023)
154	*Citrullus colocynthis*	Chromosome		Plant Biotechnology Journal	(Wu, et al. 2023)

155	*Cucumis melo*	Chromosome	Frontier in Plant Science	(Castanera, et al. 2021)
156	*C. melo var. cantalupensis*	Chromosome	iScience	(Pichot, et al. 2022)
157	*Cucumis melo* L. var. *inodorus*	Chromosome	iScience	(Zhang, et al. 2019)
158	*C. melo L. ssp. agrestis*	Chromosome	Plant Journal	(Ling, et al. 2021)
159	*C. melo var. reticulatus*	Chromosome	Communications Biology	(Yano, et al. 2020)
160	*Cucumis metuliferus*	Chromosome	Plant Journal	(Ling, et al. 2021)
161	*Cucurbita maxima*	Chromosome	Molecular Plant	(Sun, et al. 2017)
162	*Cucurbita moschata*	Chromosome	Molecular Plant	(Sun, et al. 2017)
163	*Cucurbita pepo subsp. pepo*	Chromosome	Plant Biotechnology Journal	(Montero-Pau, et al. 2018)
164	*Cucurbita argyrosperma subsp. argyrosperma*	Chromosome	Horticulture Research	(Barrera-Redondo, et al. 2021)
165	*Cucurbita argyrosperma subsp. sororia*	Chromosome	Horticulture Research	(Barrera-Redondo, et al. 2021)
166	*Momordica charantia*	Chromosome	Proc Natl Acad Sci USA	(Matsumura, et al. 2020)
167	*Momordica charantia*	Chromosome	Horticulture Research	(Cui, et al. 2020)
168	*Momordica charantia*	Chromosome	Horticulture Research	(Cui, et al. 2020)
169	*Lagenaria siceraria*	Chromosome	Plant Journal	(Wu, et al. 2017)
170	*Luffa cylindrica*	Chromosome	Molecular Ecology Resources	(Zhang, et al. 2019)
171	*Luffa cylindrica*	Chromosome	Horticulture Research	(Wu, et al. 2020)
172	*Luffa acutangula*	Chromosome	Molecular Ecology Resources	(Pootakham, et al. 2020)
173	*Benincasa hispida*	Chromosome	Nature Communications	(Xie, et al. 2019)
174	*Sechium edule*	Chromosome	Horticulture Research	(Fu, et al. 2021)
175	*Siraitia grosvenorii*	Chromosome	GigaScience	(Xia, et al. 2018)
176	*Hemsleya chinensis*	Chromosome		
177	*Begonia darthvaderiana*	Chromosome	New Phytologist	(Li, et al. 2022)

178	*Begonia loranthoides*	Chromosome		New Phytologist	(Li, et al. 2022)
179	*Begonia masoniana*	Chromosome		New Phytologist	(Li, et al. 2022)
180	*Begonia peltatifolia*	Chromosome		New Phytologist	(Li, et al. 2022)
181	*Begonia fuchsioides*	Scaffold	GCA_003255005.1	New Phytologist	(Li, et al. 2022)
182	*Begonia masoniana*	Chromosome	GCA_022432975.1	New Phytologist	(Li, et al. 2022)
183	*Datisca glomerata*	Chromosome	GCA_003255025.1	Science	(Griesmann, et al. 2018)
184	*Coriaria nepalensis*	Chromosome	GCA_027186245.1		
185	*Coriaria nepalensis*	Chromosome	GCA _027190085.1		

819 **Supplemental Table 2.** The KEGG annotations of genes in the genomes of 29-CGs and *V. vinifera*.

Species	KEGG
Begonia darthvaderiana	22,769
Benincasa hispida	31,663
Begonia loranthoides	20,077
Begonia masoniana	22,729
Begonia peltatifolia	19,834
Citrullus amarus	22,209
Cucurbita argyrosperma	27,996
Citrullus colocynthis	21,993
Citrullus ecirrhosus	21,979
Cucumis hystrix	22,200
Cucumis metuliferus	22,487
Citrullus lanatus	22,103
Citrullus lanatus cultivar	22,307
Cucurbita maxima	31,282
Cucumis melo	28,292
Cucurbita moschata	32,023
Citrullus mucsospermus	22,177

Coriaria nepalensis	20,773
Cucurbita pepo	26,225
Citrullus rehmii	22,271
Cucumis sativus	23,519
Hemsleya chinensis	22,017
Herpetospermum pedunculosum	22,094
Luffa acutangula	23,195
Luffa cylindrica	23,776
Lagenaria siceraria	22,233
Momordica charantia	22,071
Sechium edule	22,931
Trichosanthes anguina	28,853
Vitis vinifera	26,326

820 **Supplemental Table 3.** The gene pairs with different gene duplication types in the genomes of 29-CGs and *V. vinifera*.

Species	**Types**				
	WGD	**TD**	**PD**	**TRD**	**DSD**
Begonia darthvaderiana	6,546	763	1,010	6,154	15,457
Benincasa hispida	1,178	12,614	470	2,447	24,171
Begonia loranthoides	5,636	683	417	6,631	14,822
Begonia masoniana	5,188	767	1,093	6,682	15,563
Begonia peltatifolia	6,175	319	156	5,391	12,801
Citrullus amarus	2,689	1,600	660	4,133	11,765
Cucurbita argyrosperma	13,788	509	301	5,343	17,866
Citrullus colocynthis	2,769	1,490	700	4,381	12,179
Citrullus ecirrhosus	2,777	1,484	681	4,249	11,903
Cucumis hystrix	2,299	1,717	619	4,851	13,726
Cucumis metuliferus	31,300	1,793	1,153	4,498	16,718

Citrullus lanatus	2,821	1,160	685	5,795	14,734
Citrullus lanatus cultivar	2,945	1,650	776	4,396	12,240
Cucurbita maxima	13,945	1,384	589	5,115	19,684
Cucumis melo	2,107	1,952	817	4,712	17,959
Cucurbita moschata	13,564	1,668	628	5,311	20,156
Citrullus mucsospermus	3,444	2,458	1,239	5,391	14,060
Coriaria nepalensis	5,796	863	1,036	5,843	15,349
Cucurbita pepo	13,379	978	298	4,612	17,379
Citrullus rehmii	10,707	3,436	1,980	4,479	16,326
Cucumis sativus	2,385	1,597	668	5,089	14,138
Hemsleya chinensis	4,148	1,785	594	3,872	11,620
Herpetospermum pedunculosum	15,544	1,377	790	4,626	14,016
Luffa acutangula	10,708	1,676	1,565	8,540	16,732
Luffa cylindrica	2,249	1,567	914	3,614	14,677
Lagenaria siceraria	2,514	1,924	734	4,975	13,712
Momordica charantia	3,631	1,033	734	4,876	14,681
Sechium edule	9,099	12,030	480	3,843	25,618
Trichosanthes anguina	2,882	10,909	563	3,815	20,469
Vitis vinifera	2,225	2,225	1,366	4,636	14,916

821 **Supplemental Table 4.** Identified N6-methyladenosine regulators (m6As) in the genomes of 29-CGs and *V. vinifera*.

Species	Readers	Erasers	Writers			
			MTA70	WTAP	VIRILIZER	HAKAI
Begonia darthvaderiana	7	7	3	1	0	0
Benincasa hispida	7	14	2	1	0	0
Begonia loranthoides	8	7	3	2	0	0
Begonia masoniana	7	4	2	1	0	0
Begonia peltatifolia	7	5	2	1	0	0

Citrullus amarus	5	8	3	2	0	0
Cucurbita argyrosperma	10	8	4	2	0	0
Citrullus colocynthis	5	8	3	2	0	0
Citrullus ecirrhosus	5	8	3	2	0	0
Cucumis hystrix	5	8	3	0	0	0
Cucumis metuliferus	5	8	3	1	0	0
Citrullus lanatus	5	8	3	2	0	0
Citrullus lanatus cultivar	5	9	3	2	0	0
Cucurbita maxima	8	8	4	2	0	0
Cucumis melo	6	9	2	1	0	0
Cucurbita moschata	8	8	5	2	0	0
Citrullus mucsospermus	5	8	3	2	0	0
Coriaria nepalensis	31	23	5	2	2	0
Cucurbita pepo	7	8	5	2	0	0
Citrullus rehmii	5	9	3	2	0	0
Cucumis sativus	5	7	3	1	0	0
Hemsleya chinensis	5	8	3	2	0	0
Herpetospermum pedunculosum	4	8	3	1	0	0
Luffa acutangula	4	9	2	1	0	0
Luffa cylindrica	4	9	2	1	0	0
Lagenaria siceraria	4	9	3	1	0	0
Momordica charantia	4	8	3	1	0	0
Sechium edule	9	12	3	2	0	0
Trichosanthes anguina	5	9	3	1	0	0
Vitis vinifera	10	8	3	1	1	0

822 **Supplemental Table 5.** Duplicated genes generated from the specific polyploidizations in the genomes of 29-CGs and *V. vinifera*.

Species	[a] Gamma-related	[b] CCT-related	[c] CST-related	[d] BCT-related	[e] SSD-related	[f] CSD1-related	[g] CSD2-related
Begonia darthvaderiana	111/1,455/2,394	137/1,773/2,968	-	59/2,541/4,997	-	-	-
Benincasa hispida	58/814/1,603	88/1,131/2,214	-	-	-	-	-
Begonia loranthoides	106/1,850/2,826	141/1,941/3,078	-	55/2,925/5,611	-	-	-
Begonia masoniana	123/1,612/2,566	155/1,794/2,935	-	68/2,561/5,015	-	-	-
Begonia peltatifolia	127/3,617/5,983	146/3,730/6,208	-	77/2,553/4,769	-	-	-
Citrullus amarus	52/952/1,622	59/1,114/2,110	-	-	-	-	-
Cucurbita argyrosperma	127/3,094/4,658	161/3,266/4,842	71/5,624/11,030	-	-	-	-
Citrullus colocynthis	48/962/1,712	60/1,142/2,012	-	-	-	-	-
Citrullus ecirrhosus	52/974/1,796	65/1,194/1,988	-	-	-	-	-
Cucumis hystrix	43/817/1,600	58/976/1,910	-	-	-	-	-
Cucumis metuliferus	47/981/1,812	72/1,240/2,386	-	-	-	-	-
Citrullus lanatus	50/986/1,868	67/1,133/2,220	-	-	-	-	-
Citrullus lanatus cultivar	54/981/1,812	64/1,121/2,130	-	-	-	-	-
Cucurbita maxima	145/3,563/5339	186/3,742/5,486	69/6,341/12,430	-	-	-	-

Cucumis melo	52/890/1,741	72/1,043/2,049	-	-	-	-	-
Cucurbita moschata	160/2,793/3,906	202/2972/4,151	69/6,567/13,035	-	-	-	-
Citrullus mucsospermus	55/979/1,674	71/1,211/2,170	-	-	-	-	-
Coriaria nepalensis	611/4,409/3,925	-	-	-	-	200/3,571/4,514	31/5,932/11,845
Cucurbita pepo	137/5,979/9,910	170/6,238/10,435	85/3,649/6,992	-	-	-	-
Citrullus rehmii	53/913/1,766	60/1,103/1,988	-	-	-	-	-
Cucumis sativus	44/811/1,446	65/995/1,954	-	-	-	-	-
Herpetospermum pedunculosum	51/807/1,573	81/1,121/2,174	-	-	-	-	-
Luffa acutangula	58/814/1,603	87/1,091/2,072	-	-	-	-	-
Luffa cylindrica	51/779/1,436	88/1,131/2,202	-	-	-	-	-
Lagenaria siceraria	37/901/1,761	54/1,018/2,002	-	-	-	-	-
Momordica charantia	49/821/1,603	88/1,131/2,214	-	-	-	-	-
Sechium edule	143/3,761/5,014	181/6,107/11,012	-	-	79/3,117/5,992	-	-
Trichosanthes anguina	58/814/1,603	88/1,131/2,214	-	-	-	-	-
Vitis vinifera	[e]124/1,748/2,837	-	-	-	-	-	-

Supplemental Table 6. Proportion of Heat Shock Transcription Factors (HSF) gene family related to the polyploidizations in the genomes of 29-CGs.

Species name	HSF		
	Totality	WGD-related genes	Proportion
Benincasa hispida	21	10	48%
Citrullus amarus	19	8	42%
Cucurbita argyrosperma	25	18	72%
Citrullus colocynthis	20	8	40%
Citrullus ecirrhosus	19	9	47%
Cucumis hystrix	17	6	35%
Cucumis metuliferus	19	7	37%
Citrullus lanatus	17	8	47%
Citrullus lanatus cultivar	19	8	42%
Cucurbita maxima	26	22	85%
Cucumis melo	16	6	38%
Cucurbita moschata	27	27	100%
Citrullus mucsospermus	20	9	45%
Cucurbita pepo	24	18	75%
Citrullus rehmii	17	8	47%
Cucumis sativus	20	8	40%
Herpetospermum pedunculosum	16	7	44%
Luffa acutangula	17	6	35%
Luffa cylindrica	16	6	38%
Lagenaria siceraria	17	6	35%
Momordica charantia	16	7	44%
Sechium edule	37	28	76%
Trichosanthes anguina	17	9	53%

825 **Supplemental Table 7.** The 128 gene families with important biological functions in our platform.

Class	Gene family
bitter biosynthesis	C2C2-YABBY
	Cytochrome P450
	bHLH
	WRKY
Growth and development	Basic Helix-Loop-Helix Transcription Factor Gene Family
	MADS-box
	Homeobox
	TCP
	Type I MADS-box Gene Family
	GRAS
	HSF
	RAV
	GRF
	E2F-DP
	JUMONJI
	Wox
	Zinc finger-homeobox
	Orphan Transcription Factor
	AtRKD
	IDZ
	Trihelix
	MYB
	C2H2
	Basic region leucine zipper (bZIP) Transcription Factor Gene Family
	C2C2-Gata Transcription Factor Family

	MYB-Related Transcription Factor Family
	NAC Transcription Factor Family
	AP2-EREBP
	CPP
	B3
	C3H
	G2-like
	ZF-HD
	PHD Transcription Factor Family
	Single Myb Histone (SMH) Gene Family
	REM
	SBP-box
	Tify
	GeBP
	BBR/BPC
	EIL
	PHD
	C2C2-CO-like
	AS2
	C2C2-Dof
	CCAAT-HAP2
	CCAAT-HAP3
	CCAAT-HAP5
	CCAAT-DR1
	ARID
	Alfin-like
	AT-hook Motif Nuclear Localized
	Whirly

Secondary metabolite biosynthesis	HRT
	Trehalose Biosynthesis
	Monolignol Biosynthesis
	Cell Wall Biosynthesis
	Sucrose Transporter
	Carbohydrate Esterase
	Polysaccharide Lyase
	Glycoside Hydrolase
	Phytocyanin
	Xyloglucan Fucosyltransferase1
	Sulfurtransferasese
	Nicotianamine Synthase
	Aldehyde Dehydrogenase
	Glutathione S-transferase
	Plant Defensins
Biotic and abiotic stress response	VOZ-9
	Su(var)3-9
	RAD5 or RAD16-like Gene Family
	MLO family
	SUVH1/SUVH3-interacting DNAJ
	PRT Gene Family
	Receptor kinase-like
	Strubbelig Receptor
	Response Regulator Gene Family
	MAP Kinase Kinase
	MAP Kinase Kinase Kinase
	MAP Kinase Kinase Kinase Kinase
	AGC Kinase

	CBL-interacting serine-threonine protein kinases
	Antiporters
	Rapid Response to Elicitors
	Eukaryotic Initiation Factors
	Cytoplasmic Ribosomal Protein Gene Family
	Subtilisin-like Serine
	EF-hand Containing Proteins
	Calmodulin-binding Proteins
	U-box Gene Family
	Plant U-box Gene Family
	NADPH P450 reductases
	PP2C-type Phosphatase
	Protein tyrosine phosphatase
	14-3-3
	SNARE and Associated Proteins
	Expansin
	RCI2 Gene Family
	Structural Maintenance of Chromosomes
	Major Intrinsic Protein MIP
	Ion Channels
	Inorganic Solute Cotransporters
	Organic Solute Cotransporters
	Antiporter Super Family
Cold tolerance	Amino Acid or Auxin Permease AAAP
	MRS2 Gene Family
	Primary Pumps ATPases
	Primary Pumps ATPases(2)
	Cytoskeleton Gene Families

	Myosins
	IQD Protein Family
	FH2 Protein (Formin)
	COBRA
	TUB Transcription Factor
	EXO70 exocyst subunit
	Non-SMC Element
	Core Cell Cycle
	Core DNA Replication Machinery
Tolerance	Pollen Coat Proteome
	ARIADNE
	TRM Superfamily
	Leucine-rich Repeat extension Gene Families
	Miscellaneous Gene Families
	Histidine Kinase
	Mechanosensitive Ion Channel
	MAP Kinase Gene Family
	Chloroplast and Mitochondria Gene Families
	Glutamine Dumper

826 **Supplemental Table 8.** Overview of transcriptome data in the platform.

species	Tissues	Number
Hemsleya chinensis	root	3
	stem	6
	leaf	12
	tuber	12
Begonia darthvaderiana	leaf	5
	root	1

	stem	1
	seeds	24
	leaf	52
	placenta	27
Benincasa hispida	mesocarp	31
	root	16
	flower	3
	fruit	57
	pericarp	7
	bud	2
	flower	5
Begonia loranthoides	leaf	62
	root	4
	ste	5
	flower	3
Begonia masoniana	leaf	6
	flower	3
Begonia peltatifolia	leaf	6
	root	3
	stem	3
	flower	3
Cucurbita argyrosperma	leaf	2
Cucumis hystrix	leaves	10
	flesh	27
	leaf	19
Citrullus lanatus	leaves	5
	root	79
	fruit	273

Cucurbita maxima	flesh	7
	fruit	120
	leaf	13
	phloem	3
	root	27
Cucumis melo	anther	2
	carpel	13
	epicarp	5
	flesh	5
	flower	2
	fruit	263
	leaf	43
	leaves	120
	mesocarp	20
	peel	19
	root	7
	seed	3
	shoot	7
	stem	3
Cucurbita moschata	flesh	10
	fruit	117
	leaf	37
	leaves	29
	root	75
Cucurbita pepo	cotyledons	2
	fruit	61
	leaf	25
	mesocarp	31

Cucumis sativus	root	216
	seed	56
	shoot	2
	cotyledons	7
	flesh	30
	fruit	93
	leaf	202
	leaves	115
	meristem	5
	mesocarp	13
	peel	22
	pericarp	13
Lagenaria siceraria	root	184
	shoot	5
	skin	7
	flesh	6
	flower	145
	fruit	6
	leaf	8
	leaves	2
	root	30
Coriaria nepalensis	root	14
Luffa cylindrica	fruit	21
	leaf	3
Sechium edule	leaf	2
Vitis vinifera	bud	212
	flower	221
	fruit	131

	leaves	3,434
	root	53
	stem	5

827

Supplemental References

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Research **43**:W39-W49. 10.1093/nar/gkv416.

Bandi, V., Gutwin, C., Siri, J.N., Neufeld, E., Sharpe, A., and Parkin, I. (2022). Visualization Tools for Genomic Conservation. Methods in molecular biology (Clifton, N.J.) **2443**:285-308. 10.1007/978-1-0716-2067-0_16.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. Nucleic Acids Research **49**:D344-D354. 10.1093/nar/gkaa977.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**:2114-2120. 10.1093/bioinformatics/btu170.

Buchfink, B., Xie, C., and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nature Methods **12**:59-60. 10.1038/nmeth.3176.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Ami, G.O.H., and Web Presence Working, G. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics **25**:288-289. 10.1093/bioinformatics/btn615.

Caspermeyer, J. (2018). Exploring an Ancient Event in Pumpkin, Gourd, and Melon Evolution. Mol Biol Evol **35**:256-257. 10.1093/molbev/msx291.

Diesh, C., Stevens, G.J., Xie, P., De Jesus Martinez, T., Hershberg, E.A., Leung, A., Guo, E., Dider, S., Zhang, J., Bridge, C., et al. (2023). JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol **24**:74. 10.1186/s13059-023-02914-z.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol **20**:238. 10.1186/s13059-019-1832-y.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences **117**:9451-9457. 10.1073/pnas.1921046117.

Gui, S., Martinez-Rivas, F.J., Wen, W., Meng, M., Yan, J., Usadel, B., and Fernie, A.R. (2023). Going broad and deep: sequencing-driven insights into plant physiology, evolution, and crop domestication. The Plant journal : for cell and molecular biology **113**:446-459. 10.1111/tpj.16070.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59**:307-321. 10.1093/sysbio/syq010.

Guo, J., Xu, W., Hu, Y., Huang, J., Zhao, Y., Zhang, L., Huang, C.H., and Ma, H. (2020). Phylotranscriptomics in Cucurbitaceae Reveal Multiple Whole-Genome Duplications and Key Morphological and Molecular Innovations. Mol Plant **13**:1117-1133. 10.1016/j.molp.2020.05.011.

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. 10.1126/science.aat4096.

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics **31**:1296-1297. 10.1093/bioinformatics/btu817.

Hu, G., and Kurgan, L. (2019). Sequence Similarity Searching. Current protocols in protein science **95**:e71. 10.1002/cpps.71.

Jiao, Y., Li, J., Tang, H., and Paterson, A.H. (2014). Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell **26**:2792-2802. 10.1105/tpc.114.127597.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature **473**:97-100. 10.1038/nature09916.

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. 10.1093/bioinformatics/btu031.

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol **428**:726-731. 10.1016/j.jmb.2015.11.006.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology **37**:907-915. 10.1038/s41587-019-0201-4.

Kong, X., Zhang, Y., Wang, Z., Bao, S., Feng, Y., Wang, J., Yu, Z., Long, F., Xiao, Z., Hao, Y., et al. (2023a). Two-step model of paleohexaploidy, ancestral genome reshuffling and plasticity of heat shock response in Asteraceae. Hortic Res **10**:uhad073. 10.1093/hr/uhad073.

Kong, X., Zhang, Y., Wang, Z., Bao, S., Feng, Y., Wang, J., Yu, Z., Long, F., Xiao, Z., Hao, Y., et al. (2023b). Two-step model of paleohexaploidy, ancestral genome reshuffling and plasticity of heat shock response in Asteraceae. Horticulture research **10**:uhad073. 10.1093/hr/uhad073.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078-2079. 10.1093/bioinformatics/btp352.

Manna, M., Thakur, T., Chirom, O., Mandlik, R., Deshmukh, R., and Salvi, P. (2021). Transcription factors as key molecular target to strengthen the drought stress tolerance in plants. Physiol Plant **172**:847-868. 10.1111/ppl.13268.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, Gustavo A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Research **49**:D412-D419. 10.1093/nar/gkaa913.

Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J.N., and Chourrout, D. (2019). Massive Changes of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. Curr Biol **29**:1161-1168 e1166. 10.1016/j.cub.2019.01.080.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution **32**:268-274. 10.1093/molbev/msu300.

Nishihara, H. (2020). Transposable elements as genetic accelerators of evolution: contribution to genome size, gene regulatory network rewiring and morphological innovation. Genes Genet Syst **94**:269-281. 10.1266/ggs.19-00029.

One Thousand Plant Transcriptomes, I. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature **574**:679-685. 10.1038/s41586-019-1693-2.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. Nucleic Acids Res **51**:D418-D427. 10.1093/nar/gkac993.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature protocols **11**:1650-1667. 10.1038/nprot.2016.095.

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Research **46**:W200-W204. 10.1093/nar/gky448.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol **26**:1641-1650. 10.1093/molbev/msp077.

Priyam, A., Woodcroft, B.J., Rai, V., Moghul, I., Munagala, A., Ter, F., Chowdhary, H., Pieniak, I., Maynard, L.J., Gibbins, M.A., et al. (2019). Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. Mol Biol Evol **36**:2922-2924. 10.1093/molbev/msz185.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., and Paterson, A.H. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome biology **20**:38. 10.1186/s13059-019-1650-2.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res **33**:W116-120. 10.1093/nar/gki442.

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., Zeng, J., Zhou, Q., Wang, S., Gu, W., et al. (2014). Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. Science **346**:1084-1088. 10.1126/science.1259215.

Shil, P., Dudani, N., and Vidyasagar, P.B. (2006). ISHAN: sequence homology analysis package. In silico biology **6**:373-377.

Sujuan, Y., Asaithambi, A., and Liu, Y. (2008). CpGIF: an algorithm for the identification of CpG islands. Bioinformation **2**:335-338. 10.6026/97320630002335.

Tarailo‐Graovac, M., and Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Current Protocols in Bioinformatics **25**10.1002/0471250953.bi0410s25.

Teplá, M., and Klímová, H. (2015). Using Adobe Flash animations of electron transport chain to teach and learn biochemistry. Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology **43**:294-299. 10.1002/bmb.20867.

Usha, T., Middha, S.K., Babu, D., Goyal, A.K., Das, A.J., Saini, D., Sarangi, A., Krishnamurthy, V., Prasannakumar, M.K., Saini, D.K., et al. (2022). Hybrid Assembly and Annotation of the Genome of the Indian Punica granatum, a Superfood. Front Genet **13**:786825. 10.3389/fgene.2022.786825.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics **8**:77-80. 10.1016/s1672-0229(10)60008-3.

Wang, J., Ding, Y., Li, Y., Gao, X., Kong, X., Long, F., Feng, Y., Zhang, Y., Li, Y., Yu, Z., et al. (2024). Allopolyploidization events and immense paleogenome reshuffling underlying the diversification of plants and secondary metabolites in Oleaceae. Journal of Systematics and Evolution **n/a**https://doi.org/10.1111/jse.13116.

Wang, J., Yuan, M., Feng, Y., Zhang, Y., Bao, S., Hao, Y., Ding, Y., Gao, X., Yu, Z., Xu, Q., et al. (2022). A common whole-genome paleotetraploidization in Cucurbitales. Plant Physiol **190**:2430-2448. 10.1093/plphys/kiac410.

Wang, J., Yu, J., Sun, P., Li, C., Song, X., Lei, T., Li, Y., Yuan, J., Sun, S., Ding, H., et al. (2020). Paleo-polyploidization in Lycophytes. Genomics, proteomics & bioinformatics **18**:333-340. 10.1016/j.gpb.2020.10.002.

Wang, J., Yuan, J., Yu, J., Meng, F., Sun, P., Li, Y., Yang, N., Wang, Z., Pan, Y., Ge, W., et al. (2019). Recursive Paleohexaploidization Shaped the Durian Genome. Plant Physiol **179**:209-219. 10.1104/pp.18.00921.

Wang, J., Sun, P., Li, Y., Liu, Y., Yu, J., Ma, X., Sun, S., Yang, N., Xia, R., Lei, T., et al. (2017). Hierarchically Aligning 10 Legume Genomes Establishes a Family-Level Genomics Platform. Plant Physiol **174**:284-300. 10.1104/pp.16.01981.

Wang, J., Sun, P., Li, Y., Liu, Y., Yang, N., Yu, J., Ma, X., Sun, S., Xia, R., Liu, X., et al. (2018). An Overlooked Paleotetraploidization in Cucurbitaceae. Mol Biol Evol **35**:16-26. 10.1093/molbev/msx242.

Wang, X., Shi, X., Li, Z., Zhu, Q., Kong, L., Tang, W., Ge, S., and Luo, J. (2006). Statistical inference of chromosomal homology based on gene colinearity and applications to Arabidopsis and rice. BMC bioinformatics **7**:447. 10.1186/1471-2105-7-447.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research **40**:e49. 10.1093/nar/gkr1293.

Weidemuller, P., Kholmatov, M., Petsalaki, E., and Zaugg, J.B. (2021). Transcription factors: Bridge between cell signaling and gene regulation. Proteomics **21**:e2000034. 10.1002/pmic.202000034.

Wu, S., Han, B., and Jiao, Y. (2020). Genetic Contribution of Paleopolyploidy to Adaptive Evolution in Angiosperms. Mol Plant **13**:59-71. 10.1016/j.molp.2019.10.012.

Xie, D., Xu, Y., Wang, J., Liu, W., Zhou, Q., Luo, S., Huang, W., He, X., Li, Q., Peng, Q., et al. (2019). The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat Commun **10**:5158. 10.1038/s41467-019-13185-3.

Yu, J., Wu, S., Sun, H., Wang, X., Tang, X., Guo, S., Zhang, Z., Huang, S., Xu, Y., Weng, Y., et al. (2023). CuGenDBv2: an updated database for cucurbit genomics. Nucleic Acids Res **51**:D1457-D1464. 10.1093/nar/gkac921.

Yu, Y., Ouyang, Y., and Yao, W. (2018). shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics (Oxford, England) **34**:1229-1231. 10.1093/bioinformatics/btx763.

Yu, Z., Li, Y., Song, T., Gou, L., Wang, J., Ding, Y., Xiao, Z., Qin, J., Jiang, H., Zhang, Y., et al. (2025). OGRP: a comprehensive bioinformatics platform for the efficient empowerment of Oleaceae genomics research. Horticultural Plant Journal https://doi.org/10.1016/j.hpj.2024.09.006.

Yue, H., Nie, X., Yan, Z., and Weining, S. (2019). N6-methyladenosine regulatory machinery in plants: composition, function and evolution. Plant biotechnology journal **17**:1194-1208. 10.1111/pbi.13149.

Zhang, S., Tan, F.Q., Chung, C.H., Slavkovic, F., Devani, R.S., Troadec, C., Marcel, F., Morin, H., Camps, C., Gomez Roldan, M.V., et al. (2022). The control of carpel determinacy pathway leads to sex determination <b>in cucurbits</b&gt. Science **378**:543-549. 10.1126/science.add4250.

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X., and Dai, L. (2012). ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochem Biophys Res Commun **419**:779-781. 10.1016/j.bbrc.2012.02.101.

Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., et al. (2016a). iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. Molecular plant **9**:1667-1670. 10.1016/j.molp.2016.09.014.

Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., et al. (2016b). iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. Mol Plant **9**:1667-1670. 10.1016/j.molp.2016.09.014.

Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., Lin, T., Liu, Z., Zeng, K., Zhong, Y., et al. (2016). Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. Nat Plants **2**:16183. 10.1038/nplants.2016.183.