





# Genomic, transcriptomic, and metabolomic analysis of *Oldenlandia corymbosa* reveals the biosynthesis and mode of action of anti-cancer metabolites<sup>oo</sup>

Irene Julca<sup>1</sup> , Daniela Mutwil-Anderwald<sup>1</sup> , Vaishnavi Manoj<sup>1</sup>, Zahra Khan<sup>1</sup>, Soak Kuan Lai<sup>1</sup>, Lay K. Yang<sup>2</sup>, Ing T. Beh<sup>1</sup>, Jerzy Dziekan<sup>1</sup>, Yoon P. Lim<sup>3</sup>, Shen K. Lim<sup>1,3</sup>, Yee W. Low<sup>4</sup>, Yuen I. Lam<sup>1</sup>, Seth Tjia<sup>1</sup>, Yuguang Mu<sup>1</sup>, Qiao W. Tan<sup>1</sup> , Przemyslaw Nuc<sup>5</sup>, Le M. Choo<sup>4</sup>, Gillian Khew<sup>1,4</sup>, Loo Shining<sup>1</sup>, Antony Kam<sup>1</sup>, James P. Tam<sup>1</sup>, Zbynek Bozdech<sup>1</sup>, Maximilian Schmidt<sup>6</sup>, Bjoern Usadel<sup>6</sup>, Yoganathan Kanagasundaram<sup>2</sup>, Saleh Alseekh<sup>7,8</sup>, Alisdair Fernie<sup>7,8</sup>, Hoi Y. Li<sup>1</sup> and Marek Mutwil<sup>1\*</sup> 

1. School of Biological Sciences, Nanyang Technological University, Singapore 639798, Singapore
2. Shared Analytics, Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A\*STAR), Singapore 138671, Singapore
3. Department of Biochemistry, National University of Singapore, Singapore 117596, Singapore
4. Singapore Botanic Gardens, Singapore 259569, Singapore
5. Department of Gene Expression, Faculty of Biology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan 61-614, Poland
6. IBG-4 Bioinformatics, Forschungszentrum Jülich, Jülich 52428, Germany
7. Max-Planck-Institut für Molekulare Pflanzenphysiologie, Potsdam-Golm 14476, Germany
8. Center of Plant Systems Biology and Biotechnology, Plovdiv 4000, Bulgaria

\*Correspondence: Marek Mutwil ([mutwil@ntu.edu.sg](mailto:mutwil@ntu.edu.sg))



Irene Julca



Marek Mutwil

anti-cancer activity. Our results show that *O. corymbosa* extracts are active against breast cancer cell lines and that ursolic acid is responsible for this activity. Moreover, we assembled a high-quality genome and uncovered two genes involved in the biosynthesis of ursolic acid. Finally, we also revealed that ursolic acid causes mitotic catastrophe in cancer cells and identified three high-confidence protein binding targets by Cellular Thermal Shift Assay (CETSA) and reverse docking. Altogether, these results constitute a valuable resource to further characterize the biosynthesis of active metabolites in the *Oldenlandia* group, while the mode of action of ursolic acid will allow us to further develop this valuable compound.

Keywords: genome, medicinal, metabolomics, mode of action, transcriptome

Julca, I., Mutwil-Anderwald, D., Manoj, V., Khan, Z., Lai, S.K., Yang, L.K., Beh, I.T., Dziekan, J., Lim, Y.P., Lim, S.K., Low, Y. W., Lam, Y.I., Tjia, S., Mu, Y., Tan, Q.W., Nuc, P., Choo, L.M., Khew, G., Shining, L., Kam, A., Tam, J.P., Bozdech, Z., Schmidt, M., Usadel, B., Kanagasundaram, Y., Alseekh, S., Fernie, A., Li, H.Y., and Mutwil, M. (2023). Genomic, transcriptomic, and metabolomic analysis of *Oldenlandia corymbosa* reveals the biosynthesis and mode of action of anti-cancer metabolites. *J. Integr. Plant Biol.* **65**: 1442–1466.

## ABSTRACT

Plants accumulate a vast array of secondary metabolites, which constitute a natural resource for pharmaceuticals. *Oldenlandia corymbosa* belongs to the Rubiaceae family, and has been used in traditional medicine to treat different diseases, including cancer. However, the active metabolites of the plant, their biosynthetic pathway and mode of action in cancer are unknown. To fill these gaps, we exposed this plant to eight different stress conditions and combined different omics data capturing gene expression, metabolic profiles, and

## INTRODUCTION

Plants synthesize a vast array of secondary metabolites as part of their defense against pathogens or during stress conditions and are thus a natural resource for pharmaceuticals. This has been acknowledged already thousands of years ago by ancient medicinal systems, such as traditional Chinese medicine (TCM) or Ayurveda. In the past years, massive efforts were made by researchers worldwide to mine this potential for new antibacterial or anti-tumor drugs (Silva and Fernandes Júnior, 2010; Dhama et al., 2014; Yan et al., 2017; El-Seedi et al., 2021). However, most studies are limited to screening the plants for activities and identifying candidates for the active metabolites without knowing the mode of action or the enzyme-coding genes. The major reason for this shortcoming is that the genomes of medicinal plants are usually not sequenced, and even if the genome sequence is known, the successful prediction of which enzyme-coding genes synthesize a given metabolite requires advanced computational methods applied to comprehensive datasets.

Plants already constitute a significant source of anti-cancer drugs approved by government health agencies, such as the U.S. FDA (Krause and Tobin, 2013; Beutler, 2019). The most well-known ones are taxol (paclitaxel), vinca alkaloids (vinblastine, vincristine), and camptothecin derivatives. The advantage of using natural products over chemotherapeutics is their effect on multiple signaling pathways and molecular targets. For example, they induce apoptosis, inhibit proliferation, and suppress metastasis, while causing few adverse effects (Yan et al., 2017). In contrast, classical chemotherapeutics are often single-targeted and therefore more prone to drug resistance (Sauter, 2020). A recurring problem with medicinal herb extracts is that neither the active agents (and other molecular components of these mixtures) nor their modes of action are well-characterized. Additionally, the composition of the formulas is complex and not standardized, thus not fulfilling FDA guidelines for drug approval.

The *Oldenlandia-Hedyotis* complex of the family Rubiaceae comprises about 500 species found throughout the tropics. It is a highly polyphyletic group, and recent molecular phylogenetic studies suggest that it should be split into *Hedyotis* (robust, shrub-like species) and *Oldenlandia* (small herbs with paniculate or corymbose inflorescences) (Guo et al., 2013). *Oldenlandia corymbosa* is a commonly used herb in China and India for health benefits and wellbeing. The leaves are used to treat multiple ailments, such as sore eyes, anthelmintic, anti-rheumatic, depurative, digestive, diuretic, pectoral, fever, jaundice, child birth, and stomachic medicine (Pranabesh, 2019), while the roots are reported to have vermifuge properties (Nguyen, 1993). Previous pharmacological studies reported that *O. corymbosa* has hepatoprotective (Sadasivan et al., 2006; Modi and Shah, 2017), antimalarial (Mishra et al., 2009), anti-inflammatory, and antioxidant properties (Lin et al., 2002). There are a few reports on anti-tumor activities of *O. corymbosa* against skin cancer, leukemia, and liver cancer

cells (You et al., 2011; Pandey et al., 2012; Yue et al., 2012). *O. corymbosa* is often the main ingredient in TCM herbal formulations such as the popular antitumor Peh-Hue-Juwa-Chi-Cao medicine together with *Oldenlandia diffusa* and other herbs (Lin et al., 2002; Lajis and Ahmad, 2006).

Until these studies, *O. corymbosa* was long regarded as inferior when compared to the more widely studied *O. diffusa*, which has been used for the treatment of inflammation-linked diseases, such as hepatitis, appendicitis, and urethritis in TCM (Chen et al., 2016). *O. diffusa* also possesses strong anti-tumor activities against liver, lung, prostate, and ovarian cancers and is already used for clinical colon and breast cancer treatments in Taiwan (Yeh et al., 2014). For example, *O. diffusa* ethanol extracts suppress the proliferation of colorectal cancer cells and induce cell apoptosis mediated by the suppression of the STAT3 pathway in mouse xenograft models and in human HT-29 cells (Cai et al., 2012; Lin et al., 2015). In breast cancer, *O. diffusa* is often used together with *Scutellaria barbata*, and treatment with these extracts results in inhibited proliferation and migration of three types of breast cancer cells *in vitro* as well as tumor growth in nude mice by apoptosis (Ma et al., 2020). *O. diffusa* extracts also exert antiproliferative and apoptotic effects on human breast cancer cells through ER $\alpha$ /Sp1-mediated p53 activation (Gu et al., 2012). Around 171 pharmacologically active compounds of *O. diffusa* have been reported (Chen et al., 2016; Wang et al., 2017). Thus, both *O. diffusa* and *O. corymbosa* have documented medicinal properties, but the identity, biosynthesis, and mode of action of the bioactive metabolites are still largely unknown.

Considering that natural products are increasingly popular in cancer therapy, we set out to identify the main anti-cancer metabolites and their biosynthetic pathways of the *Oldenlandia* genus. To this end, we collected 11 plants comprising *O. corymbosa*, *O. biflora*, and *O. tenelliflora*, and show that *O. corymbosa* had the most potent activity against several breast cancer cell lines. We provide a gene functional resource for *O. corymbosa*, by generating a high-quality genome sequence, which revealed a highly homozygous genome caused by autogamy. We used activity-guided fractionation to reveal that ursolic acid is the main metabolite responsible for anti-cancer activity, with a minor contribution of oleanolic acid, lutein, phytol, and pheophorbide a. The gene expression and metabolomic analysis of *O. corymbosa* organs showed that the biosynthesis of these metabolites can be influenced by abiotic stress. To study the mechanism of the anti-cancer activity of ursolic acid we show that it causes a mitotic catastrophe in the breast cancer cell lines. Finally, we identified three high-confidence human protein targets of ursolic acid by a combination of Cellular Thermal Shift Assay (CETSA) and reverse docking.

## RESULTS

### Collection of *Oldenlandia* species and establishment of *O. corymbosa* in the laboratory

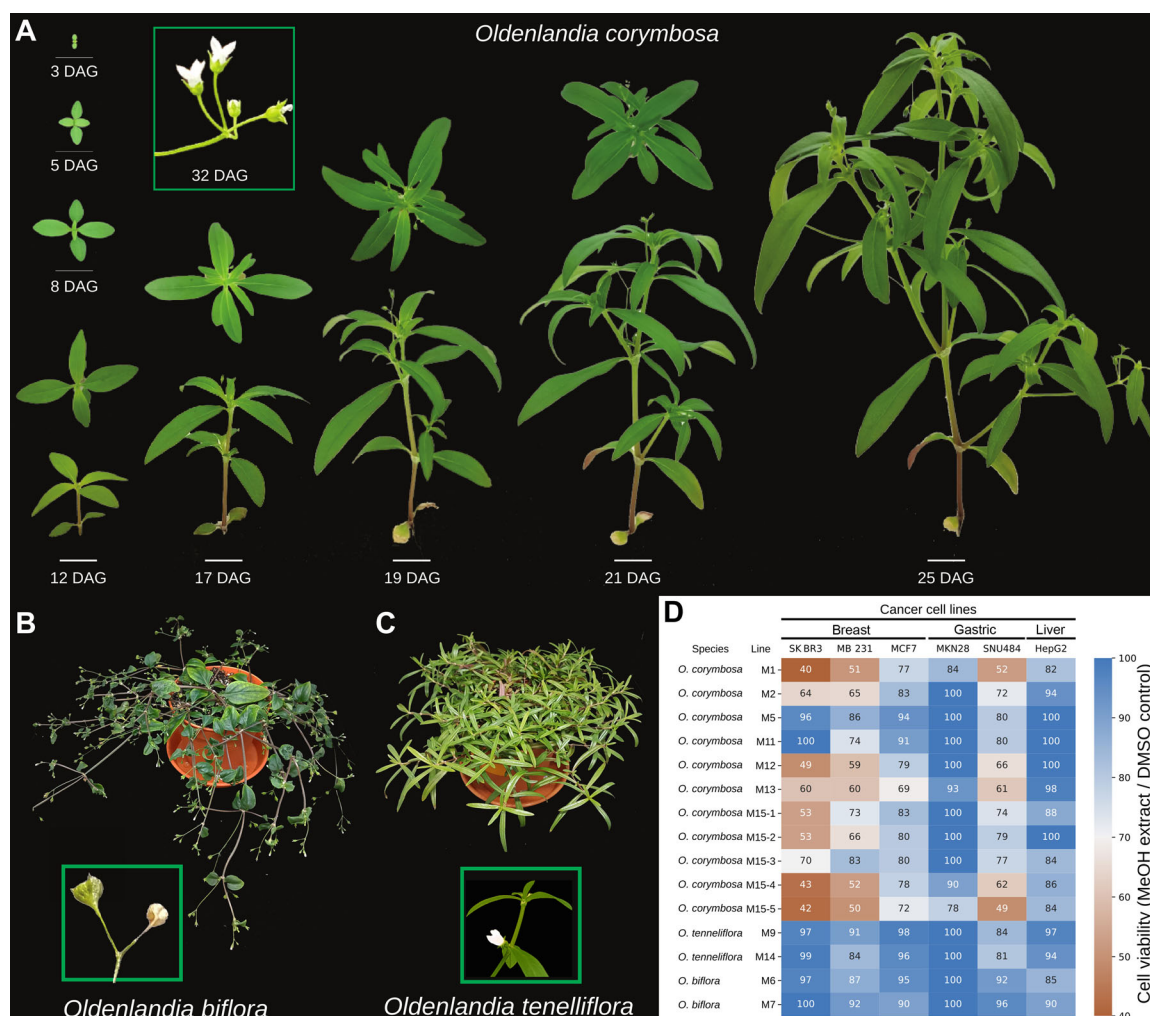
To study and compare the anti-tumor activities of different *Oldenlandia* species, we collected 10 field-grown plants from the

Bedok area in Singapore and one plant from a private residential area (Table S1). The plants were transferred to the lab, grown on a typical soil substrate (7:2:1, soil:vermiculite:perlite), and propagated by selfing or clonal propagation. Then we classified the plants by morphological characteristics (Wong et al., 2019). *O. corymbosa* (M1, M2, M5, M11-13, M15) possesses lanceolate or elliptic leaves, the stems are branching near the base, and the inflorescences have 2–5 flowered corymbs or umbels (Figure 1A). *O. biflora* (M6, M7) leaves are elliptic-ovate, the flowers are cymose, 2—many flowered and situated terminal or in the axils of uppermost leaves, and the seed capsules are oblate (Figure 1B). *O. tenelliflora* (M9, M14) is characterized by having sessile, solitary flowers, and laminar linear or linear-lanceolate leaves (Figure 1C).

To select the species with anti-tumor activity, we tested methanol leaf extracts of the 11 plants against six cancer cell lines: breast cancer (SK-BR3, MB-231, MCF7), gastric cancer

(MKN28, SNU484), and liver cancer (HepG2). To measure the cytotoxic activity of the methanol extracts we used the MTS cell viability assay which measures the metabolic activity of cancer cells, where lower metabolic activity is indicative of cell death (Cory et al., 1991). Nearly all leaf extracts from *O. corymbosa* plants showed strong activity against SK-BR3 (~40% cell viability, Figure 1D; Table S2), a somewhat lower activity (~50%) against MB231 and SNU484, whereas MKN28, HepG2, and MCF7 were not or only modestly affected. The anti-tumor activity of *O. corymbosa* persisted after selfing the plant (Figure 1D, plants M15-1 to M15-5). Conversely, the extracts from *O. tenelliflora* and *O. biflora* showed modest or no activity against all six cancer lines (Figure 1D).

In addition to the anti-cancer activity, *O. corymbosa* grew rapidly, with the first flowers emerging approximately 19 d after germination. The plant was readily self-fertilizing, able to be clonally propagated, and grew well in soil, liquid, and solid media



**Figure 1. Identification and anti-tumor activities of *Oldenlandia* species**

(A) Growth of *O. corymbosa* over 25 d after germination (DAG). The images are taken from the top (DAG 3–8) and from the side (DAG 12–25). (B) Morphological characteristics of *Oldenlandia biflora*. The bifurcating flowers are shown in the green box. (C) *O. tenelliflora*, with the flower (green box). (D) MTS assays showing the metabolic activity (proxy of viability) of six cancer cell lines (columns) and the various lines of *O. corymbosa* (M1, M2, M5, M11, M12, M13, F1 plants of M15), *O. tenelliflora* (M9, M14) and *O. biflora* (M6, M7). Values were capped at 100.

as per *Arabidopsis* protocols (Figure S1). The seeds were still viable after 3 years of storage (45% humidity, 23°C). Thus, in addition to the anti-cancer activity, *O. corymbosa* shows all characteristics of a model plant for the *Oldenlandia* genus.

To compare *O. corymbosa* to other medicinal plants regarding their activity against breast cancer, we performed 86 methanol extractions of 61 plants grown in Nanyang Technological University Herb Garden against MB231, SK-BR3, and MCF7. Some of the extracts showed no activity (*Polygonum chinense*, *Rhoeo discolor*), or indiscriminately strong activity against all three cancer cell lines (*Piper betle*, *Citrus hybrid*, *Vitex rotundifolia*, *Wedelia chinensis*, *Clerodendratum spicatum*) (Figure S2). In contrast, *O. corymbosa* M15 plants showed more selective activity against SK-BR3 (Figure S2), suggesting that the plant extracts may target specific characteristics of this cancer line. Thus, we selected *O. corymbosa* and SK-BR3 as our model plant and cancer cell line, respectively.

### The highly homozygous genome of *Oldenlandia corymbosa*

To enable molecular studies of *O. corymbosa* and to identify the biosynthetic pathways of the anti-tumor metabolites, we produced 79Gbp of long-read sequencing data from Oxford Nanopore Technologies (ONT) and 68Gbp of short reads from Illumina. The estimated genome size inferred from the *k*-mer frequency was 272.8Mbp (Figure S3A, B). The final assembly has 10 contigs (N50 = 31Mbp, Figure 2A), with a size of 273.7Mbp (Table 1), which represents 100% of the estimated genome size. The *k*-mer analysis also reveals high completeness of the genome (99.8%, Figure S3C) with a consensus quality value (QV) of 39.02. The gene completeness of the assembly as determined by BUSCO (Benchmarking Universal Single-Copy Ortholog), shows 95.7% completeness (Table 1). This indicates high completeness, comparable to other recently sequenced genomes (Guo et al., 2021; Wickell et al., 2021). Moreover, *O. corymbosa* has a haploid chromosome number  $n = 9$  (Rice et al., 2015) probably represented in the nine largest contigs, which covered the 99.5% of the genome. In addition, we assembled the plastid and mitochondrial genomes of *O. corymbosa*. The plastid genome comprised 152,414 base pairs (bp), in agreement with previously reported *Oldenlandia* and *Hedyotis* plastid sequences, which range from 152,327 to 154,560 bp (Zhang et al., 2019; Yik et al., 2021) (Table S3). The mitochondrial genome has a size of 258,274 bp, which is smaller than other species of the family such as *Damnacanthus indicus* (417,816 bp) (Han et al., 2021) and *Scyphiphora hydrophyllacea* (354,155 bp) (Chen et al., 2020).

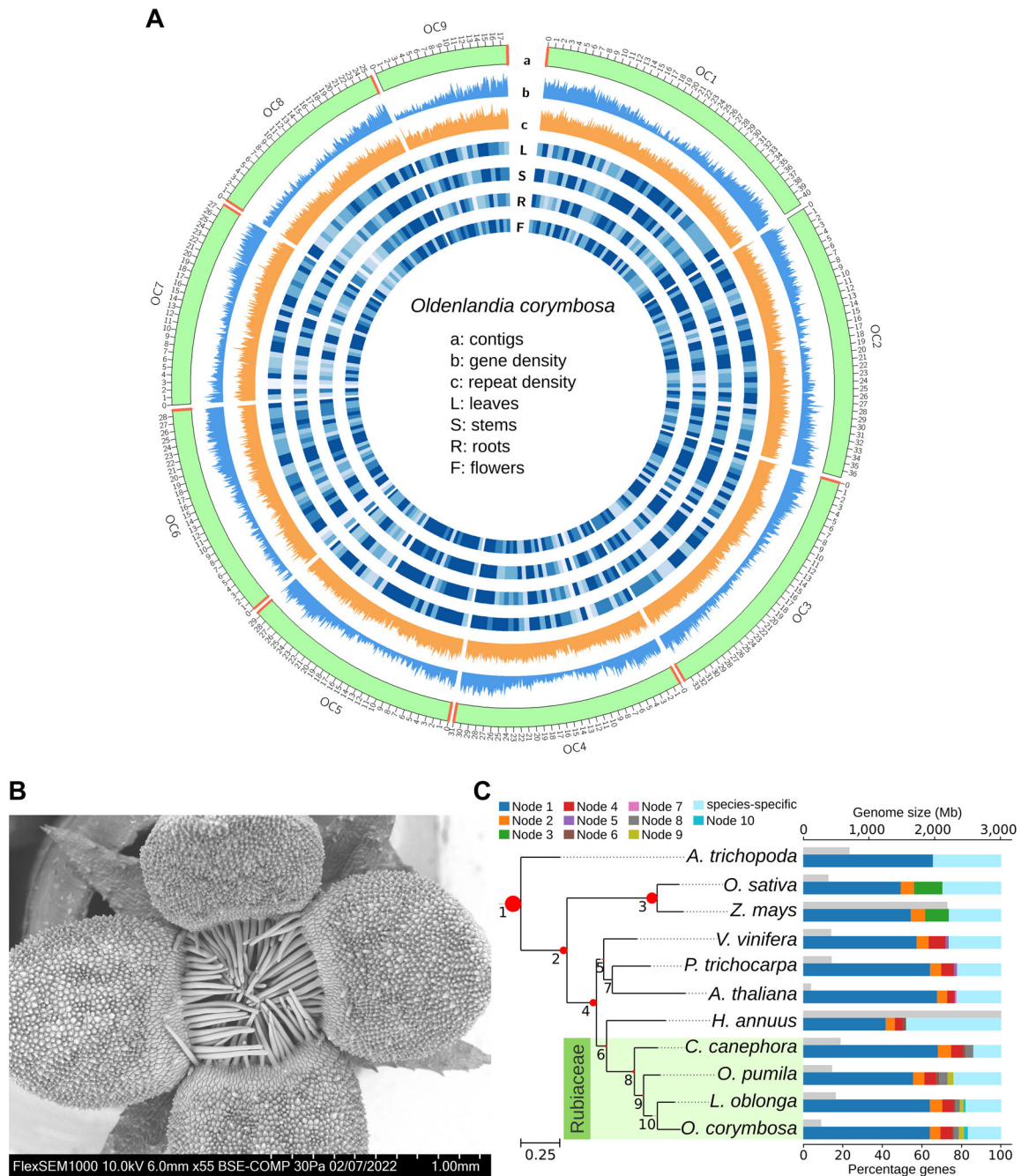
To identify fully assembled chromosomes, we searched telomere repeats (5'-CCCTAAA-3') (Richards and Ausubel, 1988) at both ends of the contigs and observed that contigs OC3-OC6, OC8 likely represent full chromosomes, as the telomeric sequences were observed on both ends (Figure 2A a, telomere sequences indicated by red lines). Conversely, most of the other contigs contained at least one terminal telomere sequence. Altogether, these results show that our genome assembly has high continuity and completeness.

Surprisingly for a weed plant that was only selfed for one generation in the lab, the genome is highly homozygous as predicted by the *k*-mer analysis (99.9%, Figure S3C). Moreover, we only detected 3,912 heterozygous SNPs (Table 1, Figure S3B, C). To investigate this further, we performed a Scanning Electron Microscopy (SEM) imaging of mature flowers and observed that the entrance to the flower is blocked by a ring of pubescent hairs (Figure 2B). This characteristic may promote autogamy, which is reflected in the homozygosity observed in the genome.

The combined size of all repeat sequences identified in the genome of *O. corymbosa* is 92.6 Mb (33.7% of the total genome length, Figure 2; Table S4). These repeats mainly consisted of retroelements (49 Mb, 53%) and unclassified interspersed repeats (24 Mb, 26%). The total number of genes annotated is 25,007 based on de novo prediction, homology search, and mRNA-seq assisted prediction (Table 1). When we compared the repeat and gene density, we observed a general trend of gene density increasing and repeat density decreasing in some regions of the complete chromosomes (Figure 2A b, c), which is typical for most flowering genomes (Schmutz et al., 2010). Additionally, we predicted functions for 20,825 genes (83%), including GO terms and MapMan bins to 15,402, and 12,505 protein-coding genes, respectively (Figure S3D). For the plastid genome, we annotated 131 genes, of which 86 are protein-coding genes, 37 are transfer RNAs (tRNAs), and eight are ribosomal RNAs (rRNAs) (Figure S4A). Finally, for the mitochondrial genome, we annotated 31 protein-coding genes, 17 tRNAs, and three rRNAs.

To study the evolutionary history of *O. corymbosa*, we compare its nuclear genome with another 10 plant species and its plastid genome with another four *Oldenlandia*-*Hedyotis* species. The number of nuclear genes identified in *O. corymbosa* is comparable to other Rubiaceae species (Table S5) and the number of plastid genes is comparable to other *Oldenlandia* species (Table S3). Also, for *O. corymbosa*, 20,798 nuclear genes (83%) have homologs in the other plant species, while for the other species this percentage ranges from 52% to 86% (Figure 2C). We reconstructed the evolutionary relationship of these 10 species using a concatenated approach with 667 single-copy orthologs, which resulted in a highly supported topology (Figure 2C), congruent with previous studies (Guo et al., 2021). Additionally, we reconstructed a phylogeny using the plastid genomes of *Oldenlandia*-*Hedyotis* species and *Coffea canephora* as an outgroup, which shows that our sample (M15) is sister to the other *O. corymbosa* (Yik et al., 2021), confirming that our plant is *O. corymbosa* (Figure S4B).

To test whether *O. corymbosa* underwent a whole genome duplication (WGD), we reconstructed gene trees for each orthogroup and obtained all orthology and paralogy relationships based on phylogenetic evidence (Gabaldón, 2008; Huerta-Cepas et al., 2016). Then, we focused on *O. corymbosa* gene duplications and found 1,969 events, from which the majority (1,737—88%) result in a moderate number of paralogs (2–5 in-paralogs, Figure S5A) and only a few (104%–105%) represent large gene family expansions



**Figure 2. Genome characteristics of *Oldenlandia corymbosa***

**(A)** Circos plot of the nine largest contigs (green graph), gene density (blue graph), and repeat density (orange graph). The red regions in the contigs show the putative telomeres. The heatmaps in blue, in order, show the expression of genes in leaves, stems, roots, and flowers. **(B)** Scanning electron microscopy (SEM) image of the opening of the flower. The scale bar (500  $\mu$ m) is indicated by white ticks. **(C)** Species tree of *O. corymbosa* and the other 10 plant species. All bootstrap values in the graph are maximal (100). The numbers in the nodes indicate the node number (e.g., 1: Node 1). The red circles in the tree show the percentage of orthogroups found in each node. Color bars on the right show the percentage of genes per species that are present in each node. Grey bars on the right show the genome size (Mbp). All species of the family Rubiaceae are highlighted in green.

( $\geq 10$  in-paralogs). To detect waves of WGDs in the evolutionary history of this species, we computed the ratios of gene duplication (average number of duplications detected in a given branch of the species tree) excluding large expansions ( $>5$  in-paralogs). This analysis showed small values (0.06–0.35) in the branches of the Rubiaceae family

(Figure S5B), which suggest that *O. corymbosa* did not undergo a recent WGD event (Julca et al., 2018; Alioto et al., 2020). Additionally, we used the MCScanX toolkit (Wang et al., 2012) to classify all *O. corymbosa* duplicated genes into WGD/segmental, tandem, proximal, and dispersed duplicates. This analysis shows that from the total duplicated

**Table 1. Genome characteristics of *Oldenlandia corymbosa***

Genome size	273,711,991
Contigs	10
Contigs >1,000 kb	10
Contigs >100 kb	10
Largest contig	40,728,208
N50	31,012,112
L50	4
GC content	0.36
Number of N	0
BUSCO C	95.70%
BUSCO S	93.70%
BUSCO D	2.00%
BUSCO F	1.40%
BUSCO M	2.90%
Number genes	25,007
Number of SNPs	3,912
Heterozygosity	0.0014%

genes (7,871) most are dispersed duplications (51%) or tandem duplications (28%) and only a few are WGD/segmental duplications (8%) (Table S6), which is typical of plants that did not undergo recent WGD (Jaillon et al., 2007; Wang et al., 2013). Moreover, if we remove large expansions (>5 in-paralogs) we cannot detect WGD/segmental duplications. Next, we searched for duplicated syntenic regions between *O. corymbosa* and two other Rubiaceae species (*Lepidodermis oblonga*, *Ophiorrhiza pumila*) using the orthologs inferred from phylogenetic evidence and MCScanX. Our results show that *O. corymbosa* did not experience a WGD event (Figure S5C, D). Moreover, when we examined the distribution of the substitutions per synonymous site (Ks) between the *O. corymbosa* paralogs and between the *O. corymbosa* and the other Rubiaceae species orthologs, we did not observe a normally distributed peak for the paralogs in the Ks plots (Figure S5E). However, the distribution of the peaks of the orthologous pairs marks speciation events, which are in agreement with the phylogeny. Additionally, the analysis of the Ks of syntenic genes between *O. corymbosa* and *L. oblonga* shows that *O. corymbosa* did not experience a recent whole-genome duplication event, but we can observe the peak corresponding to the gamma whole-genome triplication event (Jiao et al., 2012) (Figure S5F). Altogether, these results suggest that *O. corymbosa* did not experience a recent WGD.

### Identification of metabolites conferring the anti-cancer activity

Leaf extracts from *O. corymbosa* showed strong activity against the breast cancer line SKBR3 (Figures 1D, S2). While metabolites are typically responsible for the bioactivities of plants, also cyclotides, which are heat-stable macrocyclic peptides can be responsible for a wide range of activities. Cyclotides are likely

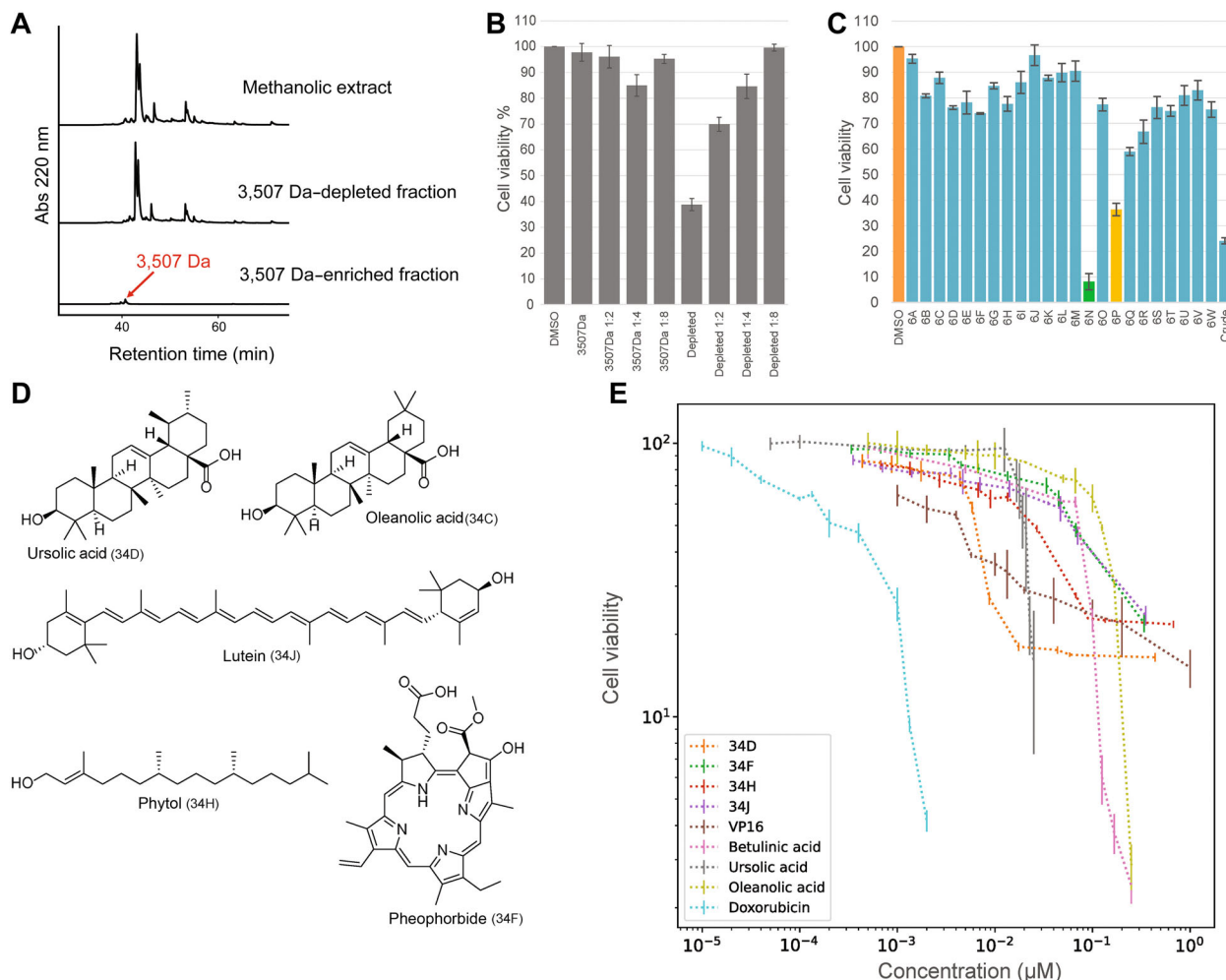
important for plant defense and show activities that are anti-HIV (Gustafson et al., 2004), anti-cancer (Hu et al., 2015), anti-microbial (Tam et al., 1999), hemolysis (Göransson et al., 1999), uterotensin (Gran et al., 2000), insecticidal (Jennings et al., 2001), and nematocidal (Colgrave et al., 2008). Since cyclotides were found in *O. diffusa* (Hu et al., 2015) and *O. biflora* (Ding et al., 2014) we investigated the presence of these peptides of molecular weight between 2,000 and 6,000 Dalton (Da) in *O. corymbosa*. We observed a peak corresponding to a 3,507 Da compound (Figures 3A, S6, peak at 1753.52 m/z), which likely represents a cyclotide. To investigate whether the putative cyclotide is responsible for the anti-cancer activity, we compared the activity of the purified 3,507 Da cyclotide to the 3,507 Da-depleted fractions. MTT assays revealed that the cyclotide did not show any anti-cancer activity and that the cyclotide-depleted fraction exhibited activity (Figure 3B; Table S7), indicating that another compound is responsible for the activity.

To identify the active compounds in *O. corymbosa* responsible for the inhibition of SKBR3 proliferation, we performed a comprehensive activity-guided fractionation and subjected 23 fractions (6A–6W) together with the crude extract control to MTT assay analyses (Figure 3C). Two fractions (6N and 6P) showed strong activity against SKBR3 cells in addition to the crude extract control (Figure 3C; Table S8). High-resolution MS analysis of the fractions identified ursolic and oleanolic acid (fraction 6N), pheophorbide a, phytol, and lutein (fraction 6P, Figure 3D) as putative active compounds. To pinpoint the active compounds, we performed a scale-up fractionation and confirmed the structures of the purified compounds by <sup>1</sup>H NMR (Figures S7, S8). The purified compounds (fractions 34D, F, H, J, see Table 2) were tested for the growth-inhibitory efficacy against SKBR3. To calculate the efficacy, we performed serial dilution to obtain the IC<sub>50</sub> values (Figure 3E; Table S9), which indicate how much of a compound is needed to decrease the cell viability (captured by MTT assay) to 50%. We additionally included commercially available ursolic acid and its structural analogs oleanolic acid, and betulinic acid. To compare the activities of these compounds to common chemotherapy drugs we also included the alkaloid etoposide (VP16) (Evans et al., 1985) and the anthracycline doxorubicin (Tacar et al., 2013).

The IC<sub>50</sub> value of fraction 34D (6.7 μM) was comparable to the IC<sub>50</sub> value of commercial ursolic acid (12.5 μM), but not to oleanolic (109.7 μM) or betulinic acid (66.65 μM), indicating that relatively minor modifications to a compound can have a large influence on the bioactivity. Phytol (20.33 μM), pheophorbide-a (74.36 μM), and lutein (57.65 μM) showed lower activities than ursolic acid. Conversely, VP16 (3.38 μM) and anthracycline doxorubicin (0.22 μM) showed the most potent activities against SKBR3 cells. Taken together, the most active compounds in *O. corymbosa* are ursolic acid (6.8 μM) and phytol (20.33 μM).

### Abiotic stress modulates the biosynthesis of anti-tumor metabolites

The content of active compounds in medicinal plants can vary greatly depending on growth conditions and organs (Yang et al., 2018; Isah, 2019). To investigate how abiotic stress can



**Figure 3. Identification of compounds of *O. corymbosa* with anti-tumor activities against SKBR3 cells**

(A) High-Performance Liquid Chromatograms of the methanolic *O. corymbosa* extract, and its 3,507 Da-depleted and 3,507 Da-enriched fractions at 220 nm. The red line corresponds to the peak containing the 3,507 Da peptide. (B) Cell viability assay of isolated peak and the depleted fraction. (C) Cell viability assay of 23 fractions and crude control. (D) Structure of the identified compounds. (E) Dilution series for each fraction to determine IC<sub>50</sub> values. Stock concentrations were 2 μg/μl for all fractions, that is, 34D (4.38 mM), 34F (3.37 mM), 34H (6.745 mM), and 34J (3.5 mM). Stock concentrations for the commercially available compounds were ursolic acid (0.5 mM), oleanolic acid (OA, 5 mM), betulinic acid (BA, 5 mM), alkaloid etoposide (VP16, 10 mM), and anthracycline doxorubicin (Doxorubicin, 0.1 mM).

influence the production of anti-tumor metabolites of *O. corymbosa*, we carried out a systematic approach involving eight abiotic stresses. We grew plants in soil under normal conditions (200 μmol·m<sup>-2</sup>·s<sup>-1</sup>, 28°C, 12 h light) until DAG 23 (±2 d, until the 4th node was developed, Figure S9A) and then transferred them to eight different stress conditions for 7 or 9 d. The 7-d long treatments included a short day (8 h light/16 h dark), long day (20 h light/4 h dark), cold (8°C and 15°C), and heat (40°C). The 9-d long treatments were drought (no watering) and two high light conditions (1,000 and 2,000 μmol·m<sup>-2</sup>·s<sup>-1</sup>).

The different stresses had impacts on overall plant morphology. The high light stresses increased the pigmentation of the plants with a stronger accumulation of the purple pigment at 2,000 μmol·m<sup>-2</sup>·s<sup>-1</sup>, particularly in stems (Figure 4A). High light also led to thick and yellow leaves, more flowers, and a larger root system (Figure S9B). Cold

treatment caused growth arrest, purple pigmentation of the stems at 15°C but not at 8°C, and rolled-in leaves (8°C, Figure S9C). 40°C heat treatment led to slow, more upright growth, few flowers, and yellow leaf color (chlorosis) of young tissues (Figure 4A). Short days resulted in light green, thin leaves (Figure 4A, 8L/16D) whereas plants grown in long days accumulated more yellow and purple pigments in leaves and stems and showed stunted growth. Finally, plants exposed to drought displayed a decrease in purple pigmentation, growth arrest, few flowers, and wrinkled leaves (Figure 4A). These results clearly show that the applied stresses have strong effects on the metabolism and growth of *O. corymbosa*, which should be reflected in metabolite changes.

To investigate the effects of the abiotic stresses on the anti-tumor potential, we performed MTT assays on root, leaf, stem, and flower extracts of the stressed plants (Table S10).

**Table 2. The computed IC50 range for the active fractions of *O. corymbosa* for SKBR3 cancer cells and commercial compounds**

Fraction/Standard	Purified amount (mg)	Best Fit IC50 (μM)	Min (95% CI)	Max (95% CI)
Ursolic acid (34D)	0.6	6.681	5.225	8.536
Pheophorbide a (34F)	0.9	74.36	63.33	88.62
Phytol (34H)	2.6	20.33	16.92	24.49
Lutein (34J)	0.6	57.65	44.24	78.11
Ursolic acid (Sigma)	-	12.5	12.32	12.69
Oleanolic acid (Sigma)	-	109.7	96.15	123.8
Betulinic acid (Sigma)	-	66.65	59.7	73.34
VP16	-	3.38	2.441	4.438
Doxorubicin	-	0.222	0.1875	0.263

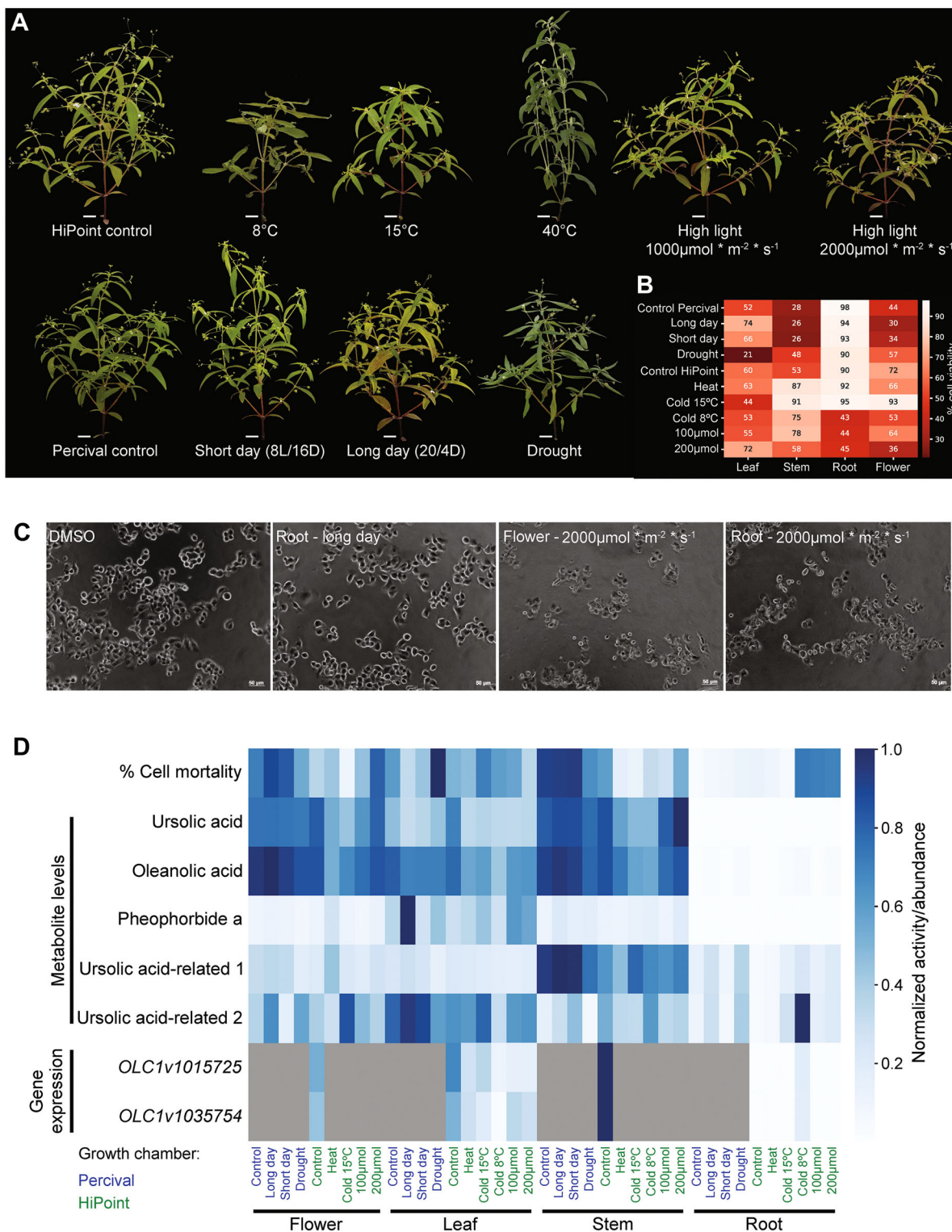
We observed that leaves, stems, and flowers showed anti-tumor activity in most conditions (Figure 4B, 20%–80% viability). Leaf extracts of control plants showed approximately 50% inhibition of SK-BR3 cells in line with our previous results (Figure 1D). Stem and flower extracts of the controls showed an inhibition range between 30% and 70% whereas root extracts did not show any activity for control conditions (>90% viability). Reduced activity of stem extracts was seen after cold and heat treatments (8°C, 15°C, and 40°C), and in high light (in four out of six biological replicates, Figure S10, Table S10). Drought increased activity (21%) of the leaf extract which could be due to water loss as suggested by the curling of leaves. In roots, after growth at 8°C, all biological replicates gained strong activity (43% viability). In the high light-treated plants half of the biological replicates gained anti-tumor activity in the roots (44%–45%). To verify the viability status of the SKBR3 cells, we investigated their morphology by microscopy (Figure 4C). While cells treated with the DMSO control and the inactive root extract showed no evidence of cell death, cells treated with the root extracts from high light plants showed cell death (Figure 4C), similarly to cells treated with the usually active flower extract. Thus, these results suggest that the different abiotic stresses can modulate the metabolic composition of plants and anti-cancer activity.

### Metabolomic and gene expression analyses of biosynthetic pathways of anti-tumor compounds

We observed that abiotic stress modulated the anti-tumor activity of different organs, which could be a consequence of the altered activity of the metabolic pathways. To better understand this phenomenon, we quantified the relative abundance of ursolic acid, oleanolic acid, and pheophorbide a in the same methanol extractions that were used for MTT assays (Table S11). Lutein and phytol could not be quantified in these

methanol extracts due to their reduced solubility in the polar phase, suggesting that these metabolites are not the main active ingredients. Pearson correlation analysis of the MTT values and the relative abundance of each metabolite shows that oleanolic acid and ursolic acid have the highest correlation (Table S12). Moreover, ursolic acid and oleanolic acid are in general abundant in all organs, except for roots, which is in line with the anti-tumor activities of the extracts (Figure 4D; Table S11). The amounts of ursolic acid are highest in the highly active stem and flower extracts of the control conditions and long- and short-day samples. In cold and heat-treated plants, the ursolic acid levels are generally lower. Interestingly, LC-MS/MS identified two compounds related to ursolic acid, of which one (ursolic acid-related 2) is elevated in root extracts of cold-treated (8°C) plants.

To better understand the molecular basis for the different physiological and metabolomic phenotypes of roots, leaves, flowers, and stems we investigated the changes in gene expression. To this end, we produced 43 RNA-seq samples from leaves, stems, flowers, and roots at one control and five stress conditions: cold at 8°C and 15°C, heat, and two high light stresses (Figure 4D; Table S13). Since the genes of the ursolic and oleanolic acid biosynthesis pathway are characterized in other plants (Aminfar et al., 2019; Hawkins et al., 2021), we carried out homology searches and expression profile analyses of the predicted genes and found two genes that putatively belong to the ursolic acid pathway: *OLC1v1035754* and *OLC1v1015725*. *OLC1v1035754* belongs to the Cytochrome P450 superfamily and is homologous to *Arabidopsis AT5G36130*. *AT5G36130* (of the CYP716A12 family) carries out the last step of ursolic biosynthesis which is the oxidation of α-amyrin to ursolic acid (Fukushima et al., 2011). The formation of α-amyrin in this pathway is catalyzed by an oxidosqualene cyclase (α-amyrin synthase). We found *OLC1v1015725*, a squalene cyclase, that also coexpresses with *OLC1v1035754* (PCC = 0.88, Table S14). Gene expression of both genes is highest in stems, leaves, and flowers but absent in roots in the control (Figure 4D), which correlates with ursolic acid metabolite levels and MTT activities. In leaves of plants treated with cold, heat, and high light, ursolic acid biosynthesis is decreased along with lower ursolic acid levels (Figure 4D). Interestingly, both genes show expression in roots of cold-treated (8°C) plants, a tissue with elevated anti-tumor activity, and an increase of an ursolic acid-related 2 compound (Figure 4D). Differential gene expression analysis revealed that 3,335 and 4,583 genes are significantly (adjusted *p*-value < 0.05) up- and down-regulated at 8°C, respectively (Table S15). Moreover, 12 and 28 genes associated with terpene biosynthesis are up- and down-regulated, respectively. Both genes associated with ursolic biosynthesis (*OLC1v1035754* and *OLC1v1015725*) are significantly up-regulated (Table S15) under cold 8°C conditions, but ursolic acid levels were not visibly up-regulated in roots (Figure 4D). Conversely, an ursolic acid-related 2 compound is strongly up-regulated in roots at 8°C, suggesting that this unknown



**Figure 4. Abiotic stress treatment of *O. corymbosa***

(A) Phenotypes of *O. corymbosa* plants grown for 7 d under different abiotic stresses and organs. (B) MTT assay activities against SKBR3 cancer cell lines. The rows and columns represent abiotic stresses and organs, respectively. (C) Phase-contrast images of SKBR3 cells untreated (DMSO control) and treated with metabolite extracts from roots from plants grown under long-day (20 L/4D), high light flowers, and high light roots. (D) Normalized activity and abundances across the different stress treatments. The rows indicate cell viability, the levels of metabolites, and gene expression, while the columns represent the different organs and abiotic stresses. The values are scaled to range from 0 (white cell color: no activity, absent metabolite, absent gene expression) to 1 (dark blue cell color: maximal activity, metabolite level, gene expression across all measurements). Green and blue letter colors represent samples grown in the Percival or HiPoint growth chamber, respectively.

compound is responsible for the gain of activity in roots at 8°C.

### Ursolic acid causes mitotic catastrophe and cell death in SKBR3

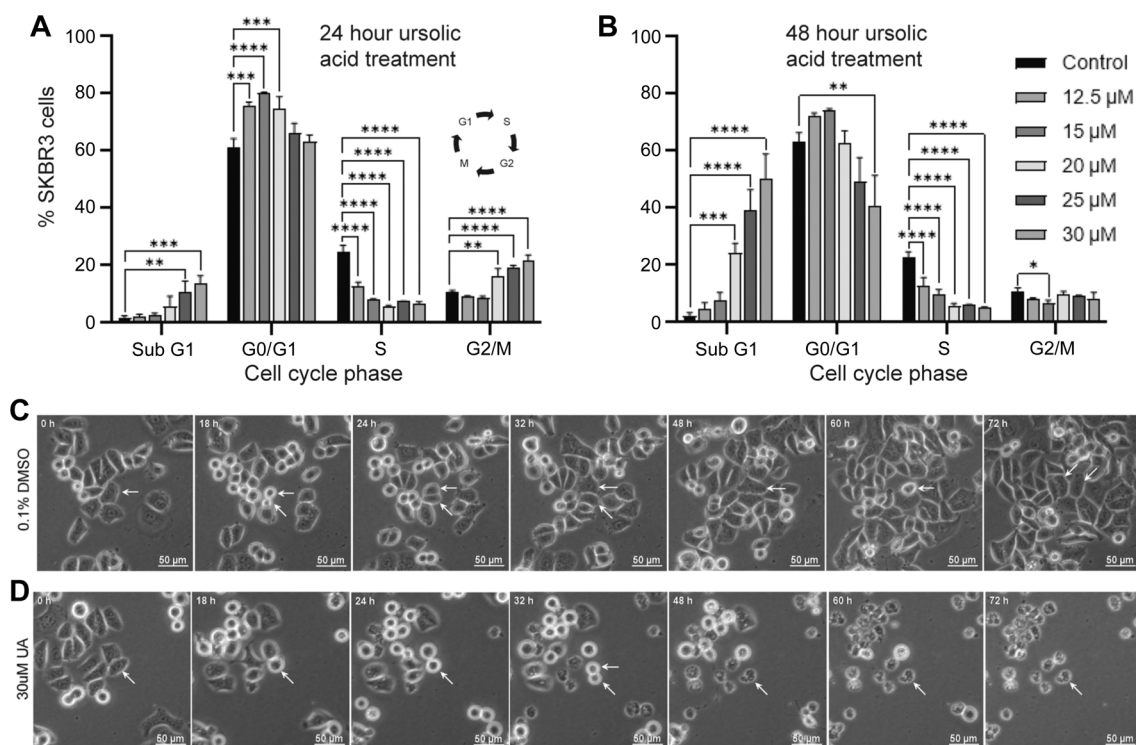
Ursolic acid has been reported to possess a wide range of anti-cancer activities (Seo et al., 2018), such as inhibition of Akt and promotion of autophagy (Lewinska et al., 2017), a decrease of ATP production (Zheng et al., 2012), and activation of adenosine monophosphate-activated protein kinase (AMPK) (Robey and Hay, 2009). Since many anti-tumor drugs interfere with the cell cycle progression, we set out to study the effect of different concentrations (12.5, 15, 20, 25, and 30  $\mu\text{M}$ ) of ursolic acid on the cell cycle progression over 24 h (Figure 5A) and 48 h (Figure 5B). At lower concentrations at 24 h of treatment, the cells accumulate at G0/G1 and S phases, while at higher concentrations, the cells no longer accumulate at G0/G1, but instead at G2/M and sub-G1 (sub-G1 represents cells that underwent apoptosis). After 48 h, we observed more cells in the sub-G1 phase at all concentrations (Figure 5B), indicating that a longer treatment causes more cell death. To track how the ursolic acid influences cell division, we performed time-lapse imaging for 72 h. In the control samples without ursolic acid treatment, SKBR3 cells

are able to undergo at least two rounds of cell division (Figure 5C, Data S1). However, 12.5–20  $\mu\text{M}$  of ursolic acid caused delays of the first or second rounds of mitosis (Figure S11), while 20–30  $\mu\text{M}$  caused cell death (Figures 5D, S11; Data S2).

The results suggest that at high concentrations, ursolic acid may inhibit the mitotic machinery, thereby causing an arrest in mitosis and cell death. At lower concentrations, cells can escape from the mitotic arrest but accumulate cellular damage. Due to the damage in the previous round of mitosis, the cells experience delays in subsequent rounds of cell division, which explains the higher percentage of cells in the G0/G1 stage.

### CETSA identifies putative human protein targets of ursolic acid

Ursolic acid shows a multi-pronged anti-cancer activity (Seo et al., 2018; Feng and Su, 2019), but its potential protein binding partners in cancer cells are poorly understood. CETSA is a method to identify protein targets of drugs (Martinez Molina et al., 2013; Jafari et al., 2014), and is based on the discovery that ligand binding to a protein increases its thermostability. CETSA combined with multiplexed mass spectrometry (MS-CETSA) represents a powerful untargeted



**Figure 5. Microscopic analysis of ursolic acid and SKBR3**

(A) Cell cycle progression analysis of SKBR3 by fluorescence-activated cell sorting (FACS) of DMSO control (black bar) or ursolic acid-treated cells (shades of gray represent ursolic acid concentrations). Each bar represents the mean  $\pm$  standard deviation from data obtained from three independent sets, where statistical significance was derived through one-way ANOVA followed by Bonferroni's multiple comparison test (\*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ ). (B) FACS analysis of 48-h treated cells. (C) Phase-contrast images of SKBR3 cells treated with DMSO control over 72 h. The white arrows are used to indicate a dividing cell (18 h). A daughter cell undergoes another division at 60–72 h. (D) Phase contrast image of a cell that undergoes division (32 h) and apoptosis (48 h).

drug-target identification methodology enabling monitoring of the entire detectable proteome for changes in protein thermostability under drug treatment. CETSA-MS does not require prior information on drug type or its mechanism of action (Dziekan et al., 2020), and its label-free nature enables direct testing of standard unmodified compounds without the need to use expensive drug probes or the generation of reporter cell lines.

To identify protein targets of ursolic acid, we leveraged the multidimensional IsoThermal Drug-dose Response (ITDR) variant of MS-CETSA, where SKBR3 cell lysate was treated with a range of ursolic acid concentrations (10 nM to 21.89  $\mu$ M) and exposed to different heat challenge temperatures (50°C, 55°C, 60°C) or a non-denaturing control (37°C) to destabilize susceptible protein subsets. The resulting soluble protein fractions were then analyzed by quantitative mass spectrometry to identify proteins exhibiting drug dose-dependent change in stability, suggestive of direct drug-protein interaction. The MS analysis detected 11,638 protein profiles (Figure 6A, Table S16), capturing abundances of 4,356 proteins in at least one thermal challenge and the 37°C reference condition. Of these, 31 proteins passed a higher confidence threshold (3\* median absolute deviation in the area under the curve normalized by the reference control,  $R^2 > 0.8$ ).

Among the 31 recorded drug-induced changes in stability, six proteins exhibited strong drug-dose dependent stabilizations, expected upon high-affinity drug-ligand interaction, and represent high confidence drug-binding candidates (Figure 6B). The proteins are *SPG21* (Maspardin), *HIBCH* (3-hydroxyisobutyryl-CoA hydrolase, mitochondrial), *DECR1* (2,4-dienoyl-CoA reductase [(3E)-enoyl-CoA-producing], mitochondrial), *RPLP1* (60 S acidic ribosomal protein P1), *CWF19-L2* (CWF19-like protein 2) and *NIPSNAP* (NipSnap homolog 1). The remaining proteins identified by CETSA exhibited only modest positive or negative shifts in their abundance profiles and could represent additional lower confidence off-targets (Figure 6A).

To investigate the binding sites of ursolic acid among the identified candidate-binding partners, we performed a reverse docking analysis using ursolic acid and the available protein structures obtained from experimental evidence for *DECR1* or AlphaFold predictions for the remaining five proteins. Ursolic acid was able to favorably bind to Maspardin (−13.0 kcal/mol), *HIBCH* (−8.9 kcal/mol), and *DECR1* (−10.6 kcal/mol) (Figure 6C, D, PyMol session file available as Data S3), but not for the other three proteins. Interestingly, for *DECR1*, the docking analysis shows that ursolic acid is potentially able to replace the natural ligands (NADP nicotinamide-adenine-dinucleotide phosphate and hexanoyl-coenzyme A, Figure S12), likely inhibiting the activity of the enzyme.

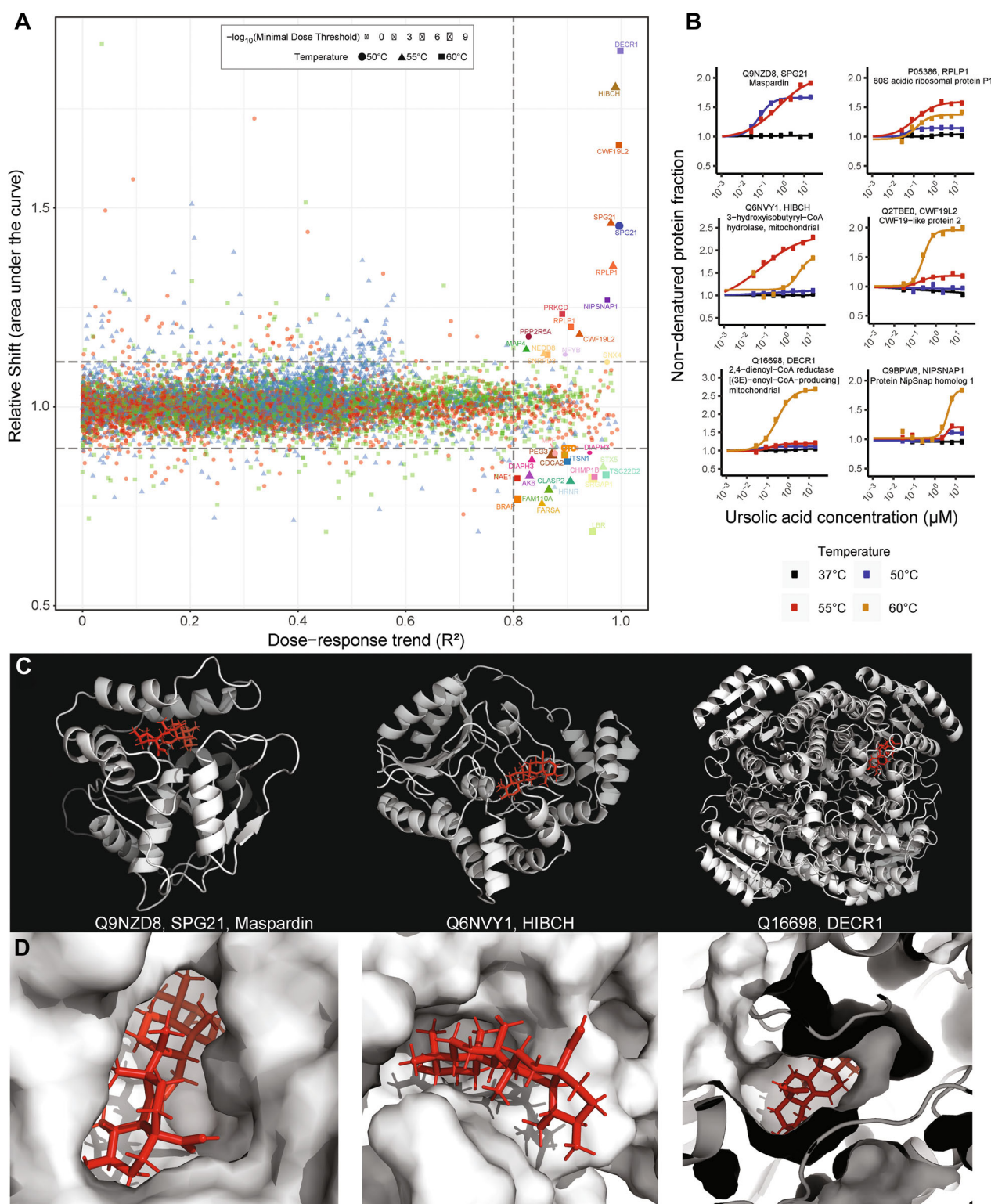
## DISCUSSION

To identify an *Oldenlandia* plant with potent anti-cancer activities, we screened 11 plants comprising *O. corymbosa*,

*O. tenelliflora*, and *O. biflora*, and observed that *O. corymbosa* showed specific activity in two breast (SKBR3, MB231) and gastric (SNU484) cancer cell lines, in contrast to the other two plant species (Figure 1). Interestingly, HepG2 and MKN28 were not affected by any of the extracts. As the molecular basis of the cancer cell lines can be different (e.g., mutations in receptors, transcription factors, copy number variations of oncogenes) (Mirabelli et al., 2019; LiuYu et al., 2020), some cancer cell lines may need higher concentrations of active compounds to be affected. Interestingly, most other herb garden plants showed either no activity or strong activity on all three cell lines at the concentration used, suggesting that *O. corymbosa* shows a more specific activity to the characteristics of the SKBR3 breast cancer line (Figure S2).

To establish a genomic resource for the *Oldenlandia* genus, and to identify the biosynthetic pathways of the anti-cancer metabolite, we produced a nearly chromosome-scale assembly comprising of nine large contigs, which likely represent chromosomes. We hypothesize that the high contiguity of the assembly is due to the extremely low heterozygosity of the genome (Table 1), which made the genome assembly easier. The low heterozygosity is comparable to that reported for the homozygous line *Arabidopsis thaliana* Col-0 (heterozygosity of 0.001%) (Ossowski et al., 2008). Electron microscopy analysis revealed that the flowers are sealed by a ring of pubescent hairs (Figure 2B), effectively forcing self-pollination (autogamy). Autogamy is thought to be reproductively beneficial in an adverse environment, as there is no need to invest energy in costly pollinating agents and guarantees a greater fruit set (Schemske, 1978; Sakai and Wright, 2007). The small, homozygous genome, no evidence of recent whole-genome duplications, short generation time (flowers appear 19 d after germination, Figure 1), and similar experimental procedures to *Arabidopsis* established (Figure S1), make *O. corymbosa* an attractive model for the *Oldenlandia* genus. However, since not all of the contigs contained telomeres at both ends, techniques such as Hi-C or Bionano optical mapping are needed to arrive at a chromosome-scale assembly.

It is known that *O. corymbosa* contains iridoids, flavonoids, saponins, triterpenes, phenols, terpenoids, sterols, and anthraquinones (Sivaprakasam et al., 2014; Li et al., 2015). To identify the compounds conferring the anti-cancer activity against SKBR3, we performed activity-guided fractionation, NMR, and high-resolution mass spectrometry, which revealed ursolic acid, oleanolic acid, pheophorbide a, phytol, and lutein in the active fractions. Ursolic and its analog oleanolic acid are found in apple peels and in many herbs (Jäger et al., 2009), while pheophorbide a and phytol are produced during chlorophyll a degradation (Kuai et al., 2018). Lutein is a xanthophyll carotenoid and antioxidant and has recently been implicated as an anti-tumor agent (Kavalappa et al., 2021). Lutein inhibited the viability of MCF-7 and MDA-MB-231 cells by inducing apoptosis resulting in elevated caspase-3 activity and downregulated



**Figure 6. Cellular Thermal Shift Assay (CETSA) analysis and reverse docking**

(A) Distribution of protein stabilization as a function of  $R^2$  value (that quantifies the adherence of protein stabilization profile to the dose-response trend) against  $\Delta\text{AUC}$  (AUC of heat-challenged sample normalized to non-denaturing 37°C control).  $3 \times \text{MAD}$  of AUC in each dataset and  $R^2 = 0.8$  cutoffs are indicated on each graph by dashed lines. Proteins above the dashed thresholds are indicated by names. (B) Protein stabilization curves of the top six candidate proteins. The y-axis shows the extent of stabilization, depicted as remaining soluble protein level after thermal challenge relative to no-drug control, while the x-axis shows drug gradients. The non-denaturing 37°C control condition is plotted in black. (C) Atomic model of Maspardin, HIBCH, and DECR1. The protein is colored white, while ursolic acid is colored red. (D) Zoom in on ursolic acid bound to the respective protein.

expression of Bcl-2 (Kavalappa et al., 2021). Phytol is an emerging candidate with anti-tumor activity, induced genotoxicity, and apoptosis in breast cancer cells. Furthermore, it showed a DNA damage repair capacity in mouse lymphocytes (de Alencar et al., 2019). Thus, it is likely that the anti-cancer effect of *O. corymbosa* is due to the synergistic action of these compounds.

A main active compound found in both *O. diffusa* and *O. corymbosa* is ursolic acid (Li et al., 2015; Chen et al., 2016). This pentacyclic triterpene shows a range of effects, including anti-tumor, anti-inflammatory, anti-oxidative, antiphytol-diabetic, and pro-apoptotic (Iqbal et al., 2018; Seo et al., 2018; Feng and Su, 2019). The IC50 value of ursolic acid against SKBR3 showed a modest activity of 6.81  $\mu\text{M}$  (Table 2), but its activity was more potent than the structurally-similar oleanolic acid (109.7  $\mu\text{M}$ ) and betulinic acid (66.65  $\mu\text{M}$ ). This indicates that the relatively minor modification to the compound can have dramatic effects on its efficacy, which makes ursolic acid-like compounds attractive targets for modification, by for example, synthetic biology approaches (Reed et al., 2017).

Plants produce a wide range of specialized metabolites to better acclimate to changing environments, such as changes in temperature, light, humidity, and developmental processes (Ramakrishna and Ravishankar, 2011; Berini et al., 2018). The specialized metabolites can serve as toxic, repellent compounds against herbivores, or may be important for functional and structural stabilization through signaling pathways (Hartmann, 2004; Isah, 2019). Thus, amounts of active compounds can be induced by stress. We show that *O. corymbosa* has strong anti-tumor activities in above-ground organs. Interestingly, we observed that the normally inactive roots gain activity in plants grown at 8°C and that this activity is likely caused by an ursolic acid-like compound accumulating in roots (Figure 4D). This is corroborated by the expression of a gene encoding a cytochrome P450 protein (*OLC1v1035754*) which is predicted to carry out the last step in the ursolic acid biosynthesis. Co-expressed with this gene is *OLC1v1015725*, a putative squalene cyclase. CYP716A (subfamily cytochrome P450) proteins and squalene cyclases are known to be key genes in triterpene biosynthesis in *Medicago truncatula* (Fukushima et al., 2011), apple (Andre et al., 2016), mints (Aminfar et al., 2019), and Arabidopsis (Yasumoto et al., 2016). Both genes are highly expressed in the stem tissue of the control, correlating with high ursolic acid amounts and strong anti-tumor activity (Figure 4D). Interestingly, our results show that genes related to terpene biosynthesis are upregulated in roots at 8°C. These data suggest that the ursolic acid-like compound found in col-treated roots is synthesized in the roots rather than transported to the roots from above-ground tissues.

Ursolic acid is mentioned in more than 2,600 articles at the time of writing, and it has been observed to have many health-promoting activities. It is a fairly common metabolite in the plant kingdom, reported in at least 23 species and in 10 plant families (Gudoityte et al., 2021). The compound has a wide range of anti-tumor activities and directly affects the

hallmark of cancer growth, which is the metabolic reprogramming of cancer cells to sustain anabolic growth (Woźniak et al., 2015). Ursolic acid can modulate the expression and function of mitochondria-associated enzymes, which promote apoptosis and lead to anti-proliferative effects in various models *in vivo* and *in vitro* (Yan et al., 2010). In three breast cancer cell lines with different growth factor receptor status (MCF7, MB-231, SKBR3), which we also used in our study, ursolic acid downregulates the Akt signaling, and this inhibition affects glycolysis and markedly decreased levels of HK2, pyruvate kinase M2, ATP, and lactate (Lewinska et al., 2017). Additionally, ursolic acid and betulinic acid (5 and 10  $\mu\text{M}$ ) caused p21-mediated G0/G1 cell cycle arrest and senescence accompanied by oxidative stress and DNA damage. Ursolic acid-induced apoptosis was achieved by targeting the glycolytic pathway and autophagy in breast cancer cells (Lewinska et al., 2017). Similarly, another study shows that ursolic acid derivatives induce cell cycle arrest at the G2/mitotic (M) phase by competing with glucose for the binding site at glucokinase (Wang et al., 2014). These results are in line with our observations for SKBR3 cells, where we observed a concentration- and time-dependent accumulation of cells at G0/G1 and S phases (low concentrations) or G2/M and sub-G1 (high concentrations), with more cells accumulating at sub-G1 phase (representing cell death) at 48 h (Figure 5). Thus, we hypothesize that ursolic acid causes mitotic cycle arrest.

Despite the numerous studies, the direct protein binding partner of ursolic acid is unknown. To address this, we performed CETSA-MS, and identified six high-confidence proteins. Three of them (*HIBCH*, *DECR1*, *RPLP1*) have been implicated in cancer. *HIBCH* is involved in L-valine degradation, and blocking *HIBCH* localization to mitochondria resulted in decreased colon cancer growth and increased autophagy *in vivo* and *in vitro* (Shan et al., 2019). *HIBCH* has also been reported to be increased in patients with prostate cancer and MDA-MB-231 breast cancer (Graff et al., 2014; Kalita-de Croft et al., 2019). *DECR1* is a short-chain dehydrogenase involved in redox homeostasis by controlling the balance between saturated and unsaturated phospholipids. *In vivo*, *DECR1* deletion impairs lipid metabolism and reduces tumor growth in prostate cancer (Blomme et al., 2020). In cervical cancer high levels of *DECR1* enhance lipolysis and the release of fatty acid energy stores to support cancer cell growth (Zhou et al., 2022). In contrast, *DECR1* was proposed to act as a tumor suppressor in HER2-positive breast cancer, however, *DECR1* overexpression was shown to protect cancer cells from glucose withdrawal-induced apoptosis (Ursini-Siegel et al., 2007). The 60S acidic ribosomal protein P1 (*RPLP1*) plays an important role in the elongation step of protein synthesis and promotes tumor metastasis in breast cancer patients, as well as in hepatocellular carcinoma cells and several gynecologic tumors (Chen et al., 2002; Artero-Castro et al., 2011; He et al., 2018; Xie et al., 2021).

The remaining three proteins have diverse roles. The first protein, Maspardin (*SPG21*), may play a role as a negative regulatory factor in CD4-dependent T-cell activation as it co-localizes with CD4 on endosomal/trans-Golgi network (Zeitlmann et al., 2001), which is in line with ursolic acid affecting T-cell activation (Kaewthawee and Brimson, 2013). The second protein, *CWF19*-like protein 2, is a cell cycle control factor located in the nucleus predicted to be involved in mRNA splicing via spliceosome. It appeared in a screen for chromosomal deletion sites in breast carcinomas together with other cell-cycle regulation proteins (*JMY*, *PTPRN2*) (Nordgard et al., 2008). In addition, it might interact with *BRCA1* (Hill et al., 2014) and *BRCA2* (Malik et al., 2016). The third protein, *NIPSNAP* homolog 1, is involved in removal of damaged mitochondria and was shown to be upregulated upon treatment with anti-tumor drugs (Khaghanzadeh et al., 2016; Abudu et al., 2019; Fathi et al., 2021).

Taken together, *HIBCH*, *DEC1*, and *RPLP1* constitute prime targets for their role in the ursolic acid-mediated activity.

## CONCLUSION

We present the first high-quality genome for the *Oldenlandia* genus and characterize the mode of action of the main active compound—ursolic acid. We envision that the genome will be valuable for researchers working on the medicinal properties of the *Oldenlandia* species, while the mode of action of ursolic acid will spur further development of this valuable compound.

## MATERIAL AND METHODS

### Establishing *Oldenlandia* plants in the laboratory

*Oldenlandia* plants (*Oldenlandia corymbosa*, *O. tenelliflora* and *O. biflora*) were collected in the field in the Bedok area of Singapore. Ten individual plants (M1, M2, M5, M6, M7, M9, M11–14) were transferred to the lab and naturally propagated by autogamy. Line M15 (*O. corymbosa* specimen (Y.W.Low et al., SING2018-1046)) was collected from a private residential area and a herbarium specimen was made following Forman and Bridson (1991). Next, we examined morphological characters of the vegetative and reproductive parts of the M15 and matched it to *Oldenlandia corymbosa* specimens collected in Singapore and Peninsular Malaysia preserved at SING (herbarium acronym follows Thiers [continuously updated]). Materials examined: SINGAPORE. Singapore: Changi Reserve, sine datum, J.S.Good-enough s.n. (SING [SING0033728]); Geylang, 4<sup>th</sup> mile, sine datum, Z.Teruya 1264 (SING [SING0037446]); Bedok, Simpang Bedok, Upper East Coast Road, near Bedok Camp, 3 Jun 2018, Y.W.Low et al., SING2018-1043 (SING), 10 Sep 2018, Y.W.Low et al., SING2018-1046 (SING). MALAYSIA. Peninsular Malaysia: Melaka, Melaka Tengah, Bukit Beruang, sine datum, R.Derry 440 (SING [SING0165554]). For all the plants, plantaflor potting soil, vermiculite, and perlite were used as a mixture (7:2:1). Standard

conditions in the growth chamber were 28°C, 50%–60% humidity and 12 h light (~200  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ).

Plants used for the initial MTS assays (Figure 1D) were grown at 23°C, 50%–60% humidity and 12 h light conditions (150  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ). MTS assays of *O. corymbosa* compared to medicinal herb garden plants (Figure S2) were done on five F1 plants of M15 grown outside at 28–35°C, 60%–80% humidity and 12 h variable light conditions (150–1,500  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ). Abiotic stress experiments were done using F2 plants of M15. We grew plants in soil under normal conditions (200  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ , 28°C, 12 h light) until DAG 23 ( $\pm 2$  d, until 4th node was developed, Figure S9A) and then transferred them to eight different stress conditions for 7 or 9 d. The 7-d long treatments included short day (8 h light/16 h dark), long day (20 h light/4 h dark), cold (8°C and 15°C), and heat (40°C). The 9-d long treatments were drought (no watering) and two high light conditions (1,000 and 2,000  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ). Drought plants were left unwatered for 9 d until leaves were rolled in. One plant was watered on the day of harvest and fully recovered the next day, thus showing that plants were still viable at the time of harvest (not shown). Flowers, leaves, and stems were harvested and directly frozen in liquid nitrogen. Roots were rinsed in water for about 5–8 min to remove soil and vermiculite, dried with paper and then frozen in liquid nitrogen.

### RNA isolation and sequencing

An RNA tissue atlas for genome annotation was prepared from flowers, leaves, stems and roots of M15 plants grown outside under low light conditions using Spectrum™ Plant Total RNA Kit (Sigma). For differential gene expression analyses of secondary metabolite genes, RNA was extracted from stressed and control plants (three biological replicates each) shown in Figure 4A using Spectrum™ Plant Total RNA Kit (Sigma). Quality control of all the extracted RNA (triplicates for each condition) was done by Novogene (Singapore) using Nanodrop and agarose gel electrophoresis (for purity and integrity) before sample quantitation and further analyses of integrity (Agilent 2100 Bioanalyzer). Library type was a eukaryotic directional mRNA library. Library construction from total RNA including eukaryotic mRNA enrichment by oligo(dT) beads, library size selection, and PCR enrichment was performed by Novogene using NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina®. The libraries for the RNA tissue atlas were then sequenced with Illumina HiSeq-4000, paired-end sequencing at 150 base pairs and at sequencing depths of approximately 51 million reads per sample (15 Gb/sample). The libraries for the abiotic stress experiments were sequenced with NovaSeq-6000, paired-end sequencing at 150 base pairs and at sequencing depths of approximately 20 million reads per sample (6 Gb/sample).

### DNA extraction and illumina sequencing

Genomic DNA was isolated from 100 mg (F.W.) of leaf tissues from M15 plants grown outside under low light conditions using the illustra Nucleon Phytopure Genomic DNA Extraction Kit (GE

Healthcare) after grinding the leaves in liquid nitrogen to a fine powder. A total amount 5.85 µg (A260/280: 1.84; A260/230: 1.36) was sent for Illumina sequencing at Novogene AIT (Singapore). Library type was a 350 bp insert DNA library (WGS). The libraries were then sequenced with Illumina HiSeq-4000, paired-end sequencing at 150 base pairs and at sequencing depths of ~102 million reads per sample (30 Gb/sample).

### High Molecular Weight DNA extraction and oxford nanopore sequencing

Leaves and stems of the clonally propagated F1 plants of M15 (*Oldenlandia corymbosa*) were harvested, cleaned and cut into smaller pieces, and care was taken to exclude flowers and fruits and seeds. Plant tissue was flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  prior to High Molecular Weight (HMW) DNA extraction. For each extraction, 10 g of plant tissue was physically homogenized in liquid nitrogen using a mortar and pestle. The BioNano Nuclear Isolation protocol was followed, in which the homogenized plant tissue was first dissolved in Isolation Buffer with Triton X-100 and  $\beta$ -mercaptoethanol (IBTB), followed by filtering to remove undissolved plant material, and then centrifugation was carried out to pellet the nuclei. Isolation Buffer (IB) consisting of 15 mM Tris, 10 mM EDTA, 130 mM KCl, 20 mM NaCl, 8% PVP-10, pH 9.4 was prepared, with 0.025% (m/v) spermine and 0.035% (m/v) spermidine added and filter-sterilized before use. 0.1% Triton X-100 and 7.5%  $\beta$ -mercaptoethanol were next added to the IB, and the resulting IBTB buffer was added to the homogenized plant tissue and mixed on ice. Cell strainers (100 µm and 40 µm) were used to remove the undissolved plant tissue. Triton X-100 was added to the cell suspension to facilitate the disruption of the cell membrane and the samples were mixed thoroughly before being spun down at 2,000 g for 10 min to pellet the nuclei. The pelleted nuclei were resuspended in cetyl trimethylammonium bromide (CTAB) solution for DNA extraction following the protocol described in Michael et al. (2018). The resulting DNA was further purified using the Qiagen Genomic Tip 500/G. The DNA pellet from the CTAB DNA extraction was solubilized in Buffer G2 from the Qiagen Genomic Tip kit, along with RNase A (2 µl/ml of buffer) and proteinase K (10 µl/ml of buffer) for further removal of contaminating RNA and proteins. The rest of the Qiagen Genomic Tip protocol was followed as per the manufacturer's instructions. The final DNA extract was quantified using Nanodrop and the DNA integrity was checked using the Genomic ScreenTape (Agilent). Finally, HMW DNA was size-selected and sequenced as previously described (van Rengs et al., 2022). In brief, DNA was size-selected using the Circulomics Short-Read Eliminator XL Kit and a sequencing library was prepared using the Oxford Nanopore LSK-110 ligation sequencing Kit. The library was sequenced using a PromethION using 9.4.1 cells. Basecalling of the ONT read data was done using guppy basecaller v3.2.2, which yielded approximately 2.7 million reads.

### Genome assembly

To assemble the genome of *O. corymbosa* we used long reads from Oxford Nanopore sequencing and short reads from Illumina sequencing. ONT reads were processed by removing adapters using porechop v0.2.3 (<https://github.com/rwick/Porechop>) with default settings, removing low quality (<85 q) and short reads (<3,000 bp) using filtlong v0.2.0 (<https://github.com/rwick/Filtlong>), and error-corrected using Canu v1.9 (Koren et al., 2017). For Illumina reads, low quality sequences and adapters were trimmed using Trimmomatic v0.39 (Bolger et al., 2014) with default parameters. To estimate the genome size of *O. corymbosa*, jellyfish v2.3.0 (Marçais and Kingsford, 2011) was run on the Illumina reads with the canonical k-mer (-C) option and k-mer size of 21. Then GenomeScope v2.0 (Ranallo-Benavidez et al., 2020) was used to analyze the k-mer distribution, which resulted in a haploid genome size of 272.8 Mb. ONT filtered reads were assembled using NextDenovo v2.5.0 (<https://github.com/Nextomics/NextDenovo>) with the estimated genome size. Then, contigs were polished for two rounds using RACON v1.4.3 (Vaser et al., 2017) with the alignments generated by MiniMap2 v2.17 (Li, 2018) using the ONT reads, and Pilon v1.24 (Walker et al., 2014) with the alignments generated by BWA v0.7.17 (Li and Durbin, 2009) using the Illumina reads. Duplicate haplotypes were removed using Purge Haplotigs (Roach et al., 2018) with default parameters and read depth thresholds of 5, 45, and 190.

### Genome annotation

The *de novo* repeat library of *O. corymbosa* was constructed following the Repeat Library Construction Advanced pipeline (Campbell et al., 2014), which uses MITE-Hunter (Han and Wessler, 2010), LTRdigest (Steinbiss et al., 2009), LTR\_harvest (Ellinghaus et al., 2008) (available in genome tools v1.6.1 [Gremme et al., 2013]) and Repeatmodeller v2.0.1 (Flynn et al., 2020). This new library was combined with the Dfam v3.5 database (Storer et al., 2021) to identify Transposable Element. The genome of *O. corymbosa* was masked using both libraries with RepeatMasker v4.1.2-p1 (Smit et al., 2015). Gene prediction was performed by combining transcript alignments, protein alignments and *ab initio* gene predictions (Alioto et al., 2020). First, the RNA-seq reads were trimmed using Trimmomatic v0.39 (Bolger et al., 2014) with default parameters. Then, a *de novo* transcriptome assembly was obtained with Trinity v2.11.0 (Haas et al., 2013) and non-redundant transcripts were obtained with PASA v2.5.1 (Haas et al., 2008). After, we detected coding regions in the transcript using TransDecoder from the PASA package. Second, we downloaded the proteomes of three Rubiaceae (*Coffea canephora*, *Ophiorrhiza pumila*, and *Leptodermis oblonga*) from NCBI database and aligned to the genome using SPALN v2.4.6 (Iwata and Gotoh, 2012). Third, the RNA-seq reads were aligned to the genome with STAR v2.7.8a (Dobin et al., 2013) and hints were generated with bam2hints from the Augustus package v3.2.3 (Stanke et al., 2006). Then, *ab initio* gene predictions were performed with Augustus v3.2.3, GeneID v1.4 (Alioto et al., 2018), and GeneMark-ES v4 (Lomsadze et al., 2014).

with and without hints. Finally, all the predictions were combined using EvidenceModeler-1.1.1 (Haas et al., 2008). Functional annotation was performed using InterProScan v5.55-88.0 (Jones et al., 2014) and Mercator4 v4.0 (Schwacke et al., 2019). Visualization of the genome characteristics was performed with Circos v0.69-8 (Krzywinski et al., 2009).

### Organellar genome assembly and annotation

The plastid genome was recovered from the Illumina reads using GetOrganelle v1.7.5 (Jin et al., 2020) and annotated using GeSeq (Tillich et al., 2017). Then, the genome map was drawn with OGDRAW (Greiner et al., 2019). The mitochondrial genome was extracted from the nuclear assembly in one single contig. Then, it was annotated by BLAST searches against previously annotated mitochondrial genomes: *Damnacanthus indicus* (Han et al., 2021) (NCBI: MZ285076) and *Scyphiphora hydrophyllacea* (Chen et al., 2020) (NCBI: MT610041). tRNA genes were annotated using tRNAscan-SE (Schattner et al., 2005).

### Genome analysis

Assembly completeness was estimated in three ways. First, gene completeness was determined by running BUSCO v5.3.1 (Simão et al., 2015) using the eudicots\_odb10 database. Second, a comparison of k-mers present in the assembly and in the Illumina reads was performed using Merqury (Rhie et al., 2020) with k-mer length of 21. Third, telomere repeats ('TTTAGGG') were searched in the contigs. Then adjacent regions were joined, and regions larger than 100 bp were included as predicted telomere regions.

To detect heterozygous positions in the genome, Illumina reads were mapped against the genome using BWA v0.7.17-r1188 (Li and Durbin, 2009), and SNPs were identified with GATK HaplotypeCaller v4.2.5.0 (McKenna et al., 2010), setting ploidy to 2, and using thresholds for mapping quality (MQ > 40), quality by depth (QD > 2), Fisher strand bias (FS < 60), mapping quality rank sum test (MQRankSum > -12.5), read pos rank sum test (ReadPosRankSum > -8), strand odds ratio (SOR < 3), read depth of coverage (DP ≥ 20), and allelic depth (AD ≥ 5).

### Prediction of orthogroups and species tree reconstruction

To study the evolution of *O. corymbosa* in the context of other nine plant species, orthogroups (orthologous gene groups) were predicted using Orthofinder v2.5.4 (Emms and Kelly, 2019) with Diamond v2.0.5.143 (Buchfink et al., 2015) as sequence aligner. A total of 667 single-copy orthogroups were selected and aligned using MAFFT v7.453 (Katoh and Standley, 2013) individually and then concatenated. The species tree was reconstructed using the amino acid substitution model LG implemented in RAxML v8.2.12 (Stamatakis, 2014) and 100 bootstrap replicates. Additionally, the assembled plastid genome of *O. corymbosa* was aligned with other five genomes (*O. corymbosa*, *O. diffusa*, *O. brachypoda*, *Hedyotis ovata*, and *C. arabica*) using MAFFT v7.453 (Katoh and Standley, 2013) and a species tree was reconstructed as described above using RAxML v8.2.12 and the GTR model.

### Phylogenomic analyses

To predict orthology and paralogy relationships, we reconstructed gene trees for all the orthogroups and used phylogenetic evidence. The protein sequences of each orthogroup were aligned using the approach described in the PhylomeDB pipeline (Huerta-Cepas et al., 2011) and phylogenetic gene trees were built using IQ-TREE v1.6.12 (Minh et al., 2020). Then, orthologs and paralogs were detected using the species overlap method (Gabaldón, 2008) as implemented in ETE v3.0 (Huerta-Cepas et al., 2016).

To search for whole genome duplications, we used three methods. First, we calculated the duplication ratios of the nodes of the species tree by mapping the gene duplications into the nodes of the species tree using a phylogeny-based phylostratigraphic approach based on a species-overlap algorithm (Huerta-Cepas and Gabaldón, 2011). For this analysis, we excluded duplication events containing more than five paralogs. Then, using the duplicate gene classifier tool from the MCScanX toolkit (Wang et al., 2012), we classified the origins of duplicated genes as dispersed, proximal, tandem, and segmental. Second, we searched for syntenic regions between *O. corymbosa* and two other Rubiaceae species (*Leptodermis oblonga* and *Ophiorrhiza pumila*) using MCScanX and the orthologs previously predicted. The syntenic plots were generated using the dot plotter tool from the MCScanX toolkit. Third, we calculated the synonymous nucleotide substitutions (Ks) of *O. corymbosa* paralogs and of orthologous gene pairs between *O. corymbosa* and *L. oblonga* and *O. corymbosa* and *O. pumila* using the add\_ka\_and\_k\_s\_to\_collinearity.pl tool from the MCScanX toolkit. Also we generated a Ks plot using SynMap from the CoGe toolbox (Lyons and Freeling, 2008; Lyons et al., 2008).

### RNA seq analysis

To obtain transcripts abundances of the RNA-seq samples under different conditions, the reads were mapped to the coding sequence (CDS) of *O. corymbosa* with Kallisto v0.46.1 (Bray et al., 2016). All samples had >8 M reads mapped. For root and stem samples, differential expression analysis was performed with the package DESeq. 2 (Love et al., 2014).

### MTT, metabolic, and RNA profiling

To detect which of the candidate metabolites is responsible of the anti-cancer activity, Pearson correlation coefficient (PCC) of the percentage of cell viability (MTT assay) profiles was calculated against the relative abundance of ursolic acid, oleanolic acid, pheophorbide a, and two compounds related to ursolic acid.

To study the genes associated with the metabolic pathway of ursolic acid, candidate genes were searched in other species in the PlantCyc database (Hawkins et al., 2021). Then homology searches in the *O. corymbosa* were done using BLAST and the orthology information. Additionally, the profiles of expression of the candidate genes were analyzed and compared with the metabolic profiles. Finally, gene co-expression analysis was performed by calculating the PCC against all the proteomes and the 50 best hits were kept.

### Metabolite extraction and LC-MS/MS analysis

For the abiotic stress experiment four major tissues were sampled. Flower tissue (open and closed flowers and green seed pods) was collected and frozen in liquid nitrogen. Leaves of >2 cm length were cut, rinsed in dH<sub>2</sub>O, pat dry on a paper tissue and flash frozen. Smaller leaves and apical tissue were spared. Stems were cut and flash frozen. Roots were carefully rinsed for 5–8 min in water to remove soil particles, dried with paper tissue and flash frozen. All samples were ground using a mortar and pestle, the frozen powder was aliquoted and stored in –80°C until use.

For metabolite extraction, 5 µL/mg F.W. of extraction buffer (80% MeOH in water, HPLC-MS grade) was added to 100 mg of frozen powder together with three stainless steel beads. The samples were homogenized in a Qiagen PowerLyser for 2 min at 2,000 r/min. Beads were removed and samples were shaken at 1,500 r/min for 10 min at 20°C in the dark followed by centrifugation at 12,000 r/min for 10 min. After transfer of supernatant, the pellets were re-extracted twice and the collected supernatants were spun again, aliquoted and dried in a cooling speed vac at 30°C. Pellets were stored in –80°C. For MTT analyses, pellets were dissolved at a concentration of 2 µL DMSO/mg frozen powder. Samples were centrifuged at 13,000 r/min for 10 min and the SUP was filtered through 0.2 µm nylon filters (Sigma CLS8169) before use.

For Figure 1D 100 mg of fresh leaf sample was homogenized in 100% MeOH in a PowerLyser (at 2,500 r/min for 30 s) and centrifuged at 12,000 r/min for 10 min. The pellet was re-extracted two times. For Figure S1 100 mg of lyophilized plant material was ground to fine powder in a PowerLyser and then extracted in 1 mL 100% MeOH for 2 h on a shaker at room temperature. After centrifugation at 12,000 r/min for 10 min, the SUP was left to dry in a fume hood. After estimating the weights of the pellets, they were dissolved in 50–200 µL DMSO and used at extract concentrations of 0.5–2.0 µL/mg

For LC-MS/MS analysis, the polar fraction was dried under vacuum, and the residue was resuspended in 250 µL of LC-grade water. After sonication of samples for 10 min in an ice-cooled sonicator bath, tubes were centrifuged for 15 min at full speed (>12,000 r/min). The supernatant (100 µL)—without any sediment from the bottom of the tube—was transferred to LC tubes for analysis. The samples were run on a UPLC-LC-MS machine as described previously (Salem et al., 2020). In brief, the UPLC system was equipped with an HSS T3 C18 reverse-phase column (100 × 2.1 mm internal diameter, 1.8 µm particle size; Waters) that was operated at a temperature of 40°C. The mobile phases consisted of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The flow rate of the mobile phase was 400 µL/min, and 2 µL of sample was loaded per injection. The UPLC instrument was connected to an Exactive Orbitrap-focus (Thermo Fisher Scientific) via a heated electrospray source (Thermo Fisher Scientific). The spectra were recorded using full-scan positive and negative ion-detection mode,

covering a mass range from m/z 100 to 1,500. The resolution was set to 70,000, and the maximum scan time was set to 250 ms. The sheath gas was set to a value of 60 while the auxiliary gas was set to 35. The transfer capillary temperature was set to 150°C while the heater temperature was adjusted to 300°C. The spray voltage was fixed at 3 kV, with a capillary voltage and a skimmer voltage of 25 V and 15 V, respectively. MS spectra were recorded from minutes 0–19 of the UPLC gradient. Processing of chromatograms, peak detection, and integration were performed using RefinerMS (version 5.3; GeneData). Metabolite identification and annotation were performed using standard compounds, tandem MS (MS/MS) fragmentation, metabolomics databases and an in-house reference compound library.

### Cell cultures

MCF7 and MDA-MB231 cells were maintained in RPMI-1640 (Thermo Scientific Hyclone) containing 10% FBS (Hyclone) and 100U penicillin/streptomycin (Biological Industries). SKBR3 was maintained in DMEM (Hyclone) containing 10% FBS and 100U penicillin/streptomycin.

### Assessing metabolic activity of cancer cells with MTS and MTT assays

For the MTT (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) cell viability assays in Figures 1D and S1 a total of 2,500 cells were seeded per well in triplicate in 96-well plates 1 d prior to the cell proliferation assay to allow for cell attachment. On the following day, cells were treated with the different extracts or vehicle control at 0.2 µL/100 µL medium for 48 h at 37°C. The medium was aspirated and fresh medium containing 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS, premixed in 5:1 ratio, Promega) was added to the treated cells. After 1 h incubation at 37°C in the dark, the optical density was read at 490 nm using a plate reader.

For MTT assays in Figures 3B, C and 4B around 11,000 SKBR3 cells were seeded per well in DMEM and incubated at 37°C overnight to allow for cell attachment. The following day cells were treated with the different extracts. After 48 h incubation at 37°C the media was aspirated and replaced with fresh DMEM media. After addition of 10 µL of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) solution (5 mg/ml, Sigma-Aldrich, USA) the cells were incubated in darkness for another 4 h. Formazan crystals were dissolved with a solubilizing agent. For Figures 3B, C and 4B 100 µL acidified SDS (10% SDS in 0.01 M HCl) was added followed by incubation at 37°C for 4 h in the dark. Alternatively, the crystals were solubilized by adding 100 µL DMSO and shaking the plate at room temperature for 15–30 min (for Figure 4E). Optical density was read at 570 nm using a plate reader.

For Figure 3B the pellet of total minus PK3507 was dissolved in 30 µL DMSO and the pellet containing the purified peak PK3507 was dissolved in 10 µL. Dilution curves were prepared for both samples and 4 µL were

used in the MTT assay. To assay the 24 fractions (Figure 3C), each fraction was resuspended in 10  $\mu$ L DMSO and 1  $\mu$ L was added to the assay. To determine IC<sub>50</sub> values all fractions were diluted as indicated in Figure 3E and 0.1% DMSO was used for control wells. IC<sub>50</sub> values of isolated compounds and standards were calculated by GraphPad Prism 4 software. All determinations were done in triplicates. For Figure 4B, 1  $\mu$ L of 100  $\mu$ L DMSO-dissolved methanol extract (F.W. 50 mg powder) was used in the assay.

#### Time-lapse imaging of SKBR3 cells treated with ursolic acid

Cells were seeded on a 6-well tissue culture plate, treated with drugs/fractions, and placed on a heat-controlled stage of a Zeiss Axiovert 200 M inverted microscope equipped with a 0.3 NA EC Plan-NeoFluar 10x lens (Carl Zeiss, Germany). The temperature was maintained at 37°C and CO<sub>2</sub> levels were maintained at 5%. Phase-contrast images were acquired at 15 min intervals for 72 h.

#### Cell cycle analysis by flow cytometry

2.75  $\times$  10<sup>5</sup> SKBR3 cells were seeded on a 6-well tissue culture plate and cultured for 72 h before treatment. Cells were treated with DMSO (0.1%) and varying concentrations of ursolic acid (12.5–30  $\mu$ M) for 24 and 48 h. Cells were harvested, washed with PBS and fixed in 70% ice-cold ethanol overnight. Fixed cells were stained in a solution containing 0.1% Triton-X 100, 10  $\mu$ g/ml RNase (Sigma-Aldrich) and 20  $\mu$ g/ml Propidium Iodide (PI) (Sigma-Aldrich). PI fluorescence was detected using BD LSRFortessa X-20 Cell Analyser (Becton, Dickinson & Company). Cell cycle analysis was performed using FlowJo software (Becton, Dickinson & Company). All experiments were performed in triplicates.

#### Identification of putative cysteine-rich peptides (cyclotides) with high-resolution LC-MS analysis

Methanolic extracts of *O. corymbosa* with anticancer activity to SK-BR3 (CIII-L2 and L3) were injected into an Orbitrap Elite mass spectrometer (Thermo Scientific Inc.) coupled with a Dionex UltiMate 3000 UHPLC (ultra-high-performance liquid chromatography) system (Thermo Scientific Inc.) to identify putative cysteine-rich peptides within the molecular weight range of 2,000–6,000 Da. Samples were sprayed using a Michrom's Thermo CaptiveSpray nanoelectrospray ion source (Bruker-Michrom Inc.) and separation was performed using a reversed-phase Acclaim PepMap RSL column (75  $\mu$ m ID  $\times$  15 cm, 2  $\mu$ m; Thermo Scientific). The mobile phase was 0.1% formic acid (FA) as eluent A and 90% ACN 0.1% FA as eluent B, with a flow rate of 0.3  $\mu$ L/min. A 60 min gradient was used for the elution as follows: 3% B for 1 min, 3%–35% B over 47 min, 35%–50% B over 4 min, 50%–80% B over 6 s, 80% for 78 s; then, it was reverted to the initial state over 6 s and maintained for 6.5 min.

#### Cyclotide purification by reversed-phased high-performance liquid chromatography (RP-HPLC) fractionation

Methanolic extracts of *O. corymbosa* with anticancer activity to SK-BR3 (CIII-L2 and L3) were fractionated using RP-HPLC (Shimadzu, Japan). A linear gradient of mobile phase A (0.05% TFA in H<sub>2</sub>O) and mobile phase B (0.05% TFA in acetonitrile [ACN]) was used with the C18 column (250  $\times$  4.6 mm, 5  $\mu$ m, 300 Å; Phenomenex, USA). MALDI-TOF MS (AB Sciex, USA) was used to identify the presence of the 3,507 Da compound in the eluted fractions which were then pooled to form the 3,507 Da-enriched fractions. The remaining fractions were pooled together to form the 3,507-Da-depleted fraction. The combined samples were then lyophilized for storage. For MTT assays, the pellet of total minus the PK3507 was dissolved in 30  $\mu$ L DMSO and the pellet containing the purified peak PK3507 was dissolved in 10  $\mu$ L. Dilution curves were prepared for both samples and 4  $\mu$ L were used in the MTT assay.

#### Identification of active metabolites with activity-guided fractionation

Frozen pulverized plant material (100 mg leaves of M15) was extracted with methanol (1 mL) in an ultrasonic bath for 20 min. The sample was centrifuged and the supernatant was dried under vacuum. The dried extract was reconstituted in methanol and fractionated on an analytical scale HPLC. 23 fractions and crude were submitted for biological testing. Two active fractions (6N and 6P) were observed.

Scale-up fractionation was carried out to obtain more 6 N and 6 P, and to further purify the major compounds. Oven-dried milled plant leaves and stems (7 g) were extracted twice with dichloromethane/methanol (1:1) in an ultrasonic bath for 20 min. The filtered extracts were combined and dried under vacuum. The dried extract (405 mg) was reconstituted in methanol and 250 mg was fractionated using a Preparative-HPLC. The fractions isolated were analyzed by <sup>1</sup>H NMR and HR-MS. Four known compounds were isolated, namely ursolic acid (1), pheophorbide a (2), phytol (3), and lutein (4). These compounds were identified by comparison of their spectroscopic data with those reported in the literature (Tsasi et al., 2015; Otaka et al., 2016; Miranda et al., 2017; Ibrahim et al., 2018).

#### Analytical scale fractionation by UPLC/QTOF

Analytical scale fractionation was carried out on an Agilent UPLC1290 coupled with a Quadrupole Time-of-Flight (Q-TOF) system. Separation was carried out with a reversed-phase C18 column (4.6  $\times$  75 mm) at 2 mL/min, using a 20 min linear gradient with 0.1% formic acid in both solvent A (water) and solvent B (acetonitrile). The typical QTOF operating parameters were as follows: positive ionization mode; sheath gas nitrogen flow, 12 L/min at 295°C; drying gas nitrogen flow, 8 L/min at 275°C; nebulizer pressure, 30 psi; nozzle voltage, 1.5 kV; capillary voltage, 4 kV. Lock masses in positive ion mode: purine ion at m/z 121.0509 and HP-921 ion at m/z 922.0098.

### Scale-up fractionation by preparative HPLC

Scale-up fractionation was performed on an Agilent HPLC1260 coupled with a 6130 Single Quadrupole (SQ) system. The extract was fractionated by a reversed-phase column C18 (30 × 100 mm) at 48 mL/min, under a gradient condition from 40% to 100% solvent B in 60 min. Ursolic acid, pheophorbide a, lutein and phytol were eluted at retention time 36, 42, 53, and 56 min respectively. Bruker DRX-400 NMR spectrometer with Cryoprobe, and 5-mm BBI probe head equipped with z-gradients was utilized to obtain <sup>1</sup>H NMR spectra of the compounds.

### Variable pressure-SEM

The flower was visualized untreated in a variable-pressure scanning electron microscope VP-SEM (Hitachi FlexSEM 1000 II) at 30 Pa and accelerating voltage of 10 kV, using a BSE detector.

### CETSA

ITDR MS-CETSA was essentially done as in [Dziekan et al. \(2020\)](#) with the following specifications. Cancer cell lysate of SKBR-3 was exposed to varying range of drug concentrations (65.67 μM–10 nM) and exposed to four different heat challenge temperatures 50°C, 55°C, 60°C, and 37°C as non-denaturing control. Remaining soluble protein fractions were analyzed by quantitative mass spectrometry to identify proteins exhibiting drug-dose dependent change in stability, suggestive of direct drug-protein interaction. During the data analysis step, two points were removed due to substantial deviations in relative protein abundances (the top concentration data point, due to a large number of single-point (likely non-specific) apparent stabilizations/interactions, as well as the no-drug control due to a likely pipetting error). Final protein stability profiles were plotted along the 10 nM–21.89 μM drug concentration gradient.

### Molecular dynamics simulation

The structure models of the proteins, Q9NZD8 Maspardin and Q6NVY1 3-hydroxyisobutyryl-CoA hydrolase, mitochondrial, were build using Alpha Fold2 server ([Mirdita et al., 2022](#)). The structure of Q16698 2,4-dienoyl-CoA reductase [(3E)-enoyl-CoA-producing], mitochondrial, was taken from the PDB 1W6U ([Alphey et al., 2005](#)). Both proteins and the ligand, ursolic acid, were constructed using the CHARMM-GUI ([Jo et al., 2008](#)) tool. Parameters of the proteins were based on the CHARMM36 force field ([Lee et al., 2016](#)) and parameters of the ligand were based on the CHARMM General Force Field ([Vanommeslaeghe and MacKerell, 2012](#); [Vanommeslaeghe et al., 2012](#)). The binding pockets of the proteins were identified by PointSite ([Yan et al., 2022](#)). The quick vina2 ([Alhossary et al., 2015](#)) was used to dock the ligand to the binding pockets. The best poses were chosen to be the initial structures for the following molecular dynamics simulations by manually checking all the vina docking output. All the topological files were converted to GROMACS ([Van Der Spoel et al., 2005](#)) format using CHARMM-GUI force field converter. The system were solvated with TIP3P ([Jorgensen et al., 1983](#)) water molecules and counterions were added to neutralize the system. The three protein-ligand systems were subjected to

molecular dynamics (MD) simulation using GROMACS v5.1.2 software ([Van Der Spoel et al., 2005](#)). The LINCS ([Hess, 2008](#)) algorithm was used to constrain bonds between heavy atoms and hydrogen to enable a time step of 2fs. A 1.2 nm cutoff was used for van der Waals interaction and short-range electrostatic interactions calculations, and the Particle Mesh Ewald method was implemented for long-range electrostatic calculations. Simulation temperature was maintained at 300 K using a V-rescale thermostat ([Bussi et al., 2007](#)) and 1 bar pressure using Parrinello-Rahman ([Parrinello, 1981](#)) barostat. Each simulation lasted 100 ns. The representative structures were taken as the center of the biggest cluster got from the gromos clustering method ([Daura et al., 1999](#)).

### Data availability statement

All the data generated or analyzed in this study are included in this article. The genome assembly, its annotation, and the sequencing data (DNA and RNA) have been deposited in the ENA (European Nucleotide Archive) under the accession number PRJEB50026.

## ACKNOWLEDGEMENTS

We thank Victor Albert and Charlotte Lindqvist for useful discussions, and the members of the Mutwil lab for feedback. I.J. is supported by Nanyang Biologics, M.M. is supported by a NTU Start-Up Grant and Singaporean Ministry of Education grant MOE2018-T2-2-053. J.M.D (NTU-PPF-2019). Y.K. is supported by Natural Product Research Laboratory Biomedical Research Council of A\*STAR (Agency for Science, Technology and Research) Transition Fund (H16/99/b0/004), the Singapore Institute of Food and Biotechnology Innovation core fund.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## AUTHOR CONTRIBUTIONS

Conceived the idea of the project: D.M.A., M.M. Designed the analysis: D.M.A., I.J., H.Y.L., M.M. Conducted the experiments and collected and analyzed the data: Y.W.L., Q.W.T., P.N., I.T.B., D.M.A., M.S., L.M.C., G.K., A.F., S.A., I.J., V.M., Z.K., S.K. Lai, S. K. Lim, H.Y.L., Y.P.L., J.D., S.T., Z.B., L.K.Y., Y.K., Y.I.L., Y.M., L. S., A.K., J.P.T., and B.U. Plant sampling and herbarium collection: Y.W.L., Q.W.T., P.N., I.T.B. Plant growth and stress experiments: D.M.A., I.T.B. RNA and DNA isolation and sequencing: Y.W.L., P.N., M.S., L.M.C., G.K. LC-MS/MS analysis: A.F., S.A. Cell cultures and anticancer activity: V.M., Z.K., S.K.L., H.Y.L., Y.P.L. Cellular Thermal Shift Assay: J.D., S.T., Z.B. Fractionation and identification of ursolic acid: L.K.Y., Y.K. Molecular dynamics simulation: Y.I.M., Y.M. Cyclotide analysis: L.S.,

A.K., J.P.T. Nanopore genome sequencing: M.S., B.U. Genome assembly and analysis: I.J. Wrote the paper with help from other co-authors: I.J., D.M.A., M.M. All authors read and approved the final manuscript.

**Edited by:** Amit Rai, Chiba University, Japan

**Received** Jul. 13, 2022; **Accepted** Feb. 18, 2023; **Published** Feb. 21, 2023

**OO:** OnlineOpen

## REFERENCES

- Abudu, Y.P., Pankiv, S., Mathai, B.J., Lamark, T., Johansen, T., and Simonsen, A. (2019). NIPSNAP1 and NIPSNAP2 act as “eat me” signals to allow sustained recruitment of autophagy receptors during mitophagy. *Autophagy* **15**: 1845–1847.
- Alhossary, A., Handoko, S.D., Mu, Y., and Kwoh, C.K. (2015). Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics* **31**: 2214–2216.
- Alioto, T., Alexiou, K.G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., Dhingra, A., Duval, H., Fernández I Martí, Á., Frias, L., Galán, B., García, J.L., Howad, W., Gómez-Garrido, J., Gut, M., Julca, I., Morata, J., Puigdomènech, P., Ribeca, P., Rubio Cabetas, M.J., and Arús, P. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: Results from the almond genome sequence. *Plant J.* **101**: 455–472.
- Alioto, T., Blanco, E., Parra, G., and Guigó, R. (2018). Using geneid to identify genes. *Curr. Protoc. Bioinformatics* **64**: e56.
- Alphey, M.S., Yu, W., Byres, E., Li, D., and Hunter, W.N. (2005). Structure and reactivity of human mitochondrial 2,4-dienoyl-CoA reductase: Enzyme-ligand interactions in a distinctive short-chain reductase active site. *J. Biol. Chem.* **280**: 3068–3077.
- Aminfar, Z., Rabiei, B., Tohidfar, M., and Mirjalili, M.H. (2019). Identification of key genes involved in the biosynthesis of triterpenic acids in the mint family. *Sci. Rep.* **9**: 15826.
- Andre, C.M., Legay, S., Deleruelle, A., Nieuwenhuizen, N., Punter, M., Brendolise, C., Cooney, J.M., Lateur, M., Hausman, J.-F., Larondele, Y., and Laing, W.A. (2016). Multifunctional oxidosqualene cyclases and cytochrome P450 involved in the biosynthesis of apple fruit triterpenic acids. *New Phytol.* **211**: 1279–1294.
- Artero-Castro, A., Castellvi, J., García, A., Hernández, J., Ramón y Cajal, S., and Leonart, M.E. (2011). Expression of the ribosomal proteins Rplp0, Rplp1, and Rplp2 in gynecologic tumors. *Hum. Pathol.* **42**: 194–203.
- Berini, J.L., Brockman, S.A., Hegeman, A.D., Reich, P.B., Muthukrishnan, R., Montgomery, R.A., and Forester, J.D. (2018). Combinations of abiotic factors differentially alter production of plant secondary metabolites in five woody plant species in the boreal-temperate transition zone. *Front. Plant Sci.* **9**: 1257.
- Beutler, J.A. (2019). Natural products as a foundation for drug discovery. *Curr. Protoc. Pharmacol.* **86**: e67.
- Blomme, A., Ford, C.A., Mui, E., Patel, R., Ntala, C., Jamieson, L.E., Planque, M., McGregor, G.H., Peixoto, P., Hervouet, E., Nixon, C., Salji, M., Gaughan, L., Markert, E., Repiscak, P., Sumpston, D., Blanco, G.R., Lilla, S., Kamphorst, J.J., Graham, D., and Leung, H.Y. (2020). 2,4-dienoyl-CoA reductase regulates lipid homeostasis in treatment-resistant prostate cancer. *Nat. Commun.* **11**: 2508.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**: 525–527.
- Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**: 59–60.
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**: 014101.
- Cai, Q., Lin, J., Wei, L., Zhang, L., Wang, L., Zhan, Y., Zeng, J., Xu, W., Shen, A., Hong, Z., and Peng, J. (2012). *Hedyotis diffusa* Willd inhibits colorectal cancer growth *in vivo* via inhibition of STAT3 signaling pathway. *Int. J. Mol. Sci.* **13**: 6117–6128.
- Campbell, M.S., Law, M., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J., Achawanantakun, R., Jiao, D., Lawrence, C.J., Ware, D., Shiu, S.-H., Childs, K.L., Sun, Y., Jiang, N., and Yandell, M. (2014). MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* **164**: 513–524.
- Chen, A., Kaganovsky, E., Rahimpour, S., Ben-Aroya, N., Okon, E., and Koch, Y. (2002). Two forms of Gonadotropin-releasing Hormone (GnRH) are expressed in human breast tissue and overexpressed in breast cancer: putative mechanism for the antiproliferative effect of GnRH by down-regulation of acidic ribosomal phosphoproteins P1 and P2. *Cancer Res.* **62**: 1036–1044.
- Chen, R., He, J., Tong, X., Tang, L., and Liu, M. (2016). The *Hedyotis diffusa* Willd. (Rubiaceae): A review on phytochemistry, pharmacology, quality control and pharmacokinetics. *Molecules* **21**: 710.
- Chen, Y., Wei, X., Lin, G., Liu, Y., and Zhang, Y. (2020). The complete mitochondrial genome of an endangered mangrove plant: *Scyphiphora hydrophyllacea*. *Mitochondrial DNA B Resour.* **5**: 2779–2780.
- Colgrave, M.L., Kotze, A.C., Huang, Y.-H., O’Grady, J., Simonsen, S.M., and Craik, D.J. (2008). Cyclotides: Natural, circular plant peptides that possess significant activity against gastrointestinal nematode parasites of sheep. *Biochemistry* **47**: 5581–5589.
- Cory, A.H., Owen, T.C., Bartrop, J.A., and Cory, J.G. (1991). Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. *Cancer Commun.* **3**: 207–212.
- Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W.F., and Mark, A.E. (1999). Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* **38**: 236–240.
- de Alencar, M.V.O.B., Islam, M.T., de Lima, R.M.T., Paz, M.F.C.J., Dos Reis, A.C., da Mata, A.M.O.F., Filho, J.W.G. de O., Cerqueira, G.S., Ferreira, P.M.P., E Sousa, J.M. de C., Mubarak, M.S., and Melo-Cavalcante, A.A.deC. (2019). Phytol as an anticarcinogenic and anti-tumoral agent: An *in vivo* study in swiss mice with DMBA-Induced breast cancer. *IUBMB Life* **71**: 200–212.
- Dhama K., Tiwari R., Chakraborty S., Saminathan M., Kumar A., Karthik K., Mohd Y.W., Amarpal, Singh V.S., Rahal A. (2014) Evidence based antibacterial potentials of medicinal plants and herbs countering bacterial pathogens especially in the era of emerging drug resistance: An integrated update. *Int. J. Pharmacol.* **10**: 1–43
- Ding, X., Bai, D., and Qian, J. (2014). Novel cyclotides from *Hedyotis biflora* inhibit proliferation and migration of pancreatic cancer cell *in vitro* and *in vivo*. *Med. Chem. Res.* **23**: 1406–1413.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
- Dziekian, J.M., Wirjanata, G., Dai, L., Go, K.D., Yu, H., Lim, Y.T., Chen, L., Wang, L.C., Puspita, B., Prabhu, N., Sobota, R.M., Nordlund, P., and Bozdech, Z. (2020). Cellular thermal shift assay for the identification of drug-target interactions in the *Plasmodium falciparum* proteome. *Nat. Protoc.* **15**: 1881–1921.
- El-Seedi, H.R., Yosri, N., Khalifa, S.A.M., Guo, Z., Musharraf, S.G., Xiao, J., Saeed, A., Du, M., Khatib, A., Abdel-Daim, M.M., Efferth, T., Göransson, U., and Verpoorte, R. (2021). Exploring natural

- products-based cancer therapeutics derived from Egyptian flora. *J. Ethnopharmacol.* **269**: 113626.
- Ellinghaus, D., Kurtz, S., and Willhoeft, U.** (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**: 18.
- Emms, D.M., and Kelly, S.** (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**: 238.
- Evans, W.K., Osoba, D., Feld, R., Shepherd, F.A., Bazos, M.J., and DeBoer, G.** (1985). Etoposide (VP-16) and cisplatin: An effective treatment for relapse in small-cell lung cancer. *J. Clin. Oncol.* **3**: 65–71.
- Fathi, E., Yarbro, J.M., and Homayouni, R.** (2021). NIPSNAP protein family emerges as a sensor of mitochondrial health. *BioEssays* **43**: e2100014.
- Feng, X.M., and Su, X.L.** (2019). Anticancer effect of ursolic acid via mitochondrial-dependent pathways. *Oncol. Lett.* **17**: 4761–4767.
- 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. *Proc. Natl. Acad. Sci. U.S.A.* **117**: 9451–9457.
- L. Forman, D. Bridson, eds.** (1991). *The herbarium handbook—R. Bot. Gard. Kew.*
- Fukushima, E.O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., Saito, K., and Muranaka, T.** (2011). CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol.* **52**: 2050–2061.
- Gabaldón, T.** (2008). Large-scale assignment of orthology: Back to phylogenetics? *Genome Biol.* **9**: 235.
- Göransson, U., Luijendijk, T., Johansson, S., Bohlin, L., and Claesson, P.** (1999). Seven novel macrocyclic polypeptides from *Viola arvensis*. *J. Nat. Prod.* **62**: 283–286.
- Graff, J.N., Puri, S., Bifulco, C.B., Fox, B.A., and Beer, T.M.** (2014). Sustained complete response to CTLA-4 blockade in a patient with metastatic, castration-resistant prostate cancer. *Cancer Immunol. Res.* **2**: 399–403.
- Gran, L., Sandberg, F., and Sletten, K.** (2000). *Oldenlandia affinis* (R&S) DC. *J. Ethnopharmacol.* **70**: 197–203.
- Greiner, S., Lehwark, P., and Bock, R.** (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**: W59–W64.
- Gremme, G., Steinbiss, S., and Kurtz, S.** (2013). GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM. Trans. Comput. Biol. Bioinform.* **10**: 645–656.
- Gu, G., Barone, I., Gelsomino, L., Giordano, C., Bonofiglio, D., Statti, G., Menichini, F., Catalano, S., and Andò, S.** (2012). *Oldenlandia diffusa* extracts exert antiproliferative and apoptotic effects on human breast cancer cells through ER $\alpha$ /Sp1-mediated p53 activation. *J. Cell. Physiol.* **227**: 3363–3372.
- Gudoityte, E., Arandarcikaite, O., Mazeikiene, I., Bendokas, V., and Liobikas, J.** (2021). Ursolic and oleanolic acids: Plant metabolites with neuroprotective potential. *Int. J. Mol. Sci.* **22**: 4599.
- Guo, X., Wang, R.-J., Simmons, M.P., But, P.P.-H., and Yu, J.** (2013). Phylogeny of the Asian *Hedyotis-Oldenlandia* complex (Spermacoceae, Rubiaceae): Evidence for high levels of polyphyly and the parallel evolution of diplophragmous capsules. *Mol. Phylogenet. Evol.* **67**: 110–122.
- Guo, X.-M., Wang, Z.-F., Zhang, Y., and Wang, R.-J.** (2021). Chromosomal-level assembly of the *Leptodermis oblonga* (Rubiaceae) genome and its phylogenetic implications. *Genomics* **113**: 3072–3082.
- Gustafson, K.R., McKee, T.C., and Bokesch, H.R.** (2004). Anti-HIV cyclotides. *Curr. Protein Pept. Sci.* **5**: 331–340.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., and Regev, A.** (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**: 1494–1512.
- Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R.** (2008). Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**: R7.
- Han, E.-K., Cho, W.-B., Tamaki, I., Choi, I.-S., and Lee, J.-H.** (2021). Comparative mitogenomic analysis reveals gene and intron dynamics in Rubiaceae and intra-specific diversification in *Damnacanthus indicus*. *Int. J. Mol. Sci.* **22**: 7237.
- Han, Y., and Wessler, S.R.** (2010). MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**: e199.
- Hartmann, T.** (2004). Plant-derived secondary metabolites as defensive chemicals in herbivorous insects: A case study in chemical ecology. *Planta* **219**: 1–4.
- Hawkins, C., Ginzburg, D., Zhao, K., Dwyer, W., Xue, B., Xu, A., Rice, S., Cole, B., Paley, S., Karp, P., and Rhee, S.Y.** (2021). Plant metabolic network 15: A resource of genome-wide metabolism databases for 126 plants and algae. *J. Integr. Plant Biol.* **63**: 1888–1905.
- Hess, B.** (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**: 116–122.
- He, Z., Xu, Q., Wang, X., Wang, J., Mu, X., Cai, Y., Qian, Y., Shao, W., and Shao, Z.** (2018). RPLP1 promotes tumor metastasis and is associated with a poor prognosis in triple-negative breast cancer patients. *Cancer Cell. Int.* **18**: 170.
- Hill, S.J., Rolland, T., Adelmant, G., Xia, X., Owen, M.S., Dricot, A., Zack, T.I., Sahni, N., Jacob, Y., Hao, T., McKinney, K.M., Clark, A.P., Reyon, D., Tsai, S.Q., Joung, J.K., Beroukhim, R., Marto, J.A., Vidal, M., Gaudet, S., Hill, D.E., and Livingston, D.M.** (2014). Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. *Genes Dev.* **28**: 1957–1975.
- Huerta-Cepas, J., and Gabaldón, T.** (2011). Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics* **27**: 38–45.
- Huerta-Cepas, J., Serra, F., and Bork, P.** (2016). ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**: 1635–1638.
- Hu, E., Wang, D., Chen, J., and Tao, X.** (2015). Novel cyclotides from *Hedyotis diffusa* induce apoptosis and inhibit proliferation and migration of prostate cancer cells. *Int. J. Clin. Exp. Med.* **8**: 4059–4065.
- Huerta-Cepas J., Capella-Gutierrez S., Pryszcz L.P., Denisov I., Kormes D., Marcet-Houben M., Gabaldón T.** (2011) PhylomeDB v3.0: An expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res.* **39**: D556–60
- Ibrahim, M.B., Sowemimo, A.A., Venables, L., Koorbanally, N.A., Awolola, G.V., Sofidiya, M.O., Odukoya, O.A., Koekemoer, T., and van de Venter, M.** (2018). Biological evaluation of phytoconstituents from *Markhamia tomentosa* ethanolic leaf extract. *S. Afr. J. Bot.* **115**: 31–36.
- Iqbal, J., Abbasi, B.A., Ahmad, R., Mahmood, T., Kanwal, S., Ali, B., Khalil, A.T., Shah, S.A., Alam, M.M., and Badshah, H.** (2018). Ursolic acid a promising candidate in the therapeutics of breast cancer: Current status and future implications. *Biomed. Pharmacother.* **108**: 752–756.
- Isah, T.** (2019). Stress and defense responses in plant secondary metabolites production. *Biol. Res.* **52**: 39.
- Iwata, H., and Gotoh, O.** (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* **40**: e161.

- Jafari, R., Almqvist, H., Axelsson, H., Ignatushchenko, M., Lundbäck, T., Nordlund, P., and Martinez Molina, D. (2014). The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* **9**: 2100–2122.
- Jäger, S., Trojan, H., Kopp, T., Laszczyk, M.N., and Scheffler, A. (2009). Pentacyclic triterpene distribution in various plants—Rich sources for a new group of multi-potent plant extracts. *Molecules* **14**: 2016–2031.
- Jaillon, O., Aury, J.-M., Noel, B., Pollicriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Huguency, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., Billault, A., and French-Italian Public Consortium for Grapevine Genome Characterization (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.
- Jennings, C., West, J., Waine, C., Craik, D., and Anderson, M. (2001). Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from *Oldenlandia affinis*. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 10614–10619.
- Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J.E., McKain, M.R., McNeal, J., Rolf, M., Ruzicka, D.R., Wafula, E., Wickett, N.J., Wu, X., Zhang, Y., Wang, J., ZhangYeting, Carpenter, E.J., Deyholos, M.K., Kutchan, T.M., Chandrabali, A.S., Soltis, P.S., Stevenson, D.W., and Depamphilis, C.W. (2012). A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* **13**: R3.
- Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C.W., Yi, T.-S., and Li, D.-Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**: 241.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjov, M., Yong, S.-Y., Lopez, R., and Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**: 926–935.
- Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**: 1859–1865.
- Julca, I., Marcet-Houben, M., Vargas, P., and Gabaldón, T. (2018). Phylogenomics of the olive tree (*Olea europaea*) reveals the relative contribution of ancient allo- and autopolyploidization events. *BMC Biol.* **16**: 15.
- Kaewthawee, N., and Brimson, S. (2013). The effects of ursolic acid on cytokine production via the MAPK pathways in leukemic T-cells. *EXCLI J.* **12**: 102–114.
- Kalita-de Croft, P., Straube, J., Lim, M., Al-Ejeh, F., Lakhani, S.R., and Saunus, J.M. (2019). Proteomic analysis of the breast cancer brain metastasis microenvironment. *Int. J. Mol. Sci.* **20**: 2524.
- Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**: 772–780.
- Kavalappa, Y.P., Gopal, S.S., and Ponesakki, G. (2021). Lutein inhibits breast cancer cell growth by suppressing antioxidant and cell survival signals and induces apoptosis. *J. Cell. Physiol.* **236**: 1798–1809.
- Khaghanzadeh, N., Nakamura, K., Kuramitsu, Y., Ghaderi, A., and Mojtahedi, Z. (2016). Immune-associated proteins with potential in vivo anti-tumor activities are upregulated in lung cancer cells treated with umbelliprenin: A proteomic approach. *Oncol. Lett.* **12**: 5295–5302.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**: 722–736.
- Krause, J., and Tobin, G. (2013). Discovery, development, and regulation of natural products. In *Using Old Solutions to New Problems: Natural Drug Discovery in the 21st Century*, M. Kulka, ed., (InTech).
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**: 1639–1645.
- Kuai, B., Chen, J., and Hörtensteiner, S. (2018). The biochemistry and molecular biology of chlorophyll breakdown. *J. Exp. Bot.* **69**: 751–767.
- Lajis, N.H.J., Ahmad, R. (2006). Phytochemical studies and pharmacological activities of plants in genus *Hedyotis/Oldenlandia*. In *Studies in natural products chemistry* (Elsevier), pp. 1057–1090.
- Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y., Jo, S., Pande, V.S., Case, D.A., Brooks, C.L., MacKerell, A.D., Klauda, J.B., and Im, W. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **12**: 405–413.
- Lewinska, A., Adamczyk-Grochala, J., Kwasniewicz, E., Deregowska, A., and Wnuk, M. (2017). Ursolic acid-mediated changes in glycolytic pathway promote cytotoxic autophagy and apoptosis in phenotypically different breast cancer cells. *Apoptosis* **22**: 800–815.
- Lin, C.-C., Ng, L.-T., Yang, J.-J., and Hsu, Y.-F. (2002). Anti-inflammatory and hepatoprotective activity of *peh-hue-juwa-chi-cao* in male rats. *Am. J. Chin. Med.* **30**: 225–234.
- Lin, J., Li, Q., Chen, H., Lin, H., Lai, Z., and Peng, J. (2015). *Hedyotis diffusa* Willd. extract suppresses proliferation and induces apoptosis via IL-6-inducible STAT3 pathway inactivation in human colorectal cancer cells. *Oncol. Lett.* **9**: 1962–1970.
- LiuYu, Li, C., Shen, S., Chen, X., Szlachta, K., Edmonson, M.N., Shao, Y., Ma, X., Hyle, J., Wright, S., Ju, B., Rusch, M.C., LiuYanling, Li, B., Macias, M., Tian, L., Easton, J., Qian, M., Yang, J.J., Hu, S., and Zhang, J. (2020). Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. *Nat. Genet.* **52**: 811–818.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li, H., Li, C., Xia, B., Zhou, Y., Lin, L., and Liao, D. (2015). A chemotaxonomic study of phytochemicals in *Hedyotis corymbosa*. *Biochem. Syst. Ecol.* **62**: 173–177.
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.
- Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**: e119.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol.* **15**: 550.
- Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* **53**: 661–673.
- Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., Lisch, D., and Freeling, M. (2008). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol.* **148**: 1772–1781.
- Malik, S., Saito, H., Takaoka, M., Miki, Y., and Nakanishi, A. (2016). BRCA2 mediates centrosome cohesion via an interaction with cytoplasmic dynein. *Cell Cycle* **15**: 2145–2156.
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764–770.

- Martinez Molina, D., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E.A., Dan, C., Sreekumar, L., Cao, Y., and Nordlund, P. (2013). Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**: 84–87.
- Ma, T.-T., Zhang, G.-L., Dai, C.-F., Zhang, B.-R., Cao, K.-X., Wang, C.-G., Yang, G.-W., and Wang, X.-M. (2020). *Scutellaria barbata* and *Hedyotis diffusa* herb pair for breast cancer treatment: Potential mechanism based on network pharmacology. *J. Ethnopharmacol.* **259**: 112929.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297–1303.
- Michael, T.P., Jupe, F., Bemm, F., Motley, S.T., Sandoval, J.P., Lanz, C., Loudet, O., Weigel, D., and Ecker, J.R. (2018). High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* **9**: 541.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**: 1530–1534.
- Mirabelli, P., Coppola, L., and Salvatore, M. (2019). Cancer cell lines are useful model systems for medical research. *Cancers* **11**: 1098.
- Miranda, N., Gerola, A.P., Novello, C.R., Ueda-Nakamura, T., de Oliveira Silva, S., Dias-Filho, B.P., Hioka, N., de Mello, J.C.P., and Nakamura, C.V. (2017). Pheophorbide a, a compound isolated from the leaves of *Arrabidaea chica*, induces photodynamic inactivation of *Trypanosoma cruzi*. *Photodiagnosis Photodyn. Ther.* **19**: 256–265.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* **19**: 679–682.
- Mishra, K., Dash, A.P., Swain, B.K., and Dey, N. (2009). *Anti-malarial activities of Andrographis paniculata and Hedyotis corymbosa* extracts and their combination with curcumin. *Malar. J.* **8**: 26.
- Modi, K., and Shah, M.B. (2017). Determination of oleanolic acid, ursolic acid, lupeol, and stigmasterol by high-performance thin-layer chromatographic method in *Oldenlandia corymbosa* Linn. *J. Planar Chrom. Modern TLC* **30**: 32–35.
- Nguyen, V.D. (1993). *Medicinal plants of Vietnam, Cambodia and Laos*. Mekong Printing, Vietnam.
- Nordgard, S.H., Johansen, F.E., Alnaes, G.I.G., Bucher, E., Syvänen, A.-C., Naume, B., Børresen-Dale, A.-L., and Kristensen, V.N. (2008). Genome-wide analysis identifies 16q deletion associated with survival, molecular subtypes, mRNA expression, and germline haplotypes in breast cancer patients. *Genes Chromosomes Cancer* **47**: 680–696.
- Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18**: 2024–2033.
- Otaka, J., Seo, S., and Nishimura, M. (2016). Lutein, a natural carotenoid, induces  $\alpha$ -1,3-glucan accumulation on the cell wall surface of fungal plant pathogens. *Molecules* **21**: 980.
- Pandey, K., Sharma, P.K., and Dudhe, R. (2012). Anticancer activity of *Parthenium hysterophorus* Linn and *Oldenlandia corymbosa* Lam by Srb Method. *Sci. Rep.* **1**: 325.
- Parrinello, M. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**: 7182–7190.
- Pranabesh, G. (2019). Botanical features, phytochemical and pharmacological overviews of *Oldenlandia corymbosa* Linn.: A brief review. *Pharma Innov. J.* **8**: 464–468.
- Ramakrishna, A., and Ravishankar, G.A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal. Behav.* **6**: 1720–1731.
- Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**: 1432.
- Reed, J., Stephenson, M.J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., Goss, R.J.M., Goossens, A., O'Connell, M.A., and Osbourn, A. (2017). A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* **42**: 185–193.
- Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**: 245.
- Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N.M., Salman-Minkov, A., Mayzel, J., Chay, O., and Mayrose, I. (2015). The Chromosome Counts Database (CCDB)—A community resource of plant chromosome numbers. *New Phytol.* **206**: 19–26.
- Richards, E.J., and Ausubel, F.M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**: 127–136.
- Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**: 460.
- Robey, R.B., and Hay, N. (2009). Is Akt the “Warburg kinase”? Akt-energy metabolism interactions and oncogenesis. *Semin. Cancer Biol.* **19**: 25–31.
- Sadasivan, S., Latha, P.G., Sasikumar, J.M., Rajashekar, S., Shyamal, S., and Shine, V.J. (2006). Hepatoprotective studies on *Hedyotis corymbosa* (L.) Lam. *J. Ethnopharmacol.* **106**: 245–249.
- Sakai, S., and Wright, S.J. (2007). Reproductive ecology of 21 coexisting *Psychotria* species (Rubiaceae): when is heterostyly lost? *Biol. J. Linn. Soc.* **93**: 125–134.
- Salem, M.A., Yoshida, T., Perez de Souza, L., Alseekh, S., Bajdzienko, K., Fernie, A.R., and Giavalisco, P. (2020). An improved extraction method enables the comprehensive analysis of lipids, proteins, metabolites and phytohormones from a single sample of leaf tissue under water-deficit stress. *Plant J.* **103**: 1614–1632.
- Sauter, E.R. (2020). Cancer prevention and treatment using combination therapy with natural compounds. *Expert Rev. Clin. Pharmacol.* **13**: 265–285.
- Schattner, P., Brooks, A.N., and Lowe, T.M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**: W686–W689.
- Schemske, D.W. (1978). Evolution of reproductive characteristics in impatiens (balsaminaceae): the significance of cleistogamy and chasmogamy. *Ecology* **59**: 596–613.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., and Jackson, S.A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* **463**: 178–183.
- Schwacke, R., Ponce-Soto, G.Y., Krause, K., Bolger, A.M., Arsova, B., Hallab, A., Gruden, K., Stitt, M., Bolger, M.E., and Usadel, B. (2019). MapMan4: A refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* **12**: 879–892.
- Seo, D.Y., Lee, S.R., Heo, J.-W., No, M.-H., Rhee, B.D., Ko, K.S., Kwak, H.-B., and Han, J. (2018). Ursolic acid in health and disease. *Korean J. Physiol. Pharmacol.* **22**: 235–248.
- Shan, Y., Gao, Y., Jin, W., Fan, M., Wang, Y., Gu, Y., Shan, C., Sun, L., Li, X., Yu, B., Luo, Q., and Xu, Q. (2019). Targeting HIBCH to reprogram valine metabolism for the treatment of colorectal cancer. *Cell Death Dis.* **10**: 618.
- Silva, N.C.C., and Fernandes Júnior, A. (2010). Biological properties of medicinal plants: A review of their antimicrobial activity. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **16**: 402–413.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: Assessing genome assembly and

- annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.
- Sivaprakasam, S., Karunakaran, K., Subburaya, U., Kuppusamy, S., and Subashini, T.** (2014). A review on phytochemical and pharmacological profile of *Hedyotis corymbosa* Linn. *Int. J. Pharm. Sci. Rev. Res.* **26**: 320–324.
- Smit, A.F.A., Hubley, R., and Green, P.** (2015). RepeatMasker Open-4.0. RepeatMasker. URL <http://www.repeatmasker.org> (accessed 10.18.21).
- Stamatakis, A.** (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B.** (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**: W435–W439.
- Steinbiss, S., Willhoeft, U., Gremme, G., and Kurtz, S.** (2009). Fine-grained annotation and classification of *de novo* predicted LTR retrotransposons. *Nucleic Acids Res.* **37**: 7002–7013.
- Storer, J., Hubley, R., Rosen, J., Wheeler, T.J., and Smit, A.F.** (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **12**: 2.
- Tacar, O., Sriamornsak, P., and Dass, C.R.** (2013). Doxorubicin: An update on anticancer molecular action, toxicity and novel drug delivery systems. *J. Pharm. Pharmacol.* **65**: 157–170.
- Tam, J.P., Lu, Y.A., Yang, J.L., and Chiu, K.W.** (1999). An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 8913–8918.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S.** (2017). GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**: W6–W11.
- Tsasi, G., Samara, P., Tsitsilonis, O., Jürgenliemk, G., and Skaltsa, H.** (2015). Isolation, identification and cytotoxic activity of triterpenes and flavonoids from green walnut (*Juglans regia* L.) Pericarps. *Rec. Nat. Prod.* **10**: 83–92.
- Ursini-Siegel, J., Rajput, A.B., Lu, H., Sanguin-Gendreau, V., Zuo, D., Papavasiliou, V., Lavoie, C., Turpin, J., Cianflone, K., Huntsman, D. G., and Muller, W.J.** (2007). Elevated expression of DecR1 impairs ErbB2/Neu-induced mammary tumor development. *Mol. Cell. Biol.* **27**: 6361–6371.
- Vanommeslaeghe, K., and MacKerell, A.D.** (2012). Automation of the CHARMM General Force Field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **52**: 3144–3154.
- Vanommeslaeghe, K., Raman, E.P., and MacKerell, A.D.** (2012). Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52**: 3155–3168.
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J.C.** (2005). GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**: 1701–1718.
- van Rengs, W.M.J., Schmidt, M.H.-W., Effgen, S., Le, D.B., Wang, Y., Zaidan, M.W.A.M., Huettel, B., Schouten, H.J., Usadel, B., and Underwood, C.J.** (2022). A chromosome scale tomato genome built from complementary PacBio and Nanopore sequences alone reveals extensive linkage drag during breeding. *Plant J.* **110**: 572–588.
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M.** (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* **27**: 737–746.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M.** (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.
- Wang, C., Zhou, X., Wang, Y., Wei, D., Deng, C., Xu, X., Xin, P., and Sun, S.** (2017). The Antitumor Constituents from *Hedyotis diffusa* Willd. *Molecules* **22**: 2101.
- Wang, J., Jiang, Z., Xiang, L., Li, Y., Ou, M., Yang, X., Shao, J., Lu, Y., Lin, L., Chen, J., Dai, Y., and Jia, L.** (2014). Synergism of ursolic acid derivative US597 with 2-deoxy-D-glucose to preferentially induce tumor cell death by dual-targeting of apoptosis and glycolysis. *Sci. Rep.* **4**: 5006.
- Wang, N., Xiang, Y., Fang, L., Wang, Y., Xin, H., and Li, S.** (2013). Patterns of gene duplication and their contribution to expansion of gene families in grapevine. *Plant Mol. Biol. Rep.* **31**: 852–861.
- Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J.C., and Paterson, A.H.** (2012). MScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**: e49.
- Wickell, D., Kuo, L.-Y., Yang, H.-P., Dhabalia Ashok, A., Irisarri, I., Dadras, A., de Vries, S., de Vries, J., Huang, Y.-M., Li, Z., Barker, M. S., Hartwick, N.T., Michael, T.P., and Li, F.-W.** (2021). Underwater CAM photosynthesis elucidated by *Isoetes* genome. *Nat. Commun.* **12**: 6348.
- Wong, K.M., Turner, I.M., Wang, R.J., Harwood, R., Seah, W.W., Ng, X. Y., Lim, R.C.J., Lua, H.K., and Mahyuni, R.** (2019). Rubiaceae. In *Flora of Singapore Vol. 13 Gentianales, Flora of Singapore* (National Parks Board), pp. 1–358.
- Woźniak, Ł., Skąpska, S., and Marszałek, K.** (2015). Ursolic acid—A pentacyclic triterpenoid with a wide spectrum of pharmacological activities. *Molecules* **20**: 20614–20641.
- Xie, C., Cao, K., Peng, D., and Qin, L.** (2021). RPLP1 is highly expressed in hepatocellular carcinoma tissues and promotes proliferation, invasion and migration of human hepatocellular carcinoma Hep3b cells. *Exp. Ther. Med.* **22**: 752.
- Yang, L., Wen, K.-S., Ruan, X., Zhao, Y.-X., Wei, F., and Wang, Q.** (2018). Response of plant secondary metabolites to environmental factors. *Molecules* **23**: 762.
- Yan, S., Huang, C., Wu, S., and Yin, M.** (2010). Oleanolic acid and ursolic acid induce apoptosis in four human liver cancer cell lines. *Toxicol. In Vitro* **24**: 842–848.
- Yan, X., Lu, Y., Li, Z., Wei, Q., Gao, X., Wang, S., Wu, S., and Cui, S.** (2022). Pointsite: A point cloud segmentation tool for identification of protein ligand binding atoms. *J. Chem. Inf. Model.* **62**: 2835–2845.
- Yan, Z., Lai, Z., and Lin, J.** (2017). Anticancer properties of traditional chinese medicine. *Comb. Chem. High Throughput Screen.* **20**: 423–429.
- Yasumoto, S., Fukushima, E.O., Seki, H., and Muranaka, T.** (2016). Novel triterpene oxidizing activity of *Arabidopsis thaliana* CYP716A subfamily enzymes. *FEBS Lett.* **590**: 533–540.
- Yeh, Y.-C., Chen, H.-Y., Yang, S.-H., Lin, Y.-i.-H.sien, Chiu, J.-H., Lin Yi-Hsuan, and Chen, J.-L.** (2014). *Hedyotis diffusa* combined with *Scutellaria barbata* are the core treatment of chinese herbal medicine used for breast cancer patients: A population-based study. *Evid. Based. Complement. Alternat. Med.* **2014**: 202378.
- Yik, M.H.-Y., Kong, B.L.-H., Siu, T.-Y., Lau, D.T.-W., Cao, H., and Shaw, P.-C.** (2021). Differentiation of *Hedyotis diffusa* and common adulterants based on chloroplast genome sequencing and DNA barcoding markers. *Plants* **10**: 161.
- You, B.-J., Wu, Y.-C., Wu, C.-Y., Bao, B.-Y., Chen, M.-Y., Chang, Y.-H., and Lee, H.-Z.** (2011). Proteomics displays cytoskeletal proteins and chaperones involvement in *Hedyotis corymbosa*-induced photokilling in skin cancer cells. *Exp. Dermatol.* **20**: 653–658.
- Yue, G.G.-L., Kin-Ming Lee, J., Cheng, L., Chung-Lap Chan, B., Jiang, L., Fung, K.-P., Leung, P.-C., and Bik-San Lau, C.** (2012). Reversal of P-glycoprotein-mediated multidrug resistance in human hepatoma cells by hedyotiscone A, a compound isolated from *Hedyotis corymbosa*. *Xenobiotica* **42**: 562–570.
- Zeitlmann, L., Sirim, P., Kremmer, E., and Kolanus, W.** (2001). Cloning of ACP33 as a novel intracellular ligand of CD4. *J. Biol. Chem.* **276**: 9123–9132.

**Zhang, X.F., Wang, J.H., Zhao, K.K., Fan, W.W., Wang, H.X., Zhu, Z. X., and Wang, H.F.** (2019). Complete plastome sequence of *Hedyotis ovata* Thunb. ex Maxim (Rubiaceae): An endemic shrub in Hainan, China Mitochondrial DNA Part B **4**: 675–676.

**Zheng, Q., Jin, F., Yao, C., Zhang, T., Zhang, G., and Ai, X.** (2012). Ursolic acid-induced AMP-activated protein kinase (AMPK) activation contributes to growth inhibition and apoptosis in human bladder cancer T24 cells. Biochem. Biophys. Res. Commun. **419**: 741–747.

**Zhou, H., Zhang, J., Yan, Z., Qu, M., Zhang, G., Han, J., Wang, F., Sun, K., Wang, L., and Yang, X.** (2022). DECR1 directly activates HSL to promote lipolysis in cervical cancer cells. Biochim. Biophys. Acta, Mol. Cell Biol. Lipids **1867**: 159090.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: <http://onlinelibrary.wiley.com/doi/10.1111/jipb.13469/supinfo>

**Data S1.** Video 1. Phase-contrast movie of SKBR3 cells treated with 0.1% DMSO over 72 h

**Data S2.** Video 2. Phase-contrast movie of SKBR3 cells treated with 30  $\mu$ M ursolic acid over 72 h

**Data S3.** Pymol session file containing the structures of *DECR1*, Maspardin and *HIBCH* bound by ursolic acid

**Figure S1.** Liquid (left) and solid (middle, right) seedling growth conditions

**Figure S2.** MTS assay comparison of *Oldenlandia corymbosa* to 61 medicinal plants using MB231, SKBR3, and MCF7 cancer cell lines

**Figure S3.** Genome analysis of *O. corymbosa*

**Figure S4.** Analysis of plastid sequence of *Oldenlandia corymbosa*

**Figure S5.** Gene duplication, synteny and Ks analysis of *O. corymbosa*

**Figure S6.** LC-MS/MS spectra of *Oldenlandia corymbosa* leaf samples identifies a potential cyclotide

**Figure S7.** 1H NMR spectra of *O. corymbosa* fractions: Fraction 34D ursolic acid in CDCl<sub>3</sub> and Fraction 34J lutein in CDCl<sub>3</sub>

**Figure S8.** 1H NMR spectra of *O. corymbosa* fractions: Fraction 34F (Pheophorbide a) in CDCl<sub>3</sub> and 34H (Phytol) in CDCl<sub>3</sub>

**Figure S9.** *O. corymbosa* phenotypes caused by abiotic stress

**Figure S10.** MTT assay activities against SKBR3 cancer cell lines

**Figure S11.** Phase contrast images of SKBR3 cells treated with ursolic acid and DMSO control over 72 h

**Figure S12.** Atomic model of *DECR1* in complex with natural ligand and ursolic acid

**Table S1.** Collected *Oldenlandia* species and locations

**Table S2.** MTT assays testing the anti-cancer activity of several *Oldenlandia* species

**Table S3.** Characteristics of the plastid and mitochondrial genome of different species of the family Rubiaceae

**Table S4.** Characterization and quantification of repetitive sequences in the *O. corymbosa* genome

**Table S5.** List of species used in this study

**Table S6.** Classification of duplicated genes of *O. corymbosa*

**Table S7.** MTT assay activities against SKBR3 cancer cell lines for the cyclotide enriched dilutions (3,507 Da), and cyclotide-depleted dilutions (Depleted—Depleted 18:8)

**Table S8.** MTT assay activities against SKBR3 cancer cell lines for the initial fractionation

**Table S9.** Dilution series for the fractions and standards

**Table S10.** MTT assay activities against SKBR3 cancer cell lines from plants exposed to abiotic stress

**Table S11.** Relative abundance of ursolic acid, oleanolic acid, and pheophorbide a in the same samples used for the MTT assay

**Table S12.** Pearson correlation coefficient (PCC) of the MTT assay profiles versus the relative abundance of five metabolites

**Table S13.** RNA-sequencing gene expression data. TPM values of each sample

**Table S14.** Pearson correlation coefficient (PCC) of the two genes of *O. corymbosa* (*OLC1v1015725* and *OLC1v1035754*) versus all the other expressed genes

**Table S15.** List of genes up- and down-regulated under cold 8°C conditions in roots

**Table S16.** CETSA-MS analysis of fraction 34D against SKBR3 cell lysate



Scan using WeChat with your smartphone to view JIPB online



Scan with iPhone or iPad to view JIPB online