Title

Single nuclei transcriptome of the Lesser Duckweed *Lemna minuta* reveals cell trajectories for an entire plant

Short title

Cell trajectories for an entire simplified plant

Authors

Bradley W. Abramson[1], Mark Novotny[2], Nolan T. Hartwick[1], Kelly Colt[1], Brian D. Aevermann[2], Richard H. Scheuermann[2,3,4], Todd P. Michael[1]

Affiliations

[1]The Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA.
[2]Department of Informatics, J. Craig Venter Institute, La Jolla, CA 92037, USA.
[3]Department of Pathology, University of California San Diego, La Jolla, CA 92093, USA.
[4]Division of Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA 92037, USA.

Sentence summary

Genome and single nuclei transcriptome of the Lesser Duckweed *Lemna minuta* enables tracing of all developmental, transitional and terminal cells of an entire plant.

Author Contributions

TPM and BWA conceived the study, BWA and MN carried out the lab work, KC performed sequencing, BWA, BDA, NH and TPM performed the analysis, RHS and TPM oversaw the project, and TPM and BWA wrote the manuscript.

Funding

This work was funded by The Tang Fund and the Pioneer Fund Trustees with their support through the Pioneer Fund Postdoctoral Scholar Award.

Corresponding Author

Todd P. Michael, tmichael@salk.edu, toddpmichael@gmail.com

Abstract

The ability to trace every cell in some model organisms has led to the fundamental understanding of development and cellular function. However, in plants the complexity of cell number, organ size and developmental times makes this a challenge even in the diminutive model plant *Arabidopsis thaliana*. Here we develop the Lesser Duckweed *Lemna minuta* as a model with a reduced body plan, small genome size and clonal growth pattern that enables simultaneous tracing of cells from the entire plant over the complete developmental cycle. We generated a chromosome-resolved genome for the 360 megabase genome and defined the growth trajectory of the entire plant with single nuclei RNA sequencing. The *L. minuta* gene complement represents a primarily non-redundant set with only the ancient *tau* whole genome

duplication shared with all monocots, and paralog expansion as a result of tandem duplications related to phytoremediation. Thirteen distinct cell types representing meristem, the leaf-stem fusion called a frond, and root-like tissues were defined using gene orthology with single cell expression from model plants, gene ontology categories, and cell trajectory analysis. Dividing meristem cells give rise to two main branches of root-transition and mesophyll cells, which then give rise to terminally differentiated parenchyma, epidermal and root cells. Mesophyll tissues express high levels of elemental transport genes consistent with this tissue playing a role in *L. minuta* wastewater detoxification. The *L. minuta* genome and cell map provide a paradigm to decipher developmental genes and pathways for an entire plant.

Introduction

Single cell sequencing ushered in a new era of biology where it is possible to characterize cell types and function with unprecedented detail. In plants, this has resulted in detailed single cell RNA-seq (scRNA-seq) and scATAC-seq (Assay for Transposase-Accessible Chromatin) datasets primarily on different organ types from well-studied model organisms (Denyer et al., 2019; Ryu et al., 2019; Dorrity et al., 2020; Gujas et al., 2020; Liu et al., 2020a; Marand et al., 2020; Song et al., 2020; Liu et al., 2020b; Lopez-Anido et al., 2021; Xu et al., 2021). All plant studies to date start with protoplast isolation, which has the potential to miss some recalcitrant cell types and often requires correction to control for transcriptional changes elicited by lengthy enzyme treatment and centrifugation steps (Ryu et al., 2019). In addition, there are different single cell approaches that generally either result in assaying many cells with lower gene coverage (10X Genomics) or fewer cells with deeper gene coverage (SMARTseq2) (Yamawaki et al., 2021). For example, SMARTseq2 sequencing coupled with Fluorescence Activated Cell Sorting (FACS) of nuclei, or single nuclei RNA sequencing (snRNA-seq), provides deep transcriptome coverage of individual nuclei (Krishnaswami et al., 2016) while avoiding loss of cells or introducing issues associated with the protoplast isolation timing. The use of snRNA-seq in plants is nascent but can extend as a broader application for studying abiotic/biotic treatments and whole plant analysis.

Most single cell studies to date have been on specific plant organ tissue, such as the root, meristem or inflorescence, that encompass limited cell trajectories. Duckweed, aquatic plants in the family Lemnaceae, provide a unique opportunity to follow cells through all developmental time scales from one population due to its minimal morphology and clonal pattern of growth (Fig. 1). These basal monocots are some of the smallest and fastest growing plants on Earth, ranging in size from under a millimeter to centimeter, with some species doubling in under a day (Michael et al., 2020). There are five genera in the Lemnaceae with three having a flat leaf-like structure called a frond and root-like structures that may act as rudders or anchors (*Spirodela*, *Lemna* and *Landoltia*); the other two lack roots and are spherical or flat (*Wolffia* and *Wolffiella*). Before the model plant *Arabidopsis thaliana* rose to prominence, Duckweed were an important system for reductionist biology leading to fundamental understanding of chronobiology, flowering time, and phytohormone action (Hillman, 1976; Hillman and Culley, 1978).

Duckweed grow primarily through an asexual budding process where a mother frond gives rise to a daughter frond, forming a dense clonal population representing all stages of development in

their lifecycle (Fig. 1, B-E). Despite being morphologically simple it has been hypothesized that the clonal population represents most complex tissues found in plants (Landolt, 1986). The mother front is an axillary shoot that gives rise to two pouches, or bracts, where the daughter frond is attached through a stipe, or internode, at the meristem and branches off in an alternating pattern similar to the growth of aerial tissue in more complex plants (Fig. 1E). Every daughter frond has 2 developing generations nested within pouches, which start to grow when the daughter is at the 18-cell stage and these new fronds start to differentiate at the 30-cell stage (Fig. 1D) (Rimon and Galun, 1968). Since the average lifespan of a duckweed plant is 30 days, and each frond is capable of generating 15-20 daughter fronds, the growth is exponential but also represents 256 (2^8) fully formed but not expanded daughter fronds per mother frond (Fig. 1B). Therefore, a Duckweed population represents a three-dimensional look at a whole plant: developmental, cellular, and structural.

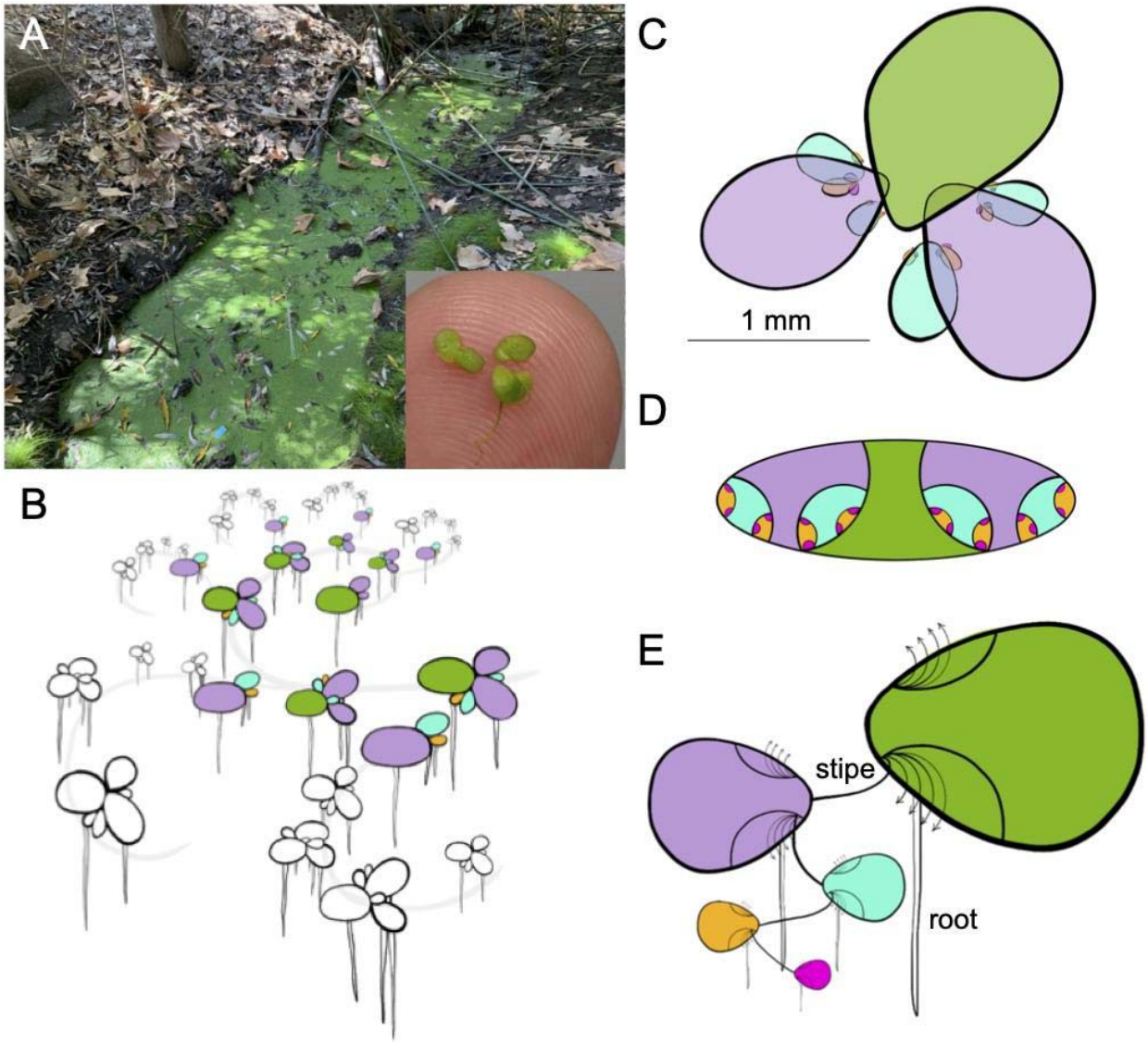

Figure 1. *Lemna minuta* growth highlighting its anatomical analogs to other plants. A) *L. minuta* (lm5633) growing as a dark green mat of fronds in a sewage slough at Cotton Creek Park, Encinitas, CA USA. The inset highlights the small size (~1 mm) of several fronds on a

finger tip. B) Cartoon of the *L. minuta* generational contribution that leads to the dense frond mat: mother (M; green), daughter (D; purple), grand-daughter (GD; blue), great grand-daughter (GGD; orange) and great, great grand-daughter fronds (GGGD; pink). The grey line represents how they are connected through exponential growth. C) A single M frond view with the attached D fronds that have GD, GGD and GGGD fronds nested in the two meristematic pockets. D) A "nested doll" view of one M frond and the maturing generations of D, GD, GGD and GGGD in the pocket. E) One interpretation of the *Lemna* frond is that the M frond is an axillary stem that has two poaches or bracts where the D frond is attached by a stipe or internode. The D, GD, GGD and GGGD progression then is similar to a branching structure of a generic plant, and the root-like structure at each subsequent axillary node is equivalent to an adventitious root. The arrows indicate that multiple internodes will emerge from the axillary stem over time (Landolt, 1986).

Duckweed also provide a compelling platform for genomic studies due to their core non-redundant gene set and relatively small genomes. The dawn of the genomics era yielded high quality genomes and transcriptomes for non-model systems, which resulted in Duckweed emerging once again as an attractive system to tackle problems of cell biology and development. *Spirodela polyrhiza*, The Greater Duckweed, was the first to be sequenced, revealing that in this 150 megabase (Mb) basal monocot genome there is a reduced complement of non-redundant protein coding genes (~19,000) representing a core plant proteome (Wang et al., 2014; Michael et al., 2017). Subsequently, genomes for *Lemna minor* (Van Hoeck et al., 2015), *Spirodela intermedia* (Hoang et al., 2020) and *Wolffia australiana* (Michael et al., 2020) have been published, revealing small genomes at 472, 160 and 375 Mb respectively. These sequenced genomes further support that Duckweed have a core plant gene set with few family expansions making them ideal for dissecting pathways and functional analysis.

Many species of Lemnaceae have dispersed widely beyond their natural ranges and are considered invasive species in their new habitats due to their rapid vegetative propagation (Moodley et al., 2016). A prime example of an alien invasive species is *Lemna minuta*, which is native throughout the temperate zones of the Americas, but has dispersed widely throughout Eurasia (Landolt, 1986). Analysis by Ceschin et al., maps the introduction of *L. minuta* in Europe to the 1950s -1960s with a dispersal rate of 40-50 km/per year (Ceschin et al., 2018). This dispersal process involved crossing of seas to reach places such as Ireland (Lucey, 2003) and Malta (Mifsud, 2010), which is consistent with a role for bird mediated dispersal, and hence the common name Duckweed (Silva et al., 2018). However, despite *L. minuta* doubling in roughly 24 hours, its growth rate does not explain its ability to outcompete native duckweed (*L. minor*) or other aquatic plants (Van Echelpoel et al., 2016; Paolacci et al., 2018).

The small organism size (~1mm), fast growth rate (~24 hrs), small genome size (~360 Mb), and ability to exploit diverse environments led us to select *L. minuta* as a model system to understand cell trajectories and function across an entire plant. *L. minuta* is emerging as a phytoremediation species due to its superior ability to remove various excess elements and toxins from wastewater (Ceschin et al., 2020; Fernández et al., 2020). This led us to

hypothesize that the invasiveness of *L. minuta* may result from a specialized set of genes or cellular function, as it is apparently not solely due to its rapid growth rate. Therefore, we collected *L. minuta* from local wastewater (Fig. 1A) in order to ensure we captured the natural diversity of a successful clone. Duckweed clones sequenced to date have all been from the Landolt Collection, which have been maintained under aseptic lab conditions for 20-50 years (~1000s of generations). We brought a single sterile clone into culture (lm5633) and generated a chromosome-resolved reference genome. Additionally, snRNA-seq was performed on a population of clonally propagating plants to understand gene expression profiles across the whole plant's cellular developmental landscape. The genome and snRNA-seq of lm5633 suggest mechanisms of invasiveness.

Results

Lemna minuta chromosome-resolved genome

Lemna minuta was collected from a wastewater run off in Cotton Creek Park, Encinitas, CA USA (33°2'58"N 117°17'29"W), sterilized and one clone was selected as a representative of the population for bulking (assigned the name lm5633 in the RDSC collection) (Fig. 1A). We estimated the genome size as 310 megabases (Mb) by k-mer (K = 31) frequency analysis using Illumina short reads sequencing (Fig. 2A), which was smaller than the reported average genome size estimated by flow cytometry of 365 Mb (Table 1) (Bog et al., 2020). The genome is highly heterozygous at 2.1% with 57% of the genome in high copy number elements (young transposable elements, centromeres, telomeres, rDNA arrays), while 28% is predicted to be repeat sequence (Fig. 2A; Table 1). The high level of heterozygosity suggests lm5633 may reproduce through outcrossing in the environment. We sequenced the genome using long-read Oxford Nanopore Technologies (ONT), and assembled reads into contigs using both our minimap/miniasm pipeline (Michael and VanBuren, 2020), as well as FlyE, the later of which produced a more contiguous assembly with a total length of 393 Mb, longest contig of 1.2 Mb and an N50 of 205 kb (Table 1).

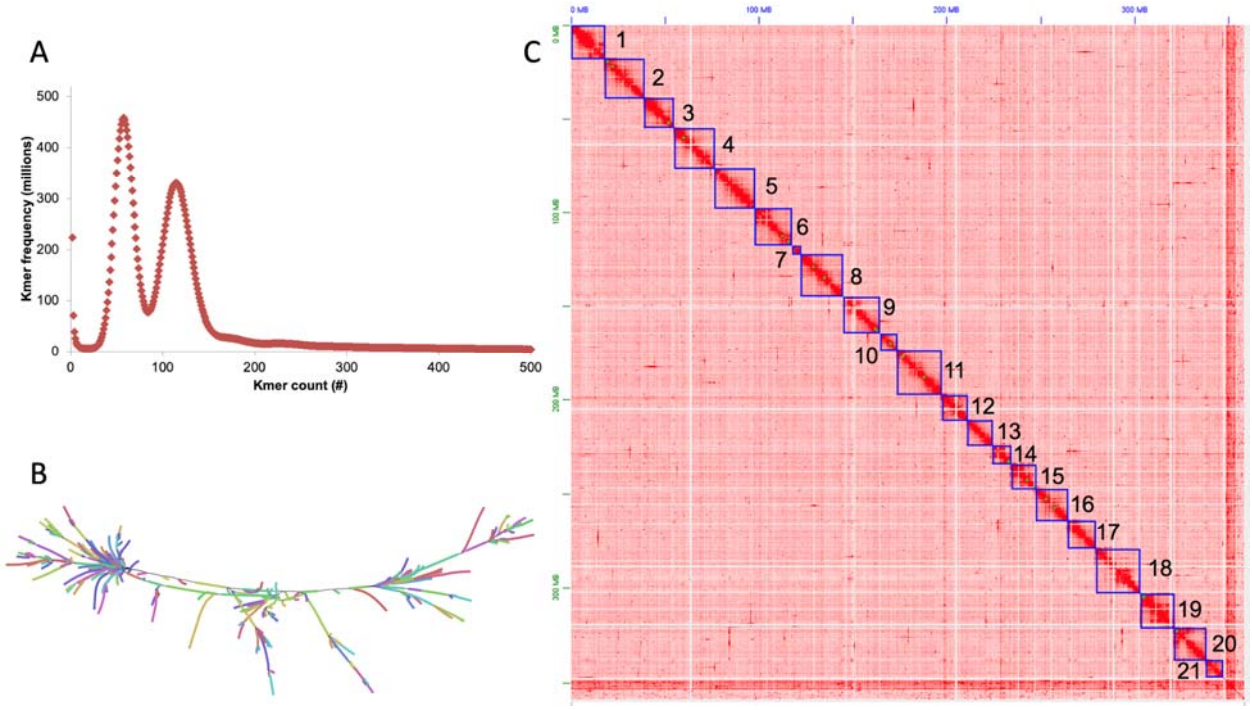

Figure 2. The highly heterozygous *Lemna minuta* (lm5633) genome resolved into 21 chromosomes. A) K-mer (k=31) frequency plot for lm5633 reveals two peaks consistent with a high level of heterozygosity (2%). B) Assembly graph of a 40 Mb region visualized with Bandage shows both the heterozygous branches as well as repeat tangles, “hairballs.” C) High throughput chromatin conformation capture (HiC) contact map resolving the polished lm5633 assembly into 21 chromosomes (darker red more contacts, lighter red to white less contacts).

Since the genome assembly was larger than the predicted genome size we reasoned that some of the genome was retained in two haplotypes. The genome assembly graph confirmed that excess haplotypes were retained appearing as branches in the graph, as well as suggesting a

complex repeat structure (transposable elements; TEs) resulting in repeat tangles, or "hairballs" in the graph (Fig. 2B). We assessed the completeness of the assembly with benchmarking universal single-copy orthologs (BUSCO) finding 72.5% complete (Table 1); the lower percent complete is consistent with other high quality chromosome resolved Duckweed genomes (Hoang et al., 2018; Harkess et al., 2020; Michael et al., 2020). In addition, we found 11.8% duplicated BUSCO genes consistent with the presence of additional haplotypes in the assembly (Table 1; Supplemental Table S1). We purged the excess overlapping haplotypes and scaffolded the contigs with Illumina-based high throughput chromatin conformation capture (HiC) that resolved the genome into 21 chromosomes consistent with published cytology (Fig. 2C) (Landolt, 1986). The final lm5633 HiC assembly was collinear with the high-quality assembly of *S. polyrhiza* clone 9509 (sp9509) chromosome structure (Fig. 3A).

Table 1. *Lemna minuta* (lm5633) genome statistics.

	lm5633
Estimated genome size flow cytometry (bp)*	365,000,000
Estimated genome size K-mer = 31 (bp)	310,069,889
Heterozygosity prediction K-mer31 (%)	2.1
High copy repeat sequence K-mer31 (%)	57
Chromosome-resolved genome size (bp)	360,454,868
Assembled genome size (bp)	392,702,877
Contig (#)	9,473
Longest contig (bp)	1,268,926
Contig N50 length (bp)	205,076
Scaffold (#)	2,407
Longest scaffold (bp)	24,571,886
Scaffold N50 length (bp)	19,492,010
Chromosomes after HiC (#)	21
BUSCO[&] complete (%)	72.5
Repeat predictions (%)	58.2
Telomere length average (bp)	10,085
Telomere length, longest (bp)	25,271
Predicted genes (#)	22,873

*Averaged over all *L. minuta* clones tested by (Bog et al., 2020)

[&]BUSCO library: liliopsida_odb10

Lm5633 gene prediction and annotation

Long read assemblies cover more repeat sequences and usually allow the identification of putative centromere sequences, definition of telomere lengths, and annotation of full-length TEs (Michael and VanBuren, 2020). Consistent with the high copy number repeat k-mer frequency estimate, we identified that 58.2% of the genome was repeat sequence, which is double that of sp9509 and similar to *W. australiana* clone 8730 (wa8730) (Table 1; Supplemental Table S2). The lm5633 genome has a gypsy/copia ratio of ~2, similar to that found in sp9509, but distinct to wa8730 where the ratio is closer to 1 (Supplemental Table S2) (Michael et al., 2017; Michael et al., 2020). The larger genome size of lm5633 compared to sp9509 was primarily due to the increase in retained TE fragments, which is seen in the number of repeats found intervening in the evolutionarily conserved linkage of core circadian clock genes (Fig. 3B) (Michael et al., 2020). Similar to *S. polyrhiza*, we did not detect high-copy number centromere repeats (Michael et al., 2017), but we did identify telomere (AAACCCT) arrays that are longer (average=10 kb; longest 25 kb) than sp9509 (average=3 kb; longest=6 kb), yet shorter than what we have found in wa8730 (average=18 kb; longest=70 kb) (Supplemental Table S3) (Michael et al., 2020).

The sequenced duckweed genomes have the fewest protein coding genes found in angiosperms to date with wa8730 and sp9509 having just 14,324 and 18,507 genes respectively (Michael et al., 2017; Michael et al., 2020). After masking the repeat sequence, we predicted 22,873 protein coding genes in the lm5633 genome (Table 1), which is similar to the 22,382 and 22,245 protein coding genes predicted in the lm5500 and si7747 assemblies, respectively (Van Hoeck et al., 2015; Hoang and Schubert, 2017). The higher gene counts for both lm5633 and si7747 compared to wa8730 and sp9509 are a result of those genomes having orthogroups with greater than 10 genes, which is similar to *Arabidopsis* and rice where gene families are much larger (Fig. 4A). Of the 99 lm5633 orthogroups that had greater than 10 genes, 50% had between 0 and 2 genes in sp9509. The expanded lm5633 families could result from whole-genome duplication (WGD), tandem duplication (TD), proximal duplication (PD), transposed duplication (TRD), or dispersed duplication (DSD) (Qiao et al., 2019). The expanded lm5633 genes were predominantly a result of TD and PD events, as compared to sp9509, leading to expansion of genes involved in pathogen defense, response to stress, and nutrient acquisition (Fig. 3D; Supplemental Table S4). For instance, there is an 11 gene TD of a boric acid channel (Fig. 3C), which may reflect the ability of lm5633 to extract large amounts of boron from its environment, improving survivability, and invasiveness (Fernández et al., 2020).

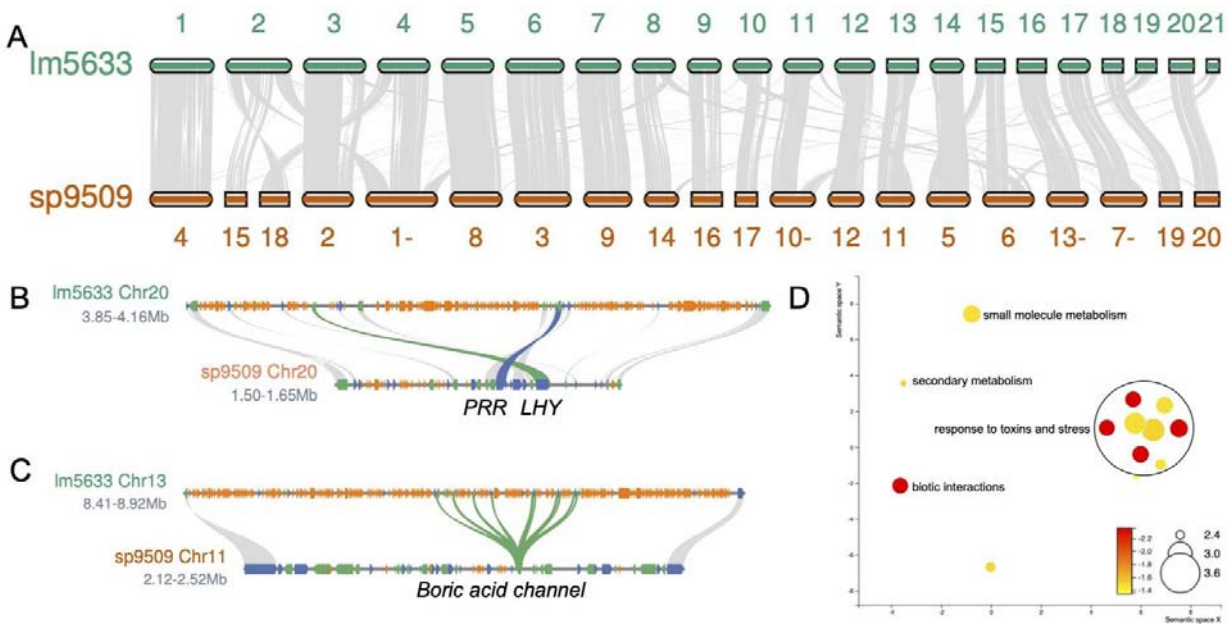

Figure 3. Gene family expansion in lm5633 is driven by tandem duplication (TD). A) Lm5633 chromosomes aligned to sp9509 chromosomes with syntenic blocks (grey lines) anchoring positions between the two genomes. Chromosomes are the correct ratio between one another but are not to scale between the two species. A minus sign after the number means the chromosome has been flipped for visualization purposes. B) The single MYB transcription factor *LATE ELONGATED HYPOCOTYL* (*LHY*; blue line) and *PSEUDO-RESPONSE REGULATOR 7* (*PRR7*; green line) are in tight linkage and TE fragments (orange) resulting in the region expansion in lm5633. Grey lines connect other syntenic genes (blue, forward; green reverse). C) Boric acid channel TD (green line) also showing the expansion of the lm5633 genome due to TE fragments (orange). D) Sympantic principle component analysis (PCA) of significant GO terms associated with TDs in lm5633. Size of the circle is the log frequency and the color (red high, and yellow low) is the log FDR.

A comparison of orthologous proteins between lm5633 and wa8730 revealed a number of expanding orthogroups relating to the ability of lm5633 to thrive in diverse environments. We found 14 significantly expanded orthogroups in lm5633 compared to the most recent common ancestor wa8730 (Fig. 4A). There are five orthogroups annotated with gene ontology (GO) terms: OG0000813, vacuolar membrane; OG0000014, signaling; OG0001260, defense response to fungus; OG0000015, detection of bacterium; OG0003377, delta14-sterol reductase activity. There are 3 significantly contracted orthogroups compared to wa8730 with one having an annotation for GO:0005787,signal peptidase complex. An additional expanded orthogroup, OG0000243, encodes a family of proteins relating to multidrug and toxic compound extrusion (MATE) proteins that have more than doubled from 2 copies in wa8730 to 5 copies in lm5633. The MATE transporters are often associated with increased plant resilience to toxic compounds and adaptability to metals including iron homeostasis (Upadhyay et al., 2019). These evolving orthogroups are consistent with the ability of lm5633 to thrive under adverse conditions of abiotic and biotic stress.

Lm5633 *chromosome evolution*

The lm5633 genome was resolved into 21 chromosomes (Fig. 2C and 3A), which is consistent with cytological studies (2n=42) (Landolt, 1986); this is one more chromosome than sp9509 and three more than si7747 (Michael et al., 2017; Hoang et al., 2020). Since *S. polyrhiza* is thought to be the basal duckweed, we looked specifically at the synteny between the lm5633 and sp9509 chromosomes and found 18 chromosomes share high sequence collinearity with at least 5 that are completely collinear (Fig. 3A). There are several lm5633 chromosomes (2,19, and 21) that are the result of fragments from several sp9509 chromosomes, whereas no sp9509 chromosomes resulted from lm5633 fusions, consistent with the basal nature of the *S. polyrhiza* genome.

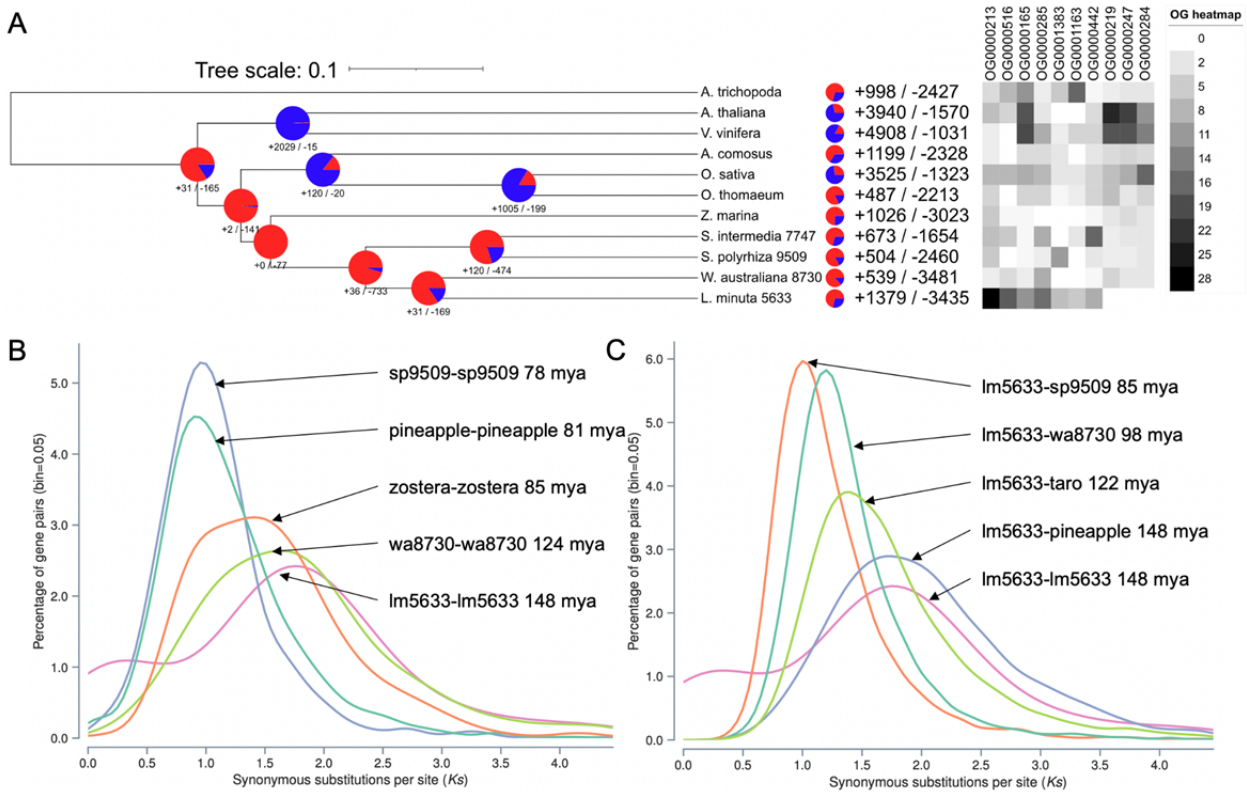

Figure 4. Whole genome evolution shows consistent gene family contractions in lm5633 ancestry. A) Gene family contractions (red) and expansions (blue) along the phylogenetic tree leading to minimized duckweed genomes are shown at each node and total protein family contractions and expansions for each species (right). Lm5633 shows the greatest similarity and gene family conservation with wa8730. A heatmap of orthogroups significantly expanded or contracted in lm5633. B) Self versus self and C) lm5633 versus other synonymous substitution (Ks) plots to elucidate the WGD history of lm5633. Dating is based on the mean Ks peak for each comparison using all paralogs/orthologs. Plant species used: *A. trichopoda* (Amborella), *A. thaliana*, *V. vinifera* (grape), *A. comosus* (pineapple), *O. sativa* (rice), *O. thomaeum* (oropetium), *Z. marina* (zostera), *C. esculenta* (taro), *S. intermedia* (si7747), *S. polyrhiza* (sp9509), *W. australiana* (wa8730), and *L. minuta* (lm5633).

Spirodela polyrhiza has experienced at least two lineage specific WGDs (B"/a") in the non-grass monocots (Wang et al., 2014; Ming et al., 2015). Consistent with this WGD history, sp9509 has

a 3:1 syntenic depth compared to the *Amborella trichopoda*, which is the progenitor basal angiosperm without a WGD (Amborella Genome Project, 2013). In contrast, lm5633 has a 2:1 syntenic depth compared to *A. trichopoda*, suggesting lm5633 may have a higher level of fractionation (WGD followed by diploidization and gene loss) compared to sp9509. Consistent with lm5633 having a higher level of fractionation, only 4.1% of its paralogs remain in syntenic blocks compared to 49% in sp9509. Moreover, lm5633 and sp9509 have a 1:1 syntenic depth, sharing 88% and 86% gene pairs respectively, indicating that despite the higher level of fractionation in lm5633 the gene content from their shared WGD history is preserved (Fig. 3).

An alternative hypothesis explaining the lower level of paralogs in syntenic blocks in lm5633 is that it did not experience the *S. polyrhiza* B"/a" WGDs, and instead only experienced the *tau* (□) WGD that is shared across most of the monocot lineage (Jiao et al., 2014). Therefore, we looked at the synonymous substitution (Ks) rates of lm5633 paralogs to estimate their age, and found that both wa8730 and lm5633 lack a Ks peak corresponding to B"/a" WGD with only a peak corresponding the to the *tau* (□) WGD (Fig. 4B). Consistent with this, the divergence of lm5633 and pineapple coincides with the *tau* (□) WGD (Fig. 4C). These results suggest that the *Wolffia* and *Lemna* genera may not have the lineage specific WGDs that are found in *Spirodela*. Additional genomes from *Lemna*, *Wolffia* and *Wolffiella* will be required to fully understand the WGD history of duckweed. However, the Ks plots are consistent with the genome innovation of lm5633 resulting from recent TDs in genes associated with abiotic and biotic stress, which provide clues as to its invasive abilities.

Defining cell types in lm5633 with snRNA-seq

The unique life cycle and growth pattern of *L. minuta* plants from mother to daughter and grand-daughter fronds covers all cell developmental stages and types providing an opportunity to follow the developmental trajectory of all cells (Fig. 1). We wanted to identify as many cell types as possible across the diverse developmental states with high gene coverage per cell. Therefore, we carried out single nuclei RNA sequencing (snRNA-seq), which enabled capturing an exact moment in development through the immediate freezing of tissue and avoiding long protoplasting steps that can result in unsampled cell types and sample preparation artifacts (Denyer et al., 2019). We isolated individual nuclei from a population of whole lm5633 plants grown under standard laboratory conditions by FACS and then prepared libraries with SMARTseq2 to achieve high gene coverage per nuclei. After quality filtering, 49% of the annotated lm5633 genes were expressed across 269 nuclei.

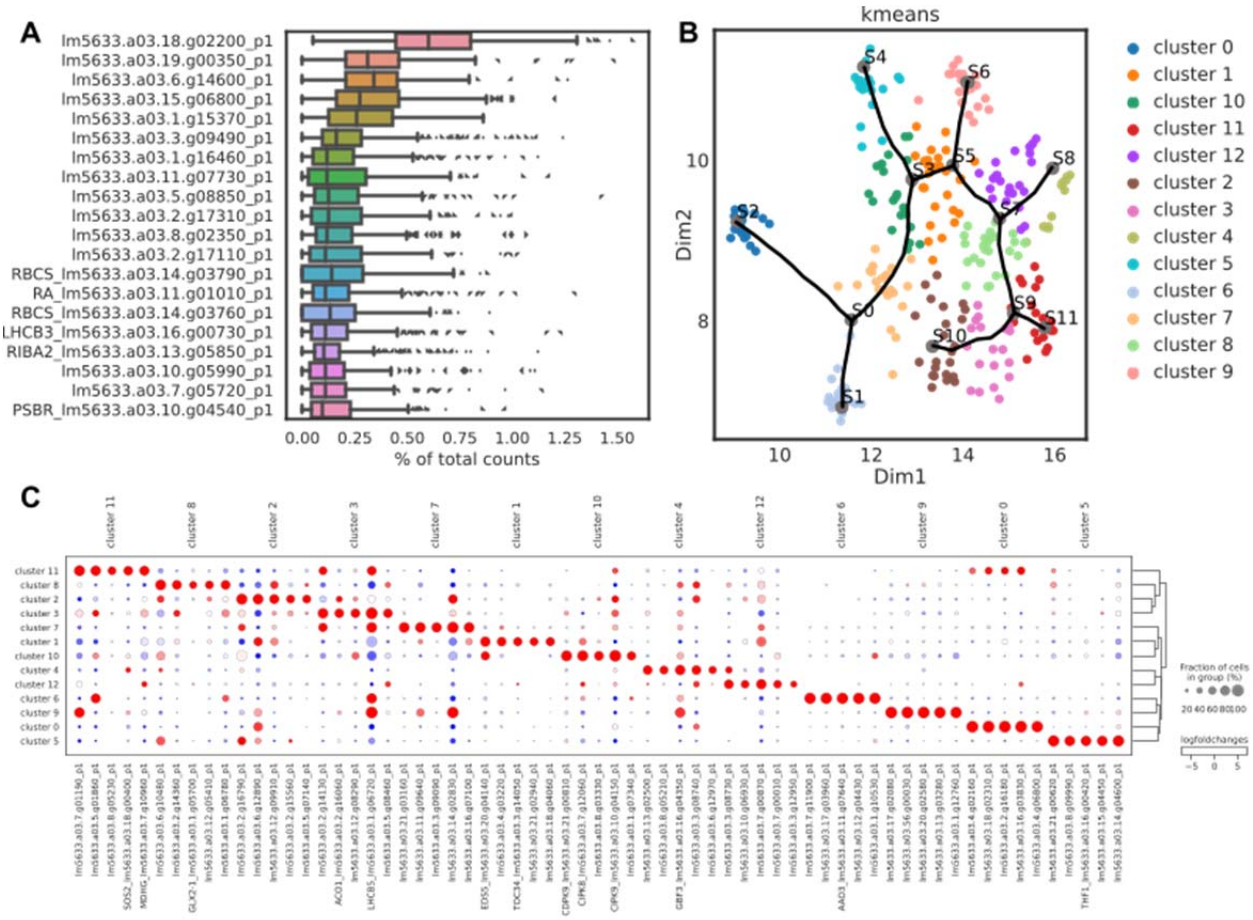

Figure 5. Single nuclei RNA sequencing (snRNA-seq) of clonally propagating whole lm5633 plants. A) The most abundant transcripts reveal that highly expressed genes are generally related to photosynthesis. B) UMAP embedding of 13 k-means clusters in the snRNA-seq dataset. Trajectory analysis (black line) reveals a complex network where multiple branches result in terminally fated cell types. C) Log fold change in marker genes for the 13 specific clusters shows robust separation of cell types for functional analysis. Expression level (red, high; blue, low; larger circle, more cells; smaller circles, fewer cells).

One problem encountered with non-model plants is the lack of annotation. For instance, only six of the twenty most highly expressed genes for all cells in this dataset had annotations, five related to photosynthesis as expected for green tissues (Fig. 5A). Therefore, we leveraged a combination of cell trajectory analysis, gene ontology (GO) annotation, and orthology prediction to define single cell marker genes from model plants for describing cell types in lm5633. Expression based dimensionality reduction resulted in 13 distinct cell clusters with a total of 1,733 significantly differentially expressed genes between clusters (Fig. 5B). To initially define the 13 clusters we identified lm5633 orthologs relative to published single cell studies in *Arabidopsis, O. sativa* (rice) and *Zea mays* (corn) (Denyer et al., 2019; Satterlee et al., 2020; Xu et al., 2021) (Supplemental Table S5). Of the 1,733 marker genes found for the 13 clusters, we found 140 marker genes that did not appear in any orthogroup and on average each marker gene was in an orthogroup with 2.3 copies per species (Supplemental Table S6). An average of

134 marker genes per cluster were determined with a total of 446 marker genes annotated based on model organism orthology, averaging to 18 annotated marker genes per cluster (Fig. 5C; Supplemental Table S7).

GO analysis of the marker genes provided additional functional support for the distinction of each cluster as well as potential functional overlap of cell types (Fig. 6A). Each cluster/cell type had unique GO terms while several had overlapping terms such as meristem-replication, meristem-axillary meristem, root epidermis-axillary meristem, root-axillary meristem, root transition I-axillary meristem, and meristem-root. These results suggest that the meristem-like tissue and different root cell types are intimately associated. Together with orthology from model plants (described below), the cell trajectory showed that meristem gives rise to root transition, meristem-like, and mesophyll cells; mesophyll then gives rise to parenchyma followed by epidermis, while root transitions gives rise to root epidermis cells followed by root cells (Fig. 7).

Meristematic-cell cluster

Following pseudo-time analysis, the gene expression profiles resulted in a large continuous cluster with multiple branches to terminally differentiated cell types (Fig. 5B and 7A). In the UMAP (Uniform Manifold Approximation and Projection) embedding we found several clusters relating to growth and division. Specifically, we defined the meristem (cluster 1) by an *TRANSLOCASE OF CHLOROPLAST 34 (TOC34)*, *ACTIN RELATED PROTEIN (ARP4*) and the *ARGONAUT-like (AGO)* marker shown to be expressed in the shoot apical meristem in maize (Xu et al., 2021). A *TOC34* promoter GUS fusion in *A. thaliana* has been reported in the meristem of green tissues and root cells, and *ARP4* has been shown to function in flower development across species (Gutensohn et al., 2000; Pandey and Chaudhary, 2016). Furthermore, we surveyed the most representative GO terms for the meristem cluster and found that the top eight most representative terms relate to response to oxidative stress or removal of superoxide radicals, which could be associated with DNA damage protection (Supplemental Table S8). However, the representative set of GO terms relate to tissue development including GO:1905330, "regulation of morphogenesis of an epithelium;" GO:0090175, "regulation of establishment of planar polarity;" GO:2000023, "regulation of lateral root development;" GO:0048831, "regulation of shoot system development;" GO:0022603, "regulation of anatomical structure morphogenesis." Finally, 12% of all GO descriptions related to development compared to terminally differentiated cell types such as “leaf epidermis” that contained only 3.6% of terms relating to development.

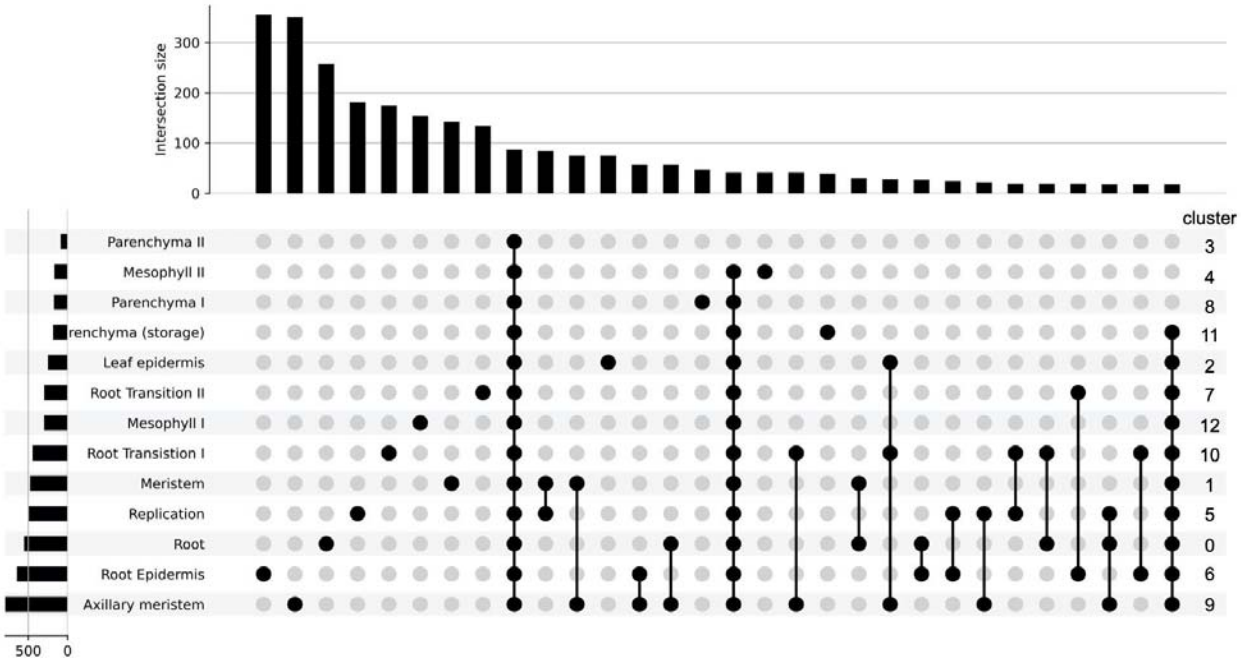

Figure 6. Cell type definitions supported by gene ontology (GO) categories. Upset plot showing the number of GO terms that are specific and overlapping across the thirteen different predicted cell types. The bars on top represent the number of intersecting GO terms per cell type, and the bars on the left represent the number of GO terms found in each cell type. The cell clusters as defined in Figure 5B are on the right hand side to highlight their correspondence with the predicted cell types.

The placement of the meristem cluster in the UMAP embedding was supported by two adjacent clusters that we described as the “axillary meristem” and “replication” (or potentially endoreplication). The axillary meristem identification is supported by the presence of *LAX PANICLE1(LAX1)*, which is mediated at the transcript level in the axillary meristem (Oikawa and Kyozuka, 2009). There are also several genes involved in ribosome biosynthesis and genes coding for the ribosomal complex *(PRPL10, RLP24, RPA1A, RPL12-C, RPS16)*, which would suggest these cells are actively translating mRNA into proteins. The DNA replication/endoreduplication cell cluster contains *ETHYLENE INSENSITIVE (EIN2)* that leads to cell expansion through ethylene signaling and an *A-TYPE CYCLIN (CYCA3-1)*, which is critical for G1-to-S phase transition having a central role in the meristematic tissue (Takahashi et al., 2010; Feng et al., 2015). It is possible that these three specific clusters are all located in the meristematic-like tissue given their close proximity in the UMAP embedding and similarity in nuclei expression profiles. Both of these clusters share common GO terms, GO:0051093, "negative regulation of developmental process;" GO:0022603, "regulation of anatomical structure morphogenesis," which suggest they are cells with defined processes similar to meristematic cells but different enough to warrant exclusive clustering.

The meristem cells give rise to all other cell types in clonally propagated plants, yet many of the transcription factors and networks leading to fated cell types are unknown. Transcription factor *WRKY32* has increased expression in the transition from meristem to green frond-like cell types,

where it has been previously shown to be involved in ethylene signaling in tomato, where RNAi repression leads to yellowing of fruits (Zhao et al., 2021). It has been shown previously that related *L. minor* clones show reduced growth from exogenous ethylene treatment (Utami et al., 2018). Conversely, WRKY transcription factors associated with root cell transition are *WRKY6* and *WRKY65*. *WRKY6* is associated with pathogen defense, phosphate translocation, and arsenate resistance and *WRKY65* induces Jasmonate and salicylic acid responses relating to pathogen response (Huang et al., 2016; Wang et al., 2020). The expression of several WRKY transcription factors in the meristematic transition leading to terminal cell types suggests hormone signaling is a major contributing factor in lm5633 cellular development.

Root-like tissues

The predicted meristem-like cells are centrally located between two branches leading to terminal cell types in the trajectory analysis (Fig. 5B and 7A). Although it has not been confirmed that duckweed roots are essential or act similarly to well developed root systems in model plants, we still define root cells by expression of two copies of the tandemly duplicated *PLANT LIPOXYGENASE 9* (*LOX9*) genes, which has been shown in soybean to be expressed in root nodules (Hayashi et al., 2008). Additional root makers are involved in metabolite transport by *HIGH-AFFINITY POTASSIUM TRANSPORTER 8* (*HAK8*) and *MANGANESE ATPASE TRANSPORTER (ECA2) (Mills et al., 2008)*. We find an additional marker gene *UBIQUITIN E2 CONJUGATING ENZYME* (*UBC13*) that has been shown to be involved in root development (Wen et al., 2014). We also find an additional paralog of *LOX9* in the root transition I cluster suggesting the paralogs are cell specific but concordant in their function progressing toward formation of root-like cells.

Visual morphologies have historically been used to define long-lived, stable cell types. Single cell/nuclei sequencing aims to capture all cell types including transitory cells. In some cases, snRNA-seq may capture transitional cell types that do not have well characterized visual properties or marker gene expressions but show progression to terminal cell types based on functional characterization of the snRNA-seq expression profile. Although difficult to define a specific cell type, here, the root transition II cluster contains a marker gene for *ASYMMETRIC LEAVES1* (*AS1*), which is essential for adaxial-abaxial leaf polarity and associated with an initial committed step towards root-like tissue development (Xu et al., 2003; Machida et al., 2015). Likewise, *SLEEPY1* (*SLY1*) is responsible for gibberellin signaling, cell growth and elongation, suggesting this cluster contains cells transitioning to the final root-like cells (McGinnis et al., 2003; Xu et al., 2003). *ELONGATED HYPOCOTYL 5 (HY5)* is also found in this cluster and is known to play a role in induction of light induced genes and root gravitropism (Srivastava et al., 2015). It is interesting to note that lm5633 roots were exposed to light in this experiment since the plants are grown in clear flasks, which results in the roots appearing visually green and expressing light harvesting genes (Fig. 7C).

Epidermal tissues

The exterior cells forming the epidermis provide a barrier from the rhizosphere and phyllosphere, and are often associated with increased production of hydrophobic waxes and cutins. Lipid biosynthesis is essential for wax and cutin production that provide extracellular

protection in epidermal cells. Fatty acid biosynthesis,known to occur in epidermal cells,is mediated through the key marker genes *3-KETOACYL-COENZYME (KCS3)* and *LONG CHAIN ACYL-CoA SYNTHETASE (LACS2)(Kim et al., 2013)* (Supplemental Table S6). Additionally, root epidermal cells were defined based on expression of *CRINKLY 4* (*CR4*), which is known to have a role in maize epidermal cell formation (Becraft et al., 1996). Interestingly, frond epidermal cells are located in close proximity to root epidermal cells in UMAP embedding consistent with these cell types sharing similar expression despite having different cell trajectories (Fig. 5B and 7). The frond epidermal cells were defined by the marker gene *ECERIFERUM-like*, which has been shown to be important for cuticle wax development (Haslam et al., 2017). Additional markers include an auxin efflux carrier in starch metabolising cells *PIN-FORMED (PIN7)* aiding root termination (Kim et al., 2013; Rosquete and Kleine-Vehn, 2018). This is consistent with this cluster's position in the UMAP embedding (Fig. 5B) in close proximity to the root epidermis but highly diverged in the complement of expressed photosynthetic genes and complete divergence in the trajectory analysis (Fig. 7A). Likewise, this cell type contains genes *ATP-DEPENDENT ZINC METALLOPROTEASE* (*FTSH2)*, two orthologues of *DEFECTIVE KERNEL 1 (DEK1)* and *CYTOCHROME C OXIDASE 15* (*COX15)*. These genes are involved in carbon storage where *FTSH2* has been shown to be involved in thylakoid biosynthesis and *COX15* shows increased expression in response to decreased cellular respiration (Vishwakarma et al., 2015; Haslam et al., 2017). Similarly, DEK1 is expressed in aleurone-like cells in maize that are involved in starch metabolism (Tian et al., 2007). It has been shown that *DEK1* also has adverse effects on leaf morphology and it is likely that this cluster contains frond epidermal cells.

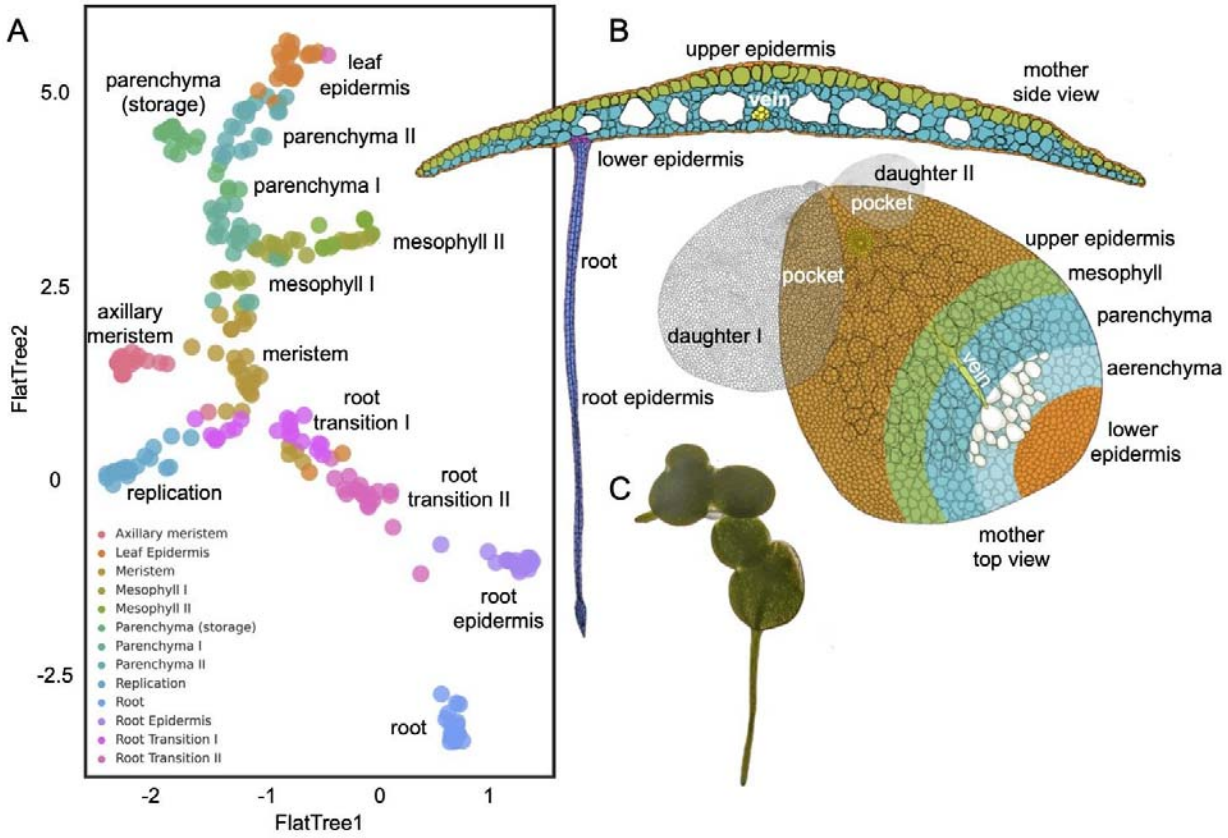

Figure 7. Trajectory analysis defines cell types of the entire *L. minuta* plant . A) UMAP tree embedding of snRNA-seq clusters describing the cell types. B) Cartoon of *L. minuta* based on (Banaszek and Musiał, 2011) coloring specific cell types per the UMAP. C) Live image of lm5633 showing the green roots.

Frond tissues

Mesophyll, located in leaf or frond tissue, are the primary cells involved in photosynthetic light and CO_2 capture for the generation of sugars to sustain the plant. Mesophyll-like cells were determined based on significant expression of multiple marker genes associated with light perception, thylakoid biogenesis, and cytokinin degradation related to stress by *YELLOW STRIPE-LIKE (YSL9)* and *THYLAKOID LUMEN PPIase* (*TLP40*). *YSL9* in rice is associated with leaf vasculature but in this case is potentially more similar to mesophyll like cells in this basal aquatic monocot (Wen et al., 2020). *TLP40* has been shown to be expressed in mesophyll and at 2-fold greater expression in bundle sheath cells (Vojta et al., 2019). Interestingly, the representative GO terms suggest these cells are carbon limited and the photorespiratory cycle is engaged where the top GO terms are mitochondrial transport of glycolate (GO:0006626, "protein targeting to mitochondrion;" GO:0072655, "establishment of protein localization to mitochondrion;" GO:1901975 , "glycerate transmembrane transport;" GO:0097339, "glycolate transmembrane transport").

Parenchyma cells in any given tissue can become specialized for a variety of purposes, yet in lm5633 it appears the specialization of parenchyma is geared toward photosynthetic metabolism and storage of photosynthate. Parenchyma I and II clusters are defined *EXORDIUM-like (EXO)*, *SUCROSE SYNTHETASE 1 (SUS1)*, *STARCH EXCESS 4 (SEX4)*, *XAP5 CIRCADIAN TIMEKEEPER (XCT)* which promotes ethylene response in aerial tissue (Ellison et al., 2011). *SUS1* is a prominent player in the accumulation of terminally synthesized photosynthate in source tissues and *SEX4* mutants in leaf tissues promote starch metabolism (Kötting et al., 2009; Ma et al., 2014; Shi et al., 2019). Mutants of *EXO* have decreased epidermis, palisade and spongy parenchyma in *A. thaliana* and suggests *EXO* may be inhibiting differentiation to a terminal cell type (Schröder et al., 2011; Florian Schröder, Janina Lisso, Peggy Lange & Carsten Müssig). Marker genes in these clusters point to photosynthetic metabolism. A terminal cluster is observed branching off the parenchyma-like cells (Fig. 7A) denoted here as “parenchyma (storage)” since cells are expressing two copies of *ALPHA AMYLASE 2 (AMY2)* and *IMBIBITION INDUCIBLE (IMB1).* The combination of these marker genes in a cell type may mean these cells are actively metabolizing starch in the vacuoles.

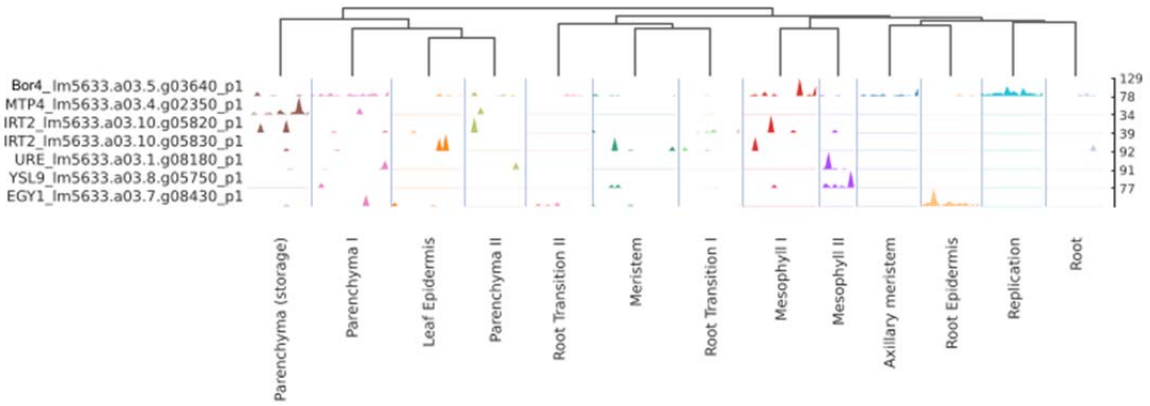

Figure 8. Metal transport and accumulation associated genes in *L. minuta* are expressed in mesophyll cells. SnRNA-seq expression profiling of metal tolerance related genes reveal a cell type specific expression pattern with greater gene expression localized to the mesophyll cell types in lm5633. Lm5633.a03.5.g03640 (top) encodes *BORON TRANSPORTER 4* (*BOR4)* localized to mesophyll and replicating cells described further in the text.

Specific cell expression associated with invasiveness

Lemna minuta can occupy a variety of ecosystems due to their adaptability in absorption and tolerance of excess micro- and macro-nutrients. The uptake and adaptability of *L. minuta* has led to its use as a wastewater detoxifying/phytoremediation species (Fernández et al., 2020). Wastewaters are often contaminated with a variety of toxic compounds yet the exact mechanisms of plants ability to adapt to these compounds, including high levels of essential elements and heavy metals, remains unclear (Frick, 1985; Davis et al., 2002; Del-Campo Marín and Oron, 2007; Gür et al., 2016; Liu et al., 2018; Türker et al., 2019). We find increased expression of *BORON TRANSPORTER 4* (*BOR4*) in the mesophyll and replicating cells (Fig. 8). We also found a large TD (11 copies; eighth largest) of a boric acid transporter (MIP aquaporin) (Fig. 3C; Supplemental Table S4). We find similar increased expression of *IRON-REGULATED TRANSPORTER 2* (*IRT2*) and *METAL-NICOTIANAMINE TRANSPORTER 9 (YSL9*) in the mesophyll cells. *IRT2* also shows a TD expansion (4 copies with 2 copies expressed here), although modest compared to *BOR4*. Overaccumulation of boron and toxic heavy metals can have an adverse effect on plant growth yet *L. minuta* grows and accumulates heavy metals readily (Gür et al., 2016). This is compelling evidence that *L. minuta*'s invasiveness and adaptability in wastewater may be, at least in part, due to the increased *BOR4* expression, as *BOR4* is essential for boron export from cells. Further snRNA-seq under micro- and macro-nutrient stress conditions would greatly improve our understanding of cell specific responses and adaptation to abiotic stressors.

Discussion

Here we endeavoured to define the trajectory of cells in a morphologically reduced, fast growing non-model plant useful for phytoremediation but also invasive in specific environments. We generated a chromosome-resolved genome for the Lesser Duckweed *L. minuta* clone Im5633 and found that it did not share a lineage specific WGD with *S. polyrhiza* yet had an increased number of Tandem Duplications (TDs) for genes involved in pathogen defence, response to stress and nutrient acquisition. Additionally, we performed snRNA-seq on a clonal population of whole *L. minor* plants that represent all developmental stages, which enabled us to discern differentiated cell types and their development from the meristematic cells. Finally, the data provided evidence that the mesophyll is a site of elemental and heavy metal transporter expression.

Duckweed are particularly suited for single cell studies due to their size, direct contact with the media, small non-redundant genome, and rapid clonal growth (Fig. 1). Most single cell studies to date have been conducted using protoplasting on root or reproductive structures, which have the advantage of having all cells in the developmental continuum (Denyer et al., 2019; Ryu et al., 2019; Liu et al., 2020a). Similarly, Duckweed provide a continuum of all cell types and developmental stages due to their rapid clonal growth and nested generations. Moreover, Duckweed are in direct contact with their media allowing for detailed and controlled manipulation of the environment, which in the the past was leveraged to understand auxin biology (Slovin and Cohen, 1988). The snRNA-seq method described here also couples well with experiments where abiotic/biotic treatments of the samples are desired because they can be immediately frozen and transcriptomic perturbations measured additionally frozen nuclei presumably capture actively transcribing mRNA. One potential downside of snRNA-seq is the loss of RNA in the cytoplasm; more experiments comparing the different techniques will be required to evaluate this limitation.

Additionally, the Duckweed genome provides a compelling platform for gene annotation and functional analysis. It was shown that *S. polyrhiza* has a reduced set of non-redundant protein coding genes even though it has had two lineage specific WGD events (Wang et al., 2014). One challenge in model species and crops is the number of paralogs (family size) that complicate gene functional analysis due to redundancy of action. Here, the Im5633 genome also has a streamlined set of genes that shows a similar reduction in paralogs compared to other plants (Fig. 4A). However, Im5633 does have a much higher level of fractionation compared to sp9509 where Im5633 has 4.1% retention of syntenic paralogs compared to 49% retained in sp9509, which means that either Im5633 is evolving at a greater rate or it has a different WGD history. We find evidence that both Im5633 ands wa8730 lack the *S. polyrhiza* lineage specific WGD and only have the shared monocot *tau* WGD (Fig. 4, B and C). Currently there are conflicting reports whether *S. polyrhiza* has the *tau* WGD (Wang et al., 2014; Ming et al., 2015; Hasing et al., 2020). One explanation is that the lineage-specific WGD has obscured the *tau* WGD in *S. polyrhiza*, and that Im5633 is more fractionated due to the very ancient (~150 mya) WGD. More Duckweed genomes will help resolve the WGD history.

All single cell studies are currently in model systems, yet as we have shown here one can go from a wild plant to a chromosome-resolved genome and leverage snRNA-seq dataset to generate functional information. While model plants and specifically *A. thaliana* have high quality annotations and marker genes for cells, there are significant challenges in defining marker genes or assigning cellular functions for even well studied plants and crops. Here we leveraged a three tiered approach to assign cell type. Our realization was that cell trajectory analysis provides rich information about cell type, especially when it is coupled with a developmental series shown here using an entire Duckweed plant similarly in roots and floral tissues (Denyer et al., 2019; Satterlee et al., 2020; Xu et al., 2021). GO and orthology comparisons are more fraught with issues due to sparse datasets (i.e. few or no GO terms associated with gene predictions) where orthology may not exist or be cryptic with model species or crops. However, when these three criteria are leveraged, annotation can be improved as well as provide support for known and new cell types, like transition cell types.

The trajectory of cell types from meristem to terminally differentiated cells provides clues as to their function. Cellular studies have suggested that the meristem is unique in Duckweed since it gives rise to daughter fronds (Landolt, 1986). We found the meristem is divided into three potential cell clusters (meristem, replication, and axillary meristem), which provides a clue as to how duckweed is dividing rapidly leading to continuous daughter frond production. The overlapping trajectory and GO terms of mersitem-like and root-like cells (Fig. 6 and 7), suggests the root may play additional roles that are unknown at this time, and may clarify its usefulness. Furthermore, root and frond epidermal cells have distinct trajectories (Fig. 7) yet arrive in a similar expression space consistent with their overall cellular function (Fig. 5B). Finally, the emergence of the photosynthetic mesophyll-like cells from the meristem then giving rise to the metabolically active and terminally differentiated parenchyma cells. We did not identify stomatal cells or vascular tissue (vein) cells and this could be a result of our methodology or the fact that *L. minuta* only has one vein and about 30 stomata (~60 cells per plant) (Landolt, 1986). Since each mother frond has 256 (each daughter frond has 2 or more generations) developing, yet unexpanded daughter/grand-daughter fronds the majority of the cells we detect will be actively dividing; deeper sequencing will be required to accurately identify small populations of terminally differentiated cells.

In model plants, nutrient uptake and transport primarily occurs in the roots and vasculature (Yan et al., 2020). However, the Duckweed frond is in direct contact with its environment, which may explain why we detect high expression of nutrient uptake genes in the cells we define as the photosynthetically active mesophyll (Fig. 8). Both *ITR2* and *BOR4*, which are also both expanded in lm5633 through TDs, are highly expressed in mesophyll. However, it has been shown that *IRT2* is expressed in the roots (Vert et al., 2001), and *BOR4* is expressed in the xylem in *A. thaliana* (Takano et al., 2002). The solitary vein of lm5633 is in close proximity to the mesophyll, which makes it formally possible that we have mis-identified some of these cells and they actually are vasculature or at an expression level these cells are too similar to separate with the current sequencing coverage. Either way, this simplified system provides an opportunity to further define the acquisition, transport and storage of toxic compounds at a single cell level.

Future studies will enable clarification of cell types as well as provide the opportunity to dissect the specific cellular mechanisms associated with phytoremediation and invasiveness.

Conclusions

Duckweed provides an unprecedented opportunity to study the cell fates across an entire morphologically reduced plant. Coupled to snRNA-seq, this system provides a new opportunity to study the cell-specific responses to environmental changes. This dataset represents a first look and will surely be refined with additional datasets in additional plants. Different plants, conditions and treatments will refine our understanding of cell type and ultimately identify new cell types and niches.

Materials and Methods

Plant collection and growth

Lemna minuta was collected on October 10, 2019 from Cotton Creek Park, Encinitas, CA, USA (33°2'58"N 117°17'29"W), which is a waste water slough close to the ocean (Fig. 1A). A population of clones were collected, surface sterilized and one clone was retained to multiply the population. The representative clone was deposited in the Rutgers Duckweed Stock Collective (RDSC; www.ruduckweed.org) and was assigned the clone number lm5633. Plants were propagated on Schenk and Hildebrandt (SH) media as described at RDSC.

Genome sequencing

High Molecular weight (HMW) DNA was extracted with modifications (Lutz et al., 2011). The resulting DNA was quality controlled for size and contamination. Unsheared HMW DNA was used to make ONT ligation-based libraries. Libraries were prepared starting with 1.5 ug of DNA and following all other steps in ONT's SQK-LSK109 protocol. Final libraries were loaded on a ONT flow cell (v9.4.1) and run on the GridION. Bases were called in real-time on the GridION using the flip-flop version of Guppy (v3.1). The resulting fastq files were concatenated and used for downstream genome assembly steps. Illumina 2x150 paired end reads were also generated for genome size estimates and polishing genome sequences. Libraries were prepared from HMW DNA using NEBNext Ultra II (NEB, Beverly, MA) and sequenced on the Illumina NovaSeq. Resulting raw sequence was only trimmed for adaptors, resulting in >60x coverage.

Genome size estimation

K-mer (k=31) frequency was estimated with Illumina paired end reads (2x150 bp) using Jellyfish (v2.3.0) (Marçais and Kingsford, 2011) and analyzed using in house scripts and GenomeScope and GenomeScope2 (Vurture et al., 2017; Ranallo-Benavidez et al., 2020). Genome size, heterozygosity and repeat content were first estimated with GenomeScope (http://qb.cshl.edu/genomescope) (Table 1). The K-mer frequency plot is consistent with a highly heterozygous diploid genome or a tetraploid genome; based on previous flow cytometry, *L. minuta* is diploid with an average genome size of 365 Mb (Table 1) (Bog et al., 2020).

Genome assembly

Resulting fastq files passing QC were assembled using our previously described pipeline (Michael et al., 2018) with the modification that the initial assembly was generated using FlyE

(Kolmogorov et al., 2019). The resulting assembly graph (gfa) was visually inspected with Bandage (v0.8.1) (Wick et al., 2015), which revealed a branching pattern consistent with a heterozygous genome with structural differences between haplotypes. Consensus was generated with three (3) iterative cycles of mapping the ONT reads back to the assembly with minimap2 followed by Racon (v1.3.1) (Vaser et al., 2017), and the final assembly was polished iteratively three times (3) using 2x150 bp paired-end Illumina reads mapped using minimap2 (v2.17-r941) (>98% mapping) followed by pilon (v1.22) (Walker et al., 2014). The resulting assembly was assessed for traditional genome statistics including assessing genome completeness with Benchmarking Universal Single-Copy Orthologs (BUSCO) (v3) liliopsida odb10 database (Table 1) (Simão et al., 2015).

High throughput Chromatin Conformation Capture (HiC) genome scaffolding

Crosslinking was performed on ground tissue and nuclei were isolated following (Colt). HiC data was generated using the Arima-HiC Kit User Guide for Plant Tissues (Link et al., 2018), according to the manufacturer's protocols (Catalog #A510008 Document Part #A160135 V00). Libraries were generated following the Arima-HiC Kit Library Preparation Guide for Swift Biosciences Accel-NGS 2S Plus DNA Library Kit (Catalog #A510008 Document Part #A160140 V00) and sequenced on Illumina NovaSeq. We used standard methods defined in the Aiden lab genome assembly cookbook (https://github.com/aidenlab/3d-dna/). The primary steps include alignment of HiC reads to the polished lm5633 assembly and creation of a contact map using the Juicer pipeline (https://github.com/aidenlab/juicer) followed by automated scaffolding using 3d-dna. The scaffolds were then inspected and manually corrected with Juicebox Assembly Tools (JBAT) before being finalized by the 3d-dna post review pipeline. The resultant scaffolds for lm5633 were output and renamed from longest to shortest.

High-copy repeat analysis

Long read ONT assemblies provide another measure of completeness through the identification of high copy repeats such as centromeres and telomeres sequences (VanBuren et al., 2015). We employed a searching strategy to identify the centromeres that leverages the idea that the highest copy number tandem repeat (TR) will be the centromere in most genomes (Melters et al., 2013). We searched the genomes using tandem repeat finder (TRF; v4.09) using modified settings (1 1 2 80 5 200 2000 -d -h) (Benson, 1999). TR were reformatted, summed and plotted to find the highest copy number TR per our previous methods (VanBuren et al., 2015). While lm5633 had robust telomere arrays (Table 1), we could not detect a high-copy number centromere repeat similar to what we have found in *S. polyrhiza* (Michael et al., 2017), which could mean lm5633 has holocentric centromeres.

Gene prediction and annotation

The chromosome resolved lm5633 genome was annotated using a pipeline consisting of four major steps: repeat masking, transcript assembly, gene model prediction, and functional annotation. Repeats were identified using EDTA (v1.9.8) (Ou et al., 2019) and these repeats were used for softmasking. ONT cDNA reads were aligned to the genomes using minimap2 and assembled into transcript models using Stringtie (v1.3.6) . Softmasked genomes and Stringtie models were then processed by Funannotate (v1.6)

(https://github.com/nextgenusfs/funannotate) to produce gene models. Predicted proteins were then functionally annotated using Eggnog-mapper (v2) (Huerta-Cepas et al., 2017). The resulting gene models were renamed reflecting the chromosome and the linear position on the chromosome.

Orthogroup analysis and synteny

Gene families and overrepresented groups were determined with Orthofinder (v2.4.0) and CAFE5 (https://github.com/hahnlab/CAFE5). Genomes were accessed from Phytozome13 (https://phytozome-next.jgi.doe.gov/) or from specific publications such as the *Colocasia esculenta* (Taro) (Yin et al., 2021), *Spirodela intermedia* (Hoang et al., 2020), *Spirodela polyrhiza clone 9509* (Hoang et al., 2018) and *Wolffia australiana* clone 8730 (Michael et al., 2020). Orthofinder results were parsed and used for CAFE5 by modifying the species tree with "make_ultrametric.py" and filtering orthocounts.tsv with "clade_and_size_filter.py". Gene trees were visualized with iTOL (Letunic and Bork, 2019). The SynMap tool on CoGe (Grover et al., 2017) and McScan python version (https://github.com/tanghaibao/jcvi/wiki/MCscan) were utilized to generate whole genome synthetic maps, identify syntenic orthologs, estimate synonymous substitutions (Ks) across genomes, and generate figures.

*Single nuclei RNAseq (*snRNA-seq*)*

We generated a snRNA-seq dataset from a population of whole lm5633 plants grown in 250 ml erlenmeyer flasks under 12 hours of light and 12 hours of dark (intermediate days) with constant 22°C temperature (LDHH). SnRNA-seq was performed using the SMARTseq2 protocol on nuclei isolated from frozen plant material as previously described (Bakken et al., 2018) with some modifications. Nuclei were extracted with a custom buffer consisting of Tris-HCL (9.5) 15mM, EDTA 10mM, KCl 130mM, NaCl 20mM, PVP-10 8%, Spermine 0.07g, Spermidine 0.05g, Triton X-100 0.10%, BME 7.50%. Individual nuclei were sorted on the FACsAria Fusion system into SMARTseq2 lysis buffers allocated in 96 wells. The resulting libraries were sequenced on the Illumina NovaSeq platform. The following QC was performed on the resulting reads: genes were filtered if they were expressed in fewer than 5 cells, cells were required to have at least 100 genes expressed and at least 20,000 reads mapped to the transcriptome; this QC resulted in 269 cells with 8,457 genes expressed. An expression matrix was generated for each cell by each gene. We leveraged Single-cell Trajectories Reconstruction, Exploration And Mapping (STREAM) to cluster and map the cell trajectories of the 269 sequenced nuclei (Chen et al., 2019). Clustering was performed with K-means (k=13) in two components and 26 nearest neighbors using the top 100 principal components based on maximized silhouette scores. A weighted centroid principle graph was generated with epg_alpha=0.01,epg_mu=0.05,epg_lambda=0.05,jobs=10 and refined with st.extend_elastic_principal_graph using epg_ext_par=0.8. To detect significant marker genes st.detect_markers with cutoff_zscore=1.0 and cutoff_pvalue=0.01 was used. Transition markers were determined with st.detect_transition_markers using cutoff_spearman=0.4 and cutoff_logfc=0.25. SCANPY was used for data visualization purposes (Wolf et al., 2018).

The 269 nuclei formed 13 clusters that we defined using gene ontology (GO), KEGG, and PFAM annotation into cell types based on orthologous marker genes. Orthology was

761 determined with Orthofinder (Emms and Kelly, 2019) as above. Marker genes from model
762 species based on existing literature were used to search the orthogroups for lm5633
763 orthologues (Supplemental Table S6). The marker genes in model organisms were compared
764 with the orthofinder gene table and a corresponding lm5633 ortholog was assigned. Generally,
765 annotations with a geneID from eggnog mapper were more reliable than observing marker
766 genes in larger gene families (>3 copies per genomes). GO terms and quantities for cluster
767 marker genes were parsed and form marker genes lists using in-house scripts. GO lists were
768 visualized using REVIGO and CirGO (Supek et al., 2011; Kuznetsova et al., 2019).
769
770 **Accession numbers**
771 The final genome assembly is available on CoGe (https://genomevolution.org/) under the
772 ID:61245 or Biosample: SAMN19243672. Genomic and snRNA-seq reads can be found in SRA
773 under Biosample: SAMN19243672.
774
775 **Competing interest statement**
776 The authors declare no competing financial interests.
777
778 **Large datasets**
779 Supplemental Table S1. BUSCO scores for lm5633.
780 Supplemental Table S2. Predicted repeat types in the lm5633, sp9509 and wa8730 genomes.
781 Supplemental Table S3. A comparison of telomere sequence lengths in Duckweeds.
782 Supplemental Table S4. Tandem duplications (TDs) in lm5633.
783 Supplemental Table S5. Previously published marker genes used for defining snRNA-seq cell
784 types.
785 Supplemental Table S6. Orthogroups across all species.
786 Supplemental Table S7. Lemna minuta (lm5633) snRNA-seq marker genes per cell type.
787 Supplemental Table S8. Gene ontology (GO) terms for the meristem cluster.
788
789 **Acknowledgements**
790 We thank Rutgers Duckweed Stock Cooperative http://www.ruduckweed.org/ for hosting the live
791 duckweed strain under lm5633.
792
793 **Table**
794 Table 1. ***Lemna minuta* (lm5633) genome statistics.**
795 **Figure legends**
796 **Figure 1. *Lemna minuta* growth highlighting its anatomical analogs to other plants.** A) *L.*
797 *minuta* (lm5633) growing as a dark green mat of fronds in a sewage slough at Cotton Creek
798 Park, Encinitas, CA USA. The inset highlights the small size (~1 mm) of several fronds on a
799 finger tip. B) Cartoon of the *L. minuta* generational contribution that leads to the dense frond
800 mat: mother (M; green), daughter (D; purple), grand-daughter (GD; blue), great grand-daughter
801 (GGD; orange) and great, great grand-daughter fronds (GGGD; pink). The grey line represents
802 how they are connected through exponential growth. C) A single M frond view with the attached
803 D fronds that have GD, GGD and GGGD fronds nested in the two meristematic pockets. D) A
804 “nested doll” view of one M frond and the maturing generations of D, GD, GGD and GGGD in

the pocket. E) One interpretation of the *Lemna* frond is that the M frond is an axillary stem that has two poaches or bracts where the D frond is attached by a stipe or internode. The D, GD, GGD and GGGD progression then is similar to a branching structure of a generic plant, and the root-like structure at each subsequent axillary node is equivalent to an adventitious root. The arrows indicate that multiple internodes will emerge from the axillary stem over time (Landolt, 1986).

Figure 2. The highly heterozygous *Lemna minuta* (lm5633) genome resolved into 21 chromosomes. A) K-mer (k=31) frequency plot for lm5633 reveals two peaks consistent with a high level of heterozygosity (2%). B) Assembly graph of a 40 Mb region visualized with Bandage shows both the heterozygous branches as well as repeat tangles, “hairballs.” C) High throughput chromatin conformation capture (HiC) contact map resolving the polished lm5633 assembly into 21 chromosomes (darker red more contacts, lighter red to white less contacts).

Figure 3. Gene family expansion in lm5633 is driven by tandem duplication (TD). A) Lm5633 chromosomes aligned to sp9509 chromosomes with syntenic blocks (grey lines) anchoring positions between the two genomes. Chromosomes are the correct ratio between one another but are not to scale between the two species. A minus sign after the number means the chromosome has been flipped for visualization purposes. B) The single MYB transcription factor *LATE ELONGATED HYPOCOTYL* (*LHY*; blue line) and *PSEUDO-RESPONSE REGULATOR 7* (*PRR7*; green line) are in tight linkage and TE fragments (orange) resulting in the region expansion in lm5633. Grey lines connect other syntenic genes (blue, forward; green reverse). C) Boric acid channel TD (green line) also showing the expansion of the lm5633 genome due to TE fragments (orange). D) Sympantic principle component analysis (PCA) of significant GO terms associated with TDs in lm5633. Size of the circle is the log frequency and the color (red high, and yellow low) is the log FDR.

Figure 4. Whole genome evolution shows consistent gene family contractions in lm5633 ancestry. A) Gene family contractions (red) and expansions (blue) along the phylogenetic tree leading to minimized duckweed genomes are shown at each node and total protein family contractions and expansions for each species (right). Lm5633 shows the greatest similarity and gene family conservation with wa8730. A heatmap of orthogroups significantly expanded or contracted in lm5633. B) Self versus self and C) lm5633 versus other synonymous substitution (Ks) plots to elucidate the WGD history of lm5633. Dating is based on the mean Ks peak for each comparison using all paralogs/orthologs. Plant species used: *A. trichopoda* (Amborella), *A. thaliana*, *V. vinifera* (grape), *A. comosus* (pineapple), *O. sativa* (rice), *O. thomaeum* (oropetium), *Z. marina* (zostera), *C. esculenta* (taro), *S. intermedia* (si7747), *S. polyrhiza* (sp9509), *W. australiana* (wa8730), and *L. minuta* (lm5633).

Figure 5. Single nuclei RNA sequencing (snRNA-seq) of clonally propagating whole lm5633 plants. A) The most abundant transcripts reveal that highly expressed genes are generally related to photosynthesis. B) UMAP embedding of 13 k-means clusters in the snRNA-seq dataset. Trajectory analysis (black line) reveals a complex network where multiple branches result in terminally fated cell types. C) Log fold change in marker genes for the 13 specific clusters shows robust separation of cell types for functional analysis. Expression level (red, high; blue, low; larger circle, more cells; smaller circles, fewer cells).

Figure 6. Cell type definitions supported by gene ontology (GO) categories. Upset plot showing the number of GO terms that are specific and overlapping across the thirteen different predicted cell types. The bars on top represent the number of intersecting GO terms per cell type, and the bars on the left represent the number of GO terms found in each cell type. The cell clusters as defined in Figure 5B are on the right hand side to highlight their correspondence with the predicted cell types.

Figure 7. Trajectory analysis defines cell types of the entire *L. minuta* plant . A) UMAP tree embedding of snRNA-seq clusters describing the cell types. B) Cartoon of *L. minuta* based on (Banaszek and Musiał, 2011) coloring specific cell types per the UMAP. C) Live image of lm5633 showing the green roots.

Figure 8. Metal transport and accumulation associated genes in *L. minuta* are expressed in mesophyll cells. SnRNA-seq expression profiling of metal tolerance related genes reveal a cell type specific expression pattern with greater gene expression localized to the mesophyll cell types in lm5633. Lm5633.a03.5.g03640 (top) encodes *BORON TRANSPORTER 4* (*BOR4)* localized to mesophyll and replicating cells described further in the text.

Parsed Citations

Amborella Genome Project (2013) The Amborella genome and the evolution of flowering plants. Science 342: 1241089
Google Scholar: Author Only Title Only Author and Title

Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, Barkan E, Bertagnolli D, Casper T, Dee N, et al (2018) Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One 13: e0209648
Google Scholar: Author Only Title Only Author and Title

Banaszek A, Musiał K (2011) The new kenophyte in Poland - Lemna minuta Humb., Bonpl. and Kunth. Acta Societatis Botanicorum Poloniae 78: 69–72
Google Scholar: Author Only Title Only Author and Title

Becraft PW, Stinard PS, McCarty DR (1996) CRINKLY4: A TNFR-like receptor kinase involved in maize epidermal differentiation. Science 273: 1406–1409
Google Scholar: Author Only Title Only Author and Title

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573–580
Google Scholar: Author Only Title Only Author and Title

Bog M, Sree KS, Fuchs J, Hoang PTN, Schubert I, Kuever J, Rabenstein A, Paolacci S, Jansen MAK, Appenroth K (2020) A taxonomic revision of Lemna sect. Uninerves (Lemnaceae). Taxon 69: 56–66
Google Scholar: Author Only Title Only Author and Title

Ceschin S, Abati S, Ellwood NTW, Zuccarello V (2018) Riding invasion waves: Spatial and temporal patterns of the invasive Lemna minuta from its arrival to its spread across Europe. Aquat Bot 150: 1–8
Google Scholar: Author Only Title Only Author and Title

Ceschin S, Crescenzi M, Iannelli MA (2020) Phytoremediation potential of the duckweeds Lemna minuta and Lemna minor to remove nutrients from treated waters. Environ Sci Pollut Res Int 27: 15806–15814
Google Scholar: Author Only Title Only Author and Title

Chen H, Albergante L, Hsu JY, Lareau CA, Lo Bosco G, Guan J, Zhou S, Gorban AN, Bauer DE, Aryee MJ, et al (2019) Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat Commun 10: 1903
Google Scholar: Author Only Title Only Author and Title

Colt K Typha (Cattail) Crosslinking and Nuclei Isolation v2 (protocols.io.bsp6ndre). protocols.io. doi: 10.17504/protocols.io.bsp6ndre

Davis SM, Drake KD, Maier KJ (2002) Toxicity of boron to the duckweed, Spirodella polyrrhiza. Chemosphere 48: 615–620
Google Scholar: Author Only Title Only Author and Title

Del-Campo Marín CM, Oron G (2007) Boron removal by the duckweed Lemna gibba: a potential method for the remediation of boron-polluted waters. Water Res 41: 4579–4584
Google Scholar: Author Only Title Only Author and Title

Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP (2019) Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. Dev Cell 48: 840–852.e5
Google Scholar: Author Only Title Only Author and Title

Dorrity MW, Alexandre C, Hamm M, Vigil A-L, Fields S, Queitsch C, Cuperus J (2020) The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. 2020.07.17.204792
Google Scholar: Author Only Title Only Author and Title

Ellison CT, Vandenbussche F, Van Der Straeten D, Harmer SL (2011) XAP5 CIRCADIAN TIMEKEEPER regulates ethylene responses in aerial tissues of Arabidopsis. Plant Physiol 155: 988–999
Google Scholar: Author Only Title Only Author and Title

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20: 238
Google Scholar: Author Only Title Only Author and Title

Feng G, Liu G, Xiao J (2015) The Arabidopsis EIN2 restricts organ growth by retarding cell expansion. Plant Signal Behav 10: e1017169
Google Scholar: Author Only Title Only Author and Title

Fernández MI, Paisio CE, González PS, Perotti R, Meringer V, Villasuso AL, Agostini E (2020) Deepening the knowledge on the removal of Cr(VI) by L. minuta Kunth: removal efficiency and mechanisms, lipid signaling pathways, antioxidant response, and toxic effects. Environ Sci Pollut Res Int 27: 14567–14580
Google Scholar: Author Only Title Only Author and Title

Florian Schröder, Janina Lisso, Peggy Lange & Carsten Müssig The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. BMC Plant Biology. doi: 10.1186/1471-2229-9-20

Frick H (1985) Boron tolerance and accumulation in the duckweed, Lemna minor. J Plant Nutr 8: 1123–1129
Google Scholar: Author Only Title Only Author and Title

Grover JW, Bomhoff M, Davey S, Gregory BD, Mosher RA, Lyons E (2017) CoGe LoadExp+: A web-based suite that integrates next-generation sequencing data analysis workflows and visualization. Plant Direct. doi: 10.1002/pld3.8
Google Scholar: Author Only Title Only Author and Title

Gujas B, Kastanaki E, Sturchler A, Cruz TMD, Ruiz-Sola MA, Dreos R, Eicke S, Truernit E, Rodriguez-Villalon A (2020) A Reservoir of Pluripotent Phloem Cells Safeguards the Linear Developmental Trajectory of Protophloem Sieve Elements. Curr Biol 30: 755–766.e4
Google Scholar: Author Only Title Only Author and Title

Gür N, Türker OC, Böcük H (2016) Toxicity assessment of boron (B) by Lemna minor L. and Lemna gibba L. and their possible use as model plants for ecological risk assessment of aquatic ecosystems with boron pollution. Chemosphere 157: 1–9
Google Scholar: Author Only Title Only Author and Title

Gutensohn M, Schulz B, Nicolay P, Flügge UI (2000) Functional analysis of the two Arabidopsis homologues of Toc34, a component of the chloroplast protein import apparatus. Plant J 23: 771–783
Google Scholar: Author Only Title Only Author and Title

Harkess A, McLoughlin F, Bilkey N, Elliott K, Emenecker R, Mattoon E, Miller K, Czymmek K, Vierstra R, Meyers BC, et al (2020) A new Spirodela polyrhiza genome and proteome reveal a conserved chromosomal structure with high abundances of proteins favoring energy production. 2020.01.23.909457
Google Scholar: Author Only Title Only Author and Title

Hasing T, Tang H, Brym M, Khazi F, Huang T, Chambers AH (2020) A phased Vanilla planifolia genome enables genetic improvement of flavour and production. Nature Food 1: 811–819
Google Scholar: Author Only Title Only Author and Title

Haslam TM, Gerelle WK, Graham SW, Kunst L (2017) The Unique Role of the ECERIFERUM2-LIKE Clade of the BAHD Acyltransferase Superfamily in Cuticular Wax Metabolism. Plants. doi: 10.3390/plants6020023
Google Scholar: Author Only Title Only Author and Title

Hayashi S, Gresshoff PM, Kinkema M (2008) Molecular analysis of lipoxygenases associated with nodule development in soybean. Mol Plant Microbe Interact 21: 843–853
Google Scholar: Author Only Title Only Author and Title

Hillman WS (1976) Calibrating duckweeds: light, clocks, metabolism, flowering. Science 193: 453–458
Google Scholar: Author Only Title Only Author and Title

Hillman WS, Culley DD (1978) The Uses of Duckweed: The rapid growth, nutritional value, and high biomass productivity of these floating plants suggest their use in water treatment, as feed crops, and in energy-efficient farming. Am Sci 66: 442–451
Google Scholar: Author Only Title Only Author and Title

Hoang PNT, Michael TP, Gilbert S, Chu P, Motley ST, Appenroth KJ, Schubert I, Lam E (2018) Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. Plant J 96: 670–684
Google Scholar: Author Only Title Only Author and Title

Hoang PTN, Fiebig A, Novák P, Macas J, Cao HX, Stepanenko A, Chen G, Borisjuk N, Scholz U, Schubert I (2020) Chromosome-scale genome assembly for the duckweed Spirodela intermedia, integrating cytogenetic maps, PacBio and Oxford Nanopore libraries. Sci Rep 10: 19230
Google Scholar: Author Only Title Only Author and Title

Hoang PTN, Schubert I (2017) Reconstruction of chromosome rearrangements between the two most ancestral duckweed species Spirodela polyrhiza and S. intermedia. Chromosoma 126: 729–739
Google Scholar: Author Only Title Only Author and Title

Huang Y, Feng C-Z, Ye Q, Wu W-H, Chen Y-F (2016) Arabidopsis WRKY6 Transcription Factor Acts as a Positive Regulator of Abscisic Acid Signaling during Seed Germination and Early Seedling Development. PLoS Genet 12: e1005833
Google Scholar: Author Only Title Only Author and Title

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P (2017) Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol 34: 2115–2122
Google Scholar: Author Only Title Only Author and Title

Jiao Y, Li J, Tang H, Paterson AH (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell 26: 2792–2802
Google Scholar: Author Only Title Only Author and Title

Kim J, Jung JH, Lee SB, Go YS, Kim HJ, Cahoon R, Markham JE, Cahoon EB, Suh MC (2013) Arabidopsis 3-ketoacyl-coenzyme a synthase9 is involved in the synthesis of tetracosanoic acids as precursors of cuticular waxes, suberins, sphingolipids, and phospholipids. Plant Physiol 162: 567–580
Google Scholar: Author Only Title Only Author and Title

Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37: 540–546
Google Scholar: Author Only Title Only Author and Title

Kötting O, Santelia D, Edner C, Eicke S, Marthaler T, Gentry MS, Comparot-Moss S, Chen J, Smith AM, Steup M, et al (2009) STARCH-EXCESS4 is a laforin-like Phosphoglucan phosphatase required for starch degradation in Arabidopsis thaliana. Plant Cell 21: 334–346
Google Scholar: Author Only Title Only Author and Title

Krishnaswami SR, Grindberg RV, Novotny M, Venepally P, Lacar B, Bhutani K, Linker SB, Pham S, Erwin JA, Miller JA, et al (2016) Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. Nat Protoc 11: 499–524
Google Scholar: Author Only Title Only Author and Title

Kuznetsova I, Lugmayr A, Siira SJ, Rackham O, Filipovska A (2019) CirGO: an alternative circular way of visualising gene ontology terms. BMC Bioinformatics 20: 84
Google Scholar: Author Only Title Only Author and Title

Landolt E (1986) Biosystematic investigations in the family of duckweeds (Lemnaceae)(Vol. 2.) The family of Lemnaceae-a monographic study. vol. 1. Veroff Geobot Inst ETH 71: 1–563
Google Scholar: Author Only Title Only Author and Title

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47: W256–W259
Google Scholar: Author Only Title Only Author and Title

Link VM, Duttke SH, Chun HB, Holtman IR, Westin E, Hoeksema MA, Abe Y, Skola D, Romanoski CE, Tao J, et al (2018) Analysis of Genetically Diverse Macrophages Reveals Local and Domain-wide Mechanisms that Control Transcription Factor Binding and Function. Cell 173: 1796–1809.e17
Google Scholar: Author Only Title Only Author and Title

Liu C, Gu W, Dai Z, Li J, Jiang H, Zhang Q (2018) Boron accumulation by Lemna minor L. under salt stress. Sci Rep 8: 8954
Google Scholar: Author Only Title Only Author and Title

Liu Q, Liang Z, Feng D, Jiang S, Wang Y, Du Z, Li R, Hu G, Zhang P, Ma Y, et al (2020a) Transcriptional landscape of rice roots at the single-cell resolution. Mol Plant. doi: 10.1016/j.molp.2020.12.014
Google Scholar: Author Only Title Only Author and Title

Liu Z, Zhou Y, Guo J, Li J, Tian Z, Zhu Z, Wang J, Wu R, Zhang B, Hu Y, et al (2020b) Global Dynamic Molecular Profiling of Stomatal Lineage Cell Development by Single-Cell RNA Sequencing. Mol Plant 13: 1178–1193
Google Scholar: Author Only Title Only Author and Title

Lopez-Anido CB, Vatén A, Smoot NK, Sharma N, Guo V, Gong Y, Anleu Gil MX, Weimer AK, Bergmann DC (2021) Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. Dev Cell 56: 1043–1055.e4
Google Scholar: Author Only Title Only Author and Title

Lucey J (2003) LEMNA MINUTA KUNTH (LEAST DUCKWEED) IN E. CORK (VC H5). IRISH BOTANICAL
Google Scholar: Author Only Title Only Author and Title

Lutz KA, Wang W, Zdepski A, Michael TP (2011) Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing. BMC Biotechnol 11: 54
Google Scholar: Author Only Title Only Author and Title

Machida C, Nakagawa A, Kojima S, Takahashi H, Machida Y (2015) The complex of ASYMMETRIC LEAVES (AS) proteins plays a central role in antagonistic interactions of genes for leaf polarity specification in Arabidopsis. Wiley Interdiscip Rev Dev Biol 4: 655–671
Google Scholar: Author Only Title Only Author and Title

Ma J, Jiang Q-T, Wei L, Yang Q, Zhang X-W, Peng Y-Y, Chen G-Y, Wei Y-M, Liu C, Zheng Y-L (2014) Conserved structure and varied expression reveal key roles of phosphoglucan phosphatase gene starch excess 4 in barley. Planta 240: 1179–1190
Google Scholar: Author Only Title Only Author and Title

Marand AP, Chen Z, Gallavotti A, Schmitz RJ (2020) A cis-regulatory atlas in maize at single-cell resolution. 2020.09.27.315499
Google Scholar: Author Only Title Only Author and Title

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27: 764–770
Google Scholar: Author Only Title Only Author and Title

McGinnis KM, Thomas SG, Soule JD, Strader LC, Zale JM, Sun T-P, Steber CM (2003) The Arabidopsis SLEEPY1 gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. Plant Cell 15: 1120–1130
Google Scholar: Author Only Title Only Author and Title

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14: R10
Google Scholar: Author Only Title Only Author and Title

Michael TP, Bryant D, Gutierrez R, Borisjuk N, Chu P, Zhang H, Xia J, Zhou J, Peng H, El Baidouri M, et al (2017) Comprehensive definition of genome features in Spirodela polyrhiza by high-depth physical mapping and short-read DNA sequencing strategies. Plant J 89: 617–635
Google Scholar: Author Only Title Only Author and Title

Michael TP, Ernst E, Hartwick N, Chu P, Bryant D, Gilbert S, Ortleb S, Baggs EL, Sree KS, Appenroth KJ, et al (2020) Genome and time-of-day transcriptome of Wolffia australiana link morphological minimization with gene loss and less growth control. Genome Res. doi: 10.1101/gr.266429.120
Google Scholar: Author Only Title Only Author and Title

Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR (2018) High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. Nat Commun 9: 541
Google Scholar: Author Only Title Only Author and Title

Michael TP, VanBuren R (2020) Building near-complete plant genomes. Curr Opin Plant Biol 54: 26–33
Google Scholar: Author Only Title Only Author and Title

Mifsud S (2010) First occurences of Lemna minuta Kunth (fam. Lemnaceae) in the Maltese islands.
Google Scholar: Author Only Title Only Author and Title

Mills RF, Doherty ML, López-Marqués RL, Weimar T, Dupree P, Palmgren MG, Pittman JK, Williams LE (2008) ECA3, a Golgi-localized P2A-type ATPase, plays a crucial role in manganese nutrition in Arabidopsis. Plant Physiol 146: 116–128
Google Scholar: Author Only Title Only Author and Title

Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang M-L, Chen J, Biggers E, et al (2015) The pineapple genome and the evolution of CAM photosynthesis. Nat Genet 47: 1435–1442
Google Scholar: Author Only Title Only Author and Title

Moodley D, Procheş Ş, Wilson JRU (2016) A global assessment of a large monocot family highlights the need for group-specific analyses of invasiveness. AoB Plants 8: lw009
Google Scholar: Author Only Title Only Author and Title

Oikawa T, Kyozuka J (2009) Two-Step Regulation of LAX PANICLE1 Protein Accumulation in Axillary Meristem Formation in Rice. Plant Cell 21: 1095–1108
Google Scholar: Author Only Title Only Author and Title

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol 20: 275
Google Scholar: Author Only Title Only Author and Title

Pandey DK, Chaudhary B (2016) Domestication-driven Gossypium profilin 1 (GhPRF1) gene transduces early flowering phenotype in tobacco by spatial alteration of apical/floral-meristem related gene expression. BMC Plant Biol 16: 112
Google Scholar: Author Only Title Only Author and Title

Paolacci S, Jansen MAK, Harrison S (2018) Competition Between Lemna minuta, Lemna minor, and Azolla filiculoides. Growing Fast or Being Steadfast? Front Chem 6: 207
Google Scholar: Author Only Title Only Author and Title

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biol 20: 38
Google Scholar: Author Only Title Only Author and Title

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 11: 1432
Google Scholar: Author Only Title Only Author and Title

Rimon D, Galun E (1968) Morphogenesis of Spirodela oligorrhiza: Ontogenesis of Fronds. Bot Gaz 129: 138–144
Google Scholar: Author Only Title Only Author and Title

Rosquete MR, Kleine-Vehn J (2018) PIN7 auxin carrier is a terminator of radial root expansion in Arabidopsis thaliana. bioRxiv 253807
Google Scholar: Author Only Title Only Author and Title

Ryu KH, Huang L, Kang HM, Schiefelbein J (2019) Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. Plant Physiol 179: 1444–1456
Google Scholar: Author Only Title Only Author and Title

Satterlee JW, Strable J, Scanlon MJ (2020) Plant stem-cell organization and differentiation at single-cell resolution. Proc Natl Acad Sci U S A 117: 33689–33699
Google Scholar: Author Only Title Only Author and Title

Schröder F, Lisso J, Müssig C (2011) EXORDIUM-LIKE1 promotes growth during low carbon availability in Arabidopsis. Plant Physiol 156: 1620–1630
Google Scholar: Author Only Title Only Author and Title

Shi Y, Xu H, Shen Q, Lin J, Wang Y, Hua X, Yao W, Yu Q, Ming R, Zhang J (2019) Comparative Analysis of SUS Gene Family between Saccharum officinarum and Saccharum spontaneum. Trop Plant Biol 12: 174–185
Google Scholar: Author Only Title Only Author and Title

Silva GG, Green AJ, Weber V, Hoffmann P, Lovas-Kiss Á, Stenert C, Maltchik L (2018) Whole angiosperms Wolffia columbiana disperse by gut passage through wildfowl in South America. Biol Lett 14: 20180703
Google Scholar: Author Only Title Only Author and Title

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212
Google Scholar: Author Only Title Only Author and Title

Slovin JP, Cohen JD (1988) Levels of Indole-3-Acetic Acid in Lemna gibba G-3 and in a Large Lemna Mutant Regenerated from Tissue Culture. Plant Physiol 86: 522–526
Google Scholar: Author Only Title Only Author and Title

Song Q, Ando A, Jiang N, Ikeda Y, Chen ZJ (2020) Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. Genome Biol 21: 178
Google Scholar: Author Only Title Only Author and Title

Srivastava AK, Senapati D, Srivastava A, Chakraborty M, Gangappa SN, Chattopadhyay S (2015) Short Hypocotyl in White Light1 Interacts with Elongated Hypocotyl5 (HY5) and Constitutive Photomorphogenic1 (COP1) and Promotes COP1-Mediated Degradation of HY5 during Arabidopsis Seedling Development. Plant Physiol 169: 2922–2934
Google Scholar: Author Only Title Only Author and Title

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6: e21800
Google Scholar: Author Only Title Only Author and Title

Takahashi I, Kojima S, Sakaguchi N, Umeda-Hara C, Umeda M (2010) Two Arabidopsis cyclin A3s possess G1 cyclin-like features. Plant Cell Rep 29: 307–315
Google Scholar: Author Only Title Only Author and Title

Takano J, Noguchi K, Yasumori M, Kobayashi M, Gajdos Z, Miwa K, Hayashi H, Yoneyama T, Fujiwara T (2002) Arabidopsis boron transporter for xylem loading. Nature 420: 337–340
Google Scholar: Author Only Title Only Author and Title

Tian Q, Olsen L, Sun B, Lid SE, Brown RC, Lemmon BE, Fosnes K, Gruis DF, Opsahl-Sorteberg H-G, Otegui MS, et al (2007) Subcellular localization and functional domain studies of DEFECTIVE KERNEL1 in maize and Arabidopsis suggest a model for aleurone cell fate specification involving CRINKLY4 and SUPERNUMERARY ALEURONE LAYER1. Plant Cell 19: 3127–3145
Google Scholar: Author Only Title Only Author and Title

Türker OC, Yakar A, Türe C, Saz Ç (2019) Boron (B) removal and bioelectricity captured from irrigation water using engineered duckweed-microbial fuel cell: effect of plant species and vegetation structure. Environ Sci Pollut Res Int 26: 31522–31536
Google Scholar: Author Only Title Only Author and Title

Upadhyay N, Kar D, Deepak Mahajan B, Nanda S, Rahiman R, Panchakshari N, Bhagavatula L, Datta S (2019) The multitasking abilities of MATE transporters in plants. J Exp Bot 70: 4643–4656
Google Scholar: Author Only Title Only Author and Title

Utami D, Kawahata A, Sugawara M, Jog RN, Miwa K, Morikawa M (2018) Effect of Exogenous General Plant Growth Regulators on the Growth of the Duckweed Lemna minor. Front Chem 6: 251
Google Scholar: Author Only Title Only Author and Title

VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Spittle K, Hall R, Gu J, Lyons E, et al (2015) Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. Nature 527: 508–511
Google Scholar: Author Only Title Only Author and Title

Van Echelpoel W, Boets P, Goethals PLM (2016) Functional Response (FR) and Relative Growth Rate (RGR) Do Not Show the Known Invasiveness of Lemna minuta (Kunth). PLoS One 11: e0166132
Google Scholar: Author Only Title Only Author and Title

Van Hoeck A, Horemans N, Monsieurs P, Cao HX, Vandenhove H, Blust R (2015) The first draft genome of the aquatic model plant Lemna minor opens the route for future stress physiology research and biotechnological applications. Biotechnol Biofuels 8: 188
Google Scholar: Author Only Title Only Author and Title

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27: 737–746
Google Scholar: Author Only Title Only Author and Title

Vert G, Briat JF, Curie C (2001) Arabidopsis IRT2 gene encodes a root-periphery iron transporter. Plant J 26: 181–189
Google Scholar: Author Only Title Only Author and Title

Vishwakarma A, Tetali SD, Selinski J, Scheibe R, Padmasree K (2015) Importance of the alternative oxidase (AOX) pathway in regulating cellular redox and ROS homeostasis to optimize photosynthesis during restriction of the cytochrome oxidase pathway in Arabidopsis thaliana. Ann Bot 116: 555–569
Google Scholar: Author Only Title Only Author and Title

Vojta L, Tomašić Paić A, Horvat L, Rac A, Lepeduš H, Fulgosi H (2019) Complex lumenal immunophilin AtCYP38 influences thylakoid remodelling in Arabidopsis thaliana. J Plant Physiol 243: 153048
Google Scholar: Author Only Title Only Author and Title

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33: 2202–2204
Google Scholar: Author Only Title Only Author and Title

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9: e112963
Google Scholar: Author Only Title Only Author and Title

Wang W, Haberer G, Gundlach H, Gläßer C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA, Shanklin J, et al (2014) The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat Commun 5: 3311
Google Scholar: Author Only Title Only Author and Title

Wang X, Li J, Guo J, Qiao Q, Guo X, Ma Y (2020) The WRKY transcription factor PlWRKY65 enhances the resistance of Paeonia lactiflora (herbaceous peony) to Alternaria tenuissima. Hortic Res 7: 57
Google Scholar: Author Only Title Only Author and Title

Wen R, Wang S, Xiang D, Venglat P, Shi X, Zang Y, Datla R, Xiao W, Wang H (2014) UBC13, an E2 enzyme for Lys63-linked ubiquitination, functions in root development by affecting auxin signaling and Aux/IAA protein stability. Plant J 80: 424–436
Google Scholar: Author Only Title Only Author and Title

Wen X, Ding Y, Tan Z, Wang J, Zhang D, Wang Y (2020) Identification and characterization of cadmium stress-related LncRNAs from Betula platyphylla. Plant Sci 299: 110601
Google Scholar: Author Only Title Only Author and Title

Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31: 3350–3352
Google Scholar: Author Only Title Only Author and Title

Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19: 15
Google Scholar: Author Only Title Only Author and Title

Xu L, Xu Y, Dong A, Sun Y, Pi L, Xu Y, Huang H (2003) Novel as1 and as2 defects in leaf adaxial-abaxial polarity reveal the requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA functions in specifying leaf adaxial identity. Development 130: 4097–4107
Google Scholar: Author Only Title Only Author and Title

Xu X, Crow M, Rice BR, Li F, Harris B, Liu L, Demesa-Arevalo E, Lu Z, Wang L, Fox N, et al (2021) Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. Dev Cell. doi: 10.1016/j.devcel.2020.12.015
Google Scholar: Author Only Title Only Author and Title

Yamawaki TM, Lu DR, Ellwanger DC, Bhatt D, Manzanillo P, Arias V, Zhou H, Yoon OK, Homann O, Wang S, et al (2021) Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling. BMC Genomics 22: 66
Google Scholar: Author Only Title Only Author and Title

Yan A, Wang Y, Tan SN, Mohd Yusof ML, Ghosh S, Chen Z (2020) Phytoremediation: A Promising Approach for Revegetation of Heavy Metal-Polluted Land. Front Plant Sci 11: 359
Google Scholar: Author Only Title Only Author and Title

Yin J, Jiang L, Wang L, Han X, Guo W, Li C, Zhou Y, Denton M, Zhang P (2021) A high-quality genome of taro (Colocasia esculenta (L.) Schott), one of the world's oldest crops. Mol Ecol Resour 21: 68–77
Google Scholar: Author Only Title Only Author and Title

Zhao W, Li Y, Fan S, Wen T, Wang M, Zhang L, Zhao L (2021) The tomato WRKY32 transcription factor affects ripe fruit color by regulating YFT1, a core component of ethylene signal transduction. J Exp Bot. doi: 10.1093/jxb/erab113
Google Scholar: Author Only Title Only Author and Title