

Genome assembly, annotation, and comparative analysis of the cattail *Typha latifolia*

Shane D. Widanagama,¹ Joanna R. Freeland,² Xinwei Xu,³ and Aaron B. A. Shafer ^{4,*}

¹Department of Computer Science, Trent University, Peterborough, ON K9L 0G2, Canada,

²Department of Biology, Trent University, Peterborough, ON K9L 0G2, Canada,

³Department of Ecology, College of Life Sciences, Wuhan University, Wuhan 430072, China, and

⁴Department of Forensic Sciences, Trent University, Peterborough, ON K9L 0G2, Canada

*Corresponding author: Department of Forensic Sciences, Trent University, DNA Building, 2140 East Bank Drive, Peterborough, ON, K9L 0G2, Canada.
Email: aaronshafer@trentu.ca

Abstract

Cattails (*Typha* species) comprise a genus of emergent wetland plants with a global distribution. *Typha latifolia* and *Typha angustifolia* are two of the most widespread species, and in areas of sympatry can interbreed to produce the hybrid *Typha* × *glauca*. In some regions, the relatively high fitness of *Typha* × *glauca* allows it to outcompete and displace both parent species, while simultaneously reducing plant and invertebrate biodiversity, and modifying nutrient and water cycling. We generated a high-quality whole-genome assembly of *T. latifolia* using PacBio long-read and high coverage Illumina sequences that will facilitate evolutionary and ecological studies in this hybrid zone. Genome size was 287 Mb and consisted of 1158 scaffolds, with an N50 of 8.71 Mb; 43.84% of the genome were identified as repetitive elements. The assembly has a BUSCO score of 96.03%, and 27,432 genes and 2700 RNA sequences were putatively identified. Comparative analysis detected over 9000 shared orthologs with related taxa and phylogenomic analysis supporting *T. latifolia* as a divergent lineage within Poales. This high-quality scaffold-level reference genome will provide a useful resource for future population genomic analyses and improve our understanding of *Typha* hybrid dynamics.

Keywords: PacBio long-read sequencing; illumina short-read; *de novo*; Typhaceae; bulrush; broadleaf cattail; hybrids

Introduction

Cattails (*Typha* spp.) are aquatic macrophytes that are essential components of wetlands around the world (reviewed in Bansal *et al.* 2019). These macrophytes often dominate ecosystems through a combination of rapid growth, large size, and sexual and asexual reproduction (Yeo 1964; Andrews and Pratt 1978; Miller and Fujii 2010). Cattails are vital to many wetlands where they cycle nutrients, provide habitat, and aid in bioremediation (Grosshans 2014; Svedarsky *et al.* 2016; Bonanno and Cirelli 2017). However, cattails can also dominate wetlands and in recent decades invasive cattails have been identified in numerous regions, often following anthropogenic changes that have altered water cycles and increased nutrient loads in wetlands (reviewed in Zedler and Kercher 2004; Bansal *et al.* 2019). Invasive cattails often form monotypic stands with deleterious effects on local plants and animals, wetland water cycling, and biogeochemical cycles (Yeo 1964; Grace and Harrison 1986; Farrer and Goldberg 2009; Lawrence *et al.* 2016).

Typha latifolia is likely the most widespread cattail species, occurring on every continent except Antarctica (Smith 1987). *Typha angustifolia* is also widespread throughout the temperate northern hemisphere (Grace and Harrison 1986), including North America where it was likely introduced from Europe several centuries ago (Ciotir *et al.* 2013, 2017). In some regions of sympatry *T. latifolia*

and *T. angustifolia* interbreed to produce the hybrid *Typha* × *glauca* (Grace and Harrison 1986; Ciotir *et al.* 2017; Bansal *et al.* 2019). In regions surrounding the Laurentian Great Lakes and St. Lawrence Seaway in North America, *Typha* × *glauca* exhibits heterosis (Bunbury-Blanchette *et al.* 2015; Zapfe and Freeland 2015), and is more abundant than its parental species (Kirk *et al.* 2011; Freeland *et al.* 2013; Pieper *et al.* 2020). In addition to displacing both parental species, invasive *Typha* × *glauca* reduces native plant and invertebrate biodiversity (Tuchman *et al.* 2009; Lawrence *et al.* 2016), and alters nutrient cycling and community structure in wetlands (Tuchman *et al.* 2009; Larkin *et al.* 2012; Geddes *et al.* 2014; Lishawa *et al.* 2014; Lawrence *et al.* 2017).

Although dominant in some regions of North America, *Typha* × *glauca* is uncommon in other regions where the parental species are sympatric, including Europe (Ciotir *et al.* 2017), eastern Canada (Freeland *et al.* 2013), and China (Zhou *et al.* 2016). It is not well understood why hybrids are dominant in some regions but not others, although Tisshaw *et al.* (2020) suggested that hybrids may be limited in coastal wetlands because they have difficulty germinating in salt-rich environments. In addition, although advanced-generation and back-crossed hybrids have been experimentally generated (Pieper *et al.* 2017) and identified in natural populations (*e.g.*, Kirk *et al.* 2011; Freeland *et al.* 2013; Pieper *et al.* 2020), it has not been possible to differentiate

Received: August 23, 2021. Accepted: November 13, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

advanced-generation and backcrossed hybrids based on the small number of species-specific molecular markers that are currently available (Snow *et al.* 2010; Kirk *et al.* 2011). Morphology-based assessments are also unreliable due to overlapping phenotypes (Tangen *et al.* 2021).

A genome-wide suite of SNPs specific to one or the other parent species would greatly facilitate investigations into the dynamics of the *Typha* × *glauca* hybrid zone, which is now expanding westwards from the Great Lakes region into the Prairie Pothole Region of Canada and the USA (Tangen *et al.* 2021). Genome characterization would also facilitate investigations into local adaptation, introgression, and hybrid dynamics; this in turn may inform future management of *T. latifolia*, which is predicted to experience a dramatically reduced distribution following climate change (Xu *et al.* 2013), likely to the detriment of wetlands throughout its widespread native distribution. Here, we report the genome assembly, annotation, and analysis of *T. latifolia*, the first fully sequenced species in the family Typhaceae.

Materials and methods

Sampling and sequencing

Leaves of a known *T. latifolia* plant (Figure 1) were taken from an individual grown at Trent University's (ON, Canada) greenhouse and ground in liquid nitrogen. DNA was immediately extracted using an E.Z.N.A Plant DNA Kit (Omega Bio-Tek, Inc. GA, USA) following the manufacturer's instructions for frozen material and eluted in a final volume of 100 µl. DNA quality was assessed on a TapeStation (Agilent Technologies, CA, USA) before being sent to The Centre for Applied Genomics, Toronto, ON, for sequencing. Paired-end reads were sequenced on a single lane of an Illumina HiSeqX system (Illumina, Inc. CA, USA). Long-reads (LR) were



Figure 1 Broadleaf cattail (*Typha latifolia*). Photo by Joanna Freeland.

sequenced on one SMRT HiFi cell using a PacBio Sequel II system (Pacific Biosciences of California, Inc. CA, USA).

Genome assembly

Adapters and low-quality bases were trimmed from paired-end short reads (SRs) using Trimmomatic -v0.36 (Bolger *et al.* 2014). The optimal k-mer and genome size from the SRs were estimated using Kmergenie -v1.7051 (Chikhi and Medvedev 2014). We then applied a four-step hybrid assembly: (1) SR assemblies were generated using multiple pipelines: ABySS -v2.2.4 (Jackman *et al.* 2017) with default settings, 32 threads, a 125 GB total memory; SOAPdenovo2 -r240 (Luo *et al.* 2012) using default settings, 16 threads, 400 GB of total memory; and Platanus -v1.2.4 (Kajitani *et al.* 2014) using default settings, 32 threads, a 343 GB memory limit, a total memory of 375 GB. We also downsampled the SR data to 20% because really high coverage can be problematic for de Bruijn graphs (Richards and Murali 2015). The SR assembly with highest quality and completeness was selected by calculating contig N50, total contig length as a percentage of estimated genome size, and the number and percentage of contigs longer than 50 Kb. (2) SR contigs were then aligned to raw LR to create scaffolds using DBG2OLC -v20160205 (options: k 17, AdaptiveTh 0.001, KmerCovTh 2, MinOverlap 20, RemoveChimera 1) (Ye *et al.* 2016). (3) LR contigs were assembled using Canu -v2.0 (Koren *et al.* 2017) using the expected genome size inferred from Kmergenie, a maximum memory of 124 GB, and a maximum of 32 threads. And (4) the hybrid assembly (Step 2) was merged with Canu LR contigs (Step 3) using QuickMerge -v0.3.0 (Chakraborty *et al.* 2016) following the 2-step strategy described by Solares *et al.* (2018). Pilon -v1.23.0 (Walker *et al.* 2014) was used to polish the genome by mapping the Illumina SR back to the genome, thereby correcting base errors and small misassemblies. The final assembly was then submitted and assessed for sequencing and assembly artifacts by NCBI.

Genome evaluation and annotation

The *T. latifolia* hybrid genome was assessed for completeness using Benchmarking Universal Single-copy Orthologs (BUSCO) -v3.0.2 (Simão *et al.* 2015). BUSCOs of the lineage Liliopsida were assessed from OrthoDB release 10 (Waterhouse *et al.* 2013). We aligned available chloroplast and mitochondrial genomes (Supplementary Table S1) to our assembly using NUCmer -v3.23 (Kurtz *et al.* 2004) to identify plastid genomes; scaffolds with long matches (>5000 bp) were extracted and further validated on NCBI Blast. Structural and functional annotation was done in the GenSAS -v6.0 annotation pipeline (Humann *et al.* 2019). We masked interspersed and simple repetitive elements throughout the genome using a database developed through repeat modeler -v2.0.1 (Smit and Hubley 2008–2015) in conjunction with repeat masker -v4.0.7 with the NCBI/rmbblast search engine, and quick sensitivity (Smit *et al.* 2013–2015). The hardmasked version of the genome sequence was used for feature prediction. Available RNA-seq data from *T. angustifolia* (SRR15541138) were mapped to the genome using hisat2 -v2.1.0 (Kim *et al.* 2019) and gene prediction with the BRAKER2 -v2.1.1 pipeline, which uses Augustus and GeneMark-ET (Lomsadze *et al.* 2014; Brůna *et al.* 2021). ncRNA was predicted using tRNAscan-SE -v2.0 (Lowe and Eddy 1997) and Infernal -v1.1.3 (Nawrocki and Eddy 2013). Refinement of the official gene set was performed with an available *T. latifolia* transcriptome (Moscou 2017) using PASA -v2.3.3 (Haas *et al.* 2003). Last, our assembly (N50 scaffolds) was aligned and compared to a recent *T. latifolia* hybrid scaffold assembly (JAAWWQ01000000) using

Mummer4 (Marçais et al. 2018) requiring a minimum alignment (l) of 10,000 bp.

Comparative genomics

Single-copy orthologous genes from five other species were identified using OrthoVenn2 with an *e*-value cutoff of 1e-5, an inflation value of 1.5, and the annotation, protein similarity network, and cluster relationship network enabled (Xu et al. 2019). OrthoVenn2 performs all-against-all genome-wide protein comparisons and groups genes into clusters with the Markov Clustering Algorithm, where a cluster is made up of orthologs and paralogs (Wang et al. 2015). Here we selected *Oryza sativa Japonica*, *Brachypodium distachyon*, and *Sorghum bicolor* from the family Poaceae, and *Ananas comosus* from the family Bromeliaceae. *Arabidopsis thaliana* from the Brassicaceae family was chosen as the dicot outgroup. All proteomes were from the Ensemble database release 104 (Howe et al. 2021). The identified orthologous sequences were then aligned using MAFFT -v7.741 (Kato and Standley 2013) and concatenated into single sequences by species using SeqKit -v0.15.0 (Shen et al. 2016). Maximum likelihood phylogenetic analysis was performed using RAxML -v8.2.12 (Stamatakis 2014) with the PROTGAMMAAUTO substitution model (Darriba et al. 2012). The phylogenetic tree divergence times were estimated using MCMCTree from the PAML package -v4.9j (Yang 1997, 2007), and was calibrated using the divergence time between *Sorghum* and *Oryza* (42–52 Mya) (Kumar et al. 2017).

Results

Sequencing data and genome assembly

A total of 138.6 Gb of raw 151 bp Illumina reads were sequenced (Supplementary Table S2). Quality filtering and trimming removed 6.5 Gb of sequence data. A recommended k-mer size of 101 bp and estimated genome size of 257 Mb was estimated from the short-read data (Supplementary Figure S1). LR data from the PacBio Sequel II generated 86.8 Gb of raw data: this included 7,244,218 subreads with a mean length of 11,978.2 bp. All sequencing reads were deposited in the NCBI Sequence Read Archive (Accession No PRJNA751759). The ABySS assembly that used 100% of Illumina reads had the most contiguous genome and was used for the assembly (Supplementary Table S3). This ABySS SR assembly had a N50 of 0.011 Mb, 365,565 contigs, and 362 contigs longer than 50 Kb (Supplementary Table S3). The DBG2OLC assemblies (ABySS contigs + raw long reads) improved the assembly statistics but resulted in a smaller than expected genome (Table 1); here we calculated N50 of 0.132 Mb, 1840

contigs, and 1445 contigs longer than 50 Kb (Supplementary Table S4). The LR Canu assembly had an N50 of 8.706 Mb, and contained 1189 contigs, 821 of which were longer than 50 Kb (95.54%) (Table 1). The final merged and polished hybrid genome assembly—DBG2OLC assembly 1+Canu—was 286.7 Mb with plastids removed. We estimated an N50 of 8.71 Mb, and consisted of 1158 scaffolds (Table 1). Step 4 resulted in the merger of three scaffolds and the subsequent removal of 30 short scaffolds by NCBI vetting. The GC content was 38.05% (Table 1). The polished genome has been deposited in the NCBI genome database (JAIOKV000000000). We observed a high degree of synteny between our alignment and similar hybrid assembly of *T. latifolia* (Supplementary Figure S2), although mapping success of unrelated individuals was significantly higher in our genome (Supplementary Figure S3).

Assembly quality and annotation

Approximately 96.03% of BUSCOs were identified (3148 of 3278) in the assembly. Of these, 3025 BUSCOs were complete (92.28%), and 123 were fragmented. Of the complete BUSCOs, 2461 were single-copies, and 564 had duplicates. Repeats represented 43.84% of the genome (see breakdown in Figure 2). The chloroplast genome was detected and split on two scaffolds, while the entire mitochondrial genome appeared to be assembled (Supplementary Table S5). Total repeat content of the genome is consistent with other Poales genomes (Kawahara et al. 2013; Ming et al. 2015; Redwan et al. 2016): long interspersed nuclear elements (LINEs) comprised 1.22% of the genome, while short interspersed nuclear elements (SINEs) were not detected in the genome (Table 2). The annotation pipeline produced 27,432 genes, which coded for 34,911 proteins and 34,974 mRNA sequences. 2095 rRNA, 502 tRNA, and 214 miRNA sequences were putatively identified, which are similar to sequence counts in related plants (Supplementary Table S6). No snRNA were identified.

Comparative genomic analyses

Ortholog clustering analysis revealed 9168 gene families were shared among *T. latifolia*, *A. comosus*, *A. thaliana*, *B. distachyon*, *O. sativa*, and *S. bicolor* (Figure 2). In total, 1806 gene families were found to be unique to *T. latifolia* (Figure 2). We aligned 1900 single-copy gene clusters to conduct phylogenomic analysis (Figure 3). The phylogenomic tree generated supports a divergent *Typha* position in Poales (Figure 3), with *Typha* forming a separate clade with pineapples (*A. comosus*). We estimated the bromeliad lineage (*Typha*-*Ananas*) to be approximately 70 million years old.

Table 1 Genome assembly statistics of our four-step hybrid assembly

	SR contig assembly—step 1	DBG2OLC contig assembly—step 2	LR contig assembly—step 3	Merged polished scaffolds—step 4
Genome size (Mb)	263.74	193.16	287.63	286.77
Contigs/scaffolds	365,565	1,840	1,190	1,158
N50/L50	11.43 Kb/5,314	132.07 Kb/412	8.71 Mb/13	8.71 Mb/13
N90/L90	201 bp/193,744	52.95 Kb/1,358	58.92 Kb/530	58.94 Kb/523
Max sequence length	154.73 Kb	934.40 Kb	18.70 Mb	18.70 Mb
Scaffolds > 10 Kb	6,048	1,833	1,140	1,132
Scaffolds > 25 Kb	1,759	1,785	1,127	1,120
Scaffolds > 50 Kb	362	1,445	821	816
% of scaffolds > 50 Kb	9.46	92.34	95.54	95.59
GC content (%)	38.40	38.50	38.11	38.05

Short read (SR) contig assembly consists of contigs assembled from 100% of the Illumina reads using the ABySS assembler. The DBG2OLC assembler combined the 100% ABySS contigs and PacBio long reads. The long read (LR) contig assembly was generated from only PacBio long reads using Canu. The merged polished scaffolds are from merging the DBG2OLC assembly and LR contigs and were polished using Pilon.

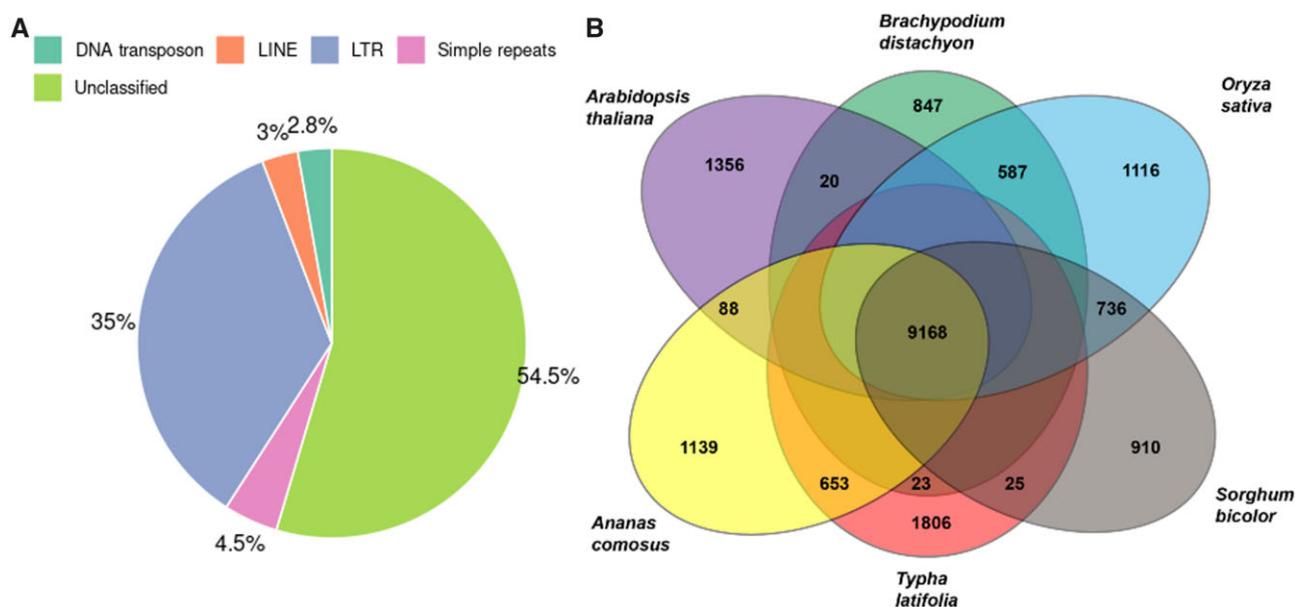


Figure 2 (A) Percentages of repeat types masked in the *T. latifolia* genome. Types of repeats include DNA transposons, long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), unclassified repeats, and simple repeats. (B) Venn diagram of orthologous gene clusters among the broadleaf cattail (*Typha latifolia*), pineapple (*Ananas comosus*), thale cress (*Arabidopsis thaliana*), stiff brome (*Brachypodium distachyon*), rice (*Oryza sativa Japonica*), and broom-corn (*Sorghum bicolor*). Only the numbers of ortholog clusters of adjacent species, those common to all species, and those unique to each species are labeled.

Table 2 Summary of repeats masked in the *Typha latifolia* genome

	Length (bp)	Percentage of genome (%)
SINE	0	0
LINE	3,499,426	1.22
LTR elements	44,172,983	15.35
DNA elements	3,735,882	1.30
Unclassified	68,736,281	23.88
Small RNA	0	0
Satellites	0	0
Simple repeats	5,689,134	1.98
Low complexity	670,871	0.23
Total	126,178,695	43.84

Discussion

The Typhaceae family (order Poales) is diverse, comprising over 50 recognized species (Christenhusz and Byng 2016). Species in this family are essential components of marshes and wetlands around the world (reviewed in Bansal et al. 2019), and play a role in both bioremediation and the production of biofuel material (Grosshans 2014; Svedarsky et al. 2016; Bonanno and Cirelli 2017). This is the first published whole-genome assembly from a member of the Typhaceae family and adds to the previously characterized chloroplast genome sequences of *T. latifolia*, *T. orientalis*, and *Sparganium stoloniferum* (Guisinger et al. 2010; Su et al. 2019; Liu et al. 2020). The assembly and annotation of the Cattail genome, which we note was a key recommendation from a diverse team of *Typha* researchers (Bansal et al. 2019), will likely be an important tool in wetland management.

We explored a variety of SR assemblers given the varying quality of outputs seen in Assemblathon metrics (Bradnam et al. 2013). Of the three SR assemblers used, ABySS produced longer, fewer contigs and scaffolds, which is consistent with previous comparisons (Bradnam et al. 2013). ABySS' de Bruijn approach did not have issues with the high coverage Illumina data (e.g., Richards and Murali 2015). We combined with SR assembly with

raw long read data using DGB2OLC (Ye et al. 2016) as previous plant genomes with similar data showed promising assembly statistics (i.e., Daccord et al. 2017; Hatakeyama et al. 2018; Zhou et al. 2019). The long reads in the DBG2OLC assembly captured more lengthy and repetitive regions, resulting in a sharp rise in contig N50 and contigs 50 Kb or longer (Table 1). However, the backbone of the final assembly was the PacBio Canu assembly (Table 1 and Supplementary Table S4), with the benefit of this strategy being the removal of chimeras with polished assemblies having very low error rates (Ye et al. 2016).

Accurate genomes are important as sequencing errors can affect both nucleotide diversity and the discovery of markers (Clark and Whittam 1992). LR sequencing further facilitates detection of structural variants (Amarasinghe et al. 2020), which might be useful in studying *Typha* dynamics given the link to hybridization and speciation (Weissensteiner et al. 2020). The final assembly consists of only 1158 scaffolds, with ~95% being 50 Kb or longer; *T. latifolia* possessed higher N50 and fewer scaffolds compared to other plant genome assemblies that used PacBio sequencing and similar hybrid assembly pipelines (Redwan et al. 2016; Hatakeyama et al. 2018; Reuscher et al. 2018). With 96.03% of BUSCOs detected and 92.28% of them complete in our assembly, this would suggest a high-quality annotation. The topology of the phylogenomic tree that we reconstructed based on whole-genome sequences of six species agrees with previously inferred evolutionary relationships (e.g., Darshetkar et al. 2019) by grouping together the two bromeliad species (*T. latifolia* and *A. comosus*) and the three graminid species (*O. sativa*, *S. bicolor*, and *B. distachyon*), and by inferring a more recent divergence date for graminids compared to bromeliads. Our estimated *Typha* lineage age of approximately 70 million years exceeds an earlier estimate of between 22.64 and 57.60 mya that was based on seven cpDNA markers (Zhou et al. 2018), although is comparable to the estimate of 69.5 Myr that was based on a combination of fossil records and cpDNA sequences (Bremer 2000). We acknowledge that the use of whole-genome sequences to infer divergence

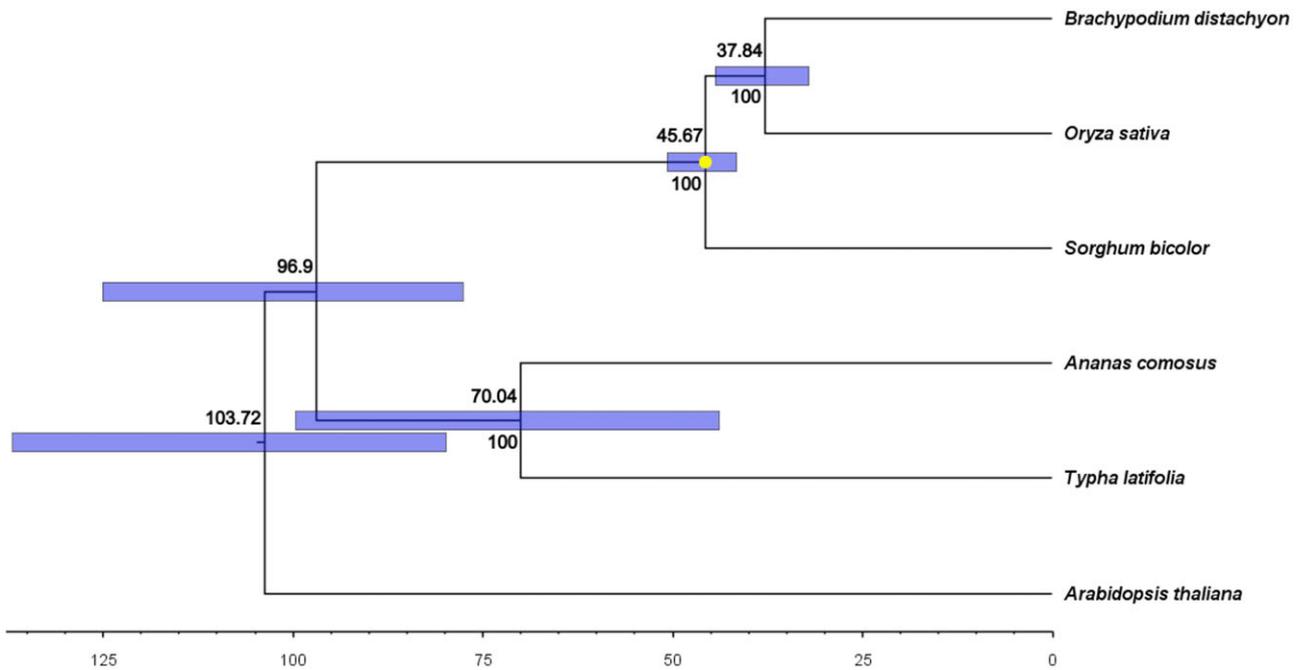


Figure 3 Phylogenetic tree with divergence times, based on the alignment of 1900 single-copy gene clusters. Ninety-five percent credible divergence times are shown as blue bars and were estimated using MCMCTree. Divergence times and bootstrap values are shown above and below the nodes, respectively. The yellow dot indicates the calibration point.

times is still in its infancy, and caution that factors such as rate heterogeneity may complicate such inferences (reviewed in Smith et al. 2018).

The *T. latifolia* reference genome now allows for studying *Typha* and hybridization at the molecular level, for example, by providing insight into the adaptation of *T. latifolia* populations that may be threatened by climate change, and facilitating research into potentially important processes such as hybrid breakdown in invasive *Typha* hybrids. In addition, genome-wide markers could provide insight into why the hybrid *Typha* × *glauca* is dominant in some areas (Kirk et al. 2011; Freeland et al. 2013) but uncommon in others (Freeland et al. 2013; Ciotir et al. 2017). For example, *T. angustifolia* in North America is thought to have arrived from Europe several centuries ago (Ciotir et al. 2013), and ancestry assessments could test the hypothesis that historical interspecific hybridization led to genetic introgression into some *T. angustifolia* populations, which may help to explain regional erosion of species barriers. Locus-specific ancestry estimation analyses, markers, and the identification of potentially adaptive genes can all now be used to find evidence of ancient hybridization in *T. angustifolia* (Goulet et al. 2017; Taylor and Larson 2019). This high-quality draft genome and its comparisons with Poales species will be an indispensable resource for ongoing research into *Typha*, a genus that both sustains and threatens wetlands around the world.

Data availability

Raw sequencing data can be found in the SR Archive (SRA) under accession number PRJNA751759. Genome assembly has been submitted to GenBank (JAIOKV000000000). Code used to generate the data can be found at https://gitlab.com/WiDGeT_TrentU/undergrad-theses/-/tree/master/Widanagama_2021. The *T. angustifolia* RNSseq data are in the SRA under SRR15541138. The *T.*

latifolia transcriptome used can be found at https://figshare.com/articles/dataset/Typha_latifolia_leaf_transcriptome/5661727/1.

Supplementary material is available at G3 online.

Acknowledgments

The authors thank Vikram Bhargav for sampling and preparing the *T. latifolia* leaves, and Eric Wootton for extracting DNA from the sample.

Funding

The CanSeq 150 project funded the PacBio sequencing; NSERC Discovery grants to J.R.F. and A.B.A.S. and two NSERC Undergraduate Student Research Awards to S.D.W. supported this research. A Compute Canada Resources for Research Groups award to A.B.A.S. and Compute Canada support staff supported the bioinformatics.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

- Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, et al. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21:1–16.
- Andrews NJ, Pratt DC. 1978. Energy potential of cattails (*Typha* spp.) and productivity in managed stands. *J Minn Acad Sci.* 44:5–8.
- Bansal S, Lishawa SC, Newman S, Tangen BA, Wilcox D, et al. 2019. *Typha* (Cattail) invasion in North American wetlands: biology, regional problems, impacts, ecosystem services, and management. *Wetlands.* 39:645–684.

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
- Bonanno G, Cirelli GL. 2017. Comparative analysis of element concentrations and translocation in three wetland congener plants: *Typha domingensis*, *Typha latifolia* and *Typha angustifolia*. *Ecotoxicol Environ Saf*. 143:92–101.
- Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, et al. 2013. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *Gigascience*. 2:2047–217X.
- Bremer K. 2000. Early cretaceous lineages of monocot flowering plants. *Proc Natl Acad Sci USA*. 97:4707–4711.
- Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 3:lqaa108.
- Bunbury-Blanchette A, Freeland JR, Dorken M. 2015. Hybrid *Typha × glauca* outperforms native *T. latifolia* under contrasting water depths in a common garden. *Basic Appl Ecol*. 16:394–402.
- Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016. Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res*. 44:e147.
- Christenhusz MJ, Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa*. 261:201–217.
- Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics*. 30:31–37.
- Ciotir C, Kirk H, Row JR, Freeland JR. 2013. Intercontinental dispersal of *Typha angustifolia* and *T. latifolia* between Europe and North America has implications for *Typha* invasions. *Biol Invasions*. 15:1377–1390.
- Ciotir C, Szabo J, Freeland JR. 2017. Genetic characterization of cattail species and hybrids (*Typha* spp.) in Europe. *Aquat Bot*. 141:51–59.
- Clark AG, Whittam TS. 1992. Sequencing errors and molecular evolutionary analysis. *Mol Biol Evol*. 9:744–752.
- Daccord N, Celton JM, Linsmith G, Becker C, Choisne N, et al. 2017. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet*. 49:1099–1106.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 9:772.
- Darshetkar AM, Datar MN, Tamhankar S, Li P, Choudhary RK. 2019. Understanding evolution in Poales: insights from Eriocaulaceae plastome. *PLoS One*. 14:e0221423.
- Farrer EC, Goldberg DE. 2009. Litter drives ecosystem and plant community changes in cattail invasion. *Ecol Appl*. 19:398–412.
- Freeland J, Ciotir C, Kirk H. 2013. Regional differences in the abundance of native, introduced, and hybrid *Typha* spp. in northeastern North America influence wetland invasions. *Biol Invasions*. 15:2651–2665.
- Geddes P, Grancharova T, Kelly JJ, Treering D, Tuchman NC. 2014. Effects of invasive *Typha × glauca* on wetland nutrient pools, denitrification, and bacterial communities are influenced by time since invasion. *Aquat Ecol*. 48:247–258.
- Goulet BE, Roda F, Hopkins R. 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiol*. 173:65–78.
- Grace JB, Harrison JS. 1986. The biology of Canadian weeds.: 73. *Typha latifolia* L., *Typha angustifolia* L. and *Typha x glauca* Godr. *Can J Plant Sci*. 66:361–379.
- Grosshans R. 2014. Cattail (*Typha* spp.) biomass harvesting for nutrient capture and sustainable bioenergy for integrated watershed management. PhD Dissertation, University of Manitoba, p. 274.
- Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK. 2010. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *J Mol Evol*. 70:149–166.
- Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Jr, et al. 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 31:5654–5666.
- Hatakeyama M, Aluri S, Balachadran MT, Sivarajan SR, Patrignani A, et al. 2018. Multiple hybrid *de novo* genome assembly of finger millet, an orphan allotetraploid crop. *DNA Res*. 25:39–47.
- Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, et al. 2021. Ensembl 2021. *Nucleic Acids Res*. 49:D884–D891.
- Humann JL, Lee T, Ficklin S, Main D. 2019. Structural and functional annotation of eukaryotic genomes with GenSAS. In: *Gene Prediction*. New York, NY: Humana. p. 29–51.
- Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, et al. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res*. 27:768–777.
- Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, et al. 2014. Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res*. 24:1384–1395.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, et al. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*. 6:1–10.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 37:907–915.
- Kirk H, Connolly C, Freeland JR. 2011. Molecular genetic data reveal hybridization between *Typha angustifolia* and *Typha latifolia* across a broad spatial scale in eastern North America. *Aquat Bot*. 95:189–193.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 27:722–736.
- Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 34:1812–1819.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol*. 5:R12.
- Larkin DJ, Freyman MJ, Lishawa SC, Geddes P, Tuchman NC. 2012. Mechanisms of dominance by the invasive hybrid cattail *Typha × glauca*. *Biol Invasions*. 14:65–77.
- Lawrence BA, Bourke K, Lishawa SC, Tuchman NC. 2016. *Typha* invasion associated with reduced aquatic macroinvertebrate abundance in northern Lake Huron coastal wetlands. *J Great Lakes Res*. 42:1412–1419.
- Lawrence BA, Lishawa SC, Hurst N, Castillo BT, Tuchman NC. 2017. Wetland invasion by *Typha × glauca* increases soil methane emissions. *Aquat Bot*. 137:80–87.
- Lishawa SC, Jankowski K, Geddes P, Larkin DJ, Monks AM, et al. 2014. Denitrification in a Laurentian Great Lakes coastal wetland invaded by hybrid cattail (*Typha × glauca*). *Aquat Sci*. 76:483–495.
- Liu ZD, Zhou XL, Ma HY, Tian YQ, Shen SK. 2020. Characterization of the complete chloroplast genome sequence of wetland

- macrophyte *Typha orientalis* (Typhaceae). Mitochondrial DNA B Resour. 5:136–137.
- Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 42:e119.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964. doi:10.1093/nar/25.5.955.
- Luo R, Liu B, Xie Y, Li Z, Huang W, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience. 1:18.
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, et al. 2018. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 14:e1005944.
- Miller RL, Fujii R. 2010. Plant community, primary productivity, and environmental conditions following wetland re-establishment in the Sacramento-San Joaquin Delta, California. Wetlands Ecol Manage. 18:1–16.
- Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, et al. 2015. The pineapple genome and the evolution of CAM photosynthesis. Nat Genet. 47:1435–1442.
- Moscou M. 2017. *Typha latifolia* Leaf Transcriptome (Version 1). doi: 10.6084/m9.figshare.5661727.v1.
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 29:2933–2935.
- Pieper S, Dorken M, Freeland J. 2020. Genetic structure in hybrids and progenitors provides insight into processes underlying an invasive cattail (*Typha × glauca*) hybrid zone. Heredity (Edinb). 124: 714–725.
- Pieper SJ, Nicholls AA, Freeland JR, Dorken ME. 2017. Asymmetric hybridization in cattails (*Typha* spp.) and its implications for the evolutionary maintenance of native *Typha latifolia*. J Hered. 108: 479–487.
- Redwan RM, Saidin A, Kumar SV. 2016. The draft genome of MD-2 pineapple using hybrid error correction of long reads. DNA Res. 23:427–439.
- Reuscher S, Furuta T, Bessho-Uehara K, Cosi M, Jena KK, et al. 2018. Assembling the genome of the African wild rice *Oryza longistaminata* by exploiting synteny in closely related *Oryza* species. Commun Biol. 1:1–10.
- Richards S, Murali SC. 2015. Best practices in insect genome sequencing: what works and what doesn't. Curr Opin Insect Sci. 7: 1–7. doi:10.1016/j.cois.2015.02.013.
- 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.
- Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 11:e0163962.
- Smit AFA, Hubley R. 2008–2015. RepeatModeler (Open-1.0). Institute for Systems Biology. <http://www.repeatmasker.org>.
- Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker (Open-4.0). Institute for Systems Biology. <http://www.repeatmasker.org>.
- Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS One. 13:e0197433.
- Smith SG. 1987. *Typha*: its taxonomy and the ecological significance of hybrids. Archiv Hydrobiol. 27:129–138.
- Snow AA, Travis SE, Wildová R, Fér T, Sweeney PM, et al. 2010. Species-specific SSR alleles for studies of hybrid cattails (*Typha latifolia* × *T. angustifolia*; Typhaceae) in North America. Am J Bot. 97:2061–2067.
- Solares EA, Chakraborty M, Miller DE, Kalsow S, Hall K, et al. 2018. Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. G3 (Bethesda). 8:3143–3154.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30: 1312–1313.
- Su T, Yang JX, Lin YG, Kang N, Hu GX. 2019. Characterization of the complete chloroplast genome of *Sparganium stoloniferum* (Poales: typhaceae) and phylogenetic analysis. Mitochondrial DNA B Resour. 4:1402–1403.
- Svedarsky D, Bruggman J, Ellis-Felege S, Grosshans R, Lane V, et al. 2016. Cattail Management in the Northern Great Plains: Implications for Wetland Wildlife and Bioenergy Harvest. University of Minnesota, Crookston, pp. 56.
- Tangen BA, Bansal S, Freeland JR, Travis SE, Wasko JD, et al. 2021. Distributions of native and invasive *Typha* (cattail) throughout the Prairie Pothole Region of North America. Wetlands Ecol Manag. 1–17. <https://link.springer.com/article/10.1007/s11273-021-09823-7#citeas>.
- Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat Ecol Evol. 3:170–177.
- Tisshaw K, Freeland J, Dorken M. 2020. Salinity, not genetic incompatibilities, limits the establishment of the invasive hybrid cattail *Typha × glauca* in coastal wetlands. Ecol Evol. 10:12091–12103.
- Tuchman NC, Larkin DJ, Geddes P, Wildova R, Jankowski K, et al. 2009. Patterns of environmental change associated with *Typha x glauca* invasion in a Great Lakes Coastal Wetland. Wetlands. 29: 964–975.
- Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 43: W78–W84.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963.
- Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 41:D358–D365.
- Weissensteiner MH, Bunikis I, Catalán A, Francoijs KJ, Knief U, et al. 2020. Discovery and population genomics of structural variation in a songbird genus. Nat Commun. 11:1–11.
- Xu L, Dong Z, Fang L, Luo Y, Wei Z, et al. 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 47: W52–W58. doi:10.1093/nar/gkz333.
- Xu Z, Feng Z, Yang J, Zheng J, Zhang F. 2013. Nowhere to invade: *Rumex crispus* and *Typha latifolia* projected to disappear under future climate scenarios. PLoS One. 8:e70728.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
- Ye C, Hill CM, Wu S, Ruan J, Ma ZS. 2016. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci Rep. 6:31900.
- Yeo RR. 1964. Life history of common cattail. Weeds. 12:284–288.

- Zapfe L, Freeland JR. 2015. Heterosis in invasive F1 cattail hybrids (*Typha* × *glauca*). *Aquat Bot.* 125:44–47.
- Zedler JB, Kercher S. 2004. Causes and consequences of invasive plants in wetlands: opportunities, opportunists, and outcomes. *Crit Rev Plant Sci.* 23:431–452.
- Zhou B, Tu T, Kong F, Wen J, Xu X. 2018. Revised phylogeny and historical biogeography of the cosmopolitan aquatic plant genus *Typha* (Typhaceae). *Sci Reports.* 8:1–7.
- Zhou B, Yu D, Ding Z, Xu X. 2016. Comparison of genetic diversity in four *Typha* species (Poales, Typhaceae) from China. *Hydrobiologia.* 770:117–128.
- Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, et al. 2019. The population genetics of structural variants in grapevine domestication. *Nat Plants.* 5:965–979.

Communicating editor: P. Morrell