Sequencing and characterization of the *Del/Tekay* Chromovirus family in *Marshallia obovata* (Asteraceae)

Nathan D. Hall* and Leslie R. Goertzen

Department of Biological Sciences, Auburn University, Auburn, AL 36849, U.S.A.
* Correspondence: ndh0004@auburn.edu

Abstract

Low coverage DNA sequencing data coupled with improved assembly methods are providing new insight into the repetitive fraction of the genomes of non-model organisms. In this study we explore the potential of these new approaches to characterize the *Del/Tekay* elements (chromodomain-containing *Ty3/Gypsy* retrotransposons) of the North American wildflower species *Marshallia obovata*. The *Marshallia Del/Tekay* chromoviruses are 8.3 Kbp in length with direct repeats of 1675 bp and 14 conserved domains across a 4506 bp gag-pol region. Read depth comparisons suggest that *Del/Tekay* elements are not highly abundant in *M. obovata* with approximately 144 unique copies per haploid genome of 6.5 Gbp (<0.02 %). Sequence variation was observed at less than 1 % of nucleotide positions overall with 60 % of variants occurring outside of any coding region. These results illustrate the utility of low coverage sequencing data for fine scale analyses of transposable elements present at low to moderate copy numbers within the genomes of non-model organisms. The *Del/Tekay* element sequences described here are the first for the plant family Asteraceae and are the first transposable elements of any kind described for the genus *Marshallia*. These initial results provide for future work on the evolutionary dynamics of *Del/Tekay* elements more broadly in the genus *Marshallia* and Asteraceae.

Introduction

Transposable elements (TEs) are a large and dynamic component of many plant genomes (Bennetzen 2002; Feschotte *et al.* 2002; Vitte & Panaud 2005; Wei *et al.* 2009). In essence they are selfish genetic elements that proliferate through copy or cut and paste mechanisms, potentially increasing genome sizes over several orders of magnitude (Bennetzen & Wang 2014; Michael 2014). Their activity in genomes can facilitate adaptive evolution in numerous ways (Kidwell & Lisch 2001; Oliver *et al.* 2013), but they are also disruptive and trigger silencing countermeasures from the host organism such as epigenetic surveillance or outright elimination (Lee & Kim 2014; Michael 2014). The consequent rise and fall of distinct TE families are notable evolutionary phenomena in their own right and can also inform host phylogeny. Transposable element family abundance, locus-specific insertions and horizontal transfer events can distinguish clades of host species and be used to infer evolutionary relationships (Shedlock *et al.* 2004; Schaack *et al.* 2010; Hertweck 2013; Piednoël *et al.* 2013; Dodsworth *et al.* 2015) but have been underutilized in general as detailed repetitive element characterizations are uncommon outside of well studied models or other species with abundant genomic data resources. Recent developments in both sequencing technology and short-read assembly and analysis methods are improving our understanding of the diversity and abundance of TEs in non-model organisms and the analyses of these data in an evolutionary framework (Novák *et al.* 2010; Muñoz-Diez *et al.* 2012; Hertweck 2013).

In this study we explore the potential of low coverage genomic sequencing to provide detailed characterization of a distinct clade of large (>8 kbp), low-abundance transposable elements within a non-model organism. We leverage paired-read information in assembly and mapping to accurately determine long terminal repeat (LTR) and flanking region sequences, clearly circumscribe the TE

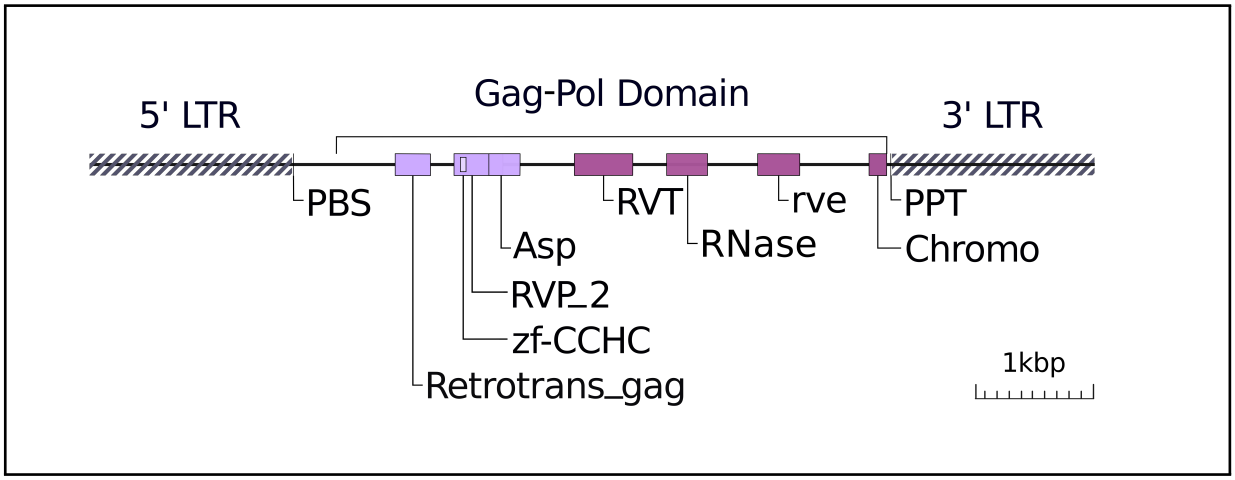

Figure 1. Schematic of *Marshallia obovata Del/Tekay* element with sequence-based annotation of characteristic retroelement features: 5′ LTR (1–1675), primer binding site (PBS, 1681–1689), Retrotransposon gag protein (Retrotrans_gag, 2522–2821), zinc finger (zf-CCHC, 3065–3113), retroviral aspartyl protease (RVP_2, 3041–3625), aspartate protease (or pepsin-like aspartate proteases, Asp_protease_2, 3293–3559), reverse transcriptase (RVT_1, 4007–4486), Ribonuclease H (RNase_HI_RT_Ty3, 4766–5110), integrase core domain (rve, 5522–5878), chromodomain (Chromo, 6437–6586), polypurine tract (PPT, 6615–6631), and 3′ LTR (6634–8308)

lineage and estimate its variability within the host genome. Structural annotation and phylogenetic analyses of the validated consensus sequence provide a starting point for additional molecular evolutionary focused studies of the element group in other individuals and related species.

Materials and Methods

Study System

This study focused on *Marshallia obovata* (Walt.) Beadle & F.E. Boynton (Asteraceae), a perennial wildflower species native to the southeastern United States. *M. obovata* is diploid ($2n = 2x = 18$) with a genome size of approximately 6.6 Gb (1 C = 6.79 pg) determined by flow cell cytometry (personal communication, T. Garnatje, Institut Botànic de Barcelona). DNA was extracted from fresh leaf material of a wild-collected plant from Macon County, Alabama (*Hansen 4956*, AUA) using the modified CTAB protocol of Doyle & Doyle (1987). Samples were submitted to the Genomic Services Lab, HudsonAlpha Institute for Biotechnology (Huntsville, AL) where paired-end libraries with a mean insert size of 281 bp were prepared and sequencing of 39,124,284 100 bp reads was performed on an Illumina Hi-Seq 2000 platform.

Assembly, Validation and Annotation

Contig assembly was performed with Ray (Boisvert *et al.* 2010, v2.3.1) using kmer length 31 and minimum 30 bp reads filtered using Sickle (Joshi & Fass 2011, v1.210) to exclude bases with a Sanger quality score less than 30. Putative TE containing contigs were identified using RepeatMasker (Smit *et al.* 2010, v3.2.7) and iteratively extended in the assembler PriceTI (Ruby *et al.* 2013, v1.0.1) with stringent (99 %) matching. The LTRharvest (Ellinghaus *et al.* 2008) and LTRdigest (Steinbiss *et al.* 2009) modules of Genome Tools (Gremme *et al.* 2013, v1.5.3) were used to identify intact Class I retrotransposons with long terminal repeats (LTRs) and complete gag-pol regions. A candidate transposon sequence was selected for further analysis based on its large size combined with moderate read coverage to explore the upper limits of repetitive element assembly and annotation in this genome skimming dataset.

The LTR regions of the partial transposon contig were extended with local assembly of read pairs until the 5′ sequence of LTR1 matched the gag-pol-LTR2 boundary and the 3′ end of LTR2 matched the LTR1-gag-pol boundary. The original genomic reads were then re-mapped to a consensus sequence in bowtie 2 (Langmead & Salzberg 2012, v.2.2.4), allowing only correctly paired reads and end-to-end alignment enforced to validate the assembly overall, determine a consensus sequence and extract variant calls. Annotation of element components was accomplished with Genome tools modules LTRdigest and sketch, sequence comparisons performed with BLAST+ (NCBI Resource Coordinators 2014) and Artemis (Rutherford *et al.* 2000, v16.0.0).

Element Classification

Sequence features of the validated consensus including LTR length and sequence domain presence/order were compared to known TEs and found to be consistent with the chromoviral branch of the *Ty3/Gypsy* LTR retroelement family. Phylogenetic analyses were also conducted for a fine-scale classification of the *Marshallia* retroelements. Representative sequences were obtained from datasets referencing chromoviruses (Neumann *et al.* 2011; Llorens *et al.* 2011; Kolano *et al.* 2013; Domingues *et al.* 2012) and

from sequence similarity searches of NCBI databases using the *Marshallia* element gag-pol region as a query.

Amino acid sequence alignment for a portion of the reverse transcriptase domain was performed in ClustalW (Thompson *et al.* 1994, v2.1) with manual adjustments where necessary. Substitution model choice and maximum likelihood analyses including 300 bootstrap replicates were carried out with RAxML (Stamatakis 2014, v8.1.5). Phylogenetic trees were visualized with FigTree (Rambaut 2010, v1.3.1).

Element Circumscription, Variation and Copy Number

A less stringent alignment of genomic reads, allowing up to 39 nucleotide differences or indels relative to the reference, was created with bowtie 2 with relaxed mapping (the minimum default score for mapping adjusted with the formula minimum score = -0.6 + -2x(length of read), with forced read pairing and no match bonus) to distinguish reads belonging to this retroelement group from more distantly related TEs. Reads were binned by the observed number of nucleotide or indel differences and plotted in R (R Core Team 2013, v.3.0.2). A high-stringency map (forced read pairing, no match bonus in default smith-waterman alignment) was produced with bowtie 2 to assess intragenomic variation. The resulting alignment was converted into pileup format with SAMtools (Li *et al.* 2009, v1.2). A custom R script (courtesy of Eric Archer, NOAA Fisheries) was used to extract nucleotide frequencies at every position, including only those single nucleotide polymorphism (SNP) variants above 5 % and supported by a minimum of 5 reads. Insertion or deletion (indel) variants coded as CIGAR strings in the SAM file produced by mapping were characterized and counted in Tablet (Milne *et al.* 2013, v1.14.10.21) using the default feature threshold of 10 CIGAR strings per site.

Average reads per kilobase (RPK) for single copy nuclear genes was calculated by local mapping (no read pairing, match bonus default value, default smith-waterman alignment) genomic reads to 897 unique COSII markers (Wu *et al.* 2006) identified and extracted from an unpublished *Marshallia obovata* transcriptome. The mean of all COSII RPKs was compared to the RPKs obtained from a high stringency mapping and a local mapping of the genomic reads to the element consensus. Number of copies was estimated by dividing the *Del/Tekay* RPK by the average COSII RPK value.

Results

The fully assembled *Marshallia obovata Del/Tekay* chromovirus consensus sequence (NCBI accession KX396599) is 8308 bp in length with a 4506 bp gag-pol region flanked by direct repeats of 1675 bp (Fig. 1). Annotation yielded 14 conserved regions or domains including a predicted 18 bp primer binding site (PBS) for tRNA(iMet) initiated replication. The gag-pol region is a single, intact open reading frame containing the expected catalytic domains providing for the retrotransposon gag protein, zinc finger, retroviral aspartyl protease, aspartate protease, reverse transcriptase, Ribonuclease H, integrase core domain and a chromodomain. A polypurine tract and uracil-rich, putative U-box were detected between the end of the gag-pol region and the start of LTR2. Additional features identified included universal minicircle sequence binding protein (UMSBP, 2996–3163), a second Zinc knuckle (zf-CCHC, 3116–3167), and Arginine methyltransfersase-interacting protein (AIR1, 3065–3175).

Table 1. Number and location of nucleotide variants in the *Marshallia obovata Del/Tekay* element family. NonORF indicates the noncoding region between the 3′ end of the LTR1 and the start of the gag-pol region. ORF indicates SNPs that occurred inside the gag-pol region but outside of conserved domains.

Region	No.	Sub-Region	No.	Syn.	N.Syn.
Non-coding	29				
		LTR	26		
		NonORF	3		
Coding	21				
		ORF	13	5	8
		Chromo	1	0	1
		RNase_H	2	2	0
		RVP_2	2	1	1
		RVT_1	1	1	0
		UMSBP	2	1	1
Total	50	All	50	10	11

The length of the terminal repeat falls within reported values (1.1–4.4kbp) for the *Del/Tekay* group of chromoviruses (Llorens *et al.* 2011), and the linear order of domains further support *Del/Tekay* assignment (Weber *et al.* 2013). Phylogenetic analyses of the reverse transcriptase domain also suggest the *Marshallia obovata* elements are members of the *Del/Tekay Ty3/Gypsy* chromoviruses and most closely related to sequences from other asterids (Fig. 2).

Low-stringency mapping identified a total of 14,946 reads that matched the *Del/Tekay* consensus sequence at a minimum of 60 nucleotides out of 100 (Fig. 3). The majority of these reads, 13,567 or 91 %, had six or fewer nucleotide differences from the consensus. Approximately 600 reads, not included in the consensus sequence but matching at the 70–85 % level were identified and localized to conserved LTR motifs and reverse transcriptase or ribonuclease H domains of other TE families.

The average read depth in the more stringent map produced for variant analysis was 85.4 reads per nucleotide with a maximum coverage of 157 reads per nucleotide. Within this map a total of 50 nucleotide polymorphisms were detected, 29 in non-coding sequence including the LTRs and 21 in the coding regions (Table 1). SNP frequencies ranged from 0.05–0.50 of total reads with roughly half of the SNP frequencies under 0.1 and nearly half above 0.25. Within the coding region, there were 13 polymorphisms in the ORF at large and 8 in the conserved domains. Of the total 11 non-synonymous variants, 8 were identified outside the conserved domains and 1 each in the Chromo, RVP2 and UMSBP domains, with frequencies

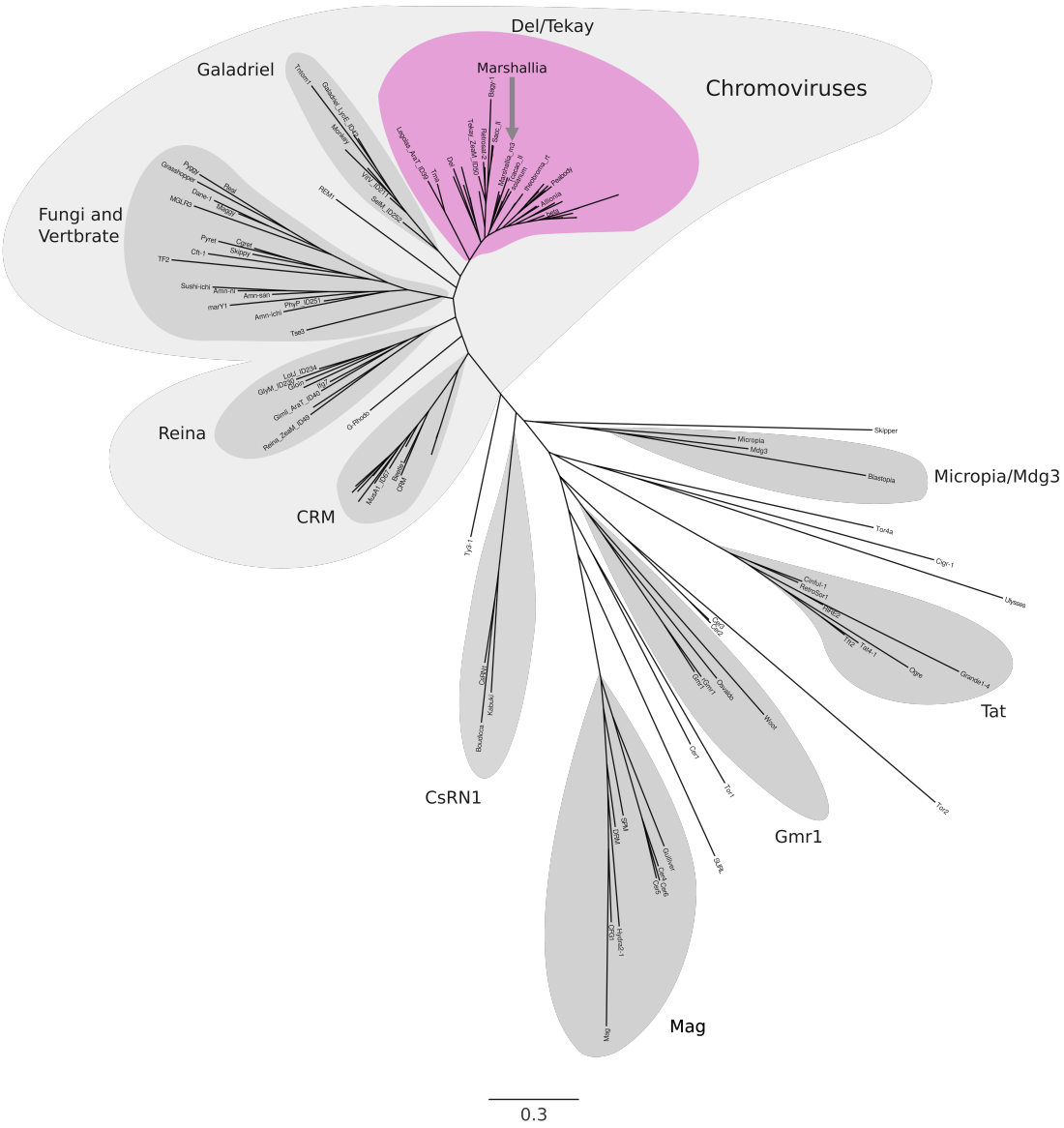

Figure 2. Maximum likelihood tree from analysis of RVT domain in *Ty3/Gypsy* elements with emphasis on chromoviruses.

of 0.062, 0.0496 and 0.440 respectively. No variants that resulted in a stop codon were detected. A total of 21 indel variants were identified, 19 in the non-coding portion of the element, 1 within the ORF at large and one within Retrotrans_gag, all consisting of slight variations in the length of mono or dinucleotide repeats.

In the maps prepared for copy number assessment, the coverage depth across the entire *Del/Tekay* element had a mean of 1633 reads per kilobase (RPK) under high-stringency conditions and 2667 RPK with local alignment allowing unpaired reads. The average read depth for 897 COSII single-copy nuclear genes (local alignment, unpaired reads allowed) was 18.55 RPK giving a *Del/Tekay* copy number estimate of 2667 RPK/18.55 RPK = 144 copies.

Discussion

Several studies have used low coverage sequencing and consensus methods to assemble and characterize the transposable element fraction of plant genomes at at the level of family and superfamily (Novák *et al.* 2010; Hertweck 2013; Staton & Burke 2015). Here we demonstrate an extension of this approach that allows for a more detailed assessment

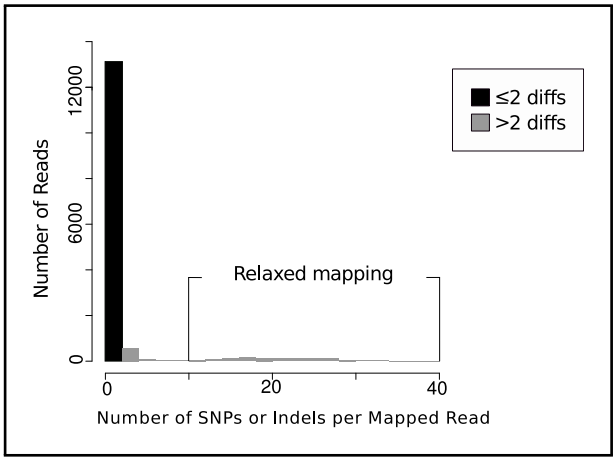

Figure 3. Number of SNP/indel variants (diffs) in each read that mapped to the reference *Marshallia obovata Del/Tekay* element with relaxed mapping parameters.

of TE composition including full-length LTR sequences and estimates of intragenomic variation and abundance.

The 8.3 kbp repetitive element described here corresponds to a uniform and clearly distinct set of LTR retrotransposons within the *Marshallia obovata* genome. Domain annotation and phylogenetic analyses identify them as members of the *Del/Tekay* lineage of *Ty3/Gypsy* chromoviruses. Although the *Ty3/Gypsy* elements are suggested to be particularly active in Asteraceae (Renaut *et al.* 2014; Staton & Burke 2015), this is the first specific report of a *Del/Tekay* retrotransposon in the family and the first transposable element of any kind described for the genus *Marshallia*.

Sequence variation for the *Del/Tekay* element family was relatively low, with slightly less than 1 % of alignment positions containing a variant nucleotide above a 0.05 frequency threshold. The majority of those variants (29/50, 59 %) occur in non-coding regions, 10 of the remaining 21 are synonymous substitutions, and 8 of the 11 non-synonymous substitutions occur outside of conserved domains. Nearly half of the SNPs occur at low-frequency (<0.1) and would be consistent with more recent proliferation of active elements. *Ty3/Gypsy* elements appear to be actively expanding in Asteraceae genomes (Staton & Burke 2015) and the *Del/Tekay* family in *Marshallia* may be experiencing similar growth. Several presumably older variants occur at a significantly higher frequency with 23 above 0.25 and 9 of those above 0.45. These markers may be useful for tracking diversification events in the *Del/Tekay* element family at the population level in *Marshallia obovata* or further back in time above the species level. Of 21 verified indel variants, 19 occur in non-coding positions and 2 in the gag-pol region. Our variant calling process used a cutoff of 0.05 and is therefore insensitive to any mutations (including missense) at individual loci but the general pattern of most variation occurring outside the ORF region would not likely change with less stringent variant calls. Additional resolution might also be obtained with a refined variant calling method such as indel realignment.

Del/Tekay elements exhibit a wide range in copy number across plant genomes from 46 distinct insertions in *Arabidopsis thaliana* (Du *et al.* 2010) to over 10,000 in *Pisum sativum* (Neumann *et al.* 2011). The comparison of mean COSII to *Del/Tekay* RPK values suggests a copy number in *Marshallia obovata* of around 144 distinct loci, assuming all insertions have been diploidized. If all occurrences are unique on individual chromosomes, the coverage depth suggests 288 distinct retrotranspositions. Given the 13 Gbp diploid genome this represents less than 0.02 % of the nuclear genome and would be considered a low abundance retroelement, although copy number may be underestimated here since *Del/Tekay* transposons are known to target regions of heterochromatin (Mlinarec *et al.* 2016) which are likely under sampled in Illumina sequencing experiments (van Dijk *et al.* 2014).

Our *Marshallia Del/Tekay* assembly illustrates how low coverage genomic sequencing data can be used for fine-scale analyses of TEs. This could be particularly valuable in non-model organisms that often contain new classes of TE and where *de novo* assembly approaches are preferable to targeted techniques that rely on known element families. The assembly and characterization of the *Marshallia Del/Tekay* element family here is the first of its kind for the Asteraceae, a large and important family of plants and for the genus *Marshallia*, an interesting group of wildflowers. This assembly also illustrates a potentially underutilized aspect of increasingly available genomic data and paves the way for broader and more detailed investigations of retrotransposon evolutionary dynamics.

Acknowledgments

We gratefully acknowledge funding from the U.S. Fish and Wildlife Service (Award F12AP00728 to LRG) and an Auburn University Cellular Molecular Biology Peak of Excellence Award to NDH.

References

Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica*, **115**, 29–36.

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual Review of Plant Biology*, **65**, 505–530.

Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *Journal of Computational Biology*, **17**, 1519–1533.

van Dijk EL, Jaszczyszyn Y, Thermes C (2014) Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research*, **322**, 12–20.

Dodsworth S, Chase MW, Kelly LJ, *et al.* (2015) Genomic repeat abundances contain phylogenetic signal. *Systematic Biology*, **64**, 112–126.

Domingues DS, Cruz GMQ, Metcalfe CJ, *et al.* (2012) Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics*, **13**, 137.

Doyle JJ, Doyle JL (1987) CTAB DNA extraction in plants. *Phytochemical Bulletin*, **19**, 11–15.

Du J, Tian Z, Hans CS, *et al.* (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *The Plant Journal*, **63**, 584–598.

Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, **9**, 18.

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics*, **3**, 329–341.

Gremme G, Steinbiss S, Kurtz S (2013) GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **10**, 645–656.

Hertweck KL (2013) Assembly and comparative analysis of transposable elements from low coverage genomic sequence data in Asparagales. *Genome*, **56**, 487–494.

Joshi NA, Fass JN (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.210). Available at https://github.com/najoshi/sickle.

Kidwell MG, Lisch DR (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution*, **55**, 1–24.

Kolano B, Bednara E, Weiss-Schneeweiss H (2013) Isolation and characterization of reverse transcriptase fragments of LTR retrotransposons from the genome of *Chenopodium quinoa* (Amaranthaceae). *Plant Cell Reports*, **32**, 1575–1588.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nature Methods*, **9**, 357–359.

Lee SI, Kim NS (2014) Transposable elements and genome size variations in plants. *Genomics and Informatics*, **12**, 87–97.

Li H, Handsaker B, Wysoker A, *et al.* (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Llorens C, Futami R, Covelli L, *et al.* (2011) The gypsy database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Research*, **39**, D70–4.

Michael TP (2014) Plant genome size variation: bloating and purging DNA. *Briefings in Functional Genomics*, **13**, 308–317.

Milne I, Stephen G, Bayer M, *et al.* (2013) Using tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, **14**, 193–202.

Mlinarec J, Franjević D, Harapin J, Besendorfer V (2016) The impact of the *Tekay* chromoviral elements on genome organisation and evolution of *Anemone* s.l. (Ranunculaceae). *Plant Biology*, **18**, 332–347.

Muñoz-Diez C, Vitte C, Ross-Ibarra J, Gaut BS, Tenaillon MI (2012) Using nextgen sequencing to investigate genome size variation and transposable element content. In *Plant Transposable Elements* (edited by MA Grandbastien, JM Casacuberta), vol. 24 of *Topics in Current Genetics*, pp. 41–58. Springer Berlin Heidelberg.

NCBI Resource Coordinators (2014) Database resources of the national center for biotechnology information. *Nucleic Acids Research*, **42**, D7–17.

Neumann P, Navrátilová A, Koblížková A, *et al.* (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA*, **2**, 1.

Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics*, **11**, 378.

Oliver KR, McComb JA, Greene WK (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biology and Evolution*, **5**, 1886–1901.

Piednoël M, Carrete-Vega G, Renner SS (2013) Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. *The Plant Journal*, **75**, 699–709.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A (2010) FigTree. Ver.1.3.1. Institute of Evolutionary Biology, University of Edinburgh.

Renaut S, Rowe HC, Ungerer MC, Rieseberg LH (2014) Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **369**, 20130345.

Ruby JG, Bellare P, Derisi JL (2013) PRICE: software for the targeted assembly of components of (meta) genomic sequence data. *G3*, **3**, 865–880.

Rutherford K, Parkhill J, Crook J, *et al.* (2000) Artemis: sequence visualization and annotation. *Bioinformatics*, **16**, 944–945.

Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends in Ecology and Evolution*, **25**, 537–546.

Shedlock AM, Takahashi K, Okada N (2004) SINEs of speciation: tracking lineages with retroposons. *Trends in Ecology Evolution*, **19**, 545–553.

Smit AFA, Hubley R, Green P (2010) 2010 RepeatMasker open-3.0. URL: http://www.repeatmasker.org.

Stamatakis A (2014) Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Staton SE, Burke JM (2015) Evolutionary transitions in the Asteraceae coincide with marked shifts in transposable element abundance. *BMC Genomics*, **16**, 623.

Steinbiss S, Willhoeft U, Gremme G, Kurtz S (2009) Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Research*, **37**, 7002–7013.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetetic and Genome Research*, **110**, 91–107.

Weber B, Heitkam T, Holtgräwe D, *et al.* (2013) Highly diverse chromoviruses of beta vulgaris are classified by chromodomains and chromosomal integration. *Mobile DNA*, **4**, 8.

Wei F, Stein JC, Liang C, *et al.* (2009) Detailed analysis of a contiguous 22-mb region of the maize genome. *PLoS Genetics*, **5**, e1000728.

Wu F, Mueller LA, Crouzillat D, Pétiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics*, **174**, 1407–1420.

Citation: Hall ND and Goertzen LR (2016) Sequencing and characterization of the *Del/Tekay* chromovirus family in *Marshallia obovata* (Asteraceae). *Paysonia* **5**, 1–7.

Keywords: *Del/Tekay*, repetitive DNA, *Marshallia*, transposable element, Asteraceae, retrotransposon, chromovirus.

Peer Review: Kate Hertweck, University of Texas at Tyler and Zdenek Kubat, Laboratory of Plant Developmental Genetics, Institute of Biophysics ASCR, Czech Republic.

ISSN: 2378-0665