American Journal of Botany

American Journal of Botany 99(2): 397–406. 2012.

Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae)[1]

Michael R. McKain[2], Norman Wickett[3,7], Yeting Zhang[3], Saravanaraj Ayyampalayam[2], W. Richard McCombie[4], Mark W. Chase[5], J. Chris Pires[6], Claude W. dePamphilis[3], and Jim Leebens-Mack[2,8]

[2]Department of Plant Biology, University of Georgia, Athens, Georgia 30602 USA; [3]Department of Biology, Intercollege Graduate Degree Program in Plant Biology, Institute of Molecular Evolutionary Genetics, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802 USA; [4]Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA; [5]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK; and [6]Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211 USA

- *Premise of the study:* The stability of the bimodal karyotype found in *Agave* and closely related species has long interested botanists. The origin of the bimodal karyotype has been attributed to allopolyploidy, but this hypothesis has not been tested. Next-generation transcriptome sequence data were used to test whether a paleopolyploid event occurred on the same branch of the Agavoideae phylogenetic tree as the origin of the *Yucca–Agave* bimodal karyotype.
- *Methods:* Illumina RNA-seq data were generated for phylogenetically strategic species in Agavoideae. Paleopolyploidy was inferred in analyses of frequency plots for synonymous substitutions per synonymous site (K_s) between *Hosta*, *Agave*, and *Chlorophytum* paralogous and orthologous gene pairs. Phylogenies of gene families including paralogous genes for these species and outgroup species were estimated to place inferred paleopolyploid events on a species tree.
- *Key results:* K_s frequency plots suggested paleopolyploid events in the history of the genera *Agave*, *Hosta*, and *Chlorophytum.* Phylogenetic analyses of gene families estimated from transcriptome data revealed two polyploid events: one predating the last common ancestor of *Agave* and *Hosta* and one within the lineage leading to *Chlorophytum.*
- *Conclusions:* We found that polyploidy and the origin of the *Yucca–Agave* bimodal karyotype co-occur on the same lineage consistent with the hypothesis that the bimodal karyotype is a consequence of allopolyploidy. We discuss this and alternative mechanisms for the formation of the *Yucca–Agave* bimodal karyotype. More generally, we illustrate how the use of next-generation sequencing technology is a cost-efficient means for assessing genome evolution in nonmodel species.

Key words: Agavoideae; bimodal karyotype; next generation sequencing; paleopolyploidy.

Karyotypes with bimodal chromosome size distributions have been described for taxa throughout the monocotyledons (e.g., Bennett et al., 1992; Talavera et al., 1993; Gitaí et al., 2005). Whereas most karyotypes exhibit a continuous range of chromosome sizes, karyotypes of some taxa are bimodal, with chromosomes falling into two distinct size classes often described as S for small and L for large ($n = S + L$). Chromosome size bimodality has arisen multiple times within Asparagales (Pires et al., 2006) including independent origins within Orchidaceae (Martínez, 1985; Giuseppina et al., 2010), Iridaceae (Goldblatt and Takei, 1997), Xanthorrhoeaceae (Taylor, 1925; Brandham and Doherty, 1998), Amaryllidaceae (Crosa, 2004) and Asparagaceae (McKelvey and Sax, 1933; Granick, 1944; Stedje, 1989). Bimodal karyotypes are most often limited to single genera (Goldblatt and Takei, 1997), small groups of closely related species (Stedje, 1996), or single species (Jones and Smith, 1967). However, bimodal chromosome size distributions are shared among multiple genera in Asphodeloideae (Xanthorrhoeaceae) and Agavoideae (Asparagaceae) (APG III, 2009; Chase et al., 2009). Within these subfamilies, bimodal karyotypes are synapomorphies for species-rich clades that may be millions of years old (Brandham and Doherty, 1998). For example, within the Asphodeloideae, *Aloe*, *Astroloba*, *Gasteria*, and *Haworthia* comprise ca. 689 species (Plant List, 2010) and all exhibit a karyotype of $n = 4S + 3L$ (Brandham, 1971). A clade composed solely of these genera is highly supported (Treutlein et al., 2003), indicating that chromosome bimodality is a synapomorphy and the ancestral karyotype for the group is $n = 4S + 3L$. The mechanism for persistence of bimodal karyotypes over millions of generations is unknown (Brandham and Doherty, 1998).

[1]Manuscript received 12 November 2011; revision accepted 4 January 2012.

We thank E. Wafula (Penn State) for help with the transcriptome assembly. We are thankful for the comments of three anonymous reviewers that greatly improved this manuscript. We are also grateful to the National Science Foundation for support this research through Assembling the Tree of Life (DEB 0732818 and 0829868) grants to J.L.-M., C.W.D., and J.C.P. and a Doctoral Dissertation Improvement Grant to M.R.M. and J.L.-M. The quality of the manuscript was greatly improved due to the suggestions of the editors of this special issue and anonymous reviewers.

[7]Present address: Chicago Botanic Garden, Glencoe, Illinois 60022 USA

[8]Author for correspondence (e-mail: jleebensmack@plantbio.uga.edu)

doi:10.3732/ajb.1100537

The clade defined as Agavaceae s.l. by Bogler et al. (2006) is composed of 15 genera and 377 species (Plant List, 2010) sharing bimodal karyotypes (Fig. 1). The APG III classification treats the group as an unnamed clade within Agavoideae, a subfamily of Asparagaceae (APG III, 2009; Chase et al., 2009). Here, we refer to this group as the Agavoideae bimodal karyotype clade or the ABK clade (Fig. 1). Within the ABK clade, 10 genera (~358 species) have a karyotype of n = 25S + 5L (Akemine, 1935; Tamura, 1995) (or some multiple of this based on ploidy) and five genera, *Camassia* Lindl., *Chlorogalum* (Lindl.) Kunth, *Hastingsia*, *Schoenolirion* Torr. Ex Durand, and *Hesperocallis* A. Gray, form a subclade of 19 species with bimodal karyotypes exhibiting varying numbers of large and small chromosomes (Gould, 1942; Cave, 1948; Sen, 1975; Fernández and Daviña, 1991; Tamura, 1995). A number of recent molecular phylogenetic analyses have placed *Camassia* Lindl., *Chlorogalum* (Lindl.) Kunth, and *Hastingsia* S.Watson, in a single clade (Bogler et al., 2006; Smith et al., 2008; Fishbein et al., 2010), and *Schoenolirion* Torr. ex Durand has long been associated with *Camassia* and *Chlorogalum* (e.g., Cronquist, 1981; Sherman and Beckling, 1991), although Halpin (2011) found support for placement of *Schoenolirion* outside of the *Camassia-Chlorogalum* clade. *Hesperocallis* has recently been placed within the ABK clade (Pires et al., 2004a; Bogler et al., 2006), and whole plastome analyses place this monotypic genus in a clade with *Camassia* and *Chlorogalum* (M. R. McKain et al., unpublished manuscript).

Current analyses support *Hosta* Tratt., with a karyotype of n = 25S + 5L (Zonneveld and Iren, 2001), as sister to the rest of the ABK clade (Bogler and Simpson, 1996; Smith et al., 2008; Steele et al., 2012 [in this issue]; M. R. McKain et al., unpublished manuscript), suggesting that the "*Yucca-Agave*" karyotype (Whitaker, 1934; Sato, 1935), was ancestral for the ABK clade. Divergence from the ancestral 25S + 5L karyotype appears to have occurred only within the clade including *Camassia*, *Chlorogalum*, *Hastingsia*, *Hesperocallis*, and *Schoenolirion.*

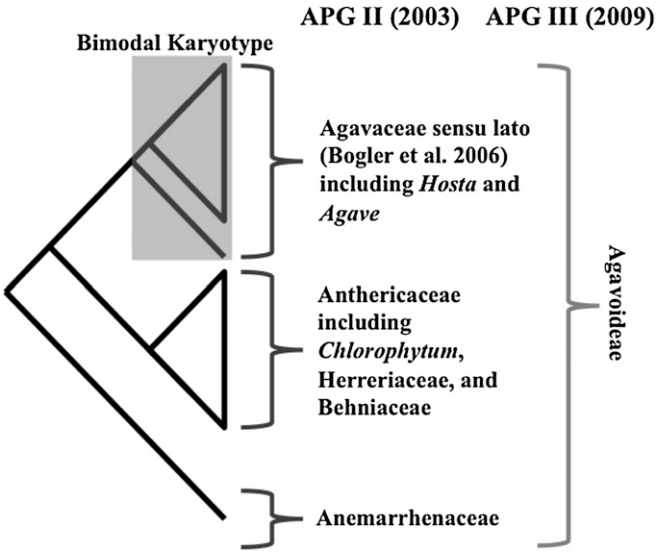

Fig. 1. The currently accepted phylogeny of Agavoideae (Asparagaceae) with both APG II and APG III nomenclature. The gray box indicates the former Agavaceae (senus Bogler et al., 2006), which we define as the Agavoideae bimodal karyotype clade. The sister clade to the ABK clade includes the former Anthericaceae, Herreriaceae, and Behniaceae.

Otherwise, the "*Yucca-Agave*" karyotype has persisted throughout the ABK clade including within polyploid series where increases in ploidy coincide with proportionate gains in the number of small and large chromosomes (Robert et al., 2008).

Chromosome sizes are uniformly distributed in karyotypes for Agavoideae species outside of the ABK clade. Relationships within the sister clade to the ABK clade within Agavoideae remain elusive. Some studies place the genera *Behnia* Didr., *Herreria* Ruiz & Pav. and *Herreriopsis* H. Perrier within a clade sister to the former Anthericaceae sensu stricto. (Fig. 1; Chase et al., 1996, 2006; Wurdack and Dorr, 2009), whereas others place *Behnia* as sister to a clade including *Herreria*, *Herreriopsis*, and the former Anthericaceae s.s. (Bogler et al., 2006; Pires et al., 2006; Kim et al., 2010). Some of the species within the former Anthericaceae s.s form a well-supported group with base chromosome numbers of x = 7,8 (Cave, 1948; Baldwin and Speese, 1951; Palomino and Romo, 1988; Bjorå et al., 2008). Karyotype information for many of the species in *Behnia*, *Herreria*, and *Herreriopsis* is unavailable; however, one study did show that the chromosomes of *Herreria salsaparilha* Mart., n = 29, exhibit a uniform but broad size distribution (1.30–10.51 μm) suggestive of fusion–fission events and potentially polyploidy (Gonçalves et al., 2007). *Anemarrhena asphodeloides* Bunge is sister to all other members of Agavoideae (Bogler et al., 2006; Pires et al., 2006; Kim et al., 2010; Steele et al., 2012; M. R. McKain et al., unpublished manuscript), and its karyotype (n = 11) exhibits a continuous range of chromosomes sizes (Rudall et al., 1998).

Two processes have been hypothesized to give rise to bimodal chromosome size distributions. The first is chromosome rearrangement involving fusion-fission events (Schubert and Lysak, 2011), which has been hypothesized for bimodal karyotypes within Asparagales (Pires et al., 2006) and specifically for some members of the genus *Ornithogalum* (Asparagaceae, Scilloideae; Vosa, 1983, 2005). In these instances, large chromosomes could be the products of fusion between two smaller chromosomes, or the small chromosomes could be the result of the fission of large chromosomes. Such fusion–fission events have been attributed to genomic shock associated with an allopolyploid event (Wendel, 2000; Comai et al., 2003; Chen and Ni, 2006) followed by chromosomal rearrangements (e.g., Song et al., 1995; Pires et al., 2004b).

A second hypothesized mechanism for bimodal karyotype formation is allopolyploidy involving parental species of different chromosome sizes. In this case, chromosomes have remained distinct following hybridization, segregating independently in the allopolyploid. Genomic in situ hybridization has been used to elucidate this mechanism in the grass (Poaceae) species *Milium montianum* Parl. (Bennett et al., 1992). This study identified *M. vernale* M. Bieb., or a closely related species, and a second, unknown species as progenitors of *M. montianum* with the large chromosomes of the bimodal karyotype being identical in number and size to those of *M. vernale*. Chromosome bimodality in Asphodeloideae and Agavoideae has been suggested to have originated through this mechanism, although this hypothesis has not been tested in either case and parental species have not been identified (Brandham, 1983; Brandham and Doherty, 1998; Vosa, 2005; Pires et al., 2006).

In this study, we used next-generation sequencing technology to sequence transcriptomes of strategically placed members of Agavoideae and test the hypothesis that the origin of chromosomal bimodality in this group coincides with a polyploid event. Numerous studies have evaluated divergence

between duplicate genes as measured by the number of synonymous substitutions per synonymous site (K_s) to identify whole genome duplications (i.e., polyploidy events) (Lynch and Conery, 2000; Blanc and Wolfe, 2004; Schlueter et al., 2004; Cui et al., 2006; Barker et al., 2008, 2009; Jiao et al., 2011) and concluded that genome duplications have been common throughout angiosperm evolution (Blanc and Wolfe, 2004; Schlueter et al., 2004; Cui et al., 2006; Soltis et al., 2009; Jiao et al., 2011; Van de Peer, 2011). Building on the approach of Jiao et al. (2011), we combined analyses of K_s plots and gene family phylogenies to test whether the origin of the bimodal karyotype on the lineage leading to the Agavoideae bimodal karyotype (ABK) clade is associated with a whole-genome duplication. In addition, the influence of polyploidy on the early diversification of the ABK clade is considered.

MATERIALS AND METHODS

***Taxon sampling*—**The taxonomy of families in "core Asparagales" including the clade investigated in this study has changed in recent years with the publication of APG II (2003) and APG III (2009). APG II included Agavaceae, Anemarrhenaceae, Anthericaeae, Behniaceae, and Herreriaceae (see Dahlgren et al., 1985) within Agavaceae, which was identified as an optional "bracketed" family within a broadly defined Asparagaceae. APG III (2009) abandoned the concept of bracketed families and applied the subfamilial classification Agavoideae (APG III, 2009; Chase et al., 2009). To simplify description of our sampling strategy, here we refer to the Agavoideae bimodal karyotype clade (ABK clade) (Fig. 1) to describe the clade that Bogler et al. (2006) named Agavaceae s.l. The ABK clade includes all descendants of the last common ancestor of *Hosta* and *Agave* L.: *Agave*, *Beschorneria* Kunth, *Camassia*, *Chlorogalum*, *Furcraea* Vent., *Hastingsia*, *Hesperaloe* Engelm. in S.Watson, *Hesperoyucca* (Engelm.) Trel., *Hesperocallis*, *Hosta*, *Manfreda* Salisb., *Polianthes* L., *Prochnyanthes* S.Watson, *Schoenolirion*, and *Yucca* L. On the basis of current phylogenetic analyses, *Hosta* is sister to the rest of Agavaceae (Givnish et al., 2006; Smith et al., 2008, Steele et al., 2012; M. R. McKain et al., unpublished manuscript) (Fig. 1). Therefore, within the ABK clade we generated transcriptome data for the diploid *Hosta venusta* F. Maek. (n = 25S + 5L) (Zonneveld and Iren, 2001) and analyzed available EST data for *Agave tequilana* F.A.C Weber (Simpson et al., 2011), also a diploid. Transcriptome data were also generated for *Chlorophytum rhizopendulum* Bjorå & Hemp (Agavoideae, classically placed in polyphyletic Anthericaceae) as an exemplar for the sister clade to the ABK clade (Fig. 1). Outside Agavoideae, transcriptome data were compiled for *Asparagus officinalis* L. (Asparagaceae) and *Leochilus labiatus* (Sw.) Kuntze (Orchidaceae) and combined with publically available EST data for *Allium cepa* L. (Amaryllidaceae) to root and identify ABK + *Chlorophytum* clades within gene trees.

***RNA isolation and sequencing*—**RNA was isolated from fresh apical meristematic tissue or very young leaves using an RNeasy Plant Mini Kit (Qiagen, Valencia, California, USA). All samples were kept on liquid nitrogen prior to RNA isolation. The optional step of heating the lysis solution to 65°C was used to maximize RNA yield. RNA was eluted into a final volume of 100 μL RNase-free water.

Total mass of RNA and quality was estimated using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, California, USA). Samples were deemed acceptable if RIN scores were greater than 8.0. A minimum of 20 μg of total RNA was required for library building and sequencing.

RNA-seq (Wang et al., 2009) paired-end libraries with average fragment lengths of 250 base pairs (bp) were constructed, and each library was sequenced on a single lane of an Illumina GAIIX sequencer flow cell (Illumina, San Diego, California, USA) at Cold Spring Harbor Laboratory to generate a minimum of 3 gigabases of 75 bp, paired-end sequences. Fastq sequence files for each taxon have been deposited in the Sequence Read Archive (SRA) database at NCBI (SRA study SRP009920).

***Transcriptome assembly*—**Illumina sequences were assembled using the CLC Genomics Workbench (CLC bio, Aarhus, Denmark). Prior to assembly, reads were trimmed to remove low quality ends, and trimmed reads shorter than 15 bp were discarded. Reads that passed the posttrimming length filter were then assembled using the default settings for de novo assembly, keeping only those contigs greater than 200 bp. Assemblies have been placed in a web-based searchable database (http://asparagalesdb.uga.edu).

***K_s plot estimation*—**K_s frequency plots were used to initially detect potential whole genome duplication (WGD) events and then to compare their origin to divergence of lineages within the ABK + *Chlorophytum* clade. Whereas studies attempting to identify more ancient genome duplications have assessed neutral sequence divergence based on transversions at 4-fold degenerate sites (4DTv; e.g., Tuskan et al., 2006), we were focused on divergence following gene duplication (paralogs) or speciation (orthologs) events within Agavoideae over the last 50 million years (Good-Avila et al., 2006). Therefore, in this study, K_s was used as a measure of neutral sequence divergence (e.g., Lynch and Conery, 2000; Blanc and Wolfe, 2004; Schlueter et al., 2004; Cui et al., 2006; Barker et al., 2008, 2009; Jiao et al., 2011). All-by-all BLASTN searches were performed, and paralogous and orthologous pairs were identified as best matches within and between species, respectively. Paralog and putative ortholog matches with minimum alignment lengths of 300 bp and at least a 40% identity were analyzed further. These cutoffs were used to provide a minimum of 100 codons for alignments used in the estimation of the number synonymous substitutions per synonymous sites (K_s). Amino acid sequences were estimated for these homologs using the program ESTscan (Iseli et al., 1999; Lottaz et al., 2003), and paired peptide sequences (orthologs or paralogs) were aligned using the program MUSCLE v3.7 (Edgar, 2004). Nucleotide sequences were then forced onto the amino acid alignments by codons (Cui et al., 2006; Suyama et al., 2006). Pairwise K_s values were then calculated for each homolog pair using codeml within the PAML 4 package (Yang, 2007) paired sequence settings (yn00; Yang and Nielsen, 2000) and the F3 × 4 model (Goldman and Yang, 1994) for estimating codon frequencies.

K_s values were normalized for among-species differences in synonymous substitution rates for *Agave*, *Hosta*, and *Chlorophytum* genes to compare K_s plots for putative paralog pairs within each species and for putative orthologs between species. Putatively single-copy genes were analyzed to estimate among-species variation in nuclear gene substitution rates. Ortholog sets for 49 single-copy genes were identified in the *Agave*, *Hosta*, *Chlorophytum*, *Asparagus*, *Allium*, and *Leochilus* transcriptome assemblies using blastx searches against a database of 970 genes inferred to be single-copy in sequenced angiosperm genomes (Wall et al., 2008; Duarte et al., 2010). Nucleotide sequences for transcripts were translated, and the amino acid sequences for each ortholog set were aligned using MUSCLE v3.7 (Edgar, 2004). Nucleotide coding sequences were then aligned to amino acid sequences using the program PAL2NAL v13 (Suyama et al., 2006). The resulting codon alignments were combined in a 56 372 column supermatrix, and a species tree was estimated using the GTR-gamma model implemented in RAxML v7.0.4 (Stamatakis, 2006). The estimated tree matched previously inferred relationships (Steele et al., 2012) (Fig. 2). K_s was estimated for each branch on the species tree using codeml (Yang, 1998; Yang and Nielsen, 1998). The cumulative K_s value for branches leading from the last common ancestor (LCA) of *Chlorophytum* and the ABK clade to the tips was lowest for *Hosta* (Fig. 2). These LCAs to tip K_s values estimated on the single-copy gene supermatrix were used to make relative rate corrections of K_s values for *Agave* and *Chlorophytum* paralog-pairs. Corrections were calculated by multiplying uncorrected *Agave* and *Chlorophytum* putative paralog pair K_s values by the ratio of *Hosta/Agave* or *Hosta/Chlorophytum* LCA to tip K_s values derived from the single-copy gene analysis (Cui et al., 2006). Similarly, K_s values for putative orthologs for each species pair were also normalized for differences in species-specific rates.

After normalization of the K_s values, frequency distributions for K_s values between 0.0 and 2.0 were plotted for putative paralogs within species and putative orthologs between each species pair. Only K_s values less than 2.0 were included in the plots because K_s estimates for more divergent gene pairs may be affected by saturation of substitution at synonymous sites.

Multivariate normal components were fit to the resulting K_s frequency distributions using the mixture model test implemented in the program EMMIX (McLachlan et al., 1999; http://www.maths.uq.edu.au/~gjm/emmix/emmix.html). The optimal number of components in the mixture model was identified using the Bayesian information criterion (BIC), and components were interpreted in terms of genome-scale duplication events and background single-gene duplications.

***Gene family circumscription and phylogeny estimation*—**Timing of WGD events relative to speciation events was deduced by comparing gene tree topologies to a species tree for taxa represented in each gene tree. Transcriptome assemblies for *H. venusta*, *A. tequilana*, *C. rhizopendulum*, *A. officinalis*, *A. cepa*, and *L. labiatus* were filtered using a cut-off of 300 bp and then translated using the program TransPipe (Barker et al., 2010). Gene family circumscriptions were estimated by clustering of inferred amino acid sequences using the

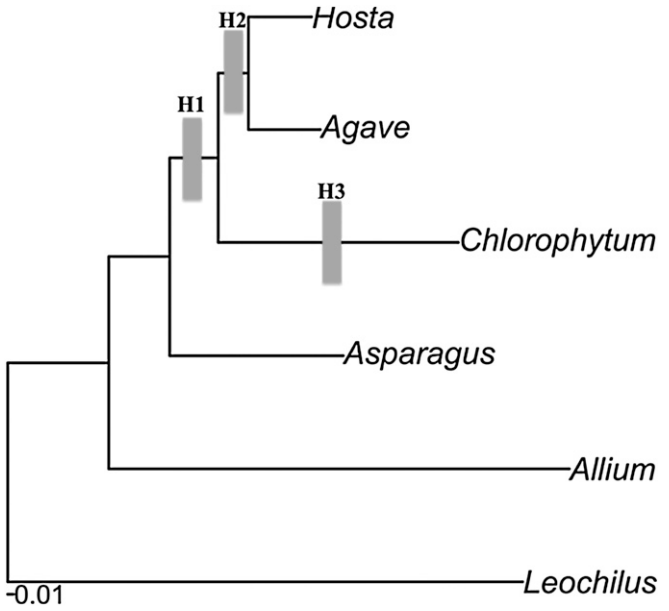

Fig. 2. The maximum likelihood species tree derived from a supermatrix analysis of putatively single-copy genes (Duarte et al., 2010) extracted from transcriptome data. All nodes have 100% bootstrap values and branch lengths are represented by K_s values. The hypothesized timings of genome duplication with the Agavoideae are marked as H1, H2, and H3 (see Methods).

program OrthoMCL v2.0 (Li et al., 2003) with suggested parameter settings. Gene family clusters were selected for gene tree estimation based on the results of K_s plot analyses of paralogs for *Hosta*, *Agave*, and *Chlorophytum*. To test our interpretation of the K_s plots, gene tree analyses focused on gene families with gene pairs having K_s values corresponding to paralog peaks in estimated frequency plots (Fig. 3). These gene family clusters included paralog pairs with K_s values between 0.1 and 0.3 for *H. venusta* and *A. tequilana,* and between 0.15 and 0.4 for *C. rhizopendulum*. When genes in paralog pairs were found in separate OrthoMCL clusters, the two clusters were combined for alignment and gene tree estimation.

Peptide sequences within each family were aligned using MUSCLE v3.7, and nucleotide sequences were aligned onto the amino acid alignments using PAL2NAL v13. Gappy alignments were filtered using two criteria. Columns in each alignment were removed if gaps were observed in more than 90% of the sequences (rows). Second, transcript assemblies (rows) were deleted if they covered less than 30% of the multiple sequence alignment's total length. Gene trees were estimated using RAxML v7.0.4 with the GTR + gamma evolutionary model and 500 bootstrap replicates. Gene trees were rooted using unigenes from the sampled outgroup taxa (*Leochilus*, *Allium*, or *Asparagus*) found in each orthogroup.

Timing of gene duplication events relative to the origin of the ABK clade and divergence of lineages leading to ABK clade and *Chlorophytum rhizopendulum* were assessed by querying estimated gene trees for the last common ancestor of focal paralog pairs and their descendant genes using in-house perl scripts (available upon request). As described above, focus was placed on paralog pairs with K_s values corresponding to hypothesized genome duplication events (Fig. 3). Gene tree topologies were inspected to determine whether individual duplication events occurred before divergence of the ABK clade and *Chlorophytum rhizopendulum* (H1), on the branch leading to the ABK clade (H2), on the branch leading to *Chlorophytum* (H3) or elsewhere on the species tree. For example, the LCA of each *Hosta* and *Agave* paralog pair with K_s values between 0.1 and 0.3 were identified in rooted gene trees. If a *Chlorophytum* gene is observed within a clade defined by a paralog pair LCA, then the duplication event was inferred to have occurred before divergence of *Chlorophytum* and the ABK clade. Alternatively, if *Chlorophytum* unigenes were placed as sister to a clade defined by the LCA of a focal paralog pair, the duplication event was inferred as having occurred after divergence of *Chlorophytum* and the ABK clade. Bootstrap percentages for the clade defined by the last common ancestor of a focal paralog pair were used to assess confidence in the inferred timing of duplication events relative to speciation events. The same approach was used to estimate the timing of *Chlorophytum* gene duplications relative to divergence of *Chlorophytum* and the ABK clade. Gene tree topologies were also inspected manually to check the results of automated gene tree queries.

RESULTS

***Assemblies*—**Assembly contig counts and lengths are shown in Table 1 for all contigs and successfully translated contigs (as estimated with TransPipe (Barker et al., 2010)). All successful translations were used for gene family analyses.

***K_s analyses of duplicate pairs*—**Two paleopolyploid events were inferred from analyses of rate-normalized K_s values for comparisons of *H. venusta*, *A. tequilana*, and *C. rhizopendulum* homolog pairs (Fig. 3). An analysis of 49 single-copy genes in a supermatrix of 56 372 base pairs was used to estimate relative silent site substitution rates on branches leading to *H. venusta*, *A. tequilana*, and *C. rhizopendulum*. K_s values for branches leading to *Agave*, *Hosta*, and *Chlorophytum* from their most recent common ancestor were 0.174, 0.159, and 0.410, respectively. To compare the K_s frequency plots, K_s values for *Agave* and *Chlorophytum* were normalized by multiplying raw values by relative rate ratios of 0.915 and 0.389, respectively. Correction factors of 0.955 and 0.578 were applied to K_s values of putatively orthologous *Agave*/*Hosta* and *Chlorophytum*/*Hosta* gene pairs, respectively, to account for differences in synonymous substitution rates on lineages leading to each species.

After applying rate corrections, maximum K_s values for paralog pairs were 77.49, 136.41, and 52.04 for *H. venusta*, *A. tequilana*, and *C. rhizopendulum*, respectively, but to avoid effects of saturation, paralog pairs with K_s values over 2.0 were not included in frequency plots (Fig. 3). K_s frequency plots include 437 gene duplicates for *H. venusta*, 2374 for *A. tequilana*, and 1704 for *C. rhizopendulum*. For comparison, uncorrected K_s frequency plots can be found in Appendix S1 (see Supplemental Data with the online version of this article). Due to its placement as the sister to the rest of the ABK clade (Fig. 1), putative ortholog sets including *Hosta* are most informative for understanding the timing of gene duplications relative to the origin of the ABK clade. After correction, K_s values for cross-species homolog pairs ranged from 0.0 to 113.54 for *Hosta*/*Agave* and 0.0 to 46.48 for *Hosta*/*Chlorophytum* gene pairs. K_s plots for putative ortholog pairs with K_s values less than 2.0 included 1656 *Hosta*-*Agave* ortholog pairs and 3639 *Hosta*-*Chlorophytum* ortholog pairs (Fig. 3D).

Mixture model analyses reveal distinct components in the K_s frequency plots that we interpret as background single-gene duplications or polyploidy associated duplication events. The Bayesian

TABLE 1. Contigs statistics for transcriptome assemblies of six study species including the filtered counts.

Species	Contig total	Contigs >300 bp	Contigs >500 bp	Contigs in gene family trees
Agave tequilana	12 972	12 972	11 087	9052
Hosta venusta	57 423	19 054	3076	9810
Chlorophytum rhizopendulum	58 766	33 369	19 770	19 879
Allium cepa	12 990	8992	6683	8992
Asparagus officinalis	107 254	62 708	43 093	31 431
Leochilus labiatus	43 860	18 316	8073	10 947

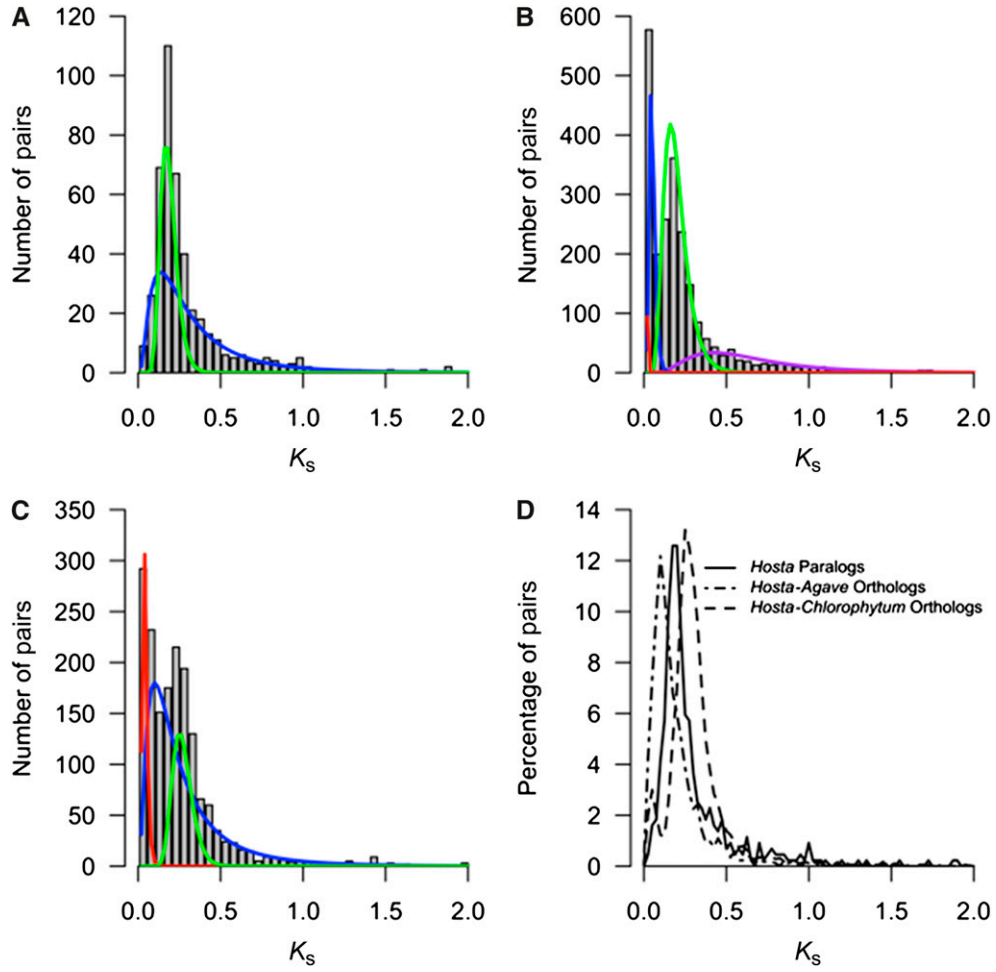

Fig. 3. Normalized K_s frequency plots (corrected for rate variation; see Results) for paralogous and orthologous gene pairs from *Hosta venusta*, *Agave tequilana*, and *Chlorophytum rhizopendulum* derived from transcriptome data. K_s distribution components estimated using EMMIX (see Methods) are superimposed on histogram for each paralog-pair K_s plot (A–C). These components are hypothesized to represent background gene duplications (blue), gene duplications associated with polyploidy events (green), alleles or sequencing errors that resulted in assembly of distinct unigenes with high sequence identity (red), and older gene duplications in the *A. tequliana* genome (purple; see Results). (A) *H. venusta* and (B) *A. tequliana* paralog plots include secondary peaks (green lines) with modal K_s value indicative of a whole genome duplication event at $K_s = 0.2$. The *C. rhizopendulum* paralog plot (C) shows a secondary peak with a mode at $K_s = 0.25$. (D) The K_s distribution for *H. venusta* paralogs (solid line) exhibits a mode between modal K_s values for putative *Hosta–Agave* ($K_s = 1.0$) and *Hosta–Chlorophyutm* ($K_s = 0.25$) orthologs.

information criterion (BIC) was used to choose the optimal number of normal distributions that fit data for each K_s plot based on the EMMIX output. The component with the smallest K_s values was interpreted as the background duplication. K_s frequency plots estimated from *Hosta* and *Agave* paralog pairs show a concentration of paralog pairs with modal K_s value around 0.2 (green lines in Fig. 3A and 3B, respectively). A slightly larger mode of 0.25 is seen in *Chlorophytum* K_s plot (green line in Fig. 3C). These peaks in the K_s distributions for all three species are suggestive of whole genome duplication events (WGD). The larger K_s value for the hypothesized duplication events in *C. rhizopendulum* may indicate an older but independent WGD on the branch leading to *C. rhizopendulum* (H3) or possibly a WGD on the lineage leading to the last common ancestor of *Chlorophytum* and members of the ABK clade (H1). Gene tree analyses were used to evaluate each of these alternative hypotheses (see below).

K_s frequency plots include at least one component representing ongoing single-gene duplication events (Fig. 3A–3C; denoted by blue lines). *Agave* and *Chlorophytum* K_s plots also show a population of putative paralog pairs with a normal distribution centered around K_s of ~0.025 (Fig. 3B, 3C; denoted by red lines). These components may represent pairs of alleles or sequencing errors that resulted in assembly of distinct unigenes with high sequence identity. When they are less common, such unigene pairs are typically combined with background duplication pairs in a single component (e.g., Fig. 3A). A separate low-K_s component in *Chlorophytum* K_s distribution was found only after the relative-rate correction increased the number of unigene pairs with the lowest K_s values (0.0–0.025) (compare Fig. 3 and Appendix S1). EMMIX identified a fourth component in the *Agave* K_s distribution centered on K_s ~0.5 (Fig. 3B; denoted by the purple line). This group of paralog pairs may represent a second, older WGD event that was not detected in the *Hosta* K_s distribution. Alternatively, background gene duplications may have been split into multiple components.

The corrected K_s frequency plot for *Hosta–Agave* and *Hosta–Chlorophytum* homolog pairs show peaks of 0.1 and 0.25,

respectively (Fig. 3D). Using normalized K_s values as a proxy for time, these values suggest that the polyploid event inferred from the *Hosta* and *Agave* paralog pairs (K_s ~0.20) occurred before divergence of lineages leading to *Hosta* and *Agave* and after divergence of the ABK clade and *Chlorophytum* (K_s ~0.25; Fig. 3D). It is important to note, however, that if this was an allopolyploid event, the estimated relative age of peaks in the paralog plots for the ABK clade would reflect timing of divergence of the parental species involved in the event rather than the actual WGD event (Doyle and Egan, 2010). This would not, however, affect our interpretation that the polyploid event occurred just before divergence of *Hosta* and *Agave*.

The peak in the *Chlorophytum* paralog K_s distribution described above, with a mode of 0.25 (Fig. 3C), largely overlaps the putative *Hosta–Chlorophytum* ortholog peak (Fig. 3D; K_s ~0.25), suggesting that the divergence of *Chlorophytum* and the ABK clade lineages occurred just before (H1) or after (H3) the WGD inferred from the *Chlorophytum* K_s plot. If H1 is correct, the paleopolyploid event in a common ancestor of the ABK clade and *Chlorophytum* may be masked by the later event (H2) inferred from the *Hosta* and *Agave* K_s plots. We analyzed nuclear gene trees to more rigorously characterize the timing of gene and genome duplications relative to speciation events in Agavoideae.

***Phylogenetic analysis of gene families*—**To further elucidate the timing of polyploidy in the species sequenced here, we conducted phylogenetic analyses of gene families that included the duplicated genes identified in the K_s analyses described above. A total of 12 724 putative gene families were circumscribed through OrthoMCL clustering of the transcripts assembled for the six species included in this study. We focused on OrthoMCL clusters containing *H. venusta*, *A. tequilana*, or *C. rhizopendulum* paralog pairs with K_s falling under the peaks interpreted as representing paleopolyploid events. These included 288, 1047, and 789 paralog pairs for *H. venusta*, *A. tequilana*, and *C. rhizopendulum*, respectively. Genes were placed in separate OrthoMCL clusters for 81 of the 2124 paralog pairs considered. In these cases, the two OrthoMCL clusters were combined before conducting phylogenetic analyses. After combining these clusters, gene sets that did not contain at least three species were removed from further consideration. The 2124 paralog pairs identified in *H. venusta*, *A. tequilana*, and *C. rhizopendulum* were distributed among 555 OrthoMCL gene sets that were aligned for gene tree estimation. All 555 alignments and ML gene trees have been deposited in the DRYAD database (http://dx.doi.org/10.5061/dryad.7pg045t2).

Relationships of genes within RAxML gene trees were analyzed within the context of species relationships (Fig. 2) to assess the placement of paleopolyploid events inferred from the ABK clade and *Chlorophytum* K_s frequency plots. The species tree estimated from a supermatrix analysis of 49 putatively single-copy nuclear genes (Fig. 2) is fully consistent with relationships inferred from analyses of plastid genes (Steele et al., 2012) and is supported with bootstrap percentages of 100% at all nodes.

We queried gene tree topologies to determine whether gene trees supported duplication events on the branch leading to the last common ancestors of the ABK clade and *Chlorophytum* (H1), the branch leading to the ABK clade after divergence from *Chlorophytum* (H2) or the branch leading to *Chlorophytum* (H3) (Fig. 2). These hypotheses are not mutually exclusive because the WGDs inferred through inspection of the K_s frequency plots could represent multiple events with homeologous gene pairs that have overlapping K_s distributions. Since our primary interest was in testing whether the origin of the *Yucca–Agave* bimodal karyotype coincided with a polyploid event, characterization of possible genome duplication events in our outgroup lineages or predating divergence of Agavoideae (including *Agave*, *Hosta*, and *Chlorophytum*) and the lineage leading to *Asparagus* (the closest outgroup used here) were outside the scope of this study. Trees were disregarded when one or more outgroup transcripts were nested within the Agavoideae paralog clade, while other outgroup transcripts rooted the clade including Agavoideae paralog pairs. These trees could reflect artifacts due to poor alignment or sparse gene sampling or duplication events predating divergence of the outgroup and ingroup (Agavoideae) taxa.

Of 555 gene trees that passed our filtering steps, 183 were informative for testing our hypotheses concerning phylogenetic placement of WGD events (i.e., >50% bootstrap support duplication events concordant with H1, H2, or H2). Of these, 102 trees contained paralog pairs from either *Hosta* or *Agave* that were used to evaluate support for H1 or H2. There were 81 trees that contained paralog pairs from *Chlorophytum*, and these were used to evaluate support for H1 or H3. Bootstrap percentages (BP) supporting clades defined by the most recent common ancestor of duplicate genes included in focal paralog pairs were used to evaluate the degree of support for one of the three hypotheses being considered. Trees were classified as providing 50–80 BP support or greater than 80 BP support for a given hypothesis.

Gene family trees that contained either *Hosta* or *Agave* paralog pairs showed clear evidence for an Agavaceae-specific ancestral WGD event. Of 102 informative trees, 54 exhibited greater than 80 BP support for H2, and another 30 had at least 50 BP for gene duplication after divergence of *Chlorophytum* and ABK clade genes (Fig. 4A). There were 18 trees that suggested gene duplication before the divergence of *Chlorophytum* and the ABK clade (consistent with H1), 13 of these with greater than 80 BP support.

Chlorophytum paralog pairs found in 81 gene trees showed clear support for duplication events after divergence of the ABK clade and *Chlorophytum*-lineage (H3; Fig. 4B). There were 52 trees that supported H3 with at least 80 BP and another 19 with at least 50 BP. Gene duplications on the lineage leading to *Chlorophytum* and the ABK clade (H1) were seen in 10 trees, 8 with at least 80 BP. In sum, we interpret these results as favoring two separate WGD events in the lineages leading to the ABK clade and *Chlorophytum*. Whereas a bimodal karyotype is associated with the WGD on the lineage leading to the ABK clade, bimodal karyotypes have not been reported in *Chlorophytum* or related taxa within Anthericaceae (Chase et al., 1996), Behniaceae, or Herreriaceae sensu APG II (2003).

DISCUSSION

Polyploidy is a ubiquitous and recurring phenomenon in angiosperms, and a recent study by Jiao et al. (2011) demonstrated that all flowering plants share at least two ancestral whole genome duplications events. Detection of ancient polyploid events can be difficult as characteristics of recent polyploids, such as doubled chromosome number or genome size, are typically lost over time (Devos et al., 2002; Leitch and Bennet, 2004), often rapidly (e.g., Song et al., 1995). K_s analyses have been used to characterize ancient polyploidization in a number of taxa (Blanc and Wolfe, 2004; Schlueter et al., 2004; Cui et al., 2006; Barker et al., 2008, 2009; Shi et al., 2010; Jiao et al.,

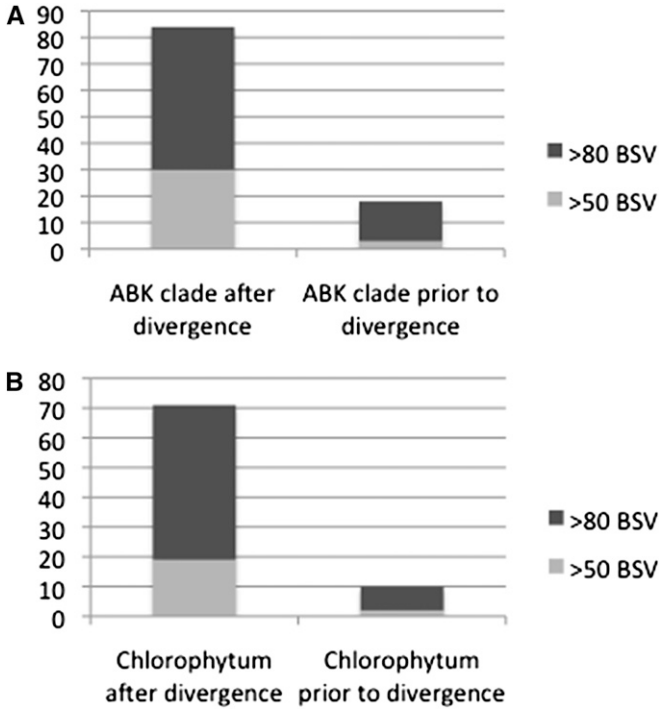

Fig. 4. Total counts for duplication events in Agavoideae inferred from gene tree topologies suggest genome-wide duplications on the branches leading to the ABK clade (H2) and *Chlorophytum* (H3). Histograms are shown for duplications observed in gene trees including (A) *A. tequilana* and *H. venusta* paralog pairs, and (B) *C. rhizopendulum* paralog pairs.

2011). Peaks observed in K_s plots have been interpreted in the context of species relationships in an effort to identify shared and independent polyploid events and their evolutionary implications. However, interpretation of cross-species comparisons of K_s plots are complicated by the fact that substitution rates may vary among lineages. Previous work has shown that with sufficient taxon sampling, analyses of gene trees can resolve the relative timing of hypothesized paleopolyploidy in the face of variable substitution rates (Pfeil et al., 2005; Cui et al., 2006; Barker et al., 2008, 2009, Jiao et al., 2011). In this study, K_s analyses of *Agave*, *Hosta*, and *Chlorophytum* transcriptomes revealed evidence of paleopolyploidy, but timing of one or more WGD events was unclear based on K_s plots alone. We resolved this uncertainty through phylogenetic analysis of gene families including paralog pairs hypothesized to represent WGD events. These analyses confirmed the existence of two paleopolyploidizations within Agavoideae, one on the branch leading to the ABK and another within the *Chlorophytum* lineage. As has been seen in previous studies (Pfeil et al., 2005; Cui et al., 2006; Barker et al., 2008, 2009), variation in substitution rates among lineages led to ambiguity in interpretation of cross-species comparisons of analysis of uncorrected K_s plots (Appendix S1). A relative rate correction was applied to K_s values in an effort to resolve this ambiguity, but analysis of gene tree topologies provided the clearest evidence for timing of gene duplications relative to speciation events.

Phylogenetic analysis of gene families constructed from the six study species supported a WGD event having occurred on the same branch as the origin of the *Yucca–Agave* bimodal karyotype, which was first described over a century ago (reviewed in Whitaker, 1934; Sato, 1935; Granick, 1944) (H2; Fig. 3A). In addition, gene trees constructed with *Chlorophytum* paralog pairs showed support for a second *Chlorophytum*-specific WGD event (H3, Fig. 3B). Gene duplications were also evident on the branch leading to the last common ancestor of *Agave*, *Hosta*, and *Chlorophytum*, but these were much less common than duplications on branches leading either to the ABK clade or *Chlorophytum* (Fig. 4).

Determining relationships of taxa, including genera, within Agavoideae has been a long-standing problem in plant systematics. The use of molecular markers in both the nuclear and plastid genomes have yielded strong support for monophyly of ABK clade but not for relationships between genera within the group (Eguiarte et al., 1994, 2000; Bogler and Simpson, 1995, 1996; Pires et al., 2004a; Bogler et al., 2006). Phylogenetic analysis of whole plastid genome alignments has resulted in a well-resolved tree with high support (M. R. McKain et al., unpublished manuscript), but short internodes separating basal nodes in the ABK clade suggest that the group diversified rapidly after its origin. The process of diploidization, including gene loss (i.e., fractionation; Freeling, 2009), following polyploidization can spur reproductive isolation and speciation (Werth and Windham, 1991; Lynch and Force, 2000; Taylor et al., 2001; Scannell et al., 2006). This process has been hypothesized as a driver of angiosperm diversification (Soltis et al., 2009). Ecological factors, including range expansion, colonization of arid habitats, and plant-pollinator interactions, are thought to have contributed to diversification of the ABK clade (Good-Avila et al., 2006), but loss of alternate homeologs following polyploidization—a special case of the Bateson–Dobzhansky–Muller speciation model (Orr, 1996)—may have also played a role in the early radiation within the ABK clade, as has been hypothesized for *Saccharomyces* (Scannell et al., 2006).

In this context, it is noteworthy that K_s values for hypothesized *Agave* and *Hosta* homeologs (~0.2; Fig. 3A, 3B) are significantly greater than those for putative orthologs (~0.1; Fig. 3D), which may be due to a gap between polyploidization and the inferred radiation in the early history of the ABK clade. Alternatively, if the radiation was spurred by an allopolyploid event, large differences between ABK clade paralog and ortholog K_s values (Fig. 3) may be due to divergence between parental genomes before hybridization.

Distinguishing between autopolyploidy and allopolyploidy, especially in ancient polyploid events, can be quite difficult and at times impossible (Doyle and Egan, 2010). Here, we have demonstrated that a paleopolyploid event occurred on the lineage leading to the ABK clade after the divergence of the *Chlorophytum* clade, consistent with the hypothesis that the *Yucca–Agave* bimodal karyotype originated with an allopolyploid event. If the last common ancestor of the ABK clade was in fact an allopolyploid, however, the progenitor species seem to be extinct, and it is not possible to definitively distinguish between autopolyploidy and allopolyploidy through gene tree analyses (Doyle and Egan, 2010). Future work will test whether homeologous gene pairs are consistently segregating on small and large chromosomes as would be expected if the last common ancestor of the ABK clade was an allopolyploid hybrid of now extinct parental species with small and large chromosomes.

The origin of chromosome bimodality in Agavoideae has been under investigation since the *Yucca–Agave* karyotype was first described, and the 5L + 25S karyotype has long been viewed as diagnostic for evolutionary affiliations with

Yucca and *Agave* (e.g., *Hosta*; Whitaker, 1934). When combined with embryological and other morphological characters, cytogenetic analyses led Cave (1948) to posit that *Hesperocallis* with four large, two medium, and 18 small chromosomes is allied with *Hosta*, *Yucca*, and *Agave*, a hypothesis that would gain molecular support 60 years later (Pires et al., 2004a). The objective of this study was to test whether a paleopolyploid event was associated with the origin of the 5L + 25S karyotype. The results are consistent with the hypothesis that the last common ancestor of the ABK clade was an allopolyploid. While the bimodal karyotype is suggestive of allopolyploidy, the possibility of chromosomal fusion and fission cannot be discounted.

The utility of next-generation sequencing for gaining insight into the genomes and evolutionary history of nonmodel species is obvious. The next-generation sequence data presented here allowed us to assess the plausibility of a long-standing hypothesis that relates the origin of chromosome bimodality to polyploidy. This work will aid in understanding the evolution of Agavoideae while providing an improved framework for future phylogenetic, ecological, and crop improvement studies. There is also great potential for investigating bimodal karyotypes across Asparagales and their implications for understanding causes and consequences of polyploidy.

LITERATURE CITED

AKEMINE, T. 1935. Chromosome studies on *Hosta* I. The chromosome numbers in various species of *Hosta. Journal of the Faculty of Science, Hokkaido Imperial University, series 5, Botany* 5: 25–32.

APG II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

BALDWIN, J. T., AND B. M. SPEESE. 1951. Cytogeography of *Chlorophytum* in Liberia. *American Journal of Botany* 38: 153–156.

BARKER, M. S., K. M. DIUGOSCH, L. DINH, R. S. CHALLA, N. C. KANE, M. G. KING, AND L. H. RIESEBERG. 2010. EvoPipes.net: Bioinformatic tools for ecological and evolutionary genomics. *Evolutionary Bioinformatics Online* 6: 143–149.

BARKER, M. S., N. C. KANE, M. MATVIENKO, A. KOZIK, R. W. MICHELMORE, S. J. KNAPP, AND L. H. RIESEBERG. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution* 25: 2445–2455.

BARKER, M. S., H. VOGEL, AND M. E. SCHRANZ. 2009. Paleopolyploidy in the Brassicales: Analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* 1: 391–399.

BENNETT, S. T., A. Y. KENTON, AND M. D. BENNETT. 1992. Genomic in situ hybridization reveals the allopolyploid nature of *Milium montianum* (Gramineae). *Chromosoma* 101: 420–424.

BJORÅ, C. S., A. HEMP, G. HOELL, AND I. NORDAL. 2008. A taxonomic and ecological analysis of two forest *Chlorophytum* taxa (Anthericaceae) on Mount Kilimanjaro, Tanzania. *Plant Systematics and Evolution* 274: 243–253.

BLANC, G., AND K. H. WOLFE. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16: 1667–1678.

BOGLER, D. J., J. C. PIRES, AND J. FRANCISCO-ORTEGA. 2006. Phylogeny of Agavaceae based on *ndhF*, *rbcL*, and ITS sequences: Implications of molecular data for classification. *Aliso* 22: 313–328.

BOGLER, D. J., AND B. B. SIMPSON. 1995. A chloroplast DNA study of the Agavaceae. *Systematic Botany* 20: 191–205.

BOGLER, D. J., AND B. B. SIMPSON. 1996. Phylogeny of Agavaceae based on ITS rDNA sequence variation. *American Journal of Botany* 83: 1225–1235.

BRANDHAM, P. E. 1971. The chromosomes of the Liliaceae. II: Polyploidy and karyotype variation in the Aloïneae. *Kew Bulletin* 25: 381–399.

BRANDHAM, P. E. 1983. Evolution in a stable chromosome system. *In* P. E. Brandham and M. D. Bennett [eds.], Proceedings of Kew Chromosome Conference II, 251–260. Allen and Unwin, London, UK.

BRANDHAM, P. E., AND M. J. DOHERTY. 1998. Genome size variation in the Aloaceae, an angiosperm family displaying karyotypic orthoselection. *Annals of Botany* 82: 67–73.

CAVE, M. S. 1948. Sporogenesis and embryo sac development of *Hesperocallis* and *Leucocrinum* in relation to their systematic position. *American Journal of Botany* 35: 343–349.

CHASE, M. W., M. F. FAY, D. S. DEVEY, O. MAURIN, N. RØNSTED, T. J. DAVIES, Y. PILLON, ET AL. 2006. Multigene analyses of monocot relationships: A summary. *Aliso* 22: 63–75.

CHASE, M. W., J. L. REVEAL, AND M. F. FAY. 2009. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Botanical Journal of the Linnean Society* 161: 132–136.

CHASE, M. W., P. J. RUDALL, AND J. G. CONRAN. 1996. New circumscriptions and a new family of Asparagoid lilies: Genera formerly included in Anthericaceae. *Kew Bulletin* 51: 667–680.

CHEN, Z. J., AND Z. NI. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* 28: 240–252.

COMAI, L., A. MADLUNG, C. JOSEFSSON, AND A. TYAGI. 2003. Do the different parental "heteromes" cause genomic shock in newly formed allopolyploids? *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 358: 1149–1155.

CRONQUIST, A. 1981. An integrated system of classification of flowering plants. Columbia University Press, New York, New York, USA.

CROSA, O. 2004. Segunda especie y justificación del género *Zoellnerallium* (Alliaceae). *Darwiniana* 42: 165–168.

CUI, L., P. K. WALL, J. H. LEEBENS-MACK, B. G. LINDSAY, D. E. SOLTIS, J. J. DOYLES, P. S. SOLTIS, ET AL. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16: 738–749.

DAHLGREN, R. M. T., H. T. CLIFFORD, AND P. F. YEO. 1985. The families of the monocotyledons: Structure, evolution, and taxonomy. Springer-Verlag, Berlin, Germany.

DEVOS, K. M., J. K. M. BROWN, AND J. L. BENNETZEN. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis. Genome Research* 12: 1075–1079.

DOYLE, J. J., AND A. N. EGAN. 2010. Dating the origins of polyploidy events. *New Phytologist* 186: 73–85.

DUARTE, J. M., P. K. WALL, P. P. EDGER, L. L. LANDHERR, H. MA, J. C. PIRES, J. LEEBENS-MACK, AND C. W. DEPAMPHILIS. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology* 10: 61.

EDGAR, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

EGUIARTE, L. E., M. R. DUVALL, G. H. LEARN JR., AND M. T. CLEGG. 1994. The systematic status of the Agavaceae and Nolinaceae and related Asparagales in the monocotyledons: An analysis based on the *rbcL* gene sequence. *Boletín de la Sociedad Botánica de México* 54: 35–56.

EGUIARTE, L. E., V. SOUZA, AND A. SILVA MONTELLANO. 2000. Evolución de la familia Agavaceae: Filogenia, biología reproductiva y genética de poblaciones. *Boletín de la Sociedad Botánica de México* 66: 131–151.

FERNÁNDEZ, A., AND J. R. DAVIÑA. 1991. Heterochromatin and genome size in *Fortunatia* and *Camassia* (Hyacinthaceae). *Kew Bulletin* 46: 307–316.

FISHBEIN, M., S. R. KEPHART, M. WILDER, K. M. HALPIN, AND S. L. DATWYLER. 2010. Phylogeny of *Camassia* (Agavaceae) inferred from plastid *rpl16* intron and *trnD-trnY-trnE-trnT* intergenic spacer DNA sequences: Implications for species delimitation. *Systematic Botany* 35: 77–85.

Freeling, M. 2009. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* 60: 433–453.

Gitaí, J., R. Horres, and A. M. Benko-Iseppon. 2005. Chromosomal features and evolution of Bromeliaceae. *Plant Systematics and Evolution* 253: 65–80.

Giuseppina, B., C. Brullo, S. Pulvirenti, A. Scrugli, M. C. Terrasi, and S. D'Emerico. 2010. Advances in chromosomal studies in Neottieae (Orchidaceae): Constitutive heterochromatin, chromosomal rearrangements and speciation. *Caryologia* 63: 184–191.

Givnish, T. J., J. C. Pires, S. W. Graham, M. A. McPherson, L. M. Prince, T. B. Patterson, H. S. Rai, et al. 2006. Phylogenetic relationships of monocots based on the highly informative plastid gene *ndhF*: Evidence for widespread concerted convergence. *Aliso* 22: 28–51.

Goldblatt, P., and M. Takei. 1997. Chromosome cytology of Iridaceae-patterns of variation, determination of ancestral base numbers, and modes of karyotype change. *Annals of the Missouri Botanical Garden* 84: 285–304.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* 11: 725–736.

Gonçalves, L. de A., W. R. Clarindo, C. R. de Carvalho, and W. C. Otoni. 2007. Cytogenetics and flow cytometry-based DNA quantification in *Herreria salsaparilha* Martius (Herreriaceae): A medicinal species. *Cytologia* 72: 295–302.

Good-Avila, S. V., V. Souza, B. S. Gaut, and L. E. Eguiarte. 2006. Timing and rate of speciation in *Agave* (Agavaceae). *Proceedings of the National Academy of Sciences, USA* 103: 9124–9129.

Gould, F. W. 1942. A systematic treatment of the genus *Camassia* Lindl. *American Midland Naturalist* 28: 712–742.

Granick, E. B. 1944. A karyosystematic study of the genus *Agave*. *American Journal of Botany* 31: 283–298.

Halpin, K. M. 2011. A chloroplast phylogeny of Agavaceae subfamily Chlorogaloideae with a focus on species relationships in *Hastingsia*. M.S. thesis, Oklahoma State University, Stillwater, Oklahoma, USA.

Iseli, C., C. V. Jongeneel, and P. Bucher. 1999. ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology* 1999: 138–148.

Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Jones, K., and J. B. Smith. 1967. The chromosomes of the Liliaceae. I: The karyotypes of twenty-five tropical species. *Kew Bulletin* 21: 31–38.

Kim, J.-H., D.-K. Kim, F. Forest, M. F. Fay, and M. W. Chase. 2010. Molecular phylogenetics of Ruscaceae sensu lato and related families (Asparagales) based on plastid and nuclear DNA sequences. *Annals of Botany* 106: 775–790.

Leitch, I. J., and M. D. Bennet. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* 82: 651–663.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research* 13: 2178–2189.

Lottaz, C., C. Iseli, C. V. Jongeneel, and P. Bucher. 2003. Modeling sequencing errors by combining hidden Markov models. *Bioinformatics* 19: 103–112.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.

Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154: 459–473.

Martínez, A. 1985. The chromosomes of orchids VIII. Spiranthinae and Cranichidinae. *Kew Bulletin* 40: 139–147.

McKelvey, S. D., and K. Sax. 1933. Taxonomic and cytological relationships of *Yucca* and *Agave*. *Journal of the Arnold Arboretum* 14: 76–81.

McLachlan, G. J., D. Peel, K. E. Basford, and P. Adams. 1999. The EMMIX software for the fitting of mixtures of normal and *t*-components. *Journal of Statistical Software* 4: 1–14.

Orr, A. H. 1996. Dobzhansky, Bateson and the genetics of speciation. *Genetics* 144: 1331–1335.

Palomino, G., and V. Romo. 1988. Karyotypic studies in two Mexican species of *Echeandia* Ort. (Liliaceae). *Southwestern Naturalist* 33: 382–384.

Pfeil, B. E., J. A. Schlueter, R. C. Shoemaker, and J. J. Doyle. 2005. Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Systematic Biology* 54: 441–454.

Pires, J. C., I. J. Maureira, T. J. Givnish, K. J. Sytsma, O. Seberg, G. Petersen, J. I. Davis, et al. 2006. Phylogeny, genome size, and chromosome evolution of Asparagales. *Aliso* 22: 287–304.

Pires, J. C., I. J. Maureira, J. P. Rebman, G. A. Salazar, L. I. Cabrera, M. F. Fay, and M. W. Chase. 2004a. Molecular data confirm the phylogenetic placement of the enigmatic *Hesperocallis* (Hesperocallidaceae) with Agave. *Madrono* 51: 307–311.

Pires, J. C., J. Zhao, M. E. Schranz, E. J. Leon, P. A. Quijada, L. N. Lukens, and T. C. Osborn. 2004b. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biological Journal of the Linnean Society* 82: 675–688.

Plant List. 2010. The plant list, version 1 [online]. Website http://www.theplantlist.org/ [accessed 28 August 2011].

Robert, M. L., K. Y. Lim, L. Hanson, F. Sanchex-Teyer, M. D. Bennett, A. R. Leitch, and I. J. Leitch. 2008. Wild and agronomically important *Agave* species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy. *Botanical Journal of the Linnean Society* 158: 215–222.

Rudall, P. J., E. M. Engleman, L. Hanson, and M. W. Chase. 1998. Embryology, cytology and systematics of *Hemiphylacus, Asparagus* and *Anemarrhena* (Asparagales). *Plant Systematics and Evolution* 211: 181–199.

Sato, D. 1935. Analysis of the karyotypes in *Yucca, Agave* and the related genera with special reference to the phylogenetic significance. *Japanese Journal of Genetics* 11: 272–278.

Scannell, D. R., K. P. Byrne, J. L. Gordon, S. Wong, and K. H. Wolfe. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440: 341–345.

Schlueter, J. A., P. Dixon, C. Granger, D. Grant, L. Clark, J. J. Doyle, and R. C. Shoemaker. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47: 868–876.

Schubert, I., and M. Lysak. 2011. Interpretation of karyotype evolution should consider structural constraints. *Genetics* 27: 207–216.

Sen, S. 1975. Cytotaxonomy of Liliales. *Feddes Repertorium* 86: 255–305.

Sherman, H., and R. Beckling. 1991. The generic distinctness of *Schoenolirion* and *Hastingsia*. *Madroño* 38: 130–138.

Shi, T., H. Huang, and M. S. Barker. 2010. Ancient genome duplications during the evolution of kiwifruit (*Actinidia*) and related Ericales. *Annals of Botany* 106: 497–504.

Simpson, J., A. Martínez Hernández, M. Jazmín Abraham Juárez, S. Delgado Sandoval, A. Sánchez Villarreal, and C. Cortés Romero. 2011. Genomic resources and transcriptome mining in *Agave tequilana*. *Global Change Biology Bioenergy* 3: 25–36.

Smith, C. I., O. Pellmyr, D. M. Althoff, M. Balcázar-Lara, J. Leebens-Mack, and K. A. Segraves. 2008. Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. *Proceedings of the Royal Society, B, Biological Sciences* 275: 249–258.

Soltis, D. E., V. A. Alber, J. Leebens-Mack, C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Song, K., P. Lu, K. Tang, and T. C. Osborn. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences, USA* 92: 7719–7723.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

STEDJE, B. 1989. Chromosome evolution within the *Ornithogalum tenuifolium* complex (Hyacinthaceae), with special emphasis on the evolution of bimodal karyotypes. *Plant Systematics and Evolution* 166: 79–89.

STEDJE, B. 1996. Karyotypes of some species of Hyacinthaceae from Ethiopia and Kenya. *Nordic Journal of Botany* 16: 121–126.

STEELE, P. R., K. L. HERTWECK, D. MAYFIELD, M. R. MCKAIN, J. LEEBENS-MACK, AND J. C. PIRES. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 331–349.

SUYAMA, M., D. TORRENTS, AND P. BORK. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34: W609–W612.

TALAVERA, S., P. GARCÍA-MURILLO, AND J. HERRERA. 1993. Chromosome numbers and a new model for karyotype evolution in *Ruppia* L. (Ruppiaceae). *Aquatic Botany* 45: 1–13.

TAMURA, M. N. 1995. A karyological review of the orders Asparagales and Liliales (Monocotyledonae). *Feddes Repertorium* 106: 83–111.

TAYLOR, J. S., Y. VAN DE PEER, AND A. MEYER. 2001. Genome duplication, divergent resolution and speciation. *Trends in Genetics* 17: 299–301.

TAYLOR, W. R. 1925. Cytological studies on *Gasteria*. II. A comparison of the chromosomes of *Gasteria, Aloë*, and *Haworthia. American Journal of Botany* 12: 219–223.

TREUTLEIN, J., G. F. SMITH, B. E. VAN WYK, AND M. WINK. 2003. Phylogenetic relationships in Asphodelaceae (subfamily Alooideae) inferred from chloroplast DNA sequences (*rbcL, matK*) and from genomic fingerprinting (ISSR). *Taxon* 52: 193–207.

TUSKAN, G. A., S. DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV, U. HELLSTEN, N. PUTANM, ET AL. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604.

VAN DE PEER, Y. 2011. A mystery unveiled. *Genome Biology* 12: 113.

VOSA, C. G. 1983. Chromosome evolution in *Ornithogalum*. *Kew Chromosome Conference II*: 370.

VOSA, C. G. 2005. On chromosome uniformity, bimodality and evolution in the tribe Aloineae (Asphodelaceae). *Caryologia* 58: 83–85.

WALL, P. K., J. LEEBENS-MACK, K. F. MÜELLER, D. FIELD, N. S. ALTMAN, AND C. W. DEPAMPHILIS. 2008. PlantTribes: A gene and gene family resource for comparative genomics in plants. *Nucleic Acids Research* 36: D970–D976.

WANG, Z., M. GERSTEIN, AND M. SNYDER. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57–63.

WENDEL, J. F. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.

WERTH, C. R., AND M. D. WINDHAM. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *American Naturalist* 137: 515–526.

WHITAKER, T. W. 1934. Chromosome constitution in certain monocotyledons. *Journal of the Arnold Arboretum* 15: 135–153.

WURDACK, K. J. J., AND L. J. DORR. 2009. The South American genera of Hemerocallidaceae (*Eccremis* and *Pasithea*): Two introductions to the New World. *Taxon* 58: 1122–1134.

YANG, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* 15: 568–573.

YANG, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.

YANG, Z., AND R. NIELSEN. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular Biology and Evolution* 15: 1600–1611.

YANG, Z., AND R. NIELSEN. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* 17: 32–43.

ZONNEVELD, B. J. M., AND F. IREN. 2001. Genome size and pollen viability as taxonomic criteria: Application to the genus *Hosta. Plant Biology* 3: 176–185.