Chapter 2

Utility of transcriptome sequencing for phylogenetic inference and character evolution

Jun Wen,[1] Ashley N. Egan,[1] Rebecca B. Dikow[2] & Elizabeth A. Zimmer[1]

1 Department of Botany, National Museum of Natural History, MRC-166, Smithsonian Institution, Washington, D.C. 20013-7012, U.S.A.
2 Center for Conservation and Evolutionary Genetics, National Zoological Park and Division of Mammals, National Museum of Natural History, MRC-108, Smithsonian Institution, Washington, D.C. 20013-7012, U.S.A.
Author for correspondence: J. Wen, wenj@si.edu

Abstract Transcriptome sequencing or RNA-Seq is one of the most efficient and cost-effective methods currently available for gene discovery in non-model organisms. Recent studies have demonstrated the utility of these data for resolving the relationships of diverse lineages of organisms, by extracting sequences of a large number of single-copy nuclear genes from transcriptomes of the taxa under study (i.e., RNA-Seq phylogenetics). Comparative transcriptomics has also been applied in several other areas in systematic biology, especially concerning polyploidy, introgression, hybridization, and horizontal gene transfer, as well as character evolution, including the identification of likely key innovations. This review focuses on the utility of transcriptomics in phylogenetic inferences and character evolution, and discusses the analytical framework, challenges, and prospects of transcriptome data in plant systematics, especially in phylogenetics. The main limitations are related to the high RNA-grade tissue quality requirement, the comparisons among expressed genes at a particular time point or developmental stage, orthology determination, and sequencing that arises from coding regions only, as well as several bioinformatics and analytical challenges. Comparative transcriptomics offers a rich set of genic sources for phylogenetic inference and single- or low-copy nuclear marker development. As whole genomes and genomic data become less costly and more prevalent, comparisons among transcriptomes will increase. With transcriptome- and genome-scale bioinformatics continuing to develop, we expect that the utility of transcriptomics will only increase in systematic biology, and that the RNA-Seq approach will offer tremendous insights into the understanding of the ontogeny and evolution of characters in the next decade.

Keywords RNA-Seq; RNA-Seq phylogenetics; systematics; transcriptome; transcriptomics

Hörandl, E. & Appelhans, M.S. (eds.), Next-Generation Sequencing in Plant Systematics.
Chapter 2, http://dx.doi.org/10.14630/000003

INTRODUCTION

Transcriptome sequencing or RNA-Seq (Wang & al., 2009) involves extracting RNAs from a specific tissue or sets of tissues of an organism, converting a population of RNA (total or fractionated, such as poly(A)+) to a library of cDNA fragments with adaptors attached to one or both ends, and sequencing the cDNAs using high-throughput sequencing platforms (Martin & Wang, 2011; Ozsolak & Milos, 2011; Wolf, 2013). The study of transcriptomes is also known as transcriptomics (Wang & al., 2009). Transcriptome sequences have been shown to provide a rich set of characters to produce phylogenies in eukaryotes, particularly non-model organisms and/or those with very large genomes. They are more efficient and cost-effective than traditional PCR- and EST-based methods (Zimmer & Wen, 2012; Lemmon E.M. & Lemmon, 2013).

Recent studies have demonstrated the utility of transcriptome data for resolving the relationships of diverse lineages of organisms, such as annelids (Struck & al., 2011; Weigert & al., 2014), mollusks (Kocot & al., 2011; Smith & al., 2011), tetrapods (including the group consisting of turtles, birds and crocodiles, Chiari & al., 2012; and rodents, Lin & al., 2014), arthropods (e.g., Hedin & al., 2012; Brewer & Bond, 2013), including several studies on various insect groups such as mosquitoes (Hittinger & al., 2010), Hymenoptera (Johnson B.R. & al., 2013) and Lepidoptera (Bazinet & al., 2013), Ecdysozoa (a large clade of arthropods, nematodes, and several smaller phyla, Borner & al., 2014), land plants (Timme & al., 2012; Wickett & al., 2014), seed plants (Xi & al., 2013); angiosperms (Xi & al., 2014; Zeng L. & al., 2014), and the grape plant family (Wen & al., 2013a) (see Table 1 below). With the speed and decreasing cost of next-generation sequencing (NGS) approaches, phylogenomics, including transcriptomics and whole-genome sequences, may soon become the standard for generating large phylogenetically informative datasets for many previously unresolvable lineages (Straub & al., 2012; McCormack & al., 2013a, b; Soltis & al., 2013; Wen & al., 2013a, b). Recent broad-scale phylogenetic studies have used both transcriptome and available whole-genome data to explore deep relationships in plants (Xi & al., 2013, 2014; Wickett & al., 2014; Zeng L. & al., 2014).

A number of studies have taken advantage of the rich resources of nuclear genes in transcriptomes and whole genomes, and mined single-copy nuclear markers for use in phylogenetic analyses at various taxonomic levels (e.g., Rothfels & al., 2013; also see Lemmon A.R. & al., 2012). In particular, transcriptome assemblies have been used to help design target enrichment probes for many exons (e.g., Mandel & al., 2014, 2015; Weitemier & al., 2014). This approach takes advantage of the power of RNA-Seq in efficiently obtain-

ing the expansive nuclear gene sampling and the stable DNA resources for enabling large taxon sampling.

Beyond its utility in phylogenetics and gene discovery, transcriptomics has been applied in other areas of systematic biology in the last few years (Fig. 1). Transcriptome data also have been widely used for exploring character evolution, and show great promise to bridge phylogeny with gene expression differences and adaptation (e.g., Garg & al., 2013; Xin & al., 2013; Hileman, 2014; Yoo & Wendel, 2014; Zhang D. & al., 2014) (see Table 2 below). Many studies have employed paralogous information from transcriptome data to explore the evolution of paleopolyploidy (e.g., Barker & al., 2008, 2009; McKain & al., 2012). Furthermore, comparative transcriptomics has provided important insights into detecting horizontal gene transfer (e.g., Li F.-W- & al., 2014b; Zhang D. & al., 2014). The 1KP (the 1000 Plants Initiative) collaborative transcriptome project has also generated large-scale gene sequence information for many different species across the entire plant tree of life (http://www.onekp.com/). This review emphasizes the utility of transcriptome data in phylogenetic inferences and character evolution, and discusses the overall challenges and prospects of transcriptome data for plant systematics.

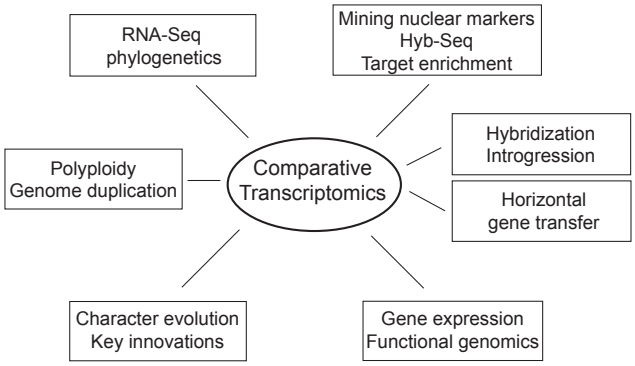

Fig. 1. Major areas of application of comparative transcriptomics in plant systematics.

RNA-SEQ PHYLOGENETICS—A SURVEY OF STUDIES

A wide range of recent studies have used transcriptome data from multiple taxa to generate nuclear marker sets conserved among these taxa with which to generate phylogenetic hypotheses (we herein refer this approach as RNA-Seq phylogenetics; methods explained in further detail in the "Analytical Strategies" section). Transcriptomes are useful for compiling nuclear marker

sets even without the comparison to published genomes. Large datasets of single-copy orthologous genes extracted from de novo transcriptome assemblies have been used to resolve phylogenetic relationships. Primarily they have been applied to questions of deep divergences and, to a lesser extent, to some evolutionary radiations in various groups of organisms. A few case studies are discussed below (see Table 1 for the available examples).

Resolving phylogenetic relationships of deep divergences and radiations. — The identification of the sister lineage to land plants has been highly controversial and represents one of the deepest evolutionary divergences in plants. Finet & al. (2010) focused on 77 ribosomal protein genes extracted from transcriptomes, which supported the charophyte order Coleochaetales as the sister lineage to land plants. A subsequent transcriptomic analysis by Wodniok & al. (2011) used an a priori set of selected genes of six species of charophyte taxa, and found support for a tree topology with Zygnematales sister to the land plants. Without the a priori gene selection, Timme & al. (2012) conducted a 160-nuclear-gene phylogenomic analysis from transcriptome data that also supported the Zygnematales as the closest living relative to land plants. The most recent phylotranscriptomic analysis of the origin and early diversification of land plants (Wickett & al., 2014) employed a much broader taxon set (transcriptome data for 92 streptophyte taxa along with 11 published plant genomes) and many more nuclear genes (up to 852 nuclear genes and 1,701,170 aligned sites) than the previous land plant studies (Timme & al., 2012), which were more algae-centric. A sister-group relationship between land plants and the green algae Zygnematophyceae still was robustly supported (Wickett & al., 2014). Mosses, liverworts and hornworts were each supported to be monophyletic, but the clade of liverworts and mosses was, however, inconsistent with the widely accepted view of early land plant evolution. This study by Wickett & al. (2014) also performed a series of analyses to assess systematic errors in phylogenetic inference caused by sampling artifacts and model misspecification for the large transcriptomic dataset.

The deep relationships of the eight major lineages of the mollusks (snails, clams, octopuses, and their close relatives) were highly controversial when molecular phylogenies used a few genes. Two independent studies (Kocot & al., 2011; Smith S.A. & al., 2011) have employed the RNA-Seq phylogenetic approach, yielding essentially the same results concerning the deep backbone relationships: Bivalvia, Gastropoda and Scaphopoda formed a clade (the B-G-S clade); Cephalopoda was sister to the B-G-S clade (the C-B-G-S clade or the Conchifera clade); Neomeniomorpha and Chaetodermomorpha constituted a clade, that was sister to Polyplacophora (the N-C-P clade or the Aculifera clade); and the Aculifera clade was sister to the Conchifera clade

Table 1. Phylogenetic case studies using transcriptome data.

Group	Subgroup	Category	Reference
Plants	Land plants (Charophytes)	77 ribosomal genes extracted from transcriptomes—RNA-Seq phylogenetics	Finet & al., 2010
Plants	Land plants (Charophytes)	Selected genes from transcriptomes—RNA-Seq phylogenetics	Wodniok & al., 2011
Plants	Land plants (Charophytes)	RNA-Seq phylogenetics	Timme & al., 2012
Plants	Land plants	RNA-Seq phylogenetics	Wickett & al., 2014
Plants	Seed plants	RNA-Seq phylogenetics and whole-genome sequencing	Xi & al., 2013
Plants	Angiosperms	RNA-Seq phylogenetics and whole-genome sequencing	Zeng L. & al., 2014
Plants	Angiosperms	RNA-Seq phylogenetics and whole-genome sequencing	Xi & al., 2014
Plants	Vitaceae (the grape plant family)	RNA-Seq phylogenetics	Wen & al., 2013a
Plants	Brassicaceae (the mustard plant family)	RNA-Seq phylogenetics	Ma & al., 2015
Plants	Betulaceae (the birch plant family)	RNA-Seq phylogenetics	Ma & al., in press
Plants	*Flaveria* (Asteraceae)	RNA-Seq phylogenetics	Lyu & al., 2015
Plants	Parasitic plant family Hydnoraceae	14 nuclear markers extracted from transcriptome data	Naumann & al., 2013
Plants	Compositae (the sunflower plant family)	Targeted sequence capture	Mandel & al., 2014
Plants	Milkweeds (*Asclepias*)	Hyb-Seq: targeted gene enrichment	Weitemier & al., 2014
Plants	Ferns	Mining single-copy nuclear markers	Rothfels & al., 2013
Arthropods	Insects: Mosquitoes	RNA-Seq phylogenetics	Hittinger & al., 2010

Table 1. Continued.

Group	Subgroup	Category	Reference
Arthropods	Insects: Advanced moths and butterflies	RNA-Seq phylogenetics	Bazinet & al., 2013
Arthropods	Insects: Hymenoptera (ants, bees, wasps)	RNA-Seq phylogenetics	Johnson B.R. & al., 2013
Arthropods	Dating deep divergences of arthropods	RNA-Seq phylogenetics	Rehm & al., 2011
Arthropods	Harvestmen (Arachnida, Opiliones)	RNA-Seq phylogenetics	Hedin & al., 2012
Arthropods	Millipedes	Ordinal-level RNA-Seq phylogenetics	Brewer & Bond, 2013
Arthropods	Centipedes	RNA-Seq phylogenetics	Fernández & al., 2014
Ecdysozoa	Arthropoda, Nematoda, and several smaller phyla	RNA-Seq phylogenetics	Borner & al., 2014
Mollusks	Deep relationships	RNA-Seq phylogenetics	Kocot & al., 2011; Smith S.A. & al., 2011
Annelids	Basal radiation of annelids	RNA-Seq phylogenetics	Struck & al., 2011; Weigert & al., 2014
Invertebrates	Ten non-model invertebrate species across five phyla (Annelida, Arthropoda, Mollusca, Nemertea, Porifera)	Comparative transcriptomics	Riesgo & al., 2012
Myzostomids	A group of symbiotic (or parasitic) protostomes that are either placed with annelids or flatworms	RNA-Seq phylogenetics	Hartmann & al., 2012
Metazoans	Sponges (Porifera): eight sponge species in four classes (Hexactinellida, Demospongiae, Homoscleromorpha, Calcarea)	RNA-Seq phylogenetics	Riesgo & al., 2014
Tetrapods	Rodents	RNA-Seq phylogenetics	Lin & al., 2014
Tetrapods	Deep relationships	RNA-Seq phylogenetics	Chiari & al., 2012

(cf. Kocot & al., 2011, and Smith S.A. & al., 2011). Furthermore, the deep relationships were consistent across analytical methods, phylogenetic inference programs, and gene number. These congruent and consistent phylogenetic results on an extremely difficult group were persuasive in assuring the systematics community of the resolving ability of RNA-Seq data for discerning relationships of deeply divergent taxa.

The deep basal radiation of the annelids was resolved based on 68,750–170,497 amino acid sites from 305–622 proteins extracted from transcriptome data of 60 annelid species, representing 39 annelid families (about one-third of the total number of approximately 125 annelid families) (Weigert & al., 2014; also see Struck & al., 2011). Chaetopteridae, Amphinomidae, Sipunculidae, Oweniidae, and Magelonidae were placed in the basal part of the annelid tree. Myzostomida was previously suggested to belong to the basal radiation, and was found to be nested within Annelida as sister group to Errantia in most analyses. This largest dataset of annelids provided a robust phylogeny overall.

The phylogenetic relationships of Ecdysozoa (arthropods, nematodes, and several smaller phyla, Dunn & al., 2008) were inferred based on a multigene dataset of 63 taxa and 24,249 amino acid positions (Borner & al., 2014). Phylogenetic analyses employing various models supported the monophyly of Ecdysozoa and a clade constituting Priapulida and Kinorhyncha (i.e., Scalidophora) as the first diverged branch within Ecdysozoa. Arthropoda were suggested to be allied with Nematoda and Tardigrada. Most clades within Arthropoda were strongly supported. However, the relationships within the Euchelicerata were largely unresolved. The Bayesian and maximum likelihood analyses using slowly evolving genes recovered Tardigrada as a sister group to Arthropoda, but analyses of the full dataset, and of subsets containing genes evolving at fast and intermediate rates identified a clade of Tardigrada and Nematoda.

Peters & al. (2014) constructed the evolutionary history of holometabolous insects using amino acid sequence data of 1343 single-copy orthologous genes derived from de novo transcriptome sequencing. Hymenoptera are placed as the sister group to all remaining holometabolan orders. The results showed several other highly supported relationships within Neuropterida and Antliophora based on the transcriptome data, although the monophyly of Coleopterida (Coleoptera and Strepsiptera) remains ambiguous in the analyses of the transcriptome data. The study generally does demonstrate the utility of well-resolved phylogenies and well-documented extensive morphological datasets for reconstructing complex morphological transformations in insects.

Bazinet & al. (2013) used RNA-Seq data to test if such data could increase the support for relationships among the advanced moths and butterflies (insect superfamilies of the Apoditrysia clade) over that found in a previous 19-gene study (Regier & al., 2013). Phylogenetic analysis of a dataset with 46 taxa and 741 genes yielded an overall increase in bootstrap support for deeper nodes within Apoditrysia as compared to results from the 19-gene analyses. Strong support was primarily restricted to the large subclade Obtectomera, in which 11 of 12 nodes subtending multiple superfamilies had bootstrap support of 100%. The strongly supported nodes showed little conflict with groupings from previous studies, and were not affected by different taxon sampling strategies, suggesting that they reflect true phylogenetic signal rather than artifacts of massive gene sampling. However, strong support was obtained for only two of the eleven deeper nodes among the "lower" non-obtectomeran apoditrysians. The results in this case may be related to the higher resolution power of transcriptome data at an intermediate depth, rather than very deep in the phylogeny.

Relationships at taxonomic levels with divergences within 100 million years. — Few attempts have been made to use RNA-Seq phylogenomic data to construct phylogenies at the intermediate or low taxonomic levels with divergences within 100 million years. With potentially more serious issues related to orthology inference at very deep levels and the potential lack of sufficient variation in expressed single-copy orthologous genes across very closely related taxa, ultimately the utility of RNA-Seq data may turn out to be most productive at these intermediate phylogenetic levels.

Hittinger & al. (2010) pioneered the use of non-normalized transcriptomes in phylogenetic inference and showed that large phylogenetic data matrices can be accurately assembled de novo from short transcript sequences of ten species of mosquitoes (the genus *Aedes*). These matrices of primarily highly expressed genes contained very few orthology assignment errors. Phylogenetic inferences based on the RNA-Seq data matrices were robust for this clade of *Aedes*, that was estimated to have diverged from its sister genus *Eretmapodites* at 64.44 Ma with the credibility interval of 94.85–38.93 Ma (Reidenbach & al., 2009).

Wen & al. (2013a) resolved the intergeneric relationships of the grape family (Vitaceae), a member of the basal rosid clade of the flowering plants, with the oldest fossil record for the family dated in the late Cretaceous around 66 Ma (Manchester & al., 2013) and the split between Vitaceae and its sister Leeaceae dated to about 95 Ma of the mid Cretaceous (Wen & al., 2013a). The study employed 417 orthologous single-copy nuclear genes from the de novo transcriptome assemblies of 15 species of the family, spanning the generic diversity

of the family. The 417-gene phylogeny provided robust support for the deep relationships, showing the phylogenetic utility of transcriptome data for plants over a time scale at least since the mid-Cretaceous. Yang Y. & Smith (2014) re-analyzed the Vitaceae dataset and used different strategies for orthology inference. They obtained much larger ortholog sets with full taxon occupancy compared to the 417 (before filtering) and 229 (after filtering) orthologs of Wen & al. (2013a) using the program Hcluster_sg (Li H. & al., 2006). Overall the topology for the grape family from the re-analysis by Yang Y. & Smith (2014) was well supported and congruent with that reported by Wen & al. (2013a), but the reanalysis suggests the need to test the position of the *Cissus* clade, which also had a long branch in the analyses by Wen & al. (2013a).

A few studies conducted comparative transcriptome analyses at even lower taxonomic levels, e.g., on *Camelina sativa* (Liang & al., 2013), *Artemisia tridentata* (Bajgain & al., 2011), *Ranunculus* (Pellino & al., 2013), and *Flaveria* (Lyu & al., 2015). RNA-Seq is potentially useful at lower taxonomic levels from a technical point of view or the data itself, but it is rather expensive in comparison to RAD-Seq or genotyping-by-sequencing (GBS) data (Zimmer & Wen, in press).

In summary, most studies utilizing the direct transcriptome approach have been at the higher taxonomic ranks or involving deep divergences. Orthology inference is one of the most critical and challenging steps in RNA-Seq phylogenetics and can be problematic in transcriptomes that include incomplete genomic data and contain errors and multiple isoforms and datasets lacking appropriate outgroups (Ebersberger & al., 2009; Kocot & al., 2013; Yang Y. & Smith, 2014). The utility of orthologous gene sequences obtained directly from transcriptome data at lower taxonomic ranks such as within a plant genus or among closely related genera needs to be rigorously explored (Lyu & al., 2015). The promising Hyb-Seq approach as described in Weitemier & al. (2014) enables targeted sequencing of hundreds or thousands low-copy nuclear exons and the flanking regions for phylogenetic inference. Nevertheless, for any particular study it is necessary to rigorously test the approximate phylogenetic depth for using RNA-Seq data, that is to determine the utility of comparative transcriptomics across a clade and its outgroup(s). Furthermore, as RNA-Seq datasets are usually much bigger than other types of phylogenetic datasets, they add more computational complications. Most commonly used phylogenetic programs will need to be improved to handle large phylogenomic datasets. There are also analytical challenges concerning the fact that large RNA-Seq datasets still may lead to erroneous but well-supported results (Wen & al., 2013a; Davis & al., 2014). See the section on Analytical Strategies in this chapter for more information.

MINING SINGLE-COPY NUCLEAR MARKERS FROM TRANSCRIPTOMICS

Phylogenetic reconstruction of evolutionary histories in plants has historically been based largely on plastid DNA markers such as *matK*, *atpB*, *rbcL*, *trnL-F* and *psbA-trnH*. In part, this has been due to the effectively non-recombining nature of the chloroplast genome that minimizes problems of paralogy and multiple amplicons, as well as ease of amplification of plastid markers due to their high copy number within the cell. These plastid sequences have been the foundation for many of the phylogenies underlying the ongoing revision of classification in angiosperms (Angiosperm Phylogeny Group, 2009). Mitochondrial genes, such as 16S, CO1, and Cytb, that evolve very rapidly have been widely used as phylogenetic markers in animalia (e.g., Nicolas & al., 2012; Mandal & al., 2014), unlike those of plant mitochondria, that evolve very slowly. Phylogenetic hypotheses based on organellar markers are not without problems, as plastid markers are usually maternally inherited (Zimmer & Wen, 2012), fail to resolve many divergences (Ren & al., 2014) and, like mitochondrial markers in animals, they are in some cases incongruent with nuclear markers due to processes such as chloroplast capture and hybrid speciation (Pillon & al., 2013; Yi & al., 2015).

The use of chloroplast DNA for phylogeny and divergence dating can be particularly problematic for non-photosynthetic, parasitic plants due to the breakdown of the chloroplast genome likely due to a decrease in selective constraint. Transcriptomics has helped overcome this barrier. Naumann & al. (2013) extracted 14 single-copy nuclear genes from the transcriptome of *Hydnora visseri* (Hydnoraceae) a highly derived, haustorial parasitic plant, by comparison with a single-copy nuclear gene dataset compiled by Duarte & al. (2010) from the *Arabidopsis*, *Vitis*, *Populus*, and *Oryza* genomes. This top-down approach allowed the placement of Hydnoraceae within Piperales and the first divergence date estimates for several parasitic plant lineages.

Until the advent of next-generation sequencing technologies, the creation of nuclear markers for phylogeny reconstruction had been slow due to the technical challenges of creating primers that target known single-copy, orthologous gene regions. This has been particularly challenging for plant biologists due to the prevalence of whole-genome duplications over plant evolutionary history. While plastid genomes have served the phylogenetic community well, they are often less variable than nuclear markers and present only a single parental history. Genomes and transcriptomes are increasingly being leveraged for phylogenomics as a source for the discovery and creation of low- or single-copy nuclear phylogenetic markers and present an unprecedented means of marker development for phylogenetics (Hughes & al., 2006;

Egan & al., 2012). Genome comparison among model organisms was the first step to developing comparative anchor-tagged sequences (CATS; Lyons & al., 1997) in mammals, or conserved ortholog sets (COS markers; Fulton & al., 2002) in plants, with comparisons among expressed sequence tags (ESTs) and model genomes following soon thereafter (e.g., in Compositae: Chapman & al., 2007).

Comparative transcriptomics has been particularly helpful in creating low- or single-copy nuclear marker sets for use in phylogenetic analyses. Rothfels & al. (2013) used a top-down approach to mining single-copy nuclear markers from fern transcriptomes from the 1 KP project (https://sites.google.com/a/ualberta.ca/onekp/). By first choosing genes of interest or markers from a series of putatively single-copy genes identified by the 1KP project team, they analyzed 62 fern transcriptomes using various scripts to optimize transcriptome alignments and screened potential markers against a set of 15 fern taxa, eventually creating 20 primer pairs spanning 10 nuclear regions for use as phylogenetic markers, effectively tripling the number of nuclear regions available for fern phylogenetics.

At the beginning of the genomics era, Hughes & al. (2006) expressed concerns over the lack of multi-locus nuclear phylogenetic analyses at the species level in plants. Transcriptomics, especially when coupled with low-coverage genome skimming or published genomes of close relatives, offers a particularly powerful resource for large-scale phylogenomic analyses that incorporate low-copy or single-copy as well as higher-copy nuclear markers and can be particularly useful at lower taxonomic levels. The Hyb-Seq technique outlined by Weitemier & al. (2014) used a transcriptome assembly and a draft genome of milkweed, *Asclepias syriaca*. It resulted in a target enrichment probe set for thousands of low-copy nuclear exons and their flanking regions from over 700 gene regions. This probe set was then used to amplify target regions across 10 *Asclepias* species and two related genera, resulting in at least partial assembly of 92.6% of targeted exons and 99.7% of the 768 gene targets, showing an incredible resolving power for lower-level phylogenetic studies. This similar technique has been shown to be successful for plants in the sunflower family Compositae (e.g., Mandel & al., 2014, 2015).

An exon-based approach is desirable for plants due to the repetitive nature of many plant genomes. For example, the *Helianthus annuus* (sunflower) genome is approximately 3 Gbp, 81% of which consists of transposable elements, 77% of which are long terminal repeat (LTR) retrotransposons (Natali & al., 2013). This not only makes genome assembly very difficult, but limits the possibilities of finding loci outside of gene regions. In vertebrates, ultraconserved elements (UCEs) have been harvested from intergenic regions

(Faircloth & al., 2013). This has been desirable for systematists looking for putatively neutral markers. UCEs are detected by pairwise whole genome alignments and subsequently probes are designed (as in the exon capture approach) and then applied to a sampling of a large number of taxa. Whole-genome alignments across plant genomes are possible, but likely to yield the most conservation in identical or semi-identical repeats, that are not appropriate phylogenetic markers. Very few UCEs have been isolated from plants up until now (Reneker & al., 2012) and it is likely that a continued focus on exons and their flanking regions will be more successful.

DETECTION AND CHARACTERIZATION OF POLYPLOID SPECIATION

Development of primers for single-copy nuclear markers can be very challenging, especially for plants, largely due to the high prevalence of polyploidy in plant evolution. It is estimated that 15% of all flowering plant speciation events result from polyploidy (Wood & al., 2009) and that all seed plants were derived from a polyploid lineage (Jiao & al., 2011). Polyploidization results in a complete duplication of the genome, initially creating duplicates of every gene region, making the idea of a single-copy nuclear gene seem almost a myth. Over time, however, the number of retained duplicates decreases due to evolutionary mechanisms such as gene conversion and loss, thereby eroding the signal of polyploidy and prospectively retaining those genes denoted as single-copy.

Because transcriptomics allows massively parallel sequencing of expressed genes within a single genome, it offers a powerful means of investigating the signals for and evolutionary implications of polyploidy within species or among taxa. The classical signal of polyploidy is a duplicated chromosome number relative to related taxa. Another possible signal of whole-genome duplication is the existence of bimodal chromosome size distributions, in which chromosomes fall into two distinct size classes, a phenomenon that has been observed in a number of plant lineages (e.g., in Goldblatt & Takei, 1997; Pires & al., 2006). McKain & al. (2012) used comparative transcriptomics to test the hypothesis that the origin of chromosomal bimodality in Agavoideae coincided with a polyploidy event. As in previous studies (e.g., Blanc & Wolfe, 2004; Egan & Doyle, 2010), they used the distribution of synonymous substitution rates (K_S) between duplicated genes within a transcriptome to detect and characterize polyploidy events present along the evolutionary lineages of selected species and combined K_S plot analyses with gene family phylogenies to show that the bimodal karyotype was associated with an ancient polyploidy

event. A well-known example of recent polyploidy has occurred within the genus *Glycine*, to which soybean (*G. max*) belongs, as well as ~30 other wild perennial species, including the *G. tomentella* species complex. Bombarely & al. (2014) sampled three *Glycine* allopolyploid "triads" (an allopolyploid species and its two putative diploid progenitors) and compared SNP data within and across their transcriptomes to help elucidate the origins of the allopolyploids, finding evidence for multiple origins thereof that occurred within the last several hundred thousand years and confirming the reticulate nature of this species complex.

Transcriptomics presents a powerful tool to investigate the evolutionary impacts of complex polyploid systems. *Tragopogon* (Asteraceae) includes a species complex created through repeated hybridization and polyploidization and presents an excellent model system for studying the impact of polyploidy on genome structure. *Tragopogon mirus* and *T. miscellus* are two allopolyploid species that have arisen repeatedly over the last ~100 years via allopolyploidy. Next-generation transcriptomics and genomics have enabled considerable advances in understanding what happens within the genome immediately following polyploidy in terms of gene loss, segregation, and conservation (Buggs & al., 2012). The hybridization of hexaploid *Spartina maritima*, a grass species indigenous to England, with the hexaploid *S. alterniflora*, a species introduced to England from the eastern United States, resulted in the sterile hybrid, *S.* ×*townsendii*. Genome duplication within *S.* ×*townsendii* subsequently produced the dodecaploid *S. anglica*, a highly invasive species (Ainouche & al., 2004). Salmon & Ainouche (chapter 3) are using transcriptomics to overcome the challenges of nested, recent polyploidy presented by the *Spartina* polyploid complex, a work that may elucidate the unusual relationship between polyploidy and invasiveness.

An investigation of polyploid speciation using transcriptomics can be difficult at the assembly level due to the duplicated nature of the polyploid genome, especially for de novo assemblies in nonmodel organisms. This has to do with the ability to unambiguously determine a read's provenance, i.e., from which homeolog it was derived (Ilut & al., 2012). This is particularly difficult for shorter read sequencing technologies because the accumulation of homeolog-specific SNPs on any given read will be few to none compared to longer reads. As NGS technologies produce significantly longer read lengths in the future, this will become less of a problem. As in nearly all phylogenetic analyses, the issue of paralog/ortholog conflation is present but magnified many-fold for polyploid speciation studies using transcriptomics, because now assemblers and ortholog detection methods are required to not only differentiate between paralogs created through tandem duplication or other pro-

cesses but also between homoeologues (paralogs duplicated via polyploidy) duplicated genome wide. The more recent the polyploidization, the more difficult this task is due to few accumulated differences over time. As assemblers and ortholog detection methods become more discriminating, these problems will lessen.

Challenges and limitations notwithstanding, transcriptomics has literally revolutionized the study of polyploidy (a search in Google Scholar for "transcriptomic and polyploidy" OR "whole genome duplication" returned 510 hits prior to 2006 and 4920 from 2006 to 15 October 2014), enabling detection, characterization, and hypothesis testing regarding polyploidy and its effects on lineage diversification. To do real justice to this topic would require a chapter in and of itself and is beyond the scope of this work. Suffice it to say that transcriptomics will continue to provide raw material (homeologs, paralogs, orthologs, gene expression data, etc.) that will enable fine- and broad-scale phylogenetic analyses to enhance our understanding of polyploidy and its impacts on speciation and evolution.

CHARACTER EVOLUTION AND KEY INNOVATIONS

In addition to providing enhanced phylogenetic resolution through generation of a large number of single-copy nuclear gene markers, transcriptome sequencing has produced information on a number of evolutionarily important variations in gene expression, thought to be the driving force in organismal change (King & Wilson, 1975). These studies include identification of: (1) genes expressed in different developmental stages and/or different tissues of a single or related sets of species; (2) identification of genes implicated in physiological changes such as increased tolerance to heavy metals, salinity or disease-causing organisms; (3) gene expression leading to changes in morphology; (4) identification of genes that may have played a role in domestication of various plants and animals; and (5) changes that impact genome evolution—tandem duplication, alternative mRNA splicing, microRNA profiles and karyotypic complexity. Examples with relevant citations for these five categories are presented here as Table 2. Some of these changes that may result in new morphological or physiological phenotypes leading to increased biological diversification have been termed "key innovations" (Ree, 2005). Among those for which transcriptome data support the existence of a key innovation are the nectar spurs of columbines (Ree, 2005; Sharma & al., 2014), the tendrils of the grape family (Zhang N. & al., 2015), the pollinator deception of orchids (Sedeek & al., 2013), the tolerance to shady environments of

Table 2. Comparative transcriptomics and character evolution.

Feature examined with comparative transcriptomics	Organism(s)	Reference
Developmental differences		
Gamete development	Ferns	Der & al., 2011
Leaf sepal petal expression	Morning glory	Guan & Lu, 2013
Developmental specialisations	Legumes	Hofer & Ellis, 2014
Leaf development	Tomatoes	Ichihashi & al., 2014
Floral initiation	*Arabidopsis*	Kaufmann & al., 2013
Floral initiation	Orchids	Liu Z. & al., 2013
Complex nervous systems	Ctenophores	Moroz & al., 2014
Petals	Cotton	Rambani & al., 2014
Tendrils and inflorescences	Grape family	Zhang N. & al., 2015
Physiological differences		
Metabolism	Cucumber	Ando & Grumet, 2012
Metabolism	Algae	Chan & al., 2012
Color variegation	Peach	Chen & al., 2014
Silk production	Black widow spiders	Clarke & al., 2014
C4/CAM photosynthesis	*Portulaca*	Christin & al., 2014
Salinity and submergence tolerance	Wild rice	Garg & al., 2013
Latex production	Various flowering plants	Hagel & al., 2008
Latex production	Rubber tree	Ko & al., 2003
Toxin production	Black widow spiders	Haney & al., 2014
Circadian rhythms	*Arabidopsis*	Hsu & al., 2013
Secondary metabolism in adventitious roots	Ginseng	Jayakodi & al., 2014
Cyclotide production	Coffee family	Kohlbach & al., 2013
Fe/Zn/Cd tolerance	Mustard family	Li J. & al., 2014
Sulfur dipeptide production	Legume family	Liao & al., 2013

Table 2. Continued.

Feature examined with comparative transcriptomics	Organism(s)	Reference
Volatiles	Strawberry	Sanchez-Sevilla & al., 2014
Metabolic pathways	Tea	Shi & al., 2011
Phosphotyrosine signaling	Holozoans	Suga & al., 2014
Iron-copper crosstalk	Melon	Waters & al., 2014
Disease resistance	Eggplant/Turkey berry	Yang X. & al., 2014
Alkaloid biosynthesis	Poppy family	Zeng J. & al., 2013
Fruit ripening	Tomato	Zouari & al., 2014
Temperature tolerance	Coffee	Combes & al., 2013
Morphology/phenotype		
Invasiveness	Yellow star thistle	Dlugosch & al., 2013
Eusociality	Wasps	Ferreira & al., 2013
Urban vs. rural forms	Deer mice	Harris & al., 2013
Bilateral symmetry	Snapdragon	Hileman, 2014
Leaf shape	Various tomato species	Ichihashi & al., 2014
Feather forms	Chicken; Zebra finch	Ng & al., 2014
Vasculature	Conifers	Raherison & al., 2012
Eye evolution	Cephalopods	Sousounis & al., 2013
Inflorescence architecture	*Arabidopsis*; rice; tomato	Teo & al., 2013
Flower whorls	Columbines vs. *Arabidopsis*	Voelckel & al., 2010
Self-incompatibility	*Petunia*	Williams & al., 2014
Fiber forms	Cotton	Yoo & Wendel, 2014
Domestication		
	Animals	Albert & al., 2012
	Beans	Bellucci & al., 2014
	Barley	Dai & al., 2014
	Pea	Franssen & al., 2011

Table 2. Continued.

Feature examined with comparative transcriptomics	Organism(s)	Reference
	Chickpea	Hiremath & al., 2011
	Sunflower family crops	Hodgins & al., 2014
	Cows	Huang & al., 2012
	Tomatoes	Koenig & al., 2013
	Wheat	Singh & al., 2014
	Rice hybrids	Xu C. & al., 2014
	Eggplant; turkey berry	Yang X. & al., 2014
Genome evolution		
Conserved noncoding elements	Eudicots	Burgess & Freeling, 2014
Transcriptomic shock	Whitefish	Dion-Cote & al., 2014
Novel microRNAs	Cichlids	Elmer & al., 2010
Heterosis	Wheat	Li A. & al., 2014
Karyotypic change	Barley	McKain & al., 2012
Genome dominance	Mustard family	Parkin & al., 2014
RNA editing	Human	Peng & al., 2012
Genomic complexity	Sponges	Riesgo & al., 2014
MicroRNA evolution	Land plants	Taylor & al., 2014
Alternative RNA splicing	Poplar; eucalyptus	Xu P. & al., 2014
Asexual genomes	Buttercups	Pellino & al., 2013
Adaptive radiations/key innovations		
Parallel coloration changes	East African cichlids	Brawand & al., 2014
Parallel jaw changes	East African cichlids	Brawand & al., 2014
Freshwater/marine species	Sticklebacks	Guo & al., 2013
Pollinator attraction	Orchids	Sedeek & al., 2013
Petal spurs	Columbines	Sharma & al., 2014

ferns (Li F.-W. & al., 2014), the feeding morphology and coloration changes in East African cichlids (Brawand & al., 2014), and the freshwater to marine transitions in sticklebacks (Guo & al., 2013).

ANALYTICAL STRATEGIES

The analytical framework for utilizing transcriptome data in phylogenetics is depicted in Fig. 2. While there are a number of software packages for each of the following steps, highlighted here are those that are most commonly used. These will undoubtedly change over time. Raw read trimming, thinning, and assembly are standard to any subsequent downstream transcriptome analysis, whether one is considering a single taxon or many taxa. These are often accomplished with Trimmomatic (Bolger & al., 2014) or PRINSEQ (Schmieder & Edwards, 2011). Trinity (which has many embedded tools: Altschul & al., 1990; Krogh & al., 2001; Lagesen & al., 2007; Finn & al., 2011; Grabherr & al., 2011; Petersen & al., 2011; Kanehisa & al., 2012; Powell & al., 2012; Punta & al., 2012), SOAPdenovo-Trans (Xie & al., 2014), Trans-ABySS (Robertson & al., 2010) are three de novo assemblers specifically designed for transcriptome data (i.e., different from whole-genome data in that one expects disjunct transcripts as results rather than long contiguous scaffolds/chromosomes). Transcriptome annotation tools include Trinotate (part of Trinity) and Blast2GO, which uses Gene Ontology terms (The Gene Ontology Consortium, 2000) to classify genes into functional categories. These predictive tools will improve as more and more transcriptomes are completed, gene function is better understood, and databases become more fully populated.

Differences between transcripts and DNA gene sequences. — In addition to choosing the appropriate analytical tools for a particular dataset, there are underlying theoretical complications owing to the difference between transcripts derived from RNA and sequences derived from genomic DNA. Transcripts, while representative of the expressed genes at a particular time point or developmental stage, are not in a 1:1 relationship to the DNA sequences from which they arise. Alternative splicing is a phenomenon by which a particular gene (DNA) sequence produces multiple proteins by differentially including or excluding exons in multiple product mRNAs (Black, 2003). This at least partly explains how organisms have kept gene count relatively low while the diversity of function and structure has increased through evolutionary time. This phenomenon is particularly common in higher eukaryotes and why RNA-Seq can produce a much larger number of transcripts than their purported genes in the genome. Even a baseline (normalized) transcriptome, in

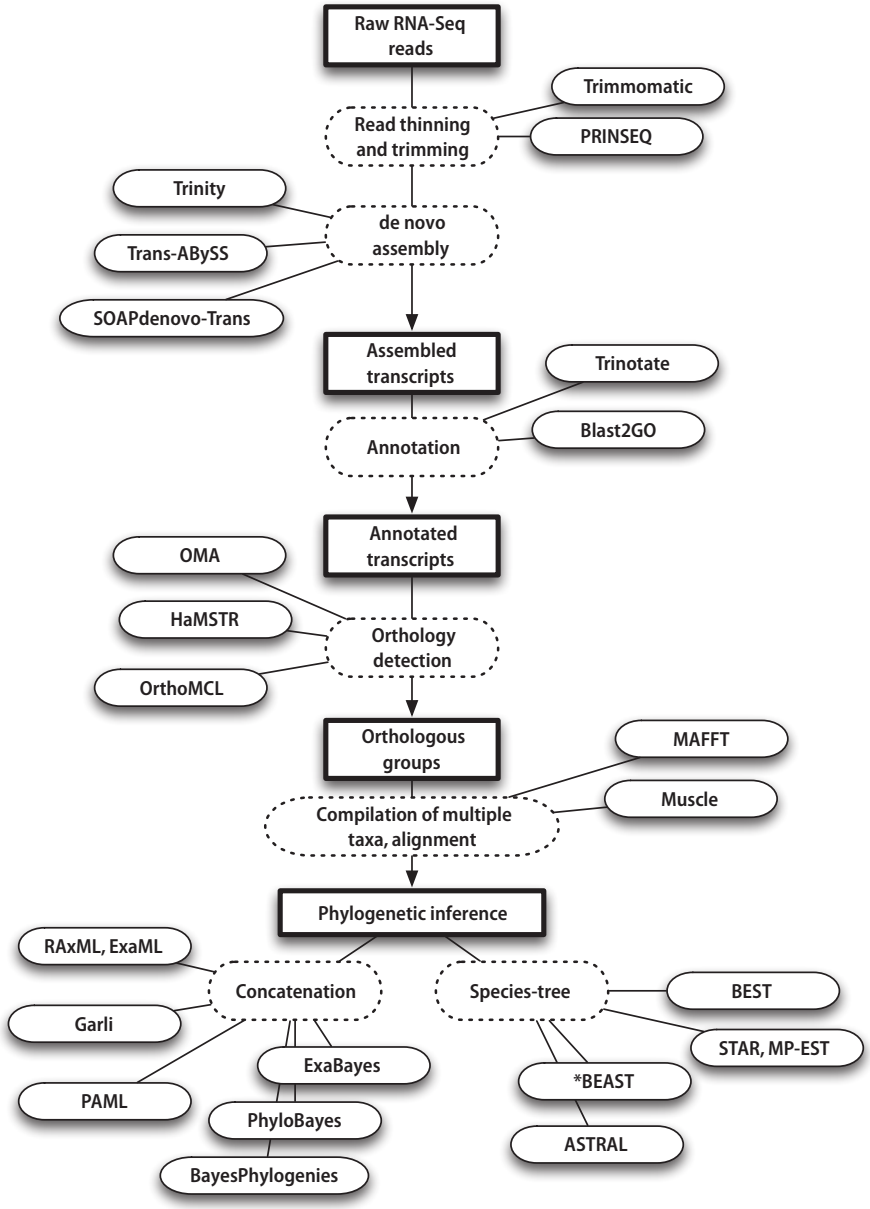

Fig. 2. Bioinformatics workflow for generating phylogenetic hypotheses from raw RNA-Seq data highlighting some of the most commonly used software packages. Citations for software programs and in depth discussion of each step are found in the "Analytical Strategies" section.

which expression level is ignored, hypothesizes a set of splice variants through a de novo assembly that cannot be "verified" by a genomic reference. This is in contrast to how transcriptomes can aid in the assembly and annotation of genomic sequence. Normalized transcriptomes are much more inexpensive to gather than complete genomes and can improve genomes without adding much to the analytical and sequencing cost.

The assembly of these transcripts, due to alternative splicing, is complicated in ways that are different for genome assembly. De novo genome assembly, while challenging in terms of repetitive DNA (particularly in plants; e.g., in sunflower, the 3 Gbp genome consists of 81% transposable elements, 77% of which are identical long terminal repeat [LTR] retrotransposons; Natali & al., 2013), and duplicated genes (a single read can map to multiple locations with equal confidence), but aims to piece together a much more static entity than transcriptome assembly. The use of short-read sequencing technology exacerbates the transcript assembly challenge. Newer technologies such as single-molecule sequencing and long-read sequencing have the promise to alleviate assembly artifacts and more accurately estimate the number and set of transcripts with the correct splicing as the reads have the capacity to span an entire transcript rather than imposing an additional bioinformatics challenge (e.g., Tilgner & al., 2014).

Orthology detection and phylogenetic analysis. — For phylogenetic and downstream evolutionary analyses, hypotheses of homology, and particularly an accurate distinction between orthology and paralogy, are standard procedure. Orthologs are descended most recently from a speciation event and paralogs are descended most recently from a duplication event (Fitch, 1970). Paralogs, thus, can produce phylogenetic hypotheses that do not reflect the pattern of evolutionary relationships of the taxa included, although they are interesting from a gene and genome evolution perspective. Commonly used orthology detection algorithms include OrthoMCL (Van Dongen, 2000; Enright & al., 2002), OMA (Roth & al., 2008), and HaMSTR (Ebersberger & al., 2009). These programs rely on functional transcript annotation (generally using Gene Ontology terms; The Gene Ontology Consortium, 2000) and comparison to a database of transcripts of known function, such as with OrthoDB (Waterhouse & al., 2013), OMA (Altenhoff & al., 2011), InParanoid (O'Brien & al., 2005). There are also a number of ways that these hypotheses of orthology can then be refined based on subsequent gene trees, such as with PhyloTreePruner (Kocot & al., 2013).

Comparing depth of coverage across transcripts can be used to separate orthologs and paralogs, particularly when a closely related reference genome does not exist. The principle is that, for example, if transcript X has double

the coverage of transcript Y, it might be indicative of misassembly caused by paralogy (Hirsch & Buell, 2013). This method relies on good overall coverage and trustworthy assemblies across the taxa being considered, however.

Protein (amino acid) sequences are generally used in phylogenetic inference based on transcriptome data rather than DNA sequences. When distantly related taxa are being considered, popular opinion is that homologies can be difficult to find at the DNA level, and protein sequences suffer less from convergence and saturation due to mutational biases. Amino acid sequences are one third of that of the corresponding DNA sequence in length, which can make alignment and phylogenetic analyses faster. When one uses this approach, however, there is a loss of phylogenetic information, which could be important at lower levels of divergence. There is a balance to be struck in the decision to analyze amino acids or nucleotides, the answer to which might depend on the evolutionary timescale of interest and on the final matrix size (of both taxa and characters).

Once a set of orthologs has been identified for the taxon sampling at hand, the next concern is whether phylogenetic analysis should be approached in a multi-gene concatenated framework or a gene-tree centric species-tree approach. Advocates of concatenation argue that hidden signal can emerge from large concatenated datasets that one would not see when analyzing genes individually (e.g., Gatesy & Baker, 2005; Hittinger & al., 2010; Salichos & Rokas, 2013). Advocates of gene-tree centric species-tree approaches (sensu Edwards, 2009; Liu L. & Yu, 2011) argue that concatenation obscures the signal and evolutionary history of individual genes (or other loci) or that concatenation can return incorrect trees with high confidence (Kubatko & Degnan, 2007). A phenomenon known as "the anomaly zone" can also occur during which the most common gene tree is not the species tree (Degnan & Rosenberg, 2006, 2009; Degnan, 2013). Discordance among gene trees can be due to a number of biological processes such as incomplete lineage sorting (ILS), deep coalescence, horizontal gene transfer and hybridization, while also possibly due to non-biological reasons such as small sample-size. Sample size is an often neglected issue that can cause incorrect gene tree specification (e.g., Dikow & Smith, 2013; Betancur & al., 2014). For example, in Dikow & Smith (2013), gene trees sampled from across 22 bacterial genomes were all unique and never the same as the concatenated species tree, even when directly adjacent within the sampled genomes. In many species-tree methods, each gene or locus is treated (weighted) equally regardless of length, rather than placing the equal weight on each nucleotide. How we can differentiate between incongruence due to an evolutionary process (e.g., ILS) and that due to small sample size or incorrect homology detection remains unclear. In addition, we are still

somewhat lacking in an understanding as to how these issues scale with different evolutionary divergences. ILS has been recognized at the species level and below for a long time (Maddison, 1997), but recently the effects of these processes across evolutionary time on a phylogenetic tree of more distantly related species is just beginning to be addressed (e.g., Springer & Gatesy, 2014).

The species-tree approach is currently implemented in two ways: summary methods and co-estimation methods. Summary methods are those in which gene trees are estimated independently and input into species-tree software programs as newick files with no associated sequence alignment matrix (as with BUCKy: Ane & al., 2007, Larget & al., 2010; ASTRAL: Mirabab & al., 2014; STAR: Liu L. & al., 2009; MP-EST: Liu L. & al., 2010). Co-estimation methods are those in which gene trees are estimated simultaneously with species trees (as with *BEAST: Heled & Drummond, 2010; Drummond & al., 2012; BEST: Liu L., 2008). The co-estimation methods are computationally prohibitive for large datasets (i.e., more than 100 loci: Bayzid & Warnow, 2013; Smith B.T. & al., 2014). Many of these methods (e.g., BUCKy, *BEAST, BEST) do not allow for any taxa to be missing any loci, which can dramatically reduce the size of a dataset depending on the evolutionary divergence among the taxon sampling of interest. A comprehensive review of ILS, and the gene-tree–species-tree debate is beyond the scope of this chapter, but the work cited above provides an introduction to these issues that have become more stark as phylogenetic datasets have grown so dramatically (see review by Liu L. & al., 2015).

The following are phylogenetic analysis programs that are able to analyze concatenated transcriptome-scale data (e.g., hundreds of thousands of amino acid or nucleotide base-pairs): Maximum likelihood approaches: RAxML (Stamatakis, 2006), ExaML (Stamatakis & al., 2012; Stamatakis, 2013), Garli (Zwickl, 2006), and PAML (Yang Z., 2007); Bayesian approaches: ExaBayes (Aberer & al., 2014), PhyloBayes (Lartillot & al., 2012), and BayesPhylogenies (Pagel & Meade, 2004); and maximum parsimony approaches: TNT (Goloboff & al., 2008). Assessing support for phylogenetic matrices is commonly done using statistical methods, but large matrices often produce high support values for incorrect topologies simply due to the large amount of data (e.g., Salichos & Rokas, 2013). Using individual gene-tree signals as measures of support is another possibility, but as mentioned above, gene-tree accuracy and meaning given small sample size and gene-level evolutionary processes is non-trivial. Transcriptome and genome-level phylogenetics is still quite young and has until now relied on the tools built for small single- or few-gene matrices. Sequencing technology, our understanding of genome evolution, systematics theory, and bioinformatics tools have to all catch up to one another in order to produce sound phylogenomic hypotheses.

CHALLENGES AND PROSPECTS

Practically, it is challenging to obtain RNA-grade tissues with a broad taxon sampling, as RNA is not preserved in dried or silica-preserved specimens; rather, it degrades very rapidly upon collection. Tissue samples for RNA extraction need to be fresh, flash frozen, or preserved in solutions like RNAlater and then placed in an ultracold freezer (Johnson & al., 2012; Wen & al., 2013a). Such requirements for tissue quality, coupled with the generally higher cost of generating transcriptomes, will limit the utility of the direct RNA-Seq phylogenomic approach in a phylogenetic study with extensive taxon sampling. In that context, the targeted or anchored sequence enrichment approach using DNAs and probes designed from transcriptomes (e.g., Lemmon A.R. & al., 2012; Weitemier & al., 2014) and genomes will be more useful and practical for large-scale phylogenetics requiring extensive taxon sampling, while also enabling the use of degraded DNA from such sources as herbarium or museum specimens. With this approach, RNA collecting efforts then can focus on using a few, easily collected reference taxa that span the evolutionary divergence of interest. Transcriptome data from these few taxa can then be used to identify nuclear markers for which probes can be designed and used for hybrid enrichment and sequencing across the divergence on specimens preserved in a variety of ways. Because these probes are designed from transcriptome data rather than genome data, they focus on exons. Nevertheless, the presence of usually more informative intron sequences in the resulting data may be produced with such methods, as sequencing of the genomic DNA may extend beyond the probe hybridization site. If the probe was close to an exon-intron boundary, some intronic sequence will be captured as well, a result that may actually be helpful in resolving close relationships.

The use of transcriptomics for gene expression studies or for mining single-copy nuclear markers for phylogenetic reconstruction is not without its limitations and challenges, not the least of which is the scope of the transcriptome itself. By definition, the transcriptome is a portion of the complement of coding genes comprising a genome that is transcribed at a given time, by a given tissue, and under the influence of a specific set of environmental conditions. The complement of genes expressed differs by tissue and can be affected by intrinsic (genetic) and extrinsic (epigenetic) factors that are influenced by environment (Jaenisch & Bird, 2003). Thus, the set of expressed genes can change across time, space, and species, as well as tissue/organ. This inevitably means that we will miss some genes within any transcriptome because not all genes are expressed at once and at the same level. This has ramifications for gene expression studies as well as for mining single-copy nuclear markers from

transcriptomic data. To maximize the genic content of a given transcriptomic sample one must take into account tissue specificity: sampling multiple tissues will yield the highest transcriptomic coverage whereas sampling the same tissues across multiple accessions may increase the shared complement of genes across accessions. To maximize the shared complement of genes across a series of transcriptomes from different accessions and to decrease the confounding impact of differing environmental variables on gene expression, sampling should be done under controlled conditions, for example, from a series of plant species grown under the same greenhouse or growth chamber environments.

The probability of missing single- or low-copy genes in the transcriptome is more likely as compared to high-copy genes. In addition, research has shown that genomic regions duplicated by polyploidy or segmental duplication show unequal contributions to the transcriptome as well as exhibiting organ-specific reciprocal silencing (Adams & al., 2003). That may lead to false positive situations in which we think we have found a truly single-copy gene when in reality we have detected one copy of a duplicated gene, the other of which was not expressed in the transcriptome. This may lead to further complications in downstream applications when primers or probes designed for this locus produce multiple amplicons.

Because the transcriptome, as a complement of coding regions, provides only exonic data to work with, primers developed for amplification of coding regions only (i.e., single exons) can produce viable options for single-copy phylogenetic markers (e.g., Weksler, 2003), but these are often less variable than non-coding or mixed-coding markers, especially at shallow evolutionary depths. Exons are the choice for anchoring primers for marker development due to their highly conserved nature, but a lack of knowledge concerning introns and flanking regions presents a challenge. Primers created solely on transcriptomic data may be problematic or even inoperable downstream due to either the existence of very long intronic regions or to the presence of complex secondary structures or long homopolymer runs. These situations, in the past, have impeded DNA amplification and Sanger sequencing through non-coding regions. Likewise, this can mean that transcripts recovered and amplified via Hyb-Seq are not combined into a single transcript later during the assembly process. Comparison with a related published genome or genome-skimmed library can help overcome this limitation by filling in the knowledge gaps for non-coding regions.

Orthologous gene determination is a critical step in RNA-Seq phylogenetics. Although the issue of paralog/ortholog conflation can inhibit the detection and development of single-copy markers for phylogeny reconstruction, this plaguing issue can be mitigated through comparison with published and

annotated genomes. These can help elucidate orthologs by detailing synteny, another level of information than can help clarify orthology vs. paralogy (e.g., Kooij & al., 2005).

Groups with a history of hybridization, introgression and allopolyploidy may be more complicated to benefit from the RNA-Seq comparative approach when constructing their phylogenies (Wen & al., 2013a). On the other hand, transcriptome data have the potential to provide the nuclear sequences to tease apart the history of reticulate evolution and genome duplications.

Selecting genes with strong phylogenetic signals and demonstrating the absence of significant incongruence among them are essential for accurately reconstructing phylogenetic relationships, especially ancient divergences (Salichos & Rokas, 2013). Approaches that can further dissect phylogenetic signals from a smaller number of large analyses rather than direct concatenations will be increasingly important for RNA-Seq data and need to be explored extensively in the coming years.

Even given the list of challenges above, comparative transcriptomics offers a veritable treasure trove of genic sources for phylogenetic inference, as well as single- and low-copy phylogenetic marker development. While we will never be able to overcome the challenges presented by the very nature of the transcriptome (i.e., coding regions only, complement of expressed genes), as published whole genomes and genomic data become less costly and more prevalent, and comparison with transcriptomic data is thus increased, many of the above mentioned limitations and challenges will be overcome. Furthermore, RNA-Seq has been widely applied in studying character development and evolution, including the exploration of possible molecular mechanisms generating key innovations. Comparative transcriptomics has already revolutionized the study of polyploidy. As transcriptome- and genome-scale bioinformatics continues to develop and mature, we expect that the utility of transcriptomics in systematics will only increase, and that the RNA-Seq approach will bring tremendous insights into understanding the ontogeny, homology and evolution of characters, which are the foundation of systematic biology.

ACKNOWLEDGEMENTS

We thank Elvira Hörandl for her invitation to contribute this paper and Sue Lutz for assistance with figure preparation. We thank one anonymous reviewer and Elvira Hörandl for their helpful comments on the manuscript. A.N. Egan was supported in part by funding from the Smithsonian Institution and the National Science Foundation (NSF DEB 1120186/1352217).

LITERATURE CITED

Aberer, A., Kobert, K. & Stamatakis, A. 2014. ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Molec. Biol. Evol.* 31: 2553–2556. http://dx.doi.org/10.1093/molbev/msu236

Adams, K.L., Cronn, R., Percifield, R. & Wendel, J.F. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. U.S.A.*100: 4649–4654. http://dx.doi.org/10.1073/pnas.0630618100

Ainouche, M.L., Baumel, A. & Salmon, A. 2004. *Spartina anglica* C.E. Hubbard: A natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biol. J. Linn. Soc.* 82: 475–484. http://dx.doi.org/10.1111/j.1095-8312.2004.00334.x

Albert, F.W., Somel, M., Carneiro, M., Aximu-Petri, A., Halbwax, M., Thalmann, O., Blanco-Aguiar, J.A., Plyusnina, I.Z., Trut, L., Villafuerte, R., Ferrand, N., Kaiser, S., Jensen, P. & Pääbo, S. 2012. A comparison of brain gene expression levels in domesticated and wild animals. *PLOS Genet.* 8(9): e1002962. http://dx.doi.org/10.1371/journal.pgen 1002962

Altenhoff, A.M., Schneider, A., Gonnet, G.H. & Dessimoz, C. 2011. OMA 2011: Orthology inference among 1000 complete genomes. *Nucl. Acids Res.* 39: D289–D294. http://dx.doi.org/10.1093/nar/gkq1238

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *J. Molec. Biol.* 215: 403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2

Ando, K. & Grumet, R. 2010. Transcriptional profiling of rapidly growing cucumber fruit by 454-pyrosequencing analysis. *J. Amer. Soc. Hort. Sci.* 135: 291–302.

Ané, C., Larget, B., Baum, D.A., Smith, S.D. & Rokas, A. 2007. Bayesian estimation of concordance among gene trees. *Molec. Biol. Evol.* 24: 412–426. http://dx.doi.org/10.1093/molbev/msl170

Angiosperm Phylogeny Group 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161: 105–121. http://dx.doi.org/10.1111/j.1095-8339.2009.00996.x

Bajgain, P., Richardson, B.A., Price, J.C., Cronn, R.C. & Udall, J.A. 2011. Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*). *B. M. C. Genomics* 12: 370. http://dx.doi.org/10.1186/1471-2164-12-370

Barker, M.S., Kane, N.C., Matvienko, M., Kozik, A., Michelmore, R.W., Knapp, S.J. & Rieseberg, L.H. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molec. Biol. Evol.* 25: 2445–2455. http://dx.doi.org/10.1093/molbev/msn187

Barker, M.S., Vogel, H. & Schranz, M.E. 2009 Paleopolyploidy in the Brassicales: Analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biol. Evol.* 1: 391–399. http://dx.doi.org/10.1093/gbe/evp040

Bayzid, M.S. & Warnow, T. 2013. Naive binning improves phylogenomic analyses. *Bioinformatics* 29: 2277–2284. http://dx.doi.org/10.1093/bioinformatics/btt394

Bazinet, A.L., Cummings, M.P., Mitter, K.T. & Mitter, C.W. 2013. Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An exploratory study. *PLOS ONE* 8: e82615. http://dx.doi.org/10.1371/journal.pone.0082615

Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., Minio, A., Rau, D., Rodriguez, M., Panziera, A., Venturini, L., Attene, G., Albertini, E., Jackson, S.A., Nanni, L., Fernie, A.R., Nikoloski, Z., Bertorelle, G., Delledonne, M. & Papa, R. 2014.

Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Pl. Cell* 26: 1901–1912. http://dx.doi.org/10.1105/tpc.114.124040

Betancur-R., R., Naylor, G. & Ortí, G. 2014. Conserved genes, sampling error, and phylogenomic inference. *Syst. Biol.* 63: 257–262. http://dx.doi.org/10.1093/sysbio/syt073

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Rev. Biochem.* 72: 291–336. http://dx.doi.org/10.1146/annurev.biochem.72.121801.161720

Blanc, G. & Wolfe, K.H. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Pl. Cell* 16: 1667–1678. http://dx.doi.org/10.1105/tpc.021345

Bolger, A.M., Lohse, M. & Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* btu170. http://dx.doi.org/10.1093/bioinformatics/btu170

Bombarely, A., Coate, J.E. & Doyle, J.J. 2014. Mining transcriptomic data to study the origins and evolution of a plant allopolyploid complex. *PeerJ* 2: e391. http://dx.doi.org/10.7717/peerj.391

Borner, J., Rehm, P., Schill, R.O., Ebersberger, I. & Burmester, T. 2014. A transcriptome approach to ecdysozoan phylogeny. *Molec. Phylogen. Evol.* 80: 79–87. http://dx.doi.org/10.1016/j.ympev.2014.08.001

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M.A., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., Maccallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O. & Di Palma, F. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513: 375–381. http://dx.doi.org/10.1038/nature13726

Brewer, M.S. & Bond, J.E. 2013. Ordinal-level phylogenomics of the Arthropod class Diplopoda (millipedes) based on an analysis of 221 nuclear protein-coding loci generated using next-generation sequence analyses. *PLOS ONE* 8(11): e79935. http://dx.doi.org/10.1371/journal.pone.0079935

Buggs, R.J., Renny-Byfield, S., Chester, M., Jordon-Thaden, I.E., Viccini, L.F., Chamala, S., Leitch, A.R., Schnable, P.S., Barbazuk, W.B., Soltis, P.S. & Soltis, D.E. 2012. Next-generation sequencing and genome evolution in allopolyploids. *Amer. J. Bot.* 99: 372–382. http://dx.doi.org/10.3732/ajb.1100395

Burgess, D. & Freeling, M. 2014. The most deeply conserved noncoding sequences in plants serve similar functions to those in vertebrates despite large differences in evolutionary rates. *Pl. Cell* 26: 946–961. http://dx.doi.org/10.1105/tpc.113.121905

Chan, C.X., Blouin, N.A.,, Zhuang, Y., Zauner, S., Prochnik, S.E., Lindquist, E., Lin, S., Benning, C., Lohr, M., Yariah, C., Gantt, E., Grossman, A.R., Lu, S., Muller, K., Stiller, J.W., Brawley, S.H. & Bhattacharya, D. 2012. *Porphyra* (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *J. Phycol.* 48: 1328–1342. http://dx.doi.org/10.1111/j.1529-8817.2012.01229.x

Chapman, M.A., Chang, J.-C., Weisman, D., Kesseli, R.V. & Burke, J.M. 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theor. Appl. Genet.* 115: 747–755. http://dx.doi.org/10.1007/s00122-007-0605-2

Chen, Y., Mao, Y., Liu, H., Yu, F., Li, S. & Yin, T. 2014. Transcriptome analysis of differentially expressed genes relevant to variegation in peach flowers. *PLOS ONE* 9: e90842. http://dx.doi.org/10.1371/journal.pone.0090842

Chiari, Y., Cahais, V., Galtier, N. & Delsuc, F. 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *B. M. C. Biol.* 10: 65. http://dx.doi.org/10.1186/1741-7007-10-65

Christin, P.A., Arakaki, M., Osborne, C.P., Brautigam, A., Sage, R.F., Hibberd, J.M., Kelly, S., Covshoff, S., Wong, G.K., Hancock, L. & Edwards, E.J. 2014. Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *J. Exp. Bot.* 65: 3609–3621. http://dx.doi.org/10.1093/jxb/eru087

Clarke, T.H., Garb, J.E., Hayashi, C.Y., Haney, R.A., Lancaster, A.K., Corbett, S. & Ayoub, N.A. 2014. Multi-tissue transcriptome of the black widow spider reveals expansions, co-options and functional processes of the silk gland gene toolkit. *B. M. C. Genomics* 15: 365. http://dx.doi.org/10.1186/1471-2164-15-365

Combes, M.-C., Dereeper, A., Severac, D., Bertrand, B. & Lashermes. 2013. Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytol.* 200: 251–260. http://dx.doi.org/10.1111/nph.12371

Dai, F., Chen, Z.-H., Wang, X., Li, Z., Jin, G., Wu, D., Cai, S., Wang, N., Wu, F., Nevo, E. & Zhang, G. 2014. Transcriptome profiling reveals mosaic genomic origins of modern cultivated barley. *Proc. Natl. Acad. Sci. U.S.A.* 111: 13403–13408. http://dx.doi.org/10.1073/pnas.1414335111

Davis, C.C., Xi, Z. & Mathews, S. 2014. Plastid phylogenomics and green plant phylogeny: Almost full circle but not quite there. *B. M. C. Biol.* 12: 11. http://dx.doi.org/10.1186/1741-7007-12-11

Degnan, J.H. 2013. Anomalous unrooted gene trees. *Syst. Biol.* 62: 574–590. http://dx.doi.org/10.1093/sysbio/syt023

Degnan, J.H. & Rosenberg, N.A. 2006. Discordance of species trees with their most likely gene trees. *PLOS Genet.* 2(5): e68. http://dx.doi.org/10.1371/journal.pgen.0020068

Degnan, J.H. & Rosenberg, N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24: 332–340. http://dx.doi.org/10.1016/j.tree.2009.01.009

Der, J.P., Barker, M.S., Wickett, N.J., dePamphilis, C.W. & Wolf, P.G. 2011. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *B. M. C. Genomics* 12: 99. http://dx.doi.org/10.1186/1471-2164-12-99

Dikow, R.B. & Smith, W.L. 2013. Complete genome sequences provide a case study for the evaluation of gene-tree thinking. *Cladistics* 29: 672–682. http://dx.doi.org/10.1111/cla.12020

Dion-Cote, A.-M., Renaut, S., Normandeau, E. & Bernatchez, L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Molec. Biol. Evol.* 31: 1188–1199. http://dx.doi.org/10.1093/molbev/msu069

Dlugosch, K.M., Lai, Z., Bonin, A., Hierro, J. & Rieseberg, L.H. 2013. Allele identification for transcriptome-based population genomics in the invasive plant *Centaurea solstitialis*. *G3-Genes-Genomes-Genetics* 3: 359–367. http://dx.doi.org/10.1534/g3.112.003871

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Molec. Biol. Evol.* 29: 1969–1973. http://dx.doi.org/10.1093/molbev/mss075

Duarte, J., Wall, P.K., Edger, P., Landherr, L., Ma, H., Pires, J.C. & Leebens-Mack, J. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *B. M. C. Evol. Biol.* 10: 61. http://dx.doi.org/10.1186/1471-2148-10-61

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E. & Rouse, G.W. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452: 745–749. http://dx.doi.org/10.1038/nature06614

Ebersberger, I., Strauss, S. & Haeseler, A. von 2009. HaMStR: Profile hidden Markov model based search for orthologs in ESTs. *B. M. C. Evol. Biol.* 9: 157. http://dx.doi.org/10.1186/1471-2148-9-157

Edwards, S.V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63: 1–19. http://dx.doi.org/10.1111/j.1558-5646.2008.00549.x

Egan, A.N. & Doyle, J.J. 2010. A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: Dating the polyploid history of soybean (*Glycine max*). *Syst. Biol.* 59: 534–547. http://dx.doi.org/10.1093/sysbio/syq041

Egan, A.N., Schlueter, J. & Spooner, D.M. 2012. Applications of next-generation sequencing in plant biology. *Amer. J. Bot.* 99: 175–185. http://dx.doi.org/10.3732/ajb.1200020

Elmer, K.R., Fan, S., Gunter, H.M., Jones, J.C., Boekhoff, S., Kuraku, S. & Meyer, A. 2010. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Molec. Ecol.* 19: 197–211. http://dx.doi.org/10.1111/j.1365-294X.2009.04488.x

Enright, A.J., Van Dongen, S. & Ouzounis, C.A. 2002. An efficient algorithm for large-scale detection of protein families. *Nucl. Acids Res.* 30: 1575–1584. http://dx.doi.org/10.1093/nar/30.7.1575

Faircloth, B.C., Sorenson, L., Santini, F. & Alfaro, M.E. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLOS ONE* 8: e65923. http://dx.doi.org/10.1371/journal.pone.0065923

Fernández, R., Laumer, C.E., Vahtera, V., Libro, S., Kaluziak, T., Sharma, P.P., Pérez-Porro, A.R., Edgecombe, G.D. & Giribet, G. 2014. Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. *Molec. Biol. Evol.* 31: 1500–1513. http://dx.doi.org/10.1093/molbev/msu108

Ferreira, P.G., Patalano, S., Chauhan, R., Ffrench-Constant, R., Gabaldon, T., Guigo, R. & Sumner, S. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* 14: R20. http://dx.doi.org/10.1186/gb-2013-14-2-r20

Finet, C., Timme, R.E., Delwiche, C.F. & Marlétaz, F. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Curr. Biol.* 20: 2217–2222. http://dx.doi.org/10.1016/j.cub.2010.11.035

Finn, R.D., Clements, J. & Eddy, S.R. 2011. HMMER web server: Interactive sequence similarity searching *Nucl. Acids Res.* 39: W29–W37. http://dx.doi.org/10.1093/nar/gkr367

Fitch, W.M. 1970. Distinguishing homologous from analogous proteins. *Syst. Zool.* 19: 99–113. http://dx.doi.org/10.2307/2412448

Franssen, S.U., Shrestha, R.P., Bräutigam, A., Bornberg-Bauer, E. & Weber, A.P.M. 2011. Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *B. M. C. Genomics* 12: 227. http://dx.doi.org/10.1186/1471-2164-12-227

Fulton, T.M., Van der Hoeven, R., Eannetta, N.T. & Tanksley, S.D. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Pl. Cell* 14: 1457–1467. http://dx.doi.org/10.1105/tpc.010479

Garg, R., Verma, M., Agrawal, S., Shankar, R., Majee, M. & Jain, M. 2013. Deep transcriptome sequencing of wild halophyte, *Porteresia coarctata*, provides novel insights into the salinity and submergence tolerance factors. *D. N. A. Res.* 69–84. http://dx.doi.org/10.1093/dnares/dst042

Gatesy, J. & Baker, R.H. 2005. Hidden likelihood support in genomic data: Can forty-five wrongs make a right? *Syst. Biol.* 54: 483–492. http://dx.doi.org/10.1080/10635150590945368

Goldblatt, P. & Takei, M. 1997. Chromosome cytology of Iridaceae: Patterns of variation, determination of ancestral base numbers, and modes of karyotype change. *Ann. Missouri Bot. Gard.* 84: 285–304. http://dx.doi.org/10.2307/2400005

Goloboff, P.A., Farris, J.S. & Nixon, K.C. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24: 774–786. http://dx.doi.org/10.1111/j.1096-0031.2008.00217.x

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29: 644–652. http://dx.doi.org/10.1038/nbt.1883

Guan, S. & Lu, Y. 2013. Dissecting organ-specific transcriptomes through RNA-sequencing. *Pl. Meth.* 9: 42. http://dx.doi.org/10.1186/1746-4811-9-42

Guo, B., Chain, F.J.J., Bornberg-Bauer, E., Leder, E.H. & Merila, J. 2013. Genomic divergence between nine- and three-spined sticklebacks. *B. M. C. Genomics* 14: 756. http://dx.doi.org/10.1186/1471-2164-14-756

Hagel, J.M., Yeung, E.C. & Facchini, P.J. 2008. Got milk? The secret life of laticifers. *Trends Pl. Sci.* 13: 631–639. http://dx.doi.org/10.1016/j.tplants

Haney, R.A., Ayoub, N.A., Clarke, T.H., Hayashi, C.Y. & Garb, J.E. 2014. Dramatic expansion of the black widow toxin arsenal uncovered by multi-tissue transcriptomics and venom proteomics. *B. M. C. Genomics* 15: 366. http://dx.doi.org/10.1186/1471-2164-15-366

Harris, S.E., Munshi-South, J., Oberfell, C. & O'Neill, R. 2013. Signatures of rapid evolution in urban and rural transcriptomes of white-footed mice (*Peromyscus leucopus*) in the New York metropolitan area. *PLOS ONE* 8: e74938. http://dx.doi.org/10.1371/journal.pone.0074938

Hartmann, S., Helm, C., Nickel, B., Meyer, M., Struck, T.H., Tiedemann, R., Selbig, J. & Bleidorn, C. 2012. Exploiting gene families for phylogenomic analysis of myzostomid transcriptome data. *PLOS ONE* 7(1): e29843. http://dx.doi.org/10.1371/journal.pone.0029843

Hedin, M., Starrett, J., Akhter, S., Schönhofer, A.L. & Shultz, J.W. 2012. Phylogenomic resolution of Paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. *PLOS ONE* 7: e428888. http://dx.doi.org/10.1371/journal.pone.0042888

Heled, J. & Drummond, A.J. 2010. Bayesian inference of species trees from multilocus data. *Molec. Biol. Evol.* 27: 570–580. http://dx.doi.org/10.1093/molbev/msp274

Hileman, L. 2014. Bilateral flower symmetry—How, when and why? *Curr. Opin. Pl. Biol.* 17: 146–152. http://dx.doi.org/10.1016/j.pbi.2013.12.002

Hiremath, P.J., Farmer, A., Cannon, S.B., Woodward, J., Kudapa, H., Tuteja, R., Kumar, A., Bhanuprakash, A., Mulaosmanovic, B., Gujaria, N., Krishnamurthy, L., Gaur, P.M., Kavikishor, P.B., Shah, T., Srinivasan, R., Lohse, M., Xiao, Y., Town, C.D., Cook, D.R., May, G.D. & Varshney, R.K. 2011. Large-scale transcriptome analysis in

chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Pl. Biotechnol. J.* 9: 922–931. http://dx.doi.org/10.1111/j.1467-7652.2011.00625.x

Hirsch, C.N. & Buell, C.R. 2013. Tapping the promise of genomics in species with complex, nonmodel genomes. *Annual Rev. Pl. Biol.* 64: 89–110. http://dx.doi.org/10.1146/annurev-arplant-050312-120237

Hittinger, C.T., Johnston, M., Tossberg, J.T. & Rokas, A. 2010. Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. *Proc. Natl. Acad. Sci. U.S.A.* 107: 1476–1481. http://dx.doi.org/10.1073/pnas.0910449107

Hodgins, K.A., Lai, Z., Oliveira, L.O., Still, D.W., Scascitelli, M., Barker, M.S., Kane, N.C., Dempewolf, H., Kozik, A., Kesseli, R.V., Burke, J.M., Michelmore, R.W. & Rieseberg, L.H. 2014. Genomics of Compositae crops: Reference transcriptome assemblies and evidence of hybridization with wild relatives. *Molec. Ecol. Resources* 14: 166–177. http://dx.doi.org/10.1111/1755-0998.12163

Hofer, J.M.I. & Ellis, T.H.N. 2014. Developmental specialisations in the legume family. *Curr. Opin. Pl. Biol.* 117: 153–158. http://dx.doi.org/10.1016/j.pbi.2013.11.014

Hsu, P.Y., Devisetty, U.K. & Harmer S.L. 2013. Accurate timekeeping is controlled by a cycling activator in *Arabidopsis*. *eLife* 2: e00473. http://dx.doi.org/10.7554/eLife.00473

Huang, W., Nadeem, A., Zhang, B., Babar, M. & Khatib, H. 2012. Characterization and comparison of the leukocyte transcriptomes of three cattle breeds. *PLOS ONE* 7(1): e30244. http://dx.doi.org/10.1371/journal.pone.0030244

Hughes, C.E., Eastwood, R.J. & Bailey, C.D. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philos. Trans., Ser. B* 361: 211–225. http://dx.doi.org/10.1098/rstb.2005.1735

Ichihashi, Y., Aguilar-Martinez, A., Farhi, M., Chitwood, D.H., Kumar, R., Millon, L.V., Peng, J., Maloof, J.N. & Sinha, N. 2014. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. U.S.A.* 111: E2626–E2621. http://dx.doi.org/10.1073/pnas.1402835111

Ilut, D.C., Coate, J.E., Luciano, A.K., Owens, T.G., May, G.D., Farmer, A. & Doyle, J.J. 2012. A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. *Amer. J. Bot.* 99: 383–396. http://dx.doi.org/10.3732/ajb.1100312

Jaenisch, R. & Bird, A. 2003. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33: 245–254. http://dx.doi.org/10.1038/ng1089

Jayakodi, M., Lee, S.-C., Park, H.-S., Jang, W., Lee, Y.S., Choi, B.-S., Nan, G.J., Kim, D.-S., Natesan, S., Sun, C. & Yang, T.-J. 2014. Transcriptome profiling and comparative analysis of *Panax ginseng* adventitious roots. *J. Ginseng Res.* 38: 278–288. http://dx.doi.org/10.1016/j.jgr.2014.05.008

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J. & dePamphilis, C.W. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: U97–U113. http://dx.doi.org/10.1038/nature09916

Johnson, B.R., Borowiec, M.L., Chiu, J.C., Lee, E.K., Atallah, J. & Ward, P.S. 2013. Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Curr. Biol.* 23: 2058–2062. http://dx.doi.org/10.1016/j.cub.2013.08.050

Johnson, M.T.J., Carpenter, E.J., Tian, Z., Bruskiewich, R., Burris, J.N., Carrigan, C.T., Chase, M.W., Clarke, N.D., Covshoff, S., Depamphilis, C.W., Edger, P.P., Goh, F., Graham, S., Greiner, S., Hibberd, J.M., Jordon-Thaden, I., Kutchan, T.M., Leebens-Mack, J., Melkonian, M., Miles, N., Myburg, H., Patterson, J., Pires, J.C., Ralph, P.,

Rolf, M., Sage, R.F., Soltis, D., Soltis, P., Stevenson, D., Stewart, C.N. Jr., Surek, B., Thomsen, C.J., Villarreal, J.C., Wu, X., Zhang, Y., Deyholos, M.K. & Wong, G.K. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLOS ONE* 7(11): e50226. http://dx.doi.org/10.1371/journal.pone.0050226

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. 2012. KEGG for integration and interpretation of large-scale molecular datasets. *Nucl. Acids Res.* 40: D109–D114. http://dx.doi.org/10.1093/nar/gkr988

Kaufmann, K., Wellmer, F., Muino, J., Ferner, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueno, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C. & Reichmann, J.L. 2010. Orchestration of floral initiation by *APETALA1*. *Science* 328: 85–89. http://dx.doi.org/10.1126/science.1185244

King, M.-C. & Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188: 107–118. http://dx.doi.org/10.1126/science.1090005

Ko, J.H., Chow, K.S. & Han, K.H. 2003. Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree). *Pl. Molec. Biol.* 53: 479–492. http://dx.doi.org/10.1023/B:PLAN.0000019119.66643.5d

Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B. & Halanych, K.M. 2011. Phylogenomics reveals deep molluscan relationships. *Nature* 477: 452–456. http://dx.doi.org/10.1038/nature10382

Kocot, K.M., Citarella, M.R., Moroz, L.R. & Halanych, K.M. 2013. PhyloTreePruner: A phylogenetic-tree based approach for selection of orthologous sequences for phylogenomics. *Evol. Bioinf.* 9: 429–435. http://dx.doi.org/10.4137/EBO.S12813

Koenig, D., Jimenez-Gomez, J., Kimura, S., Fulop, D., Chitwood, D., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., Tohge, T., Bolger, A., Schneeberger, K., Ossowski, S., Lanz, C., Xiong, G., Taylor-Teeples, M., Brady, S.M., Pauly, M., Weigel, D., Usadel, B., Fernie, A.R., Peng, J., Sinha, N.R. & Maloof, J.N. 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U.S.A.* 110: E2655–2662. http://dx.doi.org/10.1073/pnas.1309606110

Kohlbach, J., Attah, A.F. & Gruber, C.W. 2013. Cyclotide discovery in Gentianales revisited—Identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. *Peptide Sci.* 100: 438–452. http://dx.doi.org/10.1002/bip.22328

Kooij, T.W., Carlton, J.M., Bidwell, S.L., Hall, N., Ramesar, J., Janse, C.J. & Waters, A.P. 2005. A *Plasmodium* whole-genome synteny map: Indels and synteny breakpoints as foci for species-specific genes. *PLOS Pathog.* 1(4): e44. http://dx.doi.org/10.1371/journal.ppat.0010044

Krogh, A., Larsson, B., Heijne, G. von & Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Molec. Biol.* 305: 567–580. http://dx.doi.org/10.1006/jmbi.2000.4315

Kubatko, L.S. & Degnan, J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56: 17–24. http://dx.doi.org/10.1080/10635150601146041

Lagesen, K., Hallin, P.F., Rodland, E., Staerfeldt, H.H., Rognes, T. & Ussery, D.W. 2007. RNammer: Consistent annotation of rRNA genes in genomic sequences. *Nucl. Acids Res.* 35: 3100–3108. http://dx.doi.org/10.1093/nar/gkm160

Larget, B., Kotha, S.K., Dewey, C.N. & Ané, C. 2010. BUCKy: Gene tree/species tree reconciliation with the Bayesian concordance analysis. *Bioinformatics* 26: 2910–2911. http://dx.doi.org/10.1093/bioinformatics/btq539

Lartillot, N., Blanquart, S. & Lepage, T. 2012. PhyloBayes, version 3.3: A Bayesian software for phylogenetic reconstruction and molecular dating using mixture models. Available from www.phylobayes.org

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* 61: 727–744. http://dx.doi.org/10.1093/sysbio/sys049

Lemmon, E.M. & Lemmon, A.R. 2013. High-throughput genomic data in systematics and phylogenetics. *Annual Rev. Ecol. Evol. Syst.* 44: 99–121. http://dx.doi.org/10.1146/annurev-ecolsys-110512-135822

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., Yan, J., Jiang, X., Zhang, L., Wu, J., Yin, L., Zhang, R., Wu, L., Zheng, Y. & Mao, L. 2014. mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Pl. Cell* 26: 1878–1900. http://dx.doi.org/10.1105/tpc.114.124388

Li, F.-W., Villarreal, J.C., Kelly, S., Rothfels, C., Melkonian, M., Frangedakis, E., Ruhsam, M., Sigel, E.M., Der, J.P., Pittermann, J., Burge, D.O., Pokorny, L., Larsson, A., Chen, T., Weststrand, S., Thomas, P., Carpenter, E., Zhang, Y., Tian, Z., Chen, L., Yan, Z., Zhu, Y., Sun, X., Wang, J., Stevenson, D.W., Crandall-Stotler, B.J., Shaw, A.J., Deyholos, M.K., Soltis, D.E., Graham, S.W., Windham, M.D., Langdale, J.A., Wong, G. K.-S., Mathews, S. & Pryer, K.M. 2014. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc. Natl. Acad. Sci. U.S.A.* 111: 6672–6677. http://dx.doi.org/10.1073/pnas.1319929111

Li, H., Coghlan, A., Ruan, J., Coin, L., Heriche, J., Osmotherly, L., Li, R., Liu, T., Zhang, Z., Bolund, L., Wong, G.K., Zheng, W., Dehal, P., Wang, J. & Durbin, R. 2006. TreeFam: A curated database of phylogenetic trees of animal gene families. *Nucl. Acids Res.* 34: 572–580. http://dx.doi.org/10.1093/nar/gkj118

Li, J., Liu, B., Cheng, F., Wang, X., Aarts, M.G.M. & Wu, J. 2014. Expression profiling reveals functionally redundant multiple-copy genes related to zinc, iron and cadmium responses in *Brassica rapa*. *New Phytol.* 203: 182–194. http://dx.doi.org/10.1111/nph.12803

Liang, C., Liu, X., Yiu, S.-M. & Lim, B.L. 2013. De novo assembly and characterization of *Camelina sativa* transcriptome by paired-end sequencing. *B. M. C. Genomics* 14: 146. http://dx.doi.org/10.1186/1471-2164-14-146

Liao, D., Cram, D., Sharpe, A. & Marsolais, F. 2013. Transcriptome profiling identifies candidate genes associated with the accumulation of distinct sulfur gamma-glutamyl dipeptides in *Phaseolus vulgaris* and *Vigna mungo* seeds. *Frontiers Pl. Sci.* 4: 60. http://dx.doi.org/10.3389/fpls.2013.00060

Lin, G.-H., Wang, K., Deng, X.-G., Nevo, E., Zhao, F., Su, J.-P., Guo, S.-C., Zhang, T.-Z. & Zhao, H. 2014. Transcriptome sequencing and phylogenomic resolution within Spalacidae (Rodentia). *B. M. C. Genomics* 15: 32. http://dx.doi.org/10.1186/1471-2164-15-32

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* 24: 2542–2543. http://dx.doi.org/10.1093/bioinformatics/btn484

Liu, L. & Yu, L. 2011. Estimating species trees from unrooted gene trees. *Syst. Biol.* 60: 661–667. http://dx.doi.org/10.1093/sysbio/syr027

Liu, L., Yu, L., Pearl, D.K. & Edwards, S.V. 2009. Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* 58: 468–477. http://dx.doi.org/10.1093/sysbio/syp031

Liu, L., Yu, L. & Edwards, S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *B. M. C. Evol. Biol.* 10: 302. http://dx.doi.org/10.1186/1471-2148-10-302

Liu, L., Wu, S. & Yu, L. 2015. Coalescent methods for estimating species trees from phylogenomic data. *J. Syst. Evol.,* early view. http://dx.doi.org/10.1111/jse.12160

Liu, Z., Ma, L., Nan, Z. & Wang, Y. 2013. Comparative transcriptional profiling provides insights into the evolution and development of the zygomorphic flower of *Vicia sativa* (Papilionoideeae). *PLOS ONE* 8(2): e57336. http://dx.doi.org/10.1371/journal.pone.0057336

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E. & O'Brien, S.J. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* 15: 47–56. http://dx.doi.org/10.1038/ng0197-47

Lyu, M.J., Gowik, U., Kelly, S., Covshoff, S., Mallmann, J., Westhoff, P., Hibberd, J.M., Stata, M., Sage, R.F., Lu, H., Wei, X., Wong, G.K. & Zhu, X.G. 2015. RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria* (Asteraceae) with some modifications. *B. M. C. Evol. Biol.* 15: 116. http://dx.doi.org/10.1186/s12862-015-0399-9

Ma, H., Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Edger, P., Al-Shehbaz, I., Pires, J.C., Koch, M. & Tan, D.Y. 2015. Resolving Brassicaceae phylogeny using nuclear genes facilitates molecular clock estimates of major branches, character reconstruction and analysis of gene family evolution. Botany 2015 Abstracts. http://2015.botanyconference.org/engine/search/index.php?func=detail&aid=388 [accessed 1 May 2015]

Ma, H., Lu, J., Liu, B.-B., Duan, B.-B., He, X.-D. & Liu, J.-Q. In press. Phylotranscriptomic analyses in plants using Betulaceae as an example. *J. Syst. Evol.* http://dx.doi.org/10.1111/jse.12178

Maddison, W.P. 1997. Gene trees in species trees. *Syst. Biol.* 46: 523–536. http://dx.doi.org/10.1093/sysbio/46.3.523

Manchester, S.R., Kapgate, D.K. & Wen, J. 2013. Oldest fruits of the grape family (Vitaceae) from the Late Cretaceous Deccan Cherts of India. *Amer. J. Bot.* 100: 1849–1859. http://dx.doi.org/10.3732/ajb.1300008

Mandal, S.D., Chhakchhuak, L., Gurusubramanian, G. & Kumar, N.S. 2014. Mitochondrial markers for identification and phylogenetic studies in insects – A review. *D. N. A. Barcodes* 2(1). http://dx.doi.org/10.2478/dna-2014-0001

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H. & Burke, J.M. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Appl. Pl. Sci.* 2(2): 1300085. http://dx.doi.org/10.3732/apps.1300085

Mandel, J.R., Dikow, R.B. & Funk, V.A. 2015. Using phylogenomics to resolve mega-families: An example from Compositae. *J. Syst. Evol.,* early view. http://dx.doi.org/10.1111/jse.12167

Martin, J.A. & Wang, Z. 2011. Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12: 671–682. http://dx.doi.org/10.1038/nrg3068

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C. & Brumfield, R.T. 2013a. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molec. Phylogen. Evol.* 66: 526–538. http://dx.doi.org/10.1016/j.ympev.2011.12.007

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C. & Brumfield, R.T. 2013b. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLOS ONE* 8(1): e54848. http://dx.doi.org/10.1371/journal.pone.0054848

McKain, M.R., Wickett, N., Zhang, Y., Ayyampalayam, S., McCombie, W.R., Chase, M.W., Pires, J.C., dePamphilis, C.W. & Leebens-Mack, J. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *Amer. J. Bot.* 99: 397–406. http://dx.doi.org/10.3732/ajb.1100537

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. & Warnow, T. 2014. ASTRAL: Genome-scale coalescent-based species tree. *Bioinformatics* 30: i541–i548 http://dx.doi.org/10.1093/bioinformatics/btu462

Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M., Ptitsyn, A., Reshetov, D., Mukherjee, K., Moroz, T.P., Bobkova, Y., Yu, F., Kapitonov, V.V., Jurka, J., Bobkov, Y.V., Swore, J.J., Girardo, D.O., Fodor, A., Gusev, F., Sanford, R., Bruders, R., Kittler, E., Mills, C.E., Rast, J.P., Derelle, R., Solovyav, V.V., Kondrashov, F.A., Swalla, B.J., Sweedler, J.V., E., Halanych, K.M. & Kohn, A.B. 2014. The ctenophore genome and the evolutionary origins of neural systems. *Nature* 510: 109–114. http://dx.doi.org/10.1038/nature13400

Natali, L., Cossu, R.M., Barghini, E., Giordani, T., Buti, M., Mascagni, F., Morgante, M., Gill, N., Kane, N.C., Rieseberg, L. & Cavallini, A. 2013. The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads. *B. M. C. Genomics* 14: 686. http://dx.doi.org/10.1186/1471-2164-14-686

Naumann, J., Salomo, K., Der, J.P., Wafula, E.K., Bolin, J.F., Maass, E., Frenzke, L., Samain, M.S., Neinhuis, C., dePamphilis, C.W. & Wanke, S. 2013. Single-copy nuclear genes place haustorial Hydnoraceae within Piperales and reveal a Cretaceous origin of multiple parasitic angiosperm lineages. *PLOS ONE* 8(11): e79204. http://dx.doi.org/10.1371/journal.pone.0079204

Ng, C.S., Wu, P., Fan, W.-L., Yan, J., Chen, C.-K., Lai, Y.T., Wu, S.M., Mao, C.T., Chen, J.J., Lu, M.Y., Ho, M.R., Widelitz, R.B., Chen, C.F., Chuong, C.M. & Li, W.H. 2014. Genomic organization, transcriptomic analysis, and functional characterization of avian alpha- and beta-keratins in diverse feather forms. *Genome Biol.* Evol. 6: 2258–2273. http://dx.doi.org/10.1093/gbe/evu181

Nicolas, V., Schaeffer, B., Missoup, A.D., Kennis, J., Colyn, M., Denys, C., Tatard, C., Cruaud, C. & Laredo, C. 2012. Assessment of three mitochondrial genes (16S, Cytb, CO1) for identifying species in the Praomyini tribe (Rodentia: Muridae). *PLOS ONE* 7: e36586. http://dx.doi.org/10.1371/journal.pone.0036586

O'Brien, K.P., Remm, M. & Sonnhammer, E.L.L. 2005. InParanoid: A comprehensive database of eukaryotic orthologs. *Nucl. Acids Res.* 33: D476–D480. http://dx.doi.org/10.1093/nar/gki107

Ozsolak, F. & Milos, P.M. 2011. RNA sequencing: Advances, challenges and opportunities. *Nat. Rev. Genet.* 12: 87–98. http://dx.doi.org/10.1038/nrg2934

Pagel, M. & Meade, A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.* 53: 571–581. http://dx.doi.org/10.1080/10635150490468675

Parkin, I.A.P., Koh, C., Tang, H., Robinson, S.J., Kagale, S., Clarke, W.E., Town, C.D., Nixon, J., Krishnakumar, V., Bidwell, S.L., Denoeud, F., Belcram, H., Links, M.G., Just, J., Clarke, C., Bender, T., Huebert, T., Mason, A.S., Pires, J.C., Barker, G., Moore, J., Walley, P.G., Manoli, S., Batley, J., Edwards, D., Nelson, M.N., Wang, X., Paterson, A.H., King, G., Bancroft, I., Chalhoub, B. & Sharpe, A.G. 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.* 15: R77. http://dx.doi.org/10.1186/gb-2014-15-6-r77

Pellino, M., Hojsgaard, D., Schmutzer, T., Scholz, U., Hörandl, E., Vogel, H. & Sharbel, T.F. 2013. Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: Examining the effects of hybridization and mutation accumulation. *Molec. Ecol.* 22: 5908–5921. http://dx.doi.org/10.1111/mec.12533

Peng, Z., Cheng, Y., Tan, B.C., Kang, L., Tian, Z., Zhu, Y., Zhang, W., Liang, Y., Hu, X., Tan, X., Guo, J., Dong, Z., Liang, Y., Bao, L. & Yang, J. 2012. Comprehensive analysis of RNA-seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* 30: 253–260. http://dx.doi.org/10.1038/nbt.2122

Peters, R.S., Meusemann, K., Petersen, M., Mayer, C., Wilbrandt, J., Ziesmann, T., Donath, A., Kjer, K.M., Aspöck, U., Aspöck, H., Aberer, A., Stamatakis, A., Friedrich, F., Hünefeld, F., Niehuis, O., Beutel, R.G. & Misof, B. 2014. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *B. M. C. Evol. Biol.* 14: 52. http://dx.doi.org/10.1186/1471-2148-14-52

Petersen, T.N., Brunak, S., Heijne, G. von & Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature, Meth.* 8: 785–786. http://dx.doi.org/10.1038/nmeth.1701

Pillon, Y., Johansen, J.B., Sakishima, T., Roalson, E.H., Price, D.K. & Stacy, E.A. 2013. Gene discordance in phylogenomics of recent plant radiations, an example from Hawaiian *Cyrtandra* (Gesneriaceae). *Molec. Phylogen. Evol.* 69: 293–298. http://dx.doi.org/10.1016/j.ympev.2013.05.003

Pires, J.C., Maureira, I.J., Givnish, T.J., Sytsma, K.J., Seberg, O., Petersen, G., Davis, J.I., Stevenson, D.W., Rudall, P.J., Fay, M.F. & Chase, M.W. 2006. Phylogeny, genome size, and chromosome evolution of Asparagales. *Aliso* 22: 287–304.

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., Jensen, L.J., Mering, C. von & Bork, P. 2012. eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucl. Acids Res.* 40: D284–D289. http://dx.doi.org/10.1093/nar/gkr1060

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A. & Finn, R.D. 2012. The Pfam protein families database. *Nucl. Acids Res.* 40: D290–D301. http://dx.doi.org/10.1093/nar/gkr1065

Raherison, E., Rigault, P., Caron, S., Poulin, P.-L., Boyle, B., Verta, J.-P., Giquere, I., Bomal, C., Bohlmann, J. & MacKay, J. 2012. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *B. M. C. Genomics* 13: 434. http://dx.doi.org/10.1186/1471-2164-13-434

Rambani, A., Page, J.T. & Udall, J.A. 2014. Polyploidy and the petal transcriptome of *Gossypium*. *B. M. C. Pl. Biol.* 14: 3. http://dx.doi.org/10.1186/1471-2229-14-3

Ree, R.H. 2005. Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution* 59: 257–265. http://dx.doi.org/10.1111/j.0014-3820.2005.tb00986.x

Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C. & Mitter, K.T. 2013. A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLOS ONE* 8: e58568. http://dx.doi.org/10.1371/journal.pone.0058568

Rehm, P., Borner, J., Meusemann, K., Reumont, B.M. von, Simon, S., Hadrys, H., Misof, B. & Burmester, T. 2011. Dating the arthropod tree based on large-scale transcriptome data. *Molec. Phylogen. Evol.* 61: 880–887. http://dx.doi.org/10.1016/j.ympev.2011.09.003

Reidenbach, K.R., Cook, S., Bertone, M.A., Harbach, R.E., Wiegmann, B.M. & Besansky, N.J. 2009. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. *B. M. C. Evol. Biol.* 9: 298. http://dx.doi.org/10.1186/1471-2148-9-298

Ren, L., Sun, J., Chen, S., Gao, J., Dong, B., Liu, Y., Xia, X., Wang, Y.,Liao, Y., Teng, N., Fang, W., Guan, Z., Chen, F. & Jiang, J. 2014. A transcriptomic analysis of *Chrysanthemum nankingense* provides insights into the basis of low temperature tolerance. *B. M. C. Genomics* 15: 844. http://dx.doi.org/10.1186/1471-2164-15-844

Reneker, J., Lyons, E., Conant, G.C., Pires, J.C., Freeling, M., Shyu, C-R. & Korkin, D. 2012. Long identical multispecies elements in plant and animal genomes. *Proc. Natl. Acad. Sci. U.S.A.* 109: E1183–E1191. http://dx.doi.org/10.1073/pnas.1121356109

Riesgo, A., Andrade, S.C.S., Sharma, P.P., Novo, M., Pérez-Porro, A.R., Vahtera, V., González, V.L., Kawauchi, G.Y. & Giribet, G. 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Frontiers Zool.* 9: 33. http://dx.doi.org/10.1186/1742-9994-9-33

Riesgo, A., Farrar, N., Windsor, P.J., Giribet, G. & Leys, S.P. 2014. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Molec. Biol. Evol.* 31: 1102–1120. http://dx.doi.org/10.1093/molbev/msu057

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J., Hoodless, P.A. & Birol, I. 2010. De novo assembly and analysis of RNA-seq data. *Nature, Meth.* 7: 909–912. http://dx.doi.org/10.1038/nmeth.1517 http://dx.doi.org/10.1038/nature02053

Roth, A., Gonnet, G.H. & Dessimoz, C. 2008. Algorithm of OMA for large-scale orthology inference. *B. M. C. Bioinf.* 9: 518. http://dx.doi.org/10.1186/1471-2105-9-518

Rothfels, C.J., Larsson, A., Li, F.-W., Sigel, E.M., Huiet, L., Burge, D.O., Ruhsam, M., Graham, S.W., Stevenson, D., Wong, G.K., Korall, P. & Pryer, K. 2013. Transcriptome-mining for single-copy nuclear markers in ferns. *PLOS ONE* 8(10): e76957. http://dx.doi.org/10.1371/journal.pone.0076957

Salichos, L. & Rokas, A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497: 327–331. http://dx.doi.org/10.1038/nature12130

Sanchez-Sevilla, J.F., Cruz-Rus, E., Valpuesta, V., Botella, M.A. & Amaya, I. 2014. Deciphering gamma-decalactone biosynthesis in strawberry fruit using a combination of genetic mapping, RNA-Seq and eQTL analysis. *B. M. C. Genomics* 15: 218. http://dx.doi.org/10.1186/1471-2164-15-218

Schmieder, R. & Edwards, R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864. http://dx.doi.org/10.1093/bioinformatics/btr026

Sedeek, K.E.M., Qi, W., Schauer, M.A., Gupta, A.K., Poveda, L., Xu, S., Liu, Z.-J., Grossniklaus, U., Schiestl, F.P. & Schlüter, P.M. 2013. Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. *PLOS ONE* 8: e64621. http://dx.doi.org/10.1371/journal.pone.0064621

Sharma, B., Yant, L., Hodges, S.A. & Kramer, E.M. 2014. Understanding the development and evolution of novel floral form in Aquilegia. *Curr. Opin. Pl. Biol.* 17: 22–27. http://dx.doi.org/10.1016/j.pbi.2013.10.006

Shi, C.-Y., Yang, H., Wei, C.-L., Yu, O., Zhang, Z.-Z., Jiang, J., Sun, J., Li, Y.-Y., Chen, Q., Xia, T. & Wan, X.-C. 2011. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *B. M. C. Genomics* 12: 131. http://dx.doi.org/10.1186/1471-2164-12-131

Singh, A., Mantri, S., Sharma, M., Chaudhury, A., Tuli, R. & Roy, J. 2014. Genome-wide transcriptome study in wheat identified candidate genes related to processing quality,

majority of them showing interaction (quality X development) and having temporal and spatial distributions. *B. M. C. Genomics* 15: 29. http://dx.doi.org/10.1186/1471-2164-15-29

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C. & Brumfield, R.T. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary timescales. *Syst. Biol.* 63: 83–95. http://dx.doi.org/10.1093/sysbio/syt061

Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C.S., Rouse, G.W., Giribet, G. & Dunn, C.W. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* 480: 364–367. http://dx.doi.org/10.1038/nature10526

Soltis, D.E., Gitzendanner, M.A., Stull, G., Chester, M., Chanderbali, A., Chamala, S., Jordon-Thaden, I., Soltis, P.S., Schnable, P.S. & Barbazuk, W.B. 2013. The potential of genomics in plant systematics. *Taxon* 62: 886–898. http://dx.doi.org/10.12705/625.13

Sousounis, K., Ogura, A. & Tsonis, P. 2013. Transcriptome analysis of *Nautilus* and pygmy squid eye provides insights in lens and eye evolution. *PLOS ONE* 8: e78054. http://dx.doi.org/10.1371/journal.pone.0078054

Springer, M.S. & Gatesy, J. 2014. Land plant origins and coalescence confusion. *Trends Pl. Sci.* 19: 267–269. http://dx.doi.org/10.1016/j.tplants.2014.02.012

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446

Stamatakis, A. 2013. Novel parallelization schemes for large-scale likelihood-based phylogenetic inference. Pp. 1195–1204 in: *IEEE 27th International Parallel and Distributed Processing Symposium: IPDPS 2013*, 20–24 May 2013, Boston, Massachusetts. Piscataway: IEEE. http://dx.doi.org/10.1109/IPDPS.2013.70

Stamatakis, A., Aberer, A.J., Goll, C., Smith, S.A., Berger, S.A. & Izquierdo-Carrasco, F. 2012. RAxML-Light: A tool for computing TeraByte Phylogenies. *Bioinformatics* 28: 2064–2066. http://dx.doi.org/10.1093/bioinformatics/bts309

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C. & Liston, A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Amer. J. Bot.* 99: 349–364. http://dx.doi.org/10.3732/ajb.1100335

Struck, T.H., Paul, C., Hill, N., Hartmann, S., Hösel, C., Kube, M., Lieb, B., Meyer, A., Tiedemann, R., Purschke, G. & Bleidorn, C. 2011. Phylogenomic analyses unravel annelid evolution. *Nature* 471: 95–98. http://dx.doi.org/10.1038/nature09864

Suga, H., Torruella, G., Burger, G., Brown, M.W. & Ruiz-Trillo, I. 2014. Earliest holozoan expansion of phosphotyrosine signaling. *Molec. Biol. Evol.* 31: 517–528. http://dx.doi.org/10.1093/molbev/mst241

Taylor, R.S., Tarver, J.E., Hiscock, S.J. & Donoghue, P.C.J. 2014. Evolutionary history of plant microRNAs. *Trends Pl. Sci.* 19: 175–182. http://dx.doi.org/10.1016/j.tplants.2013.11.008

Teo, Z.W.N., Song, S., Wang, Y.-Q., Liu, J. & Yu, H. 2014. New insights into the regulation of inflorescence architecture. *Trends Pl. Sci.* 19: 158–165. http://dx.doi.org/10.1016/j.tplants.2013.11.001

The Gene Ontology Consortium 2000. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* 25: 25–29. http://dx.doi.org/10.1038/75556

Tilgner, H., Grubert, F., Sharon, D. & Snyder, M.P. 2014. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* 111: 9869–9874. http://dx.doi.org/10.1073/pnas.1400447111

Timme, R.E., Bachvaroff, T.R. & Delwiche, C.F. 2012. Broad phylogenetic sampling and the sister lineage of land plants. *PLOS ONE* 7: e29696. http://dx.doi.org/10.1371/journal.pone.0029696

Van Dongen, S. 2000. *Graph clustering by flow simulation*. Ph.D. thesis, University of Utrecht, Netherlands.

Voelckel, C., Borevitz, J.O., Kramer, E.M. & Hodges, S.A. 2010. Within and between whorls: Comparative transcriptional profiling of *Aquilegia* and *Arabidopsis*. *PLOS ONE* 5: e9735. http://dx.doi.org/10.1371/journal.pone.0009735

Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10: 57–63. http://dx.doi.org/10.1038/nrg2484

Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M. & Kriventseva, E.V. 2013. OrthoDB: A hierarchical catalog of animal, fungal and bacterial orthologs. *Nucl. Acids Res.* 41: D358–D365. http://dx.doi.org/10.1093/nar/gks1116

Waters, B.M., McInturf, S.A. & Amundsen, K. 2014. Transcriptomic and physiological characterization of the fefe mutant of melon (*Cucumis melo*) reveals new aspects of iron-copper crosstalk. *New Phytol.* 203: 1128–1145. http://dx.doi.org/10.1111/nph.12911

Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, D., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C. & Struck, T.H. 2014. Illuminating the base of the annelid tree using transcriptomics. *Molec. Biol. Evol.* 31: 1391–1401. http://dx.doi.org/10.1093/molbev/msu080

Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A. & Liston, A. 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl. Pl. Sci.* 2(9): 1400042. http://dx.doi.org/10.3732/apps.1400042

Weksler, M. 2003. Phylogeny of Neotropical oryzomyine rodents (Muridae: Sigmodontinae) based on the nuclear IRBP exon. *Molec. Phylogen. Evol.* 29: 331–349. http://dx.doi.org/10.1016/S1055-7903(03)00132-5

Wen, J., Xiong, Z., Nie, Z.-L., Mao, L., Zhu, Y., Kan, X.-Z., Ickert-Bond, S.M., Gerrath, J., Zimmer, E.A. & Fang, X.-D. 2013a. Transcriptome sequences resolve deep relationships of the grape family. *PLOS ONE* 8: e74394. http://dx.doi.org/10.1371/journal.pone.0074394

Wen, J., Ree, R.H., Ickert-Bond, S., Nie, Z.L. & Funk, V. 2013b. Biogeography: Where do we go from here? *Taxon* 62: 912–927. http://dx.doi.org/10.12705/625.15

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W., Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H., dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K.-S. & Leebens-Mack, J. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111: E4859–4868. http://dx.doi.org/10.1073/pnas.1323926111

Williams, J.S., Der, J.P., dePamphilis, C.W. & Kao, T.-H. 2014. Transcriptome analysis reveals the same S-Locus F-Box genes in two haplotypes of the self-incompatibility locus of *Petunia inflata*. *Pl. Cell* 26: 2873–2888. http://dx.doi.org/10.1105/tpc.114.126920

Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A.J., Philippe, H., Melkonian, M. & Becker, B. 2011. Origin of land plants: Do conjugating green algae hold the key? *B. M. C. Evol. Biol.* 11: 104. http://dx.doi.org/10.1186/1471-2148-11-104

Wolf, J.B.W. 2013. Principles of transcriptome analysis and gene expression quantification: An RNA-seq tutorial. *Molec. Ecol. Resources* 13: 559–572. http://dx.doi.org/10.1111/1755-0998.12109

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B. & Rieseberg, L.H. 2009. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 106: 13875–13879. http://dx.doi.org/10.1073/pnas.0811575106

Xi, Z., Rest, J.S. & Davis, C.C. 2013. Phylogenomics and coalescent analyses resolve extant seed plant relationships. *PLOS ONE* 8: e80870. http://dx.doi.org/10.1371/journal.pone.0080870

Xi, Z., Liu, L., Rest, J.S. & Davis, C.C. 2014. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Syst. Biol.* 63: 912–932. http://dx.doi.org/10.1093/sysbio/syu055

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Xu, X., Wang, G.K.-S. & Wang, J. 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30: 1660–1666. http://dx.doi.org/10.1093/bioinformatics/btu077

Xin, H., Zhu, W., Wang, L., Xiang, Y., Fang, L., Li, J., Sun, X., Wang, N., Londo, J.P. & Li, S. 2013. Genome wide transcriptional profile analysis of *Vitis amurensis* and *Vitis vinifera* in response to cold stress. *PLOS ONE* 8(3): e58740. http://dx.doi.org/10.1371/journal.pone.0058740

Xu, C., Bai, Y., Lin, X., Zhao, N., Hu, L., Gong, Z., Wendel, J.F. & Liu, B. 2014. Genome-wide disruption of gene expression in allopolyploids but not hybrids of rice subspecies. *Molec. Biol. Evol.* 31: 1066–1076. http://dx.doi.org/10.1093/molbev/msu085

Xu, P., Kong, Y., Song, D., Huang, C. & Li, X. 2014. Conservation and functional influence of alternative splicing in wood formation of *Populus* and *Eucalyptus*. *B. M. C. Genomics* 15: 780. http://dx.doi.org/10.1186/1471-2164-15-780

Yang, X., Cheng, Y.-F., Deng, C., Ma, Y., Wang, Z.-W., Chen, X.-H. & Xue, L.-B. 2014. Comparative transcriptome analysis of eggplant (*Solanum melongena* L.) and turkey berry (*Solanum torvum* Sw.): Phylogenomics and disease resistance analysis. *B. M. C. Genomics* 15: 412. http://dx.doi.org/10.1186/1471-2164-15-412

Yang, Y. & Smith, S.A. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molec. Biol. Evol.* 31: 3081–3092. http://dx.doi.org/10.1093/molbev/msu245

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molec. Biol. Evol.* 24: 1586–1591. http://dx.doi.org/10.1093/molbev/msm088

Yi, T.-S., Jin, G.-H. & Wen, J. 2015. Chloroplast capture and intra-and inter-continental biogeographic diversification in the Asian–New World disjunct plant genus *Osmorhiza* (Apiaceae). *Molec. Phylogen. Evol.* 85: 10–21. http://dx.doi.org/10.1016/j.ympev.2014.09.028

Yoo, M.-J. & Wendel, J.F. 2014. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLOS Genet.* 10: e1004073. http://dx.doi.org/10.1371/journal.pgen.1004073

Zeng, J., Liu, Y., Liu, W., Liu, X., Liu, F., Huang, P., Zhu, P., Chen, JH., Shi, M., Guo, F., Cheng, P., Zeng, J., Liao, Y. Gong, J., Zhang, H.M., Wang, D., Guo, A.Y. & Xiong, X. 2013. Integration of transcriptome, proteome and metabolism data reveals the alkaloids biosynthesis in *Macleaya cordata* and *Macleaya microcarpa*. *PLOS ONE* 8: e53409. http://dx.doi.org/10.1371/journal.pone.0053409

Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N. & Ma, H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature, Commun.* 5: 4956. http://dx.doi.org/10.1038/ncomms5956

Zhang, D., Qi, J., Yue, J., Huang, J., Sun, T., Li, S., Wen, J.F., Hettenhausen, C., Wu, J., Wang, L., Zhuang, H., Wu, J. & Sun, G. 2014. Root parasitic plant *Orobanche aegyptiaca* and shoot parasitic plant *Cuscuta australis* obtained Brassicaceae-specific strictosidine synthase-like genes by horizontal gene transfer. *B. M. C. Pl. Biol.* 14: 19. http://dx.doi.org/10.1186/1471-2229-14-19

Zhang, N., Wen, J. & Zimmer, E.A. 2015. Expression patterns of *API*, *FUL*, *FT* and *LEAFY* orthologs in Vitaceae support the homology of tendrils and inflorescences throughout the grape family. *J. Syst. Evol.*, early view. http://dx.doi.org/10.1111/jse.12138

Zimmer, E.A. & Wen, J. 2012. Using nuclear gene data for plant phylogenetics: Progress and prospects. *Molec. Phylogen. Evol.* 65: 774–785. http://dx.doi.org/10.1016/j.ympev.2012.07.015

Zimmer, E.A. & Wen, J. In press. Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *J. Syst. Evol.* http://dx.doi.org/10.1111/jse.12174

Zouari, I., Salvioli, A., Chialva, M., Novero, M., Miozzi, L., Tenore, G.C., Bagnaresi, P. & Bonfante, P. 2014. From root to fruit: RNA-Seq analysis shows that arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism. *B. M. C. Genomics* 15: 221. http://dx.doi.org/10.1186/1471-2164-15-221

Zwickl, D.J. 2006. *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Ph.D. thesis, University of Texas at Austin, U.S.A.