Zoological Systematics, 45(3): 151–162 (July 2020), DOI: 10.11865/zs.202021

COMMUNICATION

Phylogenomic approaches in systematic studies

Junxia Zhang*, Jiaxing Lai

The Key Laboratory of Invertebrate Systematics and Application, Institute of Life Science and Green Development, College of Life Sciences, Hebei University, Baoding 071002, Hebei, P. R. China; E-mail: jxzhang1976@163.com

Abstract Phylogenomic approaches are growingly applied in the systematic studies of animals with the rapid advancement of sequencing and analytical tools. Aiming to accelerate the systematic studies using genome-scale data in China, this paper outlines the commonly used techniques for generating phylogenomic datasets: RNA-seq based transcriptome, target-enrichment based UCEs and AHE, RAD-seq, whole genome sequencing, and briefly discusses the pros and cons of each technique. Major analytical procedures for the 1KITE transcriptome pipeline, the PHYLUCE (UCEs) and AHE pipelines and the PLWS (low-coverage whole genome sequencing) pipeline are summarized, and recent achievements valuable for phylogenomic analyses are introduced.

Key words Transcriptome, UCEs, AHE, RAD-seq, WGS.

Systematics, commonly defined as the study of biological diversity and relationships among organisms, highly depends upon phylogenetic studies to provide reliable frameworks for classifications. However, traditional phylogenetic studies using morphological characters and/or handful molecular markers from Sanger-sequencing technique, especially on the groups with megadiversity and rapid radiation, often resulted in poorly supported and inadequately resolved phylogenetic relationships due to limited number of phylogenetically informative characters, which hinder our further understanding of systematics and evolution of the groups (e.g. Munro *et al*., 2011; Heraty *et al*., 2013; Zhang & Maddison., 2013, 2015). With the rapid development of next-generation sequencing (NGS) technologies since mid-2000s, enormous amounts of genome-scale data are currently available for phylogenetic studies; the field of phylogenetic analyses has entered into the era of phylogenomics (Delsuc *et al*., 2005; Young & Gillung, 2020).

Phylogenomics, the study of evolutionary history and function based on molecular analyses of genome-scale sequence data, comprises several areas of research such as predicting putative functions of DNA or protein sequences, identifying signatures of molecular adaptation and inferring historical relationships among species (Kumar *et al*., 2012; Young & Gillung, 2020). Phylogenomic approaches have shed lights on the reconstruction of tree of life, and have been widely utilized in phylogenetic studies of a wide range of animal groups, such as birds (Prum *et al*., 2015), mammals (Chen *et al*., 2017; Árnason *et al*., 2018), frogs (Peloso *et al*., 2015), insects (Misof *et al*., 2014) and spiders (Garrison *et al*., 2016). With the great advantage of hundreds to thousands of loci from across the genome, phylogenomic methods indeed outperform traditional multi-locus approaches in resolving evolutionary history (Blaimer *et al*., 2015). One classic example is Misof *et al*. (2014), in which widely sampled transcriptomes were applied in a deep-scale phylogenomic study of major insect lineages and the results provided robust evidence to clarify the previously controversial phylogenetic relationships. We briefly surveyed the annual publications with topic on phylogenomics in Web of Science database from 2009 to 2019 and found the publications on phylogenomic studies have dramatically increased in recent years worldwide (Fig. 1).

Here we overview the major approaches currently applied in phylogenomic studies and outline the recent advances in analyses of phylogenomic datasets. A comparison of the phylogenomic methods outlined in this paper is provided in Table 1. Phylogenetic studies based on extensive mitochondrial genome sequences have also provided valuable insights on

Received 25 April 2020, accepted 15 July 2020
Executive editor: Fuqiang Chen

systematics of various animal groups (e.g. Wang *et al*., 2017; Yang *et al*., 2018). But here we would like to focus on phylogenomic studies mainly using nuclear genome sequences.

Table 1. Comparison of major phylogenomic methods.

	Transcriptome	UCEs	AHE	RAD-Seq	WGS
If alcohol-preserved specimens applicable	No, must be specimens that are fresh or flash-frozen in liquid nitrogen or preserved in RNA later	Yes, historical museum specimens often acceptable	Yes, historical museum specimens often acceptable	Yes, historical museum specimens sometimes acceptable	Yes, historical museum specimens sometimes acceptable
Template requirement	High quantity and quality of RNA	DNA (low amount and medium degradation often acceptable)	DNA (low amount and medium degradation sometimes acceptable)	DNA (medium amount and low degradation often acceptable)	High quantity and quality of DNA (low degradation often acceptable)
General lab protocols	RNA extraction, library preparation, sequencing (can be mostly done by sequencing company)	DNA extraction, library preparation, target enrichment with probe kit, sequencing	DNA extraction, library preparation, target enrichment with probe kit, sequencing	DNA extraction, enzyme digestion, library preparation, size selection, sequencing	DNA extraction, library preparation, sequencing (can be mostly done by sequencing company)
Sequence type	Coding regions	Conservative (may include both coding and non-coding) and adjacent flanking regions	Target for coding regions, adjacent non-coding (often intron) regions also obtained	May include both coding and non-coding regions	Both coding and non-coding regions
Examples of analytical methods	1KITE pipeline	PHYLUCE	Lemmons' AHE pipeline	pyRAD and Stacks	PLWS
Suitability for organisms and phylogenetic relationships	Organisms with big or small genome size; most suitable for deep level relationships	Organisms with big or small genome size; suitable for both deep and shallow level relationships	Organisms with big or small genome size; suitable for both deep and shallow level relationships	Organisms with big or small genome size; most suitable for shallow level relationships	Organisms with genome size < 1Gbp; suitable for both deep and shallow relationships
Relative cost	High	Medium (new probe-design may add on additional cost)	Medium (new probe-design may add on additional cost)	Low	High

1 Acquisition of phylogenomic data

1.1 Transcriptome from RNA-seq

Phylogenomics using transcriptome data applies RNA-seq technique to obtain the expressed sequence tags (ESTs) in the organisms for phylogenetic reconstruction. ESTs have been widely utilized in animal phylogenetic analyses and proved to be useful for resolving some difficult deep relationships (e.g. Dunn *et al*., 2008; Irisarri & Meyer, 2016; Zhang *et al*., 2016; Schwentner *et al*., 2017; Fernández *et al*., 2018a, b). The launch of the globally collaborated 1KITE (1K Insect Transcriptome Evolution) project aiming to study the transcriptomes of more than 1,000 insect species has particularly

stimulated the transcriptome-based phylogenomic studies in insects (Misof *et al*., 2014; Derst *et al*., 2016; Peters *et al*., 2017; Johnson *et al*., 2018; Zhang *et al*., 2020), and driven the development of various analytical tools and procedures as well as a comprehensive pipeline useful for not only transcriptome data but also other types of genomic data (e.g. Misof & Misof, 2009; Kück & Meusemann, 2010; Dell'Ampio *et al*., 2014; Petersen *et al*., 2017).

Phylogenomic analyses using transcriptome data enable direct exploration of protein coding sequences and have the potential of facilitating the link between evolutionary history and gene function (Zhang *et al*., 2020). As the ESTs recovered through this procedure tend to be "housekeeping" genes and thus relatively conservative, they are mostly useful for the inference of deep relationships (Cannon & Kocot, 2016). However, NGS of transcriptome often needs high sequencing depth (or coverage, i.e. the average number of times an individual base in the genome is sequenced) in order to recover a more complete set of single-copy genes in an organism, which raises the sequencing cost for a large scale phylogenomic project using transcriptomes. An even more obvious drawback of transcriptome-based phylogenomics is that the RNA-seq procedure uses the extracted RNA as template for downstream library preparation and sequencing, and then needs high-quality tissues or specimens being flash-frozen in liquid nitrogen or directly preserved in RNA later. This prohibits the inclusion of some rarely collected but important taxa for phylogeny and the direct utilization of museum collections in phylogenomic studies.

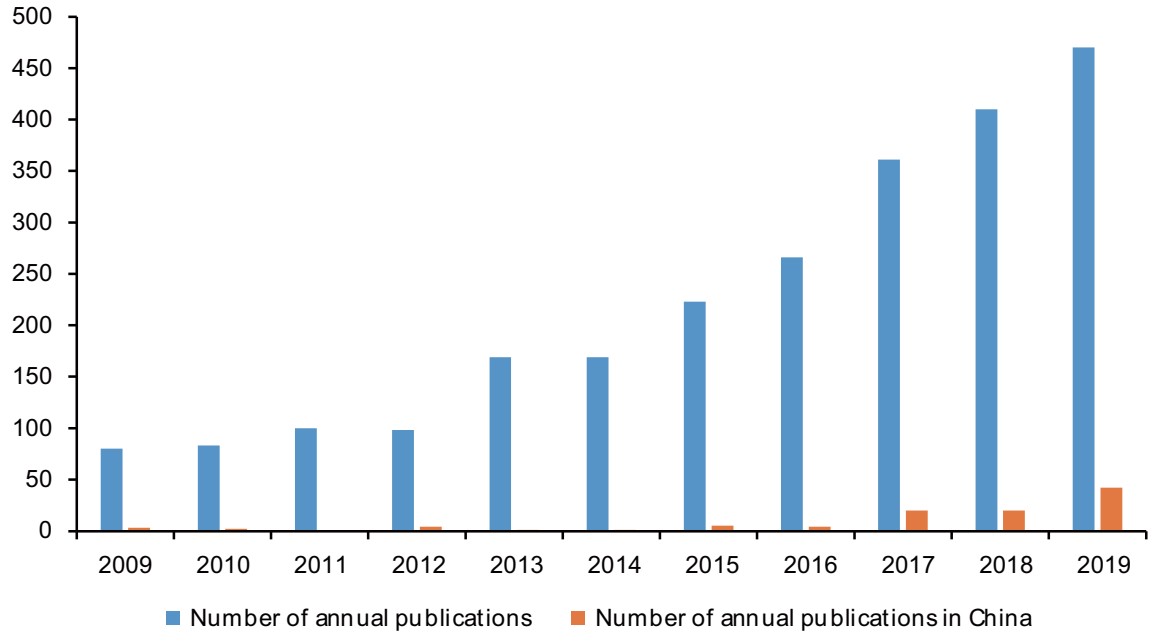

Figure 1. The number of annual publications with topic on phylogenomics in Web of Science database from 2009 to 2019.

1.2 UCEs and AHE from target enrichment

The Ultra-Conserved Elements (UCEs) and Anchored Hybrid Enrichment (AHE) methods, both rooted from the target enrichment procedure in which short nucleotide probes are first designed and then used to hybridize with the sheared genomic DNA templates to recover particular sequences of interest with high coverage (Young & Gillung, 2020), were developed independently by different research teams around the same time period (Faircloth *et al*., 2012; Lemmon *et al*., 2012). Although sharing a lot in common in the basic principles and procedures, UCEs and AHE indeed differ in the details of probe design procedure: UCEs refer to the highly conserved regions (≥80% identify and ≥100 bp) across divergent taxa (Faircloth *et al*., 2012) which could be coding regions or noncoding sequences, and the probes were designed from such UCE sequences identified from the aligned genomes of reference taxa (Faircloth *et al*., 2012; Faircloth, 2017); whereas in AHE the probes were designed mainly targeting for coding regions from alignment of genomes and transcriptomes of reference taxa (Lemmon *et al*., 2012; Baker *et al*., 2019). The analytical pipelines were also developed independently for UCE (Faircloth, 2016) and AHE (see details in Prum *et al*., 2015 and Young *et al*., 2016) data.

UCEs and AHE directly use extracted DNA for library preparation. Studies have shown that the historical museum

collections could also generate reasonable UCE and AHE data for phylogenetic inference (Blaimer *et al*., 2016; Stlaurent *et al*., 2018). In addition, because UCEs and AHE only target for about 500-2000 enriched loci, we can often multiplex a lot more samples in one sequencing lane and still generate enough reads to recover the targeted genes, so the sequencing cost is dramatically reduced. Therefore, UCEs and AHE have been widely applied in animal phylogenomic studies and have proved to be useful for resolving both deep and shallow-level relationships (e.g. Blaimer *et al*., 2015; Peloso *et al*., 2015; Prum *et al*., 2015; Branstetter *et al*., 2017; Ješovnik *et al*., 2017; Van Dam *et al*., 2017; Derkarabetian *et al*., 2018; Winterton *et al*., 2018; Baker *et al*., 2019; Blair *et al*., 2019; Hedin *et al*., 2019; Kulkarni *et al*., 2019). The AHE-based phylogenomic projects were largely conducted through collaboration with the Lemmons Lab at Florida State University (USA), where the DNA extracts were received from collaborators, the lab procedures on sequence capture were then completed using designed probes (probe set may be customized for certain project) and sent out for sequencing. The Lemmons Lab usually conduct preliminary analyses on the obtained AHE data before they are delivered to the collaborators for more comprehensive data manipulation and analyses. The UCE projects, in contrast, are usually more independent mainly benefiting from the open-source UCE pipeline PHYLUCE (Faircloth, 2016), which is user-friendly with detailed tutorials to help beginners to quickly overcome the obstacles in designing UCE probes and downstream UCE data process and analyses. Some pre-designed probe-kits ready to use are available from companies (such as Arbor Biosciences https://arborbiosci.com/genomics/targeted-sequencing/mybaits/mybaits-expert/mybaits-expert-uce/) for the laboratory protocols of sequence capture.

1.3 RAD-seq

Restriction site Associated DNA sequencing (RAD-seq) is similar to UCEs and AHE in selectively sequencing part of the whole genome, but this is accomplished in RAD-seq by utilizing one or more restriction enzymes to fragment genomic DNA and then selecting fragments that fall within a certain size distribution prior to sequencing (Miller *et al.*, 2007; Andrews *et al.*, 2016; Harvey *et al.*, 2016). One great advantage of RAD-seq is its low cost. Depending on the number of markers to be selected and sequenced, this method could be applied to large-scale projects involving hundreds or even thousands of individuals at an affordable cost (McKain *et al.*, 2018). RAD-seq does not need probes which reduces the upfront investment on designing probes and/or purchasing probe-set kit. Another strength of this approach is its flexibility: by simply choosing a more frequently cutting restriction enzyme or a wider range of fragment size, we can often obtain more genetic markers; *vice versa* (McKain *et al.*, 2018). Unlike UCEs and AHE, RAD-seq does not rely on targeting relatively conservative or coding regions of the genome and the sequenced RAD markers are often highly variable, so it can result in thousands or even millions of SNPs which are extremely suitable for resolving shallow level phylogenetic relationships of closely related species or population genetic studies (e.g. Dupuis *et al.*, 2017; Fang *et al.*, 2018; Theodorou *et al.*, 2018).

The RAD-seq data could be analyzed using reference-based or de novo approach with available tools such as pyRAD (Eaton, 2014) or Stacks (Catchen *et al.*, 2013). They could either output SNP-based sequences or full sequence data for traditional phylogenetic analyses (McKain *et al.*, 2018). A big challenge in RAD-seq based study is missing data. RAD-seq relies on the conservation of restriction recognition sites across samples to recover homologous markers, and thus disruption of these sites by mutations results in missing data and more divergent taxa are expected to share fewer RAD markers (McKain *et al.*, 2018). This may hinder the application of RAD-seq approach to deeper phylogenetic scales.

1.4 Whole genome sequencing

Compared with transcriptomes, UCEs and AHE, fewer phylogenomic studies have utilized the whole genome sequencing (WGS) method probably due to its high demands for computational resources and sequencing costs (but see Nater *et al*., 2015; Árnason *et al*., 2018; Olofsson *et al*., 2019; Zhang *et al*., 2019). However, the rapid development of gene sequencing technologies has prompted the initiation of many genome sequencing projects, e.g. the Genome 10 K project, the Bird 10 K project and the "i5k" project aiming to sequence the genomes of 5,000 arthropods with important biological significance or economic value (Li *et al*., 2019), which will dramatically increase the number of genomes available for building the tree of life.

The laboratory protocols for WGS procedure are much simpler than UCEs, AHE and RAD-Seq, in which the researchers usually only need to extract high-quality DNA in the lab, and then send the DNA samples to sequencing companies for all the downstream library preparation and sequencing. In contrast, the UCEs and AHE protocols involve complicated procedures for generating the sequence capture libraries which often need to be completed by the researchers in the lab before sequencing. Because WGS needs DNA as template for library preparation, it is less stringent on the specimen/tissue preservation than transcriptomes. Although WGS at high sequencing depth is still relatively expensive, low-coverage WGS for organisms with small genome size is often affordable and sufficient to generate enough genome-scale

data for phylogenetic inference (Olofsson *et al*., 2019; Zhang *et al*., 2019). However, the application of low-coverage WGS for organisms with large genome size of several giga base pairs is still challenging.

Two kinds of analytical approaches are available for phylogenomic studies of low-coverage WGS data: One is an assembly-free and reference-based approach, in which the sequenced raw reads are directly mapped to a known reference genome to call SNPs and extract consensus sequences (Árnason *et al*., 2018; Olofsson *et al*., 2019). This method is more suitable for the inference of shallow nodes with closely related taxa, but may be difficult to apply in phylogenetic studies of deep relationships because more distantly related taxa will tend to have trouble to map to the reference genome. The other approach first conducts de novo assembly of genomes and then extracts protein-coding-gene dataset and UCE dataset (Zhang *et al*., 2019; Sun *et al*., 2020) for phylogenetic inference. Because it is not dependent on a reference genome, this method has a wide application in phylogenomic studies of both deep and shallow relationships.

2 Analytical procedures on phylogenomic data

Phylogenomic projects usually involve a large amount of sequence data which are often challenging to analyze compared to Sanger-based multi-locus data. Fortunately, along the years many analytical tools and methods have been developed and multiple sophisticated pipelines are currently available for manipulation and analyses of massive phylogenomic datasets. The analytical pipelines often contain complex steps and involve lots of computer programs. The 1KITE phylotrancriptomic pipeline, as an example, is summarized in Fig. 2 (for the detailed explanation of the pipeline see the supplementary material in Misof *et al*., 2014 and Peters *et al*., 2017). Major analytical steps for phylogenomic procedures are briefly introduced below, with an emphasis on the 1KITE phylotrancriptomic pipeline (Misof *et al*., 2014; Peters *et al*., 2017), the UCE (PHYLUCE; Faircloth, 2016) and AHE (see details in Prum *et al*., 2015 and Young *et al*., 2016) pipelines, and the low-coverage WGS pipeline (PLWS; Zhang *et al*., 2019; Sun *et al*., 2020).

2.1 Quality control and reads assembly

Quality control removes low-quality bases and adapter contamination in raw reads obtained from NGS and can be completed by several available programs such as Illumiprocessor (Faircloth, 2013), Trimmomatic (Bolger *et al*., 2014) and BBTools (Bushnell, 2014). The assembly of reads into contigs in 1KITE, PLYLUCE and PLWS pipelines applies de novo assembly technique with programs such as Velvet (Zerbino & Birney, 2008), Trinity (Grabherr, 2011), SOAPdenovo (Li *et al*., 2010), ABYSS (Jackman *et al*., 2017) or Minia3 (Chikhi *et al*., 2016). The AHE pipeline uses a so-called Quasi-de-novo assembler (java script, Prum *et al*., 2015), which assembles the conservative probe regions using a divergent reference assembly approach and the more variable flanking regions using de novo assembly approach (Prum *et al*., 2015; Young *et al*., 2016).

2.2 Orthology assignment

An essential step in phylogenomic pipelines is the determination of orthologous groups, in which the assembled sequences are sorted into orthologues, i.e. sequences with common ancestor diverged from speciation event rather than duplication event (paralogues). Accurate orthology assignment is crucial for downstream phylogenetic inference. Both PHYLUCE and AHE pipelines have customized scripts for orthology assignment. The PHYLUCE pipeline assigns the assembled contigs to corresponding UCE loci and remove potential paralogues by aligning the contigs to probes and then finding matches (see PHYLUCE online instruction); whereas in the AHE pipeline the orthology assignment is achieved by computing pairwise distances of assembled contigs and then clustering the sequences based on the distance matrix (see details in Prum *et al*., 2015).

The orthology assignment in the 1KITE pipeline is accomplished through HaMStR (Ebersberger *et al*., 2009) or Orthograph (Petersen *et al*., 2017), both of which are similarity or graph-based methods applying a best reciprocal hit search strategy with profile hidden Markov models and mapping nucleotide sequences to the globally best matching cluster of orthologous genes (Petersen *et al*., 2017). Orthology assignment in these programs requires pre-defined orthologous groups (or Core Orthologue Set, COS) which contain sequences of a collection of single-copy protein coding genes in reference species. In addition to some pre-compiled COSs used in previous studies (Misof *et al*., 2014; Chesters, 2017; Peters *et al*., 2017; Zhang *et al*., 2020), the COS could be customized for specific project based on the orthologous protein coding gene information for the reference taxa of interest available in the public databases such as OrthoDB (Kriventseva *et al*., 2015),

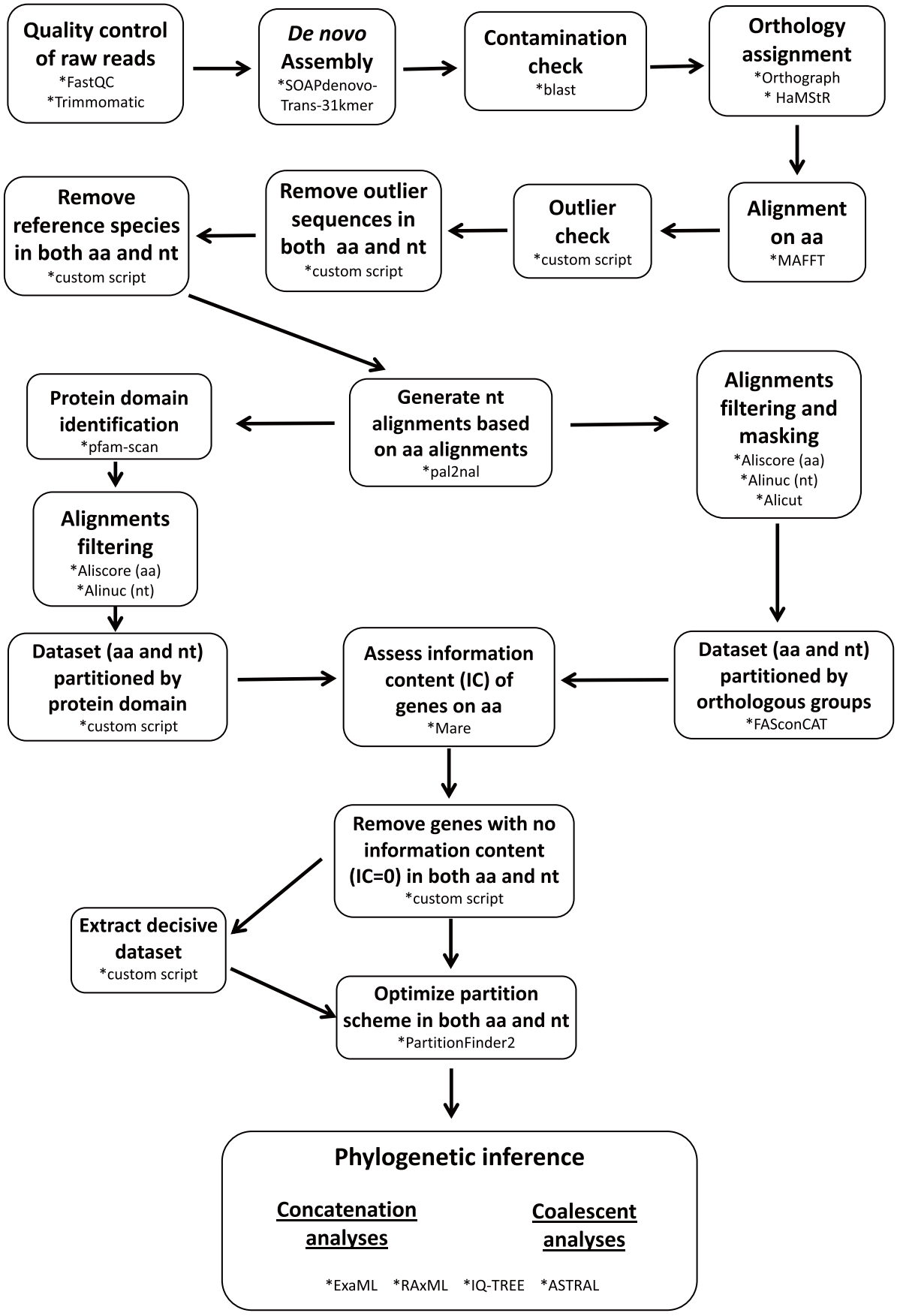

Figure 2. Summary of 1KITE phylotranscriptomic pipeline (aa—amino acid; nt—nucleotide).

but will take some manipulation on the database output. An alternative to these two programs is BUSCO (Benchmarking Universal Single-Copy Orthologs; Seppey *et al*., 2019). BUSCO was originally developed to assess the completeness of genome assembly and annotation but holds great potential for orthology assignment by identifying near-universal single-copy orthologs based on the OrthoDB database, and has been directly applied in some phylogenomic studies (e.g. Fernández *et al*., 2018a, b) and implemented in the PLWS pipeline (Zhang *et al*., 2019). A collection of pre-compiled sets suitable for orthology assignment of various lineages, such as Bacteria, Metazoa, Arthropoda, Insecta and Vetebrata, are currently available in BUSCO and could be directly downloaded (https://busco.ezlab.org/busco_v4_data.html). We expect that with more thoroughly annotated genomes and comprehensive understanding of genealogical relationship of genes, an increasing number of sets will be available in BUSCO for orthology assignment of diverse lineages on the tree of life.

2.3 Multiple-sequence alignment and alignment masking/trimming

Several programs for multiple-sequence alignment (MSA) developed for Sanger-based multi-locus data, such as MAFFT (Katoh & Standley, 2013) and MUSCLE (Edgar, 2004), can be directly implemented in phylogenomic pipelines. A masking or trimming step is usually conducted to remove the poorly aligned regions or errors in the alignments following the multiple-sequence alignment. The alignment masking/trimming programs, such as Alicore and Alicut (Misof & Misof, 2009), Gblocks (Castresana, 2000) and trimAl (Capella-Gutiérrez *et al*., 2009), often target and remove blocks of columns in the sequence alignment and only retain the relatively conservative and well-aligned regions for downstream phylogenetic inference. While removing the noise in alignments some informative sites are often trimmed as well inevitably. An alternative is the site-based masking approach using SeqTools package (Barson & Griffiths, 2016) which particularly removes the sites in an alignment with high proportion of missing characters and very short sequences. The site-based trimming procedure may lower the removal of informative sites during alignment masking. Another useful tool for cleaning MSA is Spruceup, which removes the poorly aligned fragments in each sequence of MSA (Borowiec, 2019).

2.4 Post-pipeline proof-checking

Phylogenomic analyses have become more and more dependent on automated bioinformatic pipelines due to the rapid increase of the amount of data in a typical phylogenomic dataset. The standard practice of manually screening out contaminations and paralogous loci and curating gene alignments in the PCR and Sanger-sequencing age of molecular systematics are often overlooked in phylogenomic approaches. Personal experience working with multiple phylogenomic pipelines and previous studies suggest that the pipeline-processed datasets are not immune to errors even though some outlier-check or error-check steps are implemented in the pipelines (e.g. 1KITE pipeline in Fig. 2; Gatesy & Springer, 2017; Shen *et al*., 2017; Zhang *et al*., 2020). Even a small proportion of contamination and homology errors in a phylogenomic dataset could result in mistaken inference of phylogenetic relationships (Zhang *et al*., 2020). Therefore, a post-pipeline proof-checking (careful inspection of alignments and gene trees) may be necessary or even critical for building a robust phylogeny in phylogenomic studies.

2.5 Phylogenetic inference

Inference of phylogeny using genome-scale data can be accomplished through concatenation-based methods in which multiple gene alignments are concatenated to build a supermatrix for phylogenetic reconstruction. Several programs, such as IQ-TREE (Nguyen *et al*., 2015), RAxML (Stamatakis, 2014) and FastTree (Price *et al*., 2010), have been fine-tuned to handle very large phylogenomic datasets. In addition to the concatenation-based methods, phylogenomic datasets are often analyzed using coalescent methods (e.g. ASTRAL; Mirarab *et al*., 2014), in which gene tree is built for each gene alignment and then the species tree is estimated from the set of gene trees accounting for potential gene tree heterogeneity and discordance. Shallow-scale studies such as species delimitation often apply other methods accounting for multispecies coalescent, e.g. the quartet-based SVD-quartets (Chifman & Kubatko, 2014); the Bayesian-based SNAPP (in BEAST2; Bouckaert *et al*., 2014), *BEAST (Heled & Drummond, 2010) and BPP (Yang, 2015). However, the Bayesian approaches are usually computationally intensive and may be difficult to apply in large-scale phylogenomic analyses.

To reduce the artefact of missing data in phylogenetic inference, the phylogenomic data are often filtered to compose a dataset with more complete sequences for all or most of the sampled taxa. A useful strategy for reducing the non-random distribution of missing data in phylogenomic studies is to build the “decisive dataset”, in which only data blocks that contain sequences with at least one representative in each of the predefined taxonomic groups are retained (Dell’Ampio *et al*., 2014). Partition optimization and model selection are essential for phylogenetic reconstruction and can be achieved by commonly

used programs such as PartitionFinder (Lanfear *et al*., 2016) and ModelFinder (Kalyaanamoorthy *et al*., 2017), which are also implemented in the program IQ-TREE. Recent studies have suggested that the site-heterogeneous models accounting for site heterogeneity of frequencies, such as CAT (Lartillot & Philippe, 2004), are helpful for resolving some difficult phylogenetic questions (Wang *et al*., 2019). Methods that model such heterogeneity of site frequency have been developed and implemented in IQ-TREE to alleviate the computational burden of applying CAT model in phylogenomic analyses by employing either fixed empirical amino acid frequency profiles (Wang *et al*., 2008) or a posterior mean site frequency (PMSF) model (Wang *et al*., 2018). The assessment of branch support using standard bootstrap scheme usually takes prohibitively long computational time and is often completed by alternative methods such as rapid bootstrap in RAxML (Stamatakis *et al*., 2008) and ultrafast bootstrap in IQ-TREE (Hoang *et al*., 2018) for a large phylogenomic dataset. In addition, the quartet sampling method (Pease *et al*., 2018) and concordance factors (Minh *et al*., 2018) provide alternative measures of branch support that are valuable to detect potential underlying conflict in the phylogenomic dataset.

In summary, phylogenomic approaches have shown great potential in studies on systematics and evolution of animals, and the rapid development of high throughput sequencing technology and analytical tools are strongly promoting researches on such areas. In China we also see growing interest in applying genome-scale data in phylogenetic studies (Fig. 1). Through this paper as well as several published reviews on this topic (Delsuc *et al*., 2005; Posada, 2016; Young & Gillung, 2020), we hope that more and more Chinese researchers will quickly grasp the basic principles and recent advances in this field and start to apply the phylogenomic approaches in their projects. With increased funding support from our government, development of more user-friendly phylogenomic pipelines and platforms and relentless effort of Chinese scientists, we are confident of catching up with the speed of phylogenomic studies in the world.

Acknowledgements This work was supported by the Advanced Talents Incubation Program of the Hebei University to Junxia Zhang (521000981324). We would like to express our gratitude to Dr. Shuqiang Li, Dr. Zhisheng Zhang, Dr. Chao Zhang and two anonymous reviewers for their valuable comments on the manuscript.

References

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2): 81–92. doi: 10.1038/nrg.2015.28.

Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M.A., Janke, A. 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science Advances*, 4(4): 1–10. doi: 10.1126/sciadv.aap9873.

Baker, A.J., Heraty, J.M., Mottern, J., Zhang, J., Hines, H.M., Lemmon, A.R., Lemmon, E.M. 2019. Inverse dispersal patterns in a group of ant parasitoids (Hymenoptera: Eucharitidae: Oraseminae) and their ant hosts. *Systematic Entomology*, 45(1): 1–19. doi: 10.1111/syen.12371.

Barson, G., Griffiths, E.D. 2016. SeqTools: visual tools for manual analysis of sequence alignments. *BMC Research Notes*, 9: 1–10. doi: 10.1186/s13104-016-1847-3.

Blaimer, B.B., Brady, S.G., Schultz, T.R., Lloyd, M.W., Fisher, B.L., Ward, P.S. 2015. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. *BMC Evolutionary Biology*, 15: 271–285. doi: 10.1186/s12862-015-0552-5.

Blaimer, B.B., Lloyd, M.W., Guillory, W.X., Brady, S.G. 2016. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLOS ONE*, 11(8): e0161531–e0161551. doi: 10.1371/journal.pone.0161531.

Blair, C., Bryson Jr, R.W., Linkem, C.W., Lazcano, D., Klicka, J., McCormack, J.E. 2019. Cryptic diversity in the Mexican highlands: thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: *Crotalus*). *Molecular Ecology Resources*, 19(2): 349–365. doi: 10.1111/1755-0998.12970.

Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114–2120. doi: 10.1093/bioinformatics/btu170.

Borowiec, M.L. 2019. Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. *Journal of Open Source Software*, 4(42): 1635–1641. doi: 10.21105/joss.01635.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J. 2014. BEAST2: a software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 10: e1003537. doi: 10.1371/journal.pcbi.1003537.

Branstetter, M.G., Danforth, B.N., Pitts, J.P., Faircloth, B.C., Ward, P.S., Buffington, M.L., Gates, M.W., Kula, R.R., Brady, S.G. 2017. Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Current Biology*, 27(7): 1019–1025.

doi: 10.1016/j.cub.2017.03.027.
Bushnell, B. 2014. BBtools. Available from https://sourceforge.net/ projects/bbmap/ (accessed 21 April 2020).
Cannon, J.T., Kocot, K.M. 2016. Phylogenomics using transcriptome data. *Methods in Molecular Biology*, 1452: 65–80. doi: 10.1007/978-1-4939-3774-5_4.
Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T. 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15): 1972–1973. doi: 10.1093/bioinformatics/btp348.
Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4): 540–552. doi: 10.1093/oxfordjournals.molbev.a026334.
Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11): 3124–3140. doi: 10.1111/mec.12354.
Chen, M.Y., Liang, D., Zhang, P. 2017. Phylogenomic resolution of the phylogeny of laurasiatherian mammals: exploring phylogenetic signals within coding and noncoding sequences. *Genome Biology and Evolution*, 9(8): 1998–2012. doi: 10.1093/gbe/evx147.
Chesters, D. 2017. Construction of a species-level tree of life for the insects and utility in taxonomic profiling. *Systematic Biology*, 66(3): 426–439. doi: 10.1093/sysbio/syw099.
Chifman, J., Kubatko, L.S. 2014. Quartet inference from SNP data under the coalescence model. *Bioinformatics*, 30(23): 3317–3324. doi: 10.1093/bioinformatics/btu530.
Chikhi, R., Limasset, A., Medvedev, P. 2016. Compacting de Bruijn graphs from sequencing data quickly and in low memory. *Bioinformatics*, 32(12): i201–i208. doi: 10.1093/bioinformatics/btw279.
Dell'Ampio, E., Meusemann, K., Szucsich, N.U., Peters, R.S., Meyer, B., Borner, J., Petersen, M., Aberer, A.J., Stamatakis, A., Walzl, M.G., Minh, B.Q., von Haeseler, A., Ebersberger, I., Pass, G., Misof, B. 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Molecular Biology and Evolution*, 31(1): 239–249. doi: 10.1093/molbev/mst196.
Delsuc, F., Brinkmann, H., Philippe, H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5): 361–375. doi: 10.1038/nrg1603.
Derkarabetian, S., Starrett, J., Tsurusaki, N., Ubick, D., Castillo, S., Hedin, M. 2018. A stable phylogenomic classification of Travunioidea (Arachnida, Opiliones, Laniatores) based on sequence capture of ultraconserved elements. *ZooKeys*, 2018(760): 1–36. doi: 10.3897/zookeys.760.24937.
Derst, C., Dircksen, H., Meusemann, K., Zhou, X., Liu, S., Predel, R. 2016. Evolution of neuropeptides in non-pterygote hexapods. *BMC Evolutionary Biology*, 16: 51–61. doi: 10.1186/s12862-016-0621-4.
Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sørensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452: 745–749. doi: 10.1038/nature06614.
Dupuis, J.R., Brunet, B.M.T., Bird, H.M., Lumley, L.M., Fagua, G., Boyle, B., Levesque, R., Cusson, M., Powell, J.A., Sperling, F.A.H. 2017. Genome-wide SNPs resolve phylogenetic relationships in the North American spruce budworm (*Choristoneura fumiferana*) species complex. *Molecular Phylogenetics and Evolution*, 111: 158–168. doi: 10.1016/j.ympev.2017.04.001.
Eaton, D.A.R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13): 1844–1849. doi: 10.1093/bioinformatics/btu121.
Ebersberger, I., Strauss, S., von Haeseler, A. 2009. HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evolutionary Biology*, 9: 157–166. doi:10.1186/1471-2148-9-157.
Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5): 1792–1797. doi: 10.1093/nar/gkh340.
Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5): 717–726. doi: 10.1093/sysbio/sys004.
Faircloth, B.C. 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. Available from http://dx.doi.org/10.6079/J9ILL (accessed 22 April 2020).
Faircloth, B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5): 786–788. doi: 10.1093/bioinformatics/btv646.
Faircloth, B.C. 2017. Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, 8(9): 1103–1112. doi: 10.1111/2041-210X.12754.
Fang, B., Merilää, J., Ribeirob, F., Alexandrec, C.M., Momigliano, P. 2018. Worldwide phylogeny of three-spined sticklebacks. *Molecular Phylogenetics and Evolution*, 127: 613–625. doi: 10.1016/j.ympev.2018.06.008.
Fernández, R., Edgecombe, G.D., Giribet, G. 2018a. Phylogenomics illuminates the backbone of the myriapoda tree of life and reconciles morphological and molecular phylogenies. *Scientific Reports*, 8(1): 83–90. doi: 10.1038/s41598-017-18562-w.
Fernández, R., Kallal, R.J., Dimitrov, D., Ballesteros, J.A., Arnedo, M.A., Giribet, G. Hormiga, G. 2018b. Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Current Biology*, 28(9): 1489–1497. doi: 10.1016/j.cub.2018.03.064.

Garrison, N.L., Rodriguez, J., Agnarsson, I., Coddington, J.A., Griswold, C.E., Hamilton, C.A., Hedin, M., Kocot, K.M., Ledford, J.M., Bond, J.E. 2016. Spider phylogenomics: untangling the Spider Tree of Life. *PeerJ*, 4: e1719–e1754. doi: 10.7717/peerj.1719.

Gatesy, J., Springer, M.S. 2017. Phylogenomic red flags: homology errors and zombie lineages in the evolutionary diversification of placental mammals. *Proceedings of the National Academy of Sciences*, 114(45): e9431–e9432. doi: 10.1073/pnas.1715318114.

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. Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nature Biotechnology*, 29(7): 644–652. doi: 10.1038/nbt.1883.

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, 65(5): 910–924. doi: 10.7717/peerj.6864.

Hedin, M., Derkarabetian, S., Alfaro, A., Ramírez, M.J., Bond, J.E. 2019. Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci. *PeerJ*, 7: e6864–e6888. doi: 10.7717/peerj.6864.

Heled, J., Drummond, A.J. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, 27(3): 570–580. doi: /10.1093/molbev/msp274.

Heraty, J.M., Burks, R.A., Cruaud, A., Gibson, G.A.P., Liljeblad, J., Munro, J., Rasplus, J-Y., Delvare, G., Janšta, P., Gumovsky, A., Huber, J., Woolley, J.B., Krogmann, L., Heydon, S., Polaszek, A., Schmidt, S., Darling, D.C., Gates, M.W., Mottern, J., Murray, E., Molin, A.D., Triapitsyn, S., Baur, H., Pinto, J.D., van Noort, S., George, J., Yoder, M. 2013. A phylogenetic analysis of the megadiverse Chalcidoidea (Hymenoptera). *Cladistics*, 29(5): 466–542. doi: 10.1111/cla.12006.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2): 518–522. doi: 10.1093/molbev/msx281.

Irisarri, I., Meyer, A. 2016. The identification of the closest living relative(s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. *Systematic Biology*, 65(6): 1057–1075. doi: 10.1093/sysbio/syw057.

Jackman, S.D., Vandervalk, B.P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S.A., Jahesh, G., Khan, H., Coombe, L., Warren, R.L., Birol, I. 2017. Abyss 2.0: resource-efficient assembly of large genomes using a bloom filter. *Genome Research*, 27(5): 768–777. doi: 10.1101/gr.214346.116.

Ješovnik, A., Sosa-Calvo, J., Lloyd, M.W., Branstetter, M.G., FernÁndez, F., Schultz, T.R. 2017. Phylogenomic species delimitation and host-symbiont coevolution in the fungus-farming ant genus *Sericomyrmex* Mayr (Hymenoptera: Formicidae): ultraconserved elements (UCEs) resolve a recent radiation. *Systematic Entomology*, 42(3): 523–542. doi: 10.1111/syen.12228.

Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B., Peters, R.S., Allen, J.M., Petersen, M., Donath, A., Walden, K.K.O., Kozlov, A.M., Podsiadlowski, L., Mayer, C., Meusemann, K., Vasilikopoulos, A., Waterhouse, R.M., Cameron, S.L., Weirauch, C., Swanson, D.R., Percy, D.M., Hardy, N.B., Terry, I., Liu, S., Zhou, X., Misof, B., Robertson, H.M., Yoshizawa, K. 2018. Phylogenomics and the evolution of hemipteroid insects. *PNAS*, 115(50): 12775–12780. doi: 10.1073/pnas.1815820115.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6): 587–589. doi: 10.1038/nmeth.4285.

Katoh, K., Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4): 772–780. doi: 10.1093/molbev/mst010.

Kriventseva, E.V., Tegenfeldt, F., Petty, T.J., Waterhouse, R.M., Simão, F.A., Pozdnyakov, I.A., Ioannidis, P., Zdobnov, E.M., 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Research*, 43(D1): D250–D256. doi: 10.1093/nar/gku1220.

Kück, P., Meusemann, K. 2010. FASconCAT: Convenient handling of data matrices. *Molecular Phylogenetics and Evolution*, 56(3): 1115–1118. doi: 10.1016/j.ympev.2010.04.024.

Kulkarni, S., Wood, H., Lloyd, M., Hormiga, G. 2019. Spider-specific probe set for ultraconserved elements offers new perspectives on the evolutionary history of spiders (Arachnida, Araneae). *Molecular Ecology Resources*, 20(1): 185–203. doi: 10.1111/1755-0998.13099.

Kumar, S., Filipski, A.J., Battistuzzi, F.U., Kosakovsky Pond, S.L., Tamura, K. 2012. Statistics and truth in phylogenomics. *Molecular Biology and Evolution*, 29(2): 457–472. doi: 10.1093/molbev/msr202.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3): 772–773. doi: 10.1093/molbev/msw260.

Lartillot, N., Philippe, H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, 21(6): 1095–1109. doi: 10.1093/molbev/msh112.

Lemmon, A.R., Emme, S.A., Lemmon, E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, 61(5): 727–744. doi: 10.1093/sysbio/sys049.

Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., Walters, J.R. 2019. Insect genomes: progress and challenges. *Insect Molecular Biology*, 28(6): 739–758. doi: 10.1111/imb.12599.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., Wang, J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, 20(2): 265–272.

doi: 10.1101/gr.097261.109.

McKain, M.R., Johnson, M.G., Uribe-Convers, S., Eaton, D., Yang, Y. 2018. Practical considerations for plant phylogenomics. *Applications in Plant Sciences*, 6(3): e1038. doi: 10.1002/aps3.1038.

Miller, M.R., Dunham, J., Amores, A., Cresko, W.A., Johnson, E.A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17(10): 240–248. doi: 10.1101/gr.5681207.

Minh, B.Q., Hahn, M.W., Lanfear, R. 2018. New methods to calculate concordance factors for phylogenomic datasets. doi: 10.1101/487801.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. Warnow, T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, 30(6): 541–548. doi: 10.1093/bioinformatics/btu462.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong. T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang. Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210): 763–767. doi: 10.1126/science.1257570.

Misof, B., Misof, K. 2009. A monte carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology*, 58(1): 21–34. doi: 10.1093/sysbio/syp006.

Munro, J.B., Heraty, J.M., Burks, R.A., Hawks, D., Mottern, J., Cruaud, A., Rasplus, J-Y., Jansta, P. 2011. A molecular phylogeny of the chalcidoidea (Hymenoptera). *PLoS ONE*, 6(11): e27023–e27050. doi: 10.1371/journal.pone.0027023.

Nater, A., Burri, R., Kawakami, T., Smeds, L., Ellegren, H. 2015. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Systematic Biology*, 64(6): 1000–1017. doi: 10.1093/sysbio/syv045.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1): 268–274. doi: 10.1093/molbev/msu300.

Olofsson, J.K., Cantera, I., Van de Paer, C., Hong-Wa, C., Zedane, L., Dunning, L.T., Alberti, A., Christin, P.-A., Besnard, G. 2019. Phylogenomics using low-depth whole genome sequencing: a case study with the olive tribe. *Molecular Ecology Resources*, 19(4): 877–892. doi: 10.1111/1755-0998.13016.

Pease, J.B., Brown, J.W., Walker, J.F., Hinchliff, C.E., Smith, S.A. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany*, 105(3): 385–403. doi: 10.1002/ajb2.1016.

Peloso, P.L.V., Frost, D. R., Richards, S.J., Rodrigues, M.T., Donnellan, S., Matsui, M., Raxworthy, C.J., Biju, S.D., Lemmon, E.M., Lemmon, A.R., Wheeler, W.C. 2015. The impact of anchored phylogenomics and taxon sampling on phylogenetic inference in narrow-mouthed frogs (Anura, Microhylidae). *Cladistics*, 32(2): 113–140. doi: 10.1111/cla.12118.

Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R.S., Podsiadlowsk, L., Vasilikopoulos, A., Zhou, X., Misof, B., Niehuis, O. 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics*, 18(1): 111–110. doi: 10.1186/s12859-017-1529-8.

Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P.A., Heraty, J., Kjer, K.M., Klopfstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S., Zhou, X., Wappler, T., Rust, J., Misof, B., Niehuis, O. 2017. Evolutionary history of the Hymenoptera. *Current Biology*, 27(7): 1013–1018. doi: 10.1016/j.cub.2017.01.027.

Price, M.N., Dehal, P.S., Arkin, A.P. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLOS ONE*, 5(3): e9490–e9502. doi: 10.1371/journal.pone.0009490.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generationn, DNA sequencing. *Nature*, 526(7574): 569–573. doi: 10.1038/nature15697.

Posada, D. 2016. Phylogenomics for systematic biology. *Systematic Biology*, 65(3): 353–356. doi: 10.1093/sysbio/syw027.

Schwentner, M., Combosch, D.J., Nelson, J.P., Giribet, G. 2017. A phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships. *Current Biology*, 27(12): 1818–1824. doi: 10.1016/j.cub.2017.05.040.

Seppey, M., Manni, M., Zdobnov, E.M. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, 1962: 227–245. doi: 10.1007/978-1-4939-9173-0_14.

Shen, X.X., Hittinger, C.T., Rokas, A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology and Evolution*, 1(5): 0126–0136. doi: 10.1038/s41559-017-0126.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9): 1312–1313. doi: 10.1093/bioinformatics/btu033.

Stamatakis, A., Hoover, P., Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, 57(5): 758–771. doi: 10.1080/10635150802429642.

Stlaurent, R.A., Hamilton, C.A., Kawahara, A.Y. 2018. Museum specimens provide phylogenomic data to resolve relationships of sack-bearer moths (Lepidoptera, Mimallonoidea, Mimallonidae). *Systematic Entomology*, 43(4): 729–761. doi: 10.1111/syen.12301.

Sun, X., Ding, Y., Orr, M.C., Zhang, F. 2020. Streamlining universal single-copy orthologue and ultraconserved element design: a case study in Collembola. *Molecular Ecology Resources*, 00: 1–12. doi: 10.1111/1755-0998.13146.

Theodorou, P., Radzeviciute, R., Kahnt B., Soro, A., Grosse, I., Paxton, R.J. 2018. Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proceedings of the Royal Society B*, 285(1877): 20172806. doi: 10.1098/rspb.2017.2806.

Van Dam, M.H., Lam, A.W., Sagata, K., Gewa, B., Laufa, R., Balke, M., Faircloth, B.C., Riedel, A. 2017. Ultraconserved elements (UCEs) resolve the phylogeny of australasian smurf-weevils. *PLOS ONE*, 12(11): 1–21. doi: 10.1371/journal.pone.018804.

Wang, H.C., Edward, S., Roger, A.J. 2019. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. *Systematic Biology*, 68(6): 1003–1019. doi: 10.1093/sysbio/syz021.

Wang, H.C., Li, K., Susko, E., Roger, A.J. 2008. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC evolutionary biology*, 8(331): 1–13. doi: 10.1186/1471-2148-8-331.

Wang, H.C., Minh, B.Q., Susko, E., Roger, A.J. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Systematic Biology*, 67(2): 216–235. doi: 10.1093/sysbio/syx068.

Wang, Y., Liu, X., Garzón-Orduña, I.J., Winterton, S.L., Yan, Y., Aspöck, U., Aspöck, H., Yang, D. 2017. Mitochondrial phylogenomics illuminates the evolutionary history of neuropterida. *Cladistics*, 33(6): 617–636. doi: 10.1111/cla.12186.

Winterton, S.L., Lemmon, A.R., Gillung, J.P., Garzon, I.J., Badano, D., Bakkes, D.K., Breitkreuz, L.C.V., Engel, M.S., Lemmon, E.M., Liu, X., Machado, R.J.P., Skevington, J.H., Oswald, J.D. 2018. Evolution of lacewings and allied orders using anchored phylogenomics (Neuroptera, Megaloptera, Raphidioptera). *Systematic Entomology*, 43(2): 330–354. doi: 10.1111/syen.12278.

Yang, H., Li, T., Dang, K., Bu, W. 2018. Compositional and mutational rate heterogeneity in mitochondrial genomes and its effect on the phylogenetic inferences of Cimicomorpha (Hemiptera: Heteroptera). *BMC Genomics*, 19(1): 264–277. doi: 10.1186/s12864-018-4650-9.

Yang, Z. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology*, 61(5): 854–865. doi: 10.1093/czoolo/61.5.854.

Young, A.D., Gillung, J.P. 2020. Phylogenomics – principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45: 225–247. doi: 10.1111/syen.12406.

Young, A.D., Lemmon, A.R., Skevington, J.H., Mengual, X., Ståhls, G., Reemer, M., Jordaens, K., Kelso, S., Lemmon, E.M., Hauser, M., Meyer, M.D., Misof, B., Wiegmann, B.M. 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). *BMC evolutionary biology*, 16(1): 143–156. doi: 10.1186/s12862-016-0714-0.

Zerbino, D.R., Birney, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18: 821–829. doi: 10.1101/gr.074492.107.

Zhang, F., Ding, Y., Zhu, C.D., Zhou, X., Orr, M.C., Scheu, S., Luan, Y.X. 2019. Phylogenomics from low-coverage whole-genome sequencing. *Methods in Ecology and Evolution*, 10(4): 1–11. doi: 10.1111/2041-210X.13145.

Zhang, J., Gordon, E.R.L., Forthman, M., Hwang, W.S., Walden, K., Swanson, D.R., Johnson, K.P., Meier, R., Weirauch, C. 2016. Evolution of the assassin's arms: insights from a phylogeny of combined transcriptomic and ribosomal DNA data (Heteroptera: Reduvioidea). *Scientific Reports*, 6: 22177–22185. doi: 10.1038/srep22177.

Zhang, J., Lindsey, A.R.I., Peters, R.S., Heraty, J.M., Hopper, K.R., Werren, J.H., Martinson, E.O., Woolley, J.B., Yoder, M.J., Krogmann, L. 2020. Conflicting signal in transcriptomic markers leads to a poorly resolved backbone phylogeny of chalcidoid wasps. *Systematic Entomology*, 1–20. doi: 10.1111/syen.12427.

Zhang, J., Maddison, W.P. 2013. Molecular phylogeny, divergence times and biogeography of spiders of the subfamily Euophryinae (Araneae: Salticidae). *Molecular Phylogenetics and Evolution*, 68(1): 81–92. doi: 10.1016/j.ympev.2013.03.017.

Zhang, J., Maddison, W.P. 2015. Genera of euophryine jumping spiders (Araneae: Salticidae), with a combined molecular-morphological phylogeny. *Zootaxa*, 3938(1): 001–147. doi: 10.11646/zootaxa.3938.1.1.