**International Journal of Zoology Studies**
**www.zoologyjournals.com**
**ISSN: 2455-7269**
**Received: 21-10-2023, Accepted: 07-11-2023, Published: 23-11-2023**
**Volume 8, Issue 4, 2023, Page No. 36-42**

# A review on challenges and advances in DNA barcoding of insect phylogenomics

**Meena Poonja**
Assistant Professor, Department of Zoology, Smt. Chandibai Himathmal Mansukhani College, Ulhasnagar, University of Mumbai, Maharashtra, India

## Abstract

**Background:** DNA barcoding is a novel and prevalent approach to molecular categorization and identification of insect species using short genomic sequences. It supplementally solves current difficulties of classical taxonomy and phylogenetics allowing quick identification, effective taxonomic discrimination, and validated categorization. This groundbreaking technology is being applied to insect phylogenetics in a wide range of studies worldwide.

**Objective:** The current literature critically reviews insect barcoding, its challenges, recent advancements, the role of bioinformatics, and novel computational methodologies and software tools involved.

**Methods:** A thorough search for manuscripts was conducted using a variety of platforms such as Google Scholar, PubMed, ResearchGate, Science Direct, NCBI, and Springer Link.

**Results and Discussion:** Although DNA barcoding in phylogenetics is quickly gaining eminence, it poses certain challenges when applied to insect phylogeny because of the sluggish rate of genetic evolution in insect mitochondrial genomes. In closely related insect species, there is a minimal variation between COI. Insect hybridization and polyploidy are projected to have an impact on insect species identification by DNA barcoding. Insect materials rich in polysaccharides, polyphenols, and other secondary metabolites increase DNA destruction, as shown in museum specimens. Amplification and sequencing of DNA might be difficult with such degraded samples.

**Conclusion:** In a large-scale effort to address these difficulties with conventional insect barcoding, advances in possible markers, sequencing, and computational technologies will reshape DNA barcoding in the future, making it an extensively utilized and useful tool.

**Keywords:** Insects, DNA barcoding, Super barcoding, Ultra-barcoding, ME Barcoding, Next-generation sequencing, Bioinformatics

## Introduction

Because species share a common history through their ancestry, reconstructing evolutionary history provides information about the systematic relationships among species. Traditionally, phylogenetic reconstruction has used morphological or ultrastructural characters. The major animal and plant groups on Earth have been delineated based on the comparative anatomy of fossils and living animals. However, a major drawback of this approach is the number of reliable morphological features as homologous features between species. With the advent of DNA sequencing in the early 1970s, the use of molecular data in the reconstruction of phylogenies has gained enormous popularity. The small subunit gene of ribosomal RNA has been used as a reference for constructing phylogenetic trees because of considerable sequence conservation between species (Olsen and Woese 1993). Subsequently, extensive studies have been conducted to combine multiple genes and/or morphology data to infer phylogenetic relationships among species. Since the last decade, the genomic revolution has heralded the application of large-scale multigene data in phylogenetics. The use of genome sequences to infer species phylogenies is the central theme of phylogenomics. It is the intersection of the fields of evolution and genomics to reconstruct the evolutionary history of species.

This is an emerging field that integrates extensive sequence data, computational tools, and evolutionary principles to study phylogenetics. Phylogenomics differs from phylogenetics primarily in the input data. Unlike phylogenetics, phylogenomic studies use large-scale genomic or transcriptomic data to generate input data based on orthology predictions of genes or whole-genome alignments.

For many insect strains, resolving the true phylogeny has been challenging. One example is the phylum Polyneoptera. The branching patterns of Polyneoptera insects such as cockroaches, mantids, earwigs, grasshoppers, and phasmids are highly ambiguous (Wipfler *et al*. 2014) [45]. The Polyneoptera consist of 11 orders: Blattodea, Dermaptera, Embiodea, Grylloblattodea, Isoptera, Mantodea, Manto Phasmatodea, Orthoptera, Phasmatodea, Plecoptera, and Zoraptera. It has been difficult to determine the relationships among these orders (Kristensen 1991) [21]. With the help of phylogenomic approaches, new insights into the internal relationships of these insects are gradually emerging (Wipfler *et al*. 2014) [45].

The application of phylogenomics to insects is also useful because insects are very diverse. With approximately one million identified species, insects comprise the largest proportion (~75%) of all animals (May 1988, Foottit and Adler 2009). Thus, phylogenomic analyses are important in entomology because they can help us better understand insect evolutionary relationships relevant to health, environment, and agriculture. In addition, sequencing of the genomes of numerous insect species has begun in recent years, and hundreds of insect genomes have been completed. This opens new opportunities for identifying orthologs for large-scale phylogenetic studies involving multiple genes. With the increasing trends in insect genomics and transcriptomic studies, phylogenomic studies are becoming more popular with the potential to provide new insights into the true phylogeny of many insect strains.

**Phylogenomic Methods**

Phylogenomic analyses are performed using three main methods: 1) sequence alignment, 2) comparison of the occurrence of 'DNA strings', and 3) comparison of gene content or sequence in the genome. The sequence alignment method is the most commonly used approach in phylogenomic studies. This approach relies on the accurate prediction of orthologous single copies representing genes in different species that evolved from a common ancestor gene by speciation. Multiple alignments of orthologous genes are performed using either single orthologs or their concatenated sequences. Phylogenomic inferences are made by constructing a supertree from the individual genes (Bininda-Emonds and Sanderson 2001) [2] or from supermatrix data generated by aligning concatenated sequences (Delsuc *et al*. 2005, Savard *et al*. 2006) [39]. The supertree-based method relies on genes from overlapping taxa, whereas the supermatrix method treats genes from non-overlapping taxa as missing data. The independent evolutionary rates of multigene sequences are accounted for by methods such as partitioned likelihoods to infer phylogenetic relationships among species. In addition to the sequence alignment-based method, DNA strings (distribution patterns of short sequences between genomes) and random anchors (a longer stretch of nucleotides) are also used in phylogenomic studies, especially in closely related species (Vishnoi *et al*. 2010) [43].

The DNA string method does not require sequence alignment and has an advantage over the sequence alignment method. It is based on the abundance of short oil oligonucleotide sequence combinations in genomes used to directly compare higher-order features of non-homologous DNA sequences (Edwards *et al*. 2002). DNA stranding is also a suitable method for making phylogenetic inferences across deeper taxonomic levels where genomic homology is difficult to detect (Qi *et al*. 2004). Gene content and gene order information are also useful for phylogenetic inference (Murphy *et al*. 2004, Bourque and Tesler 2008) [28]. Gene order information is determined by assessing the presence or absence of pairs of orthologous genes, which are then used as a measure of genetic distance between species (Wolf *et al*. 2001, Zhang *et al*. 2009) [46]. Gene rearrangements and duplications are critical factors affecting the accuracy of phylogenies based on gene orders. Recently, Hu *et al*. (2014) developed a maximum likelihood approach that considers gene rearrangements, insertions, deletions, and duplications for phylogeny reconstruction from gene order data. Phylogenomic studies using DNA strings or gene sequence methods are rare compared to those using sequence alignment methods. Although most phylogenomic studies of insects have been conducted using sequence alignment methods, the derivation of phylogenies based on the rearrangement of genomic regions, also known as 'chromosomal phylogeny,' has provided useful information on the evolution of certain insect lineages such as the species complex of *Anopheles gambiae* (Kamali *et al*. 2012, Sharakhov *et al*. 2013).

**Current Status of Insect Phylogenomics**

In recent years, phylogenomics studies have provided valuable insights into the evolutionary relationships of insects. Numerous phylogenomic studies have been performed in diverse lineages of insects that have addressed several fundamental issues of the classification and evolution of insect species. A non-comprehensive list of recent phylogenomic studies of insects is provided in Table 1.

**Table 1**

**Table 1: Recent developments in insect phylogenomics.**

| Study objective | No. of Species/taxa | No. of genes | Reference |
|---|---|---|---|
| Pattern and time of insect evolution | 144 | 1478 | Misof *et al*. 2014 |
| Phylogeny of aculeate Hymenoptera | 18 | 308 | Johnson *et al*. 2013 |
| Phylogeny of malaria mosquitoes | 6 | 49 | Kamali *et al*. 2014 [18] |
| Phylogeny of malaria mosquitoes | 43 | 1085 | Neafsey *et al*. 2015 |
| Phylogeny of malaria mosquitoes | 8 | whole-genome alignment | Fontaine *et al*. 2015 |
| Phylogeny of lower neopteran orders | 48 | 229 | Letsch and Simon 2013 |
| Phylogeny of holometabolan insects | 13 | 1,343 | Peters *et al*. 2014 |
| Phylogeny of Lepidoptera | 46 | 2,696 | Kawahara and Breinholt 2014 |
| Phylogeny of wingless insects | 73 | 1,866 | Dell'Ampio *et al*. 2014 |
| Phylogeny of Neuropteroidea | 36 | 668 | Boussau *et al*. 2014 |

Along with the study objectives. The expressed sequence tags (ESTs) (Behura 2006) and whole transcriptome data (Trautwein *et al*. 2012) [42] are predominantly used to develop data for phylogenomic investigations of many insects.

Although the application of EST datasets has been useful for phylogenetic analysis (Theodorides *et al*. 2002, Hughes *et al*. 2006, Parkinson and Blaxter 2009, Meusemann *et al*. 2010) [16, 33, 26], they may often bias the results. This is because the representation of these data to the coding sequences may vary between species. It has been shown that EST-based phylogenies may also lead to inconsistent results when compared with the known speciation events and life history of the organisms (Andrew 2011) [1].

Currently, multigene datasets and whole-transcriptome data are extensively used in phylogenomic analyses (Chan and Ragan 2013) [7]. In a recent publication, Misof *et al* (2014) conducted a large-scale phylogenomic analysis of 144 insect species that provided a holistic view of the origin and evolution of insects. Using 1478 orthologous genes, this study suggested that insects originated approximately 479 million years ago. By partitioning transcriptome sequence data into protein domains, Misof *et al.* (2014) used maximum likelihood models to show the diversification of insects into four groups: Palaeoptera, Polyneoptera, Condylognatha, and Holometabola. The study further showed that diversification within modern winged insects started in the Paleozoic era, and suggested that insect flight occurred approximately 406 million years ago. This study has clarified our understanding of insect origin and evolution and arguably has laid the foundation for building the insect tree of life in finer detail (Jones 2015) [17].

Phylogenomics has also helped resolve species relationships of specific lineages of insects. Species phylogenies of mosquitoes are one of the ongoing goals of vector biologists. Over 3,500 species, belonging to at least 43 genera, of the Anophelinae and Culicinae mosquitoes are known. Several species of these mosquitoes act as global vectors of malaria, dengue, West Nile virus, and others. In addition to a rich literature on classical and molecular systematics of mosquitoes (Munstermann and Conn 1997) [27], evolutionary investigations of *Anopheles gambiae* (malaria vector), *Aedes aegypti* (dengue vector), and *Culex quinquefasciatus* (lymphatic filariasis and West Nile virus vector) have gained a lot of momentum in the recent years after the availability of genome sequences of these species (Severson and Behura 2012, Neafsey *et al*. 2013, Kamali *et al*. 2014, Fontaine *et al*. 2015) [29, 18].

Using a combined dataset of six nuclear protein-coding genes and an array of morphological characters (n =80), Reidenbach *et al* (2009) [35] performed maximum parsimony and maximum likelihood analyses to understand phylogenetic radiation among mosquitoes representing 25 genera. Their analysis provided renewed and stronger evidence for the basal position of the Anophelinae subfamily. It was further suggested from this study that divergence times for major culicid lineages might date back to the early Cretaceous. The recent study by Kamali *et al* (2014) [18] further shows that *Anopheles nili* occupies the basal clade that diversified from other studied malaria mosquito species some 47.6 million years ago. Recently, 16 Anopheles species have been sequenced (Neafsey *et al* 2015), and this has set the stage for phylogenomic investigation of malaria vector mosquitoes.

Fontaine *et al* (2015) performed a phylogenomic analysis among eight Anopheles species and that lineages leading to the principal vectors of human malaria were among the first to split within the species complex. Their data further revealed extensive introgression in autosomes that may have important implications in the vectorial capacity of Anopheles to malaria transmission. Recent studies also demonstrate the confounding effects of introgression and shared mutations on applying genomic data in inferring true speciation histories of *Anopheles gambiae* species complex (O'Loughlin *et al*. 2014, Crawford *et al*. 2014) [9]. Introgression, an important source of genetic variation in natural populations, is caused by the stable integration of genetic material from one species into another through repeated back-crossing. Introgression is particularly a problem in phylogeny reconstruction of closely related lineages partly because commonly sequenced genetic markers often lack sufficient phylogenetic signal at the lowest taxonomic levels. When introgression occurs, a substantial fraction of their genomes can be permeable to alleles from related species. Discordant genealogies of closely related species can also occur by incomplete lineage sorting where two lineages fail to coalesce within a population. Incomplete lineage sorting generally arises from stochastic coalescence that leads to mask the signatures of true phylogeny. Distinguishing introgression from incomplete lineage sorting is important in evolutionary studies of closely related species.

Topology-based phylogeny discordance, a test of isolation model (speciation with no gene flow), or detection of gene flow are commonly used methods to distinguish incomplete lineage sorting from genetic introgression. Phylogenomics of Polyneoptera has been gaining a lot of interest in recent times. Employing large-scale EST datasets, Letsch *et al* (2012) performed phylogenomics analysis among Polyneopteran and Paraneopteran insects and showed that Polyneoptera and Eumetabola (Paraneoptera + Holometabola) have a monophyletic origin. In parallel to this study, Simon *et al* (2012) performed further phylogenomic analyses among three polyneopteran orders; Dermaptera, Plecoptera, and Zoraptera, the results of which provided conclusive support for monophyletic Polyneoptera. However, later results of Letsch and Simon (2013) rejected the previous classification of Parametabola (= Zoraptera + Paraneoptera), Mystroptera (= Embioptera + Zoraptera) and Orthopterida (= Orthoptera + Phasmatodea) and indicated that several polyneopteran orders still show unstable positions within a monophyletic Polyneoptera.

Phylogenomics investigations have aided the resolution of holometabolous insects. Savard *et al.* (2006) [39] conducted a study using 185 orthologous genes among six species (*Drosophila melanogaster*, *Anopheles gambiae*, *Bombyx mori*, *Tribolium castaneum*, *Apis mellifera*, *Nasonia vitripennis,* and *Nasonia giraulti*) representing four insect orders (Diptera, lepidoptera, coleopteran and hymenoptera). The results of this phylogenomics study suggested that bees and wasps retain the basal position of holometabolous insects' phylogeny.

Later, the maximum-likelihood tree generated from concatenated sequences of 1,150 single-copy orthologs among 10 metazoan species (Richards *et al*. 2008) also confirmed the results of Savard *et al*. (2006) [39] a recent phylogenomics study (Peters *et al*.2014) exploited *de novo* transcriptome data to study holometabolous insects. By analyzing 1,343 single-copy orthologous genes among 13 species, this study suggested that hymenopteran insects belong to a sister group of other holometabolan insects that comprises Mecopterida and Neuropteroidea. The results of this study strongly supported the relations of Raphidioptera + (Neuroptera + monophyletic Megaloptera), and Diptera + (Siphonaptera+ Mecoptera) within Neuropterida and Antliophora.

The order Lepidoptera has shown several ambiguities in the phylogenetic placement of specific species such as moths and butterflies. Kawahara and Breinholt (2014) performed a phylogenomics study based on 2,696 single-copy ortholog genes among 46 taxa that provided strong evidence that butterflies and moths have shared evolutionary relationships. The study revealed the monophyly of butterflies with Hesperiidae (skippers) and Hedylidae (moth-butterflies) and provided support for placing butterflies sister to the obtectomeran Lepidoptera. Similarly, the aculeate Hymenoptera are extensively investigated for the eusocial behavior of many species such as ants, bees, and stinging insects which have unclear phylogenies. Johnson *et al*. (2013) performed phylogenomic analysis among 18 species by exploiting gene portioning from transcriptome data. They found that the eusocial insects are contained within two major groups which are interpolated among three other clades of wasps. It also showed that ants are the sister group of spheciform wasps and bees (Apoidea).

**Major Challenges**

Although significant progress has been made in the field of phylogenomics, several factors have been identified that impose challenges in these studies. Here, I discuss three broad areas that are often problematic in phylogenomic studies: A) accurate prediction of gene orthology, B) gene tree heterogeneity, and C) assumption and statistics. A prerequisite step in preparing phylogenomic data generally involves the identification of 1:1 orthologous genes among the species. This has been a challenging step particularly when genes are present in duplicated copies within the genome (Jensen 2001, Dalquen *et al*. 2013) [10], and also for the reason that building orthology models relies upon all-against-all gene comparisons (Sonnhammer *et al.* 2014) [41]. Similarly, the presence of horizontally transferred genes further confounds the problem of identification of orthologs (Dalquen *et al*. 2013) [10]. Furthermore, the number of single-copy ortholog groups sharply decreases as we consider more diverse species (Figure 1).

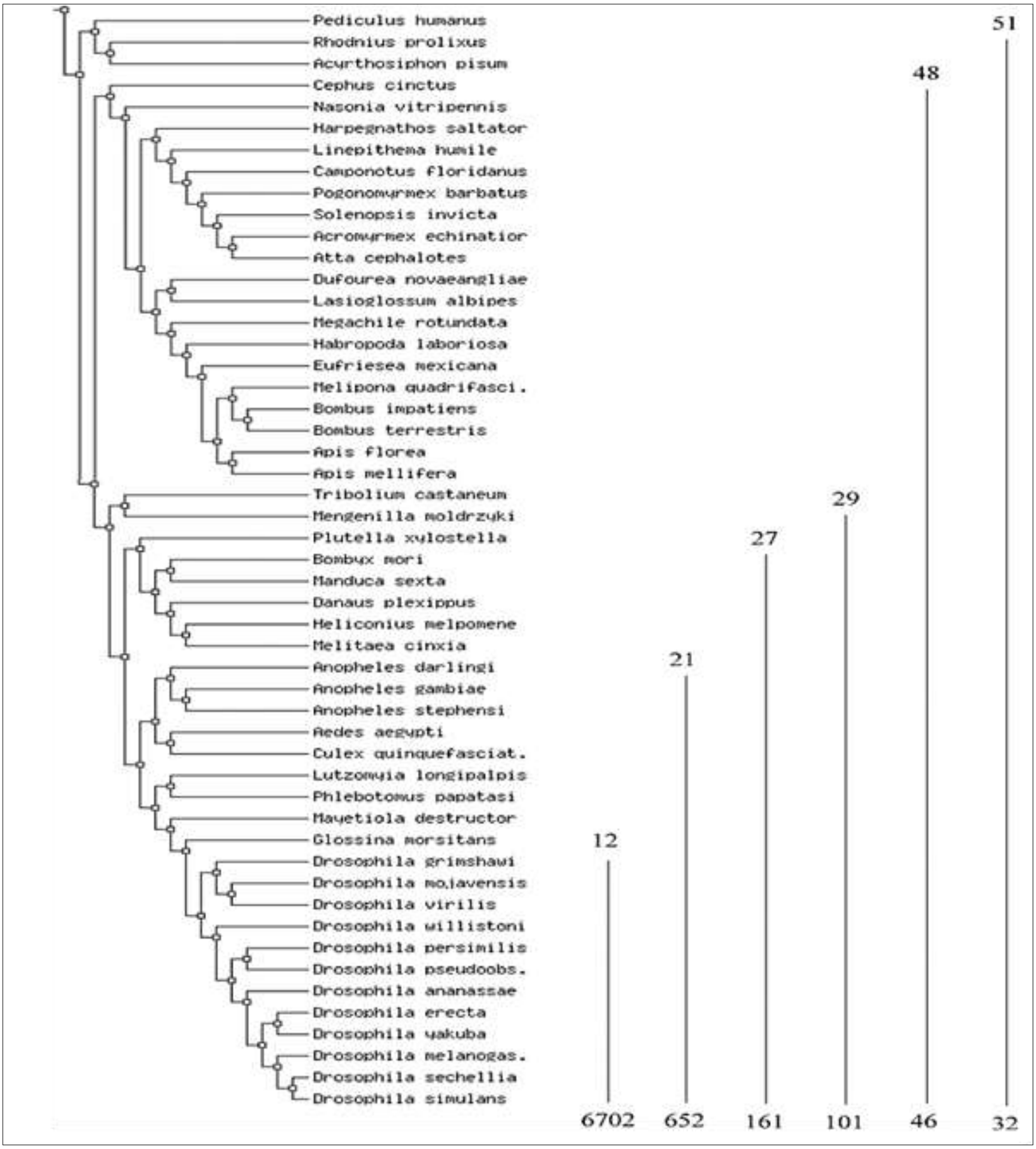

**Fig 1:** Number of single-copy orthologs among different insect species. The screenshot image shows a phylogenetic grouping of different insect species where genome sequences are available (OrthoDB7, http://cegg.unige.ch/orthodb7). The vertical lines on the right represent different groups with several species on the top and the number of single-copy orthologous genes on the bottom

This is based on known orthologues predicted among different insect species where genome sequences are available (Waterhouse *et al.* 2013) [41]. Thus, the number of single-copy orthologous genes that can be used in the phylogenomic analysis becomes a limiting factor in carrying out phylogenomic investigations among distantly related species. Besides single-copy orthologs, phylogenomics studies are also carried out by partitioning the data into predicted protein domains. This approach may also have a limitation in the accurate partitioning of sequence data into orthologous domains when the number of species compared is large and the size of protein domains is small (Storm and Sonnhammer 2003). Furthermore, differential evolution and architecture of protein domains (Buljan and Bateman 2009) [6] may impact the identification of orthologous protein domains.

Accounting for heterogeneity among gene trees is a major challenge in phylogenomics investigations. Discrepancies between gene trees and species trees are well known which imposes uncertainty in predicting species phylogenies based on multigene datasets (Maddison 1997). Several factors such as recombination, hybridization, and introgression (Siepel 2009), gene duplication (Page and Charleston 1997) [30], horizontal gene transfer (Doolittle 1999), and incomplete lineage sorting (Pamilo and Nei 1988) [32] may contribute to heterogeneity among gene trees. Gene tree heterogeneity is a problem in phylogenomics because genealogical histories often vary among different genes throughout the genome (Degnan and Rosenberg 2009). The discrepancy in the inferred gene trees and the actual species tree stems from different coalescent events among genes due to differential lineage sorting, selection, and drift of ancestral polymorphisms in the orthologous genes (Rannala and Yang 2008) [34]. Incongruence, the topological conflict between different gene trees, is often a critical issue of phylogenomics studies. Several recent studies have addressed the issue of incongruence and suggested approaches to overcome it (Salichos and Rokas 2013, Salichos *et al*. 2014, Dell' Ampio *et al.* 2014) [36].

Long-branch attraction (where species with high evolutionary rates tend to group) and heterarchy (site-specific variation of evolutionary rate) can also be potential causes of inaccurate phylogenies (Zhang *et al*. 2009). In a recent study, Boussau *et al*. (2014) [4] showed that correcting long-branch attraction by usage of appropriate models accurately places Strepsiptera as a sister group to Coleoptera. Although the long-branch attraction problem was initially thought to be associated with parsimony methods (Felsenstein 1978), a study (Huelsenbeck and Lander 2003) has suggested that this problem is frequent, and may arise with other methods as well (Swofford *et al.* 2001). Heterotachy is a potential problem in phylogenomics when an appropriate model is not implemented to account for site-specific changes in evolutionary rates across the tree. The homogamous techniques may produce inaccurate phylogenies as these models assume that relatively fast-evolving sites are fast across the entire tree, whereas slower sites always evolve at relatively slower rates.

Kolaczkowski and Thornton (2008) [20] used simulation to show that the mixed branch length model is more accurate than homotachous techniques. In addition to selection of appropriate evolutionary models, insufficient sequence data (Philippe *et al*. 2004) and sampling bias (Driskell *et al.* 2004) [11] may also lead to conflicting phylogenetic inferences. By analyzing ~300,000 protein sequences across eukaryotes and prokaryotes, Driskell *et al*. (2004) [11] showed that the ‘supermatrix’ approach may provide useful insights into broad sections of the Tree of Life even with incomplete datasets.

The third but serious concern in phylogenomics study stems from statistical analyses. The inaccurate results primarily originate from unrealistic assumptions and statistical methods of multigene phylogenies. For example, when genes are acted upon by episodic positive selection in specific lineages, the statistical inference of evolutionary relationships can be misleading as the amount of sequence data required to accommodate the evolutionary model is invariably underestimated (Kumar *et al*. 2012) [22]. Similarly, bias in data matrix composition is known to mislead phylogenomic inferences. The phylogenomic investigation by Letsch *et al* (2012) [25], using expressed sequence tags of different Polyneoptera and Paraneoptera species, clearly demonstrated the effect of matrix composition on the phylogenetic placement of *Pediculus humanus*, the human body louse. It was observed that biased gene overlap resulting from orthology predictions from transcript data was the reason for the inaccurate placement of the human louse and Holometabola. This study suggested that when transcript sequences used in phylogenomic analyses are not comprehensive for each genome, the results may have potential pitfalls.

## Prospects and Concluding Remarks

Despite the inherent challenges, phylogenomic approaches have provided valuable information about systematics and the evolutionary history of insects. Within the last two years, several phylogenomic investigations have provided new insights into insect evolution of different lineages. The evolution of eusociality (Johnson *et al*. 2013) and the vectorial ability of malaria mosquitoes (Fontaine *et al.* 2015) are recent examples of this progress.

In recent years, phylogenomics tools have been applied in diverse domains of molecular evolutionary studies ranging from predicting gene function, studying evolutionary patterns of macromolecules and molecular adaptation, and resolving relationships and divergence times of genes and species (Kumar *et al*. 2012) [22]. The emerging concept of ‘integrative phylogenomics’ is now gaining popularity in diverse domains of evolutionary biology (Bapteste and Burian 2010). The integration of anatomical data generated by the non-invasive three-dimensional reconstruction techniques along with fossil and genomic data (Giribet and Edgecombe 2012) [14] is a clear example of that concept. Combining genomic data from protein-coding and non-coding genes is another way of integrating data for phylogenomics analyses (Rota-Stabelli *et al*. 2010). One example would be the application of microRNA gene sequences in phylogenomics studies (Rota-Stabelli *et al.* 2010, Campbell *et al*. 2011). MicroRNA genes are appropriate for phylogenomics investigations because 1) their stem-loop precursors are associated with differential nucleotide diversity among different species groups (Behura 2007), 2) they mostly lack convergent evolution (Sperling and Peterson 2009), 3) their birth rate is higher than the death rate in most metazoan taxa (microRNA families are continually added to metazoan genomes), and 4) generating genome-wide microRNA sequences is becoming very routine these days via next-generation sequencing.

The application of phylogenomic tools to predict gene functions (Eisen 1998, Sjölander 2004, Brown and Sjölander 2006) is an important area of research that may find utility in future phylogenomic studies of insects. The phylogenomic approaches to studying the functional adaption of a large number of genes may be beneficial compared to the phylogenetic approaches that rely on the correlation between a single gene tree and the associated trait (gene function) (Pagel 1999). Besides, phylogenomic tools are also applicable to studying the evolutionary histories of endosymbionts (Comas *et al.* 2007, Kembel *et al.* 2011). They may also find utility in cell biology to define the evolutionary origins of cell lineages. This is a particular area that holds exceptional promise in cell fate mapping projects of various organisms. While phylogenetics approaches have been successfully applied in mapping cell fates (Salipante and Horwitz 2007), recent advances in single-cell sequencing (Shapiro *et al*. 2013, Liang *et al*. 2014) are expected to aid phylogenomic approaches to reconstruct cell lineages. These specific examples not only indicate the broad utilities of phylogenomic tools but also attest that they can be harnessed in a deeper understanding of biology and evolution holistically. In particular, the explosion of genome sequencing efforts of 5,000 insects and other arthropods (Robinson *et al*. 2011, i5K consortium 2013) is expected to aid the integration of phylogenomics tools to provide new insights into systematics and evolutionary relationships of several unresolved or poorly resolved insect lineages. Without a doubt, insect phylogenomics holds huge promise in understanding insect biology in evolutionary terms; and sooner or later, they may unravel some of the many secrets of why insects are so diverse and adaptive in nature.

## References

1. Andrew DR. A new view of insect-crustacean relationships II. Inferences from expressed sequence tags and comparisons with neural cladistics. Arthropod Struct Dev,2011:40:289–302. [PubMed: 21315832]
2. Bininda-Emonds OR, Sanderson MJ. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. Syst Biol,2001:50:565-579. [PubMed: 12116654]
3. Blair C, Murphy RW. Recent trends in molecular phylogenetic analysis: where to next? J Hered,2011:102:130-138. [PubMed: 20696667]
4. Boussau B, Walton Z, Delgado JA, Collantes F, Beani L, Stewart IJ, *et al*. Strepsiptera, phylogenomics, and the long branch attraction problem. PLoS One,2014:9: e107709. [PubMed: 25272037]
5. Brown D, Sjölander K. Functional classification using phylogenomic inference. PLoS Comput Biol,2006:2: e77. [PubMed: 16846248]
6. Buljan M, Bateman A. The evolution of protein domain families. Biochem Soc Trans,2009:37:751– 755. [PubMed: 19614588]
7. Chan CX, Ragan MA. Next-generation phylogenomics. Biol Direct, 2013, 8:3. [PubMed: 23339707]
8. Comas I, Moya A, González-Candelas F. From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic gamma-Proteobacteria as a test case. Syst Biol,2007:56:1-16. [PubMed: 17366133]
9. Crawford J, Riehle MM, Guelbeogo WM, Gneme A, Sagnon N, Vernick KD, *et al*. Reticulate speciation, and adaptive introgression in the Anopheles gambiae species complex. Bio Rxiv. 2014, 009837.
10. Dalquen DA, Altenhoff AM, Gonnet GH, Dessimoz C. The impact of gene duplication, insertion, deletion, lateral gene transfer, and sequencing error on orthology inference: a simulation study. PLoS One,2013:8: e56925. [PubMed: 23451112]
11. Driskell AC, Ane C, Burleigh JG, McMahon MM, O'Meara BC, Sanderson MJ. Prospects for building the Tree of Life from large sequence databases. Science,2004:306:1172–1174. [PubMed: 15539599]
12. Eisen JA. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res,1998:8:163-167. [PubMed: 9521918]
13. Friend WG, Smith JJ. Factors affecting feeding by bloodsucking insects. Annu Rev Entomol,1977:22:309-331. [PubMed: 319741]
14. Giribet G, Edgecombe GD. Reevaluating the arthropod tree of life. Annu Rev Entomol.2012:57:167–186. [PubMed: 21910637]
15. Hu F, Lin Y, Tang J. MLGO: phylogeny reconstruction and ancestral inference from gene-order data. BMC Bioinformatics,2014:15:354. [PubMed: 25376663]
16. Hughes J, Longhorn SJ, Papadopoulou A, Theodorides K, de Riva A, Mejia-Chang M, *et al*. Dense taxonomic EST sampling and its applications for molecular systematics of the Coleoptera (beetles). Mol Biol Evol,2006:23:268–278. [PubMed: 16237206]
17. Jones B. Building the insect tree-of-life. Nat Rev Genet,2015:16:2-3. [PubMed: 25404110]
18. Kamali M, Marek PE, Peery A, Antonio-Nkondjio C, Ndo C, Tu Z, *et al*. Multigene phylogenetics reveals temporal diversification of major African malaria vectors. PLoS One,2014:9:e93580. [PubMed: 24705448]
19. Kembel SW, Eisen JA, Pollard KS, Green JL. The Phylogenetic diversity of metagenomes. PLoS One,2011:6: e23214. [PubMed: 21912589]
20. Kolaczkowski B, Thornton JW. A mixed branch length model of heterotactic improves phylogenetic accuracy. Mol Biol Evol,2008:25:1054–1066. [PubMed: 18319244]
21. Kristensen NP. Phylogeny of extant hexapods. In: Naumann, ID.; Carne, PB.; Lawrence, JF., *et al*., editors. The Insects of Australia. A textbook for students and research workers. 2nd ed. Vol. I. Melbourne University Press; Carlton, 1991, 125-140.
22. Kumar S, Filipski AJ, Battistuzzi FU, Kosakovsky Pond SL, Tamura K. Statistics, and truth in phylogenomics. Mol Biol Evol,2012:29:457–472. [PubMed: 21873298]
23. Legg DA, Sutton MD, Edgecombe GD. Arthropod fossil data increase the congruence of morphological and molecular phylogenies. Nat Commun,2013:4:2485. [PubMed: 24077329]
24. Letsch H, Simon S. Insect phylogenomics: new insights on the relationships of lower neopteran orders (Polyneoptera). Sys Entomol,2013:38:783–793.
25. Letsch HO, Meusemann K, Wipfler B, Schütte K, Beutel R, Misof B. Insect phylogenomics: results, problems and the impact of matrix composition. Proc Biol Sci,2012:279:3282-3290. [PubMed: 22628473]
26. Meusemann K, von Reumont BM, Simon S, Roeding F, Strauss S, Kück P, *et al.* A phylogenomic approach to resolve the arthropod tree of life. Mol Biol Evol,2010:27:2451-2464. [PubMed: 20534705]

27. Munstermann LE, Conn JE. Systematics of mosquito disease vectors (Diptera, Culicidae): impact of molecular biology and cladistic analysis. Annu Rev Entomol,1997:42:351-369. [PubMed: 9017898]
28. Murphy WJ, Pevzner PA, O'Brien SJ. Mammalian phylogenomics comes of age. Trends Genet,2004:20:631-639. [PubMed: 15522459]
29. Neafsey DE, Christophides GK, Collins FH, Emrich SJ, Fontaine MC, Gelbart W, *et al*. The evolution of the Anopheles 16 genomes project. G3,2013:3:1191-1194. [PubMed: 23708298]
30. Page RDM, Charleston MA. From gene to organismal phylogeny: reconciled trees and the gene tree species tree problem. Mol Phylogenet Evol,1997:7:231-240. [PubMed: 9126565]
31. Pagel M. Inferring the historical patterns of biological evolution. Nature,1999:401:877-884. [PubMed: 10553904]
32. Pamilo P, Nei M. Relationships between gene trees and species trees. Mol Biol Evol,1988:5:568-583. [PubMed: 3193878]
33. Parkinson J, Blaxter M. Expressed sequence tags: an overview. Methods Mol Biol,2009:533:1-12. [PubMed: 19277571]
34. Rannala B, Yang Z. Phylogenetic inference using whole genomes. Annu Rev Genomics Hum Genet,2008:9:217-231. [PubMed: 18767964]
35. Reidenbach KR, Cook S, Bertone MA, Harbach RE, Wiegmann BM, Besansky NJ. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol Biol,2009:9:298. [PubMed: 20028549]
36. Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature,2013:497:327-331. [PubMed: 23657258]
37. Salichos L, Stamatakis A, Rokas A. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol Biol Evol,2014:31:1261–1271. [PubMed: 24509691]
38. Salipante SJ, Horwitz MS. A phylogenetic approach to mapping cell fate. Curr Top Dev Biol,2007:79:157–184. [PubMed: 17498550]
39. Savard J, Tautz D, Richards S, Weinstock GM, Gibbs RA, Werren JH, *et al*. Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. Genome Res,2006:16:1334-1338. [PubMed: 17065606]
40. Sjölander K. Phylogenomic inference of protein molecular function: advances and challenges. Bioinformatics,2004:20:170-179. [PubMed: 14734307]
41. Sonnhammer EL, Gabaldón T, Sousa da Silva AW, Martin M, Robinson-Rechavi M, Boeckmann B, *et al*. Quest for Orthologs consortium. Big data and other challenges in the quest for orthologs. Bioinformatics,2014:30:2993-2998. [PubMed: 25064571]
42. Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK. Advances in insect phylogeny at the dawn of the postgenomic era. Annu Rev Entomol,2012:57:449-468. [PubMed: 22149269]
43. Vishnoi A, Roy R, Prasad HK, Bhattacharya A. Anchor-based whole genome phylogeny (ABWGP): a tool for inferring evolutionary relationship among closely related microorganisms. PLoS One,2010:5:e14159. [PubMed: 21152403]
44. Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. Ortho DB: a hierarchical catalog of animal, fungal, and bacterial orthologs. Nucleic Acids Res, 2013, 41. (Database issue): D358– D365. [PubMed: 23180791]
45. Wipfler B, Bai M, Schoville S, Dallai R, Uchifune T, Machida R, *et al*. Ice crawlers (Grylloblattodea) are the history of the investigation of a highly unusual group of insects. J Insect Biodiversity,2014:2:1-25.
46. Wolf YI, Rogozin IB, Grishin NV, Tatusov RL, Koonin EV. Genome trees constructed using five different approaches suggest new major bacterial clades. BMC Evol Biol, 2001, 1:8. [PubMed: 11734060].