Published in final edited form as:

Insect Mol Biol. 2015 August ; 24(4): 403–411. doi:10.1111/imb.12174.

Insect Phylogenomics

Susanta K. Behura

Eck Institute for Global Health and Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA

Abstract

With the advent of next-generation sequencing methods, phylogenetics has taken a new turn in the recent years. Phylogenomics, the integration of phylogenetics with genome data, has emerged as a powerful approach to study systematics and evolution of species. Recently, breakthrough researches employing phylogenomic tools have provided better insights into the timing and pattern of insect evolution. The next-generation sequencing methods are now increasingly used by entomologists to generate genomic and transcript sequences of various insect species and strains. These data provide opportunities for comparative genomics and large-scale multigene phylogenies of diverse lineages of insects. Phylogenomic investigations help us better understand systematic and evolutionary relationships of insect species that play important roles as herbivores, predators, detritivores, pollinators, or disease vectors. It is important that we critically assess the prospects and limitations of phylogenomic methods. In this review, I describe the current status, outline the major challenges, and remark on potential future applications of phylogenomic tools in studying insect systematics and evolution.

Keywords

Phylogeny; Insects; Phylogenetics; Next-generation sequencing; Phylogenomics

Introduction

As species share a common history through their ancestry, reconstructing evolutionary history provides information about systematic relationships among species. Traditionally, morphological or ultrastructural characters were exploited in phylogenetic reconstruction. The major groups of animals and plants on earth have been delineated by using comparative anatomy of fossils and extant. However, the number of reliable morphological features as homolog characters among species remains a major disadvantage of this approach. With the introduction of DNA sequencing in the early 1970s, the application of molecular data in reconstructing phylogenies has gained huge popularity. The small subunit ribosomal RNA gene is used as reference for constructing phylogenetic trees due to considerable sequence conservation among 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 (Legg 2013). Since the last decade or so, genomic revolution

has ushered in application of large scale multigene data in phylogenetics. The use of genome sequences to infer species phylogenies constitutes the central theme of phylogenomics. It is the intersection of the fields of evolution and genomics for reconstructing evolutionary histories among species. This is an emerging area that integrates large scale sequence data, computational tools and evolutionary principles to study phylogenetics (Blair and Murphy 2011). Phylogenomics separates from phylogenetics primarily in the input data. Unlike phylogenetics, phylogenomics investigations make use of large scale genomic or transcriptomic data to generate the input data based on orthology prediction of genes or whole-genome alignments.

For many lineages of insects, the resolution of the true phylogeny has been a challenge. One example is the Polyneopteran lineage. The branching patterns of the Polyneopteran insects such as cockroaches, mantids, earwigs, grasshoppers, and phasmids are highly ambiguous (Wipfler *et al.* 2014). The Polyneoptera are composed of 11 orders: Blattodea, Dermaptera, Embiodea, Grylloblattodea, Isoptera, Mantodea, Mantophasmatodea, Orthoptera, Phasmatodea, Plecoptera and Zoraptera. Establishing relationships among these orders has been difficult (Kristensen 1991). Using phylogenomic approaches, new insights are beginning to emerge on the internodal relationships of these insects (Wipfler *et al.* 2014). Moreover, application of phylogenomics in insects is appropriate because of the reason that insects are highly diverse. With about one million identified species, insects represent the highest proportion (~75%) of all animals (May 1988, Footit and Adler 2009). Hence, phylogenomics analyses are important in entomology because they can help us better understand evolutionary relationships of insects that are relevant to health, environment and agriculture. Moreover, during the last several years, genome sequencing of numerous insect species has been initiated, in addition to completion of almost hundreds of insect genomes. These are opening up new avenues for identification of orthologs for large-scale multigene phylogenetic studies. With the rising trends in insect genomics and transcriptomic studies, phylogenomic studies are gaining popularity that has potential to provide new insights into true phylogenies of many insect lineages.

Phylogenomic methods

Phylogenomic analyses are broadly carried out using three methods: 1) sequence alignment, 2) comparison of occurrences of 'DNA strings', and 3) comparison of gene content or order in the genome. The sequence alignment method is the most widely used approach in phylogenomic investigations. This approach relies on an accurate prediction of single copy orthologs representing genes in different species that have evolved from a common ancestral gene via speciation. Multiple alignment of orthologous genes are carried out either using individual orthologs or their concatenated sequences. The phylogenomic inferences are made from constructing a supertree of the individual genes (Bininda-Emonds and Sanderson 2001) or from supermatrix data that is generated by the alignment of concatenated sequences (Delsuc *et al.* 2005, Savard *et al.* 2006). The supertree based method relies on genes from overlapping taxa, whereas in the supermatrix method genes from non-overlapping taxa are treated as missing data. The independent evolutionary rates of multigene sequences are accommodated by methods such as partitioned-likelihoods to infer the phylogenetic relationships among the species. In addition to the sequence alignment based method, DNA

strings (distribution patterns of short sequences among genomes) and random anchor (a longer stretch of nucleotides) are also used in phylogenomics studies, particularly among the closely related species (Vishnoi *et al.* 2010).

The DNA string method does not require sequences to be aligned, and hence has an advantage over sequence alignment method. It is based on the frequency of short-oligonucleotide sequence combinations in the genomes that is used to directly compare higher order features of non-homologous DNA sequences (Edwards *et al.* 2002). The DNA string is also an appropriate method to delineate phylogenetic inference across deeper taxonomic levels where genomic homology is difficult to establish (Qi *et al.* 2004). The gene content and gene order information are also applicable in performing phylogenomic inferences (Murphy *et al.* 2004, Bourque and Tesler 2008). Gene order information are determined from scoring the presence or absence of pairs of orthologous genes that is then used as the measure of genetic distance between species (Wolf *et al.* 2001, Zhang *et al.* 2009). Gene rearrangement and duplications are critical factors that influence the accuracy of phylogenies based on gene orders. Recently, Hu *et al.* (2014) developed a maximum likelihood approach that takes into account of gene rearrangements, insertions, deletions and duplications for reconstruction of phylogeny from gene-order data. Phylogenomic studies using DNA strings or gene order methods are infrequent compared to those using sequence alignment methods. Although the majority of insect phylogenomics studies have been carried out using the sequence alignment methods, inference of phylogenies based on rearrangement of genomic regions, also known as ‘chromosomal phylogeny’, have provided useful information about evolution of specific insect lineages such as the species complex of *Anopheles gambiae* (Kamali *et al.* 2012, Sharakhov *et al.* 2013).

Current status of insect phylogenomics

In the recent years, phylogenomics studies have provided valuable insights into evolutionary relationships of insects. Numerous phylogenomic studies have been performed in diverse lineages of insects that have addressed several fundamental issues of classification and evolution of insect species. A non-comprehensive list of recent phylogenomic studies of insects is provided in Table 1 along with the study objectives. The expressed sequence tags (ESTs) (Behura 2006) and whole transcriptome data (Trautwein *et al.* 2012) are predominantly used to develop data for phylogenomic investigations of many insects. Although the application of EST datasets have been useful for phylogenetic analysis (Theodorides *et al.* 2002, Hughes *et al.* 2006, Parkinson and Blaxter 2009, Meusemann *et al.* 2010), 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).

Currently, multigene datasets and whole-transcriptome data are extensively used in phylogenomic analyses (Chan and Ragan 2013). In a recent publication, Misof *et al.* (2014) conducted a large-scale phylogenomic analysis of 144 insect species that provided a holistic view about 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 diversification of insects into four groups: Palaeoptera, Polyneoptera, Condylgnatha 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 foundation towards building the insect tree-of-life in finer detail (Jones 2015).

Phylogenomics has also been helpful in resolving species relationship of specific lineages of insects. Species phylogenies of mosquitoes is 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), 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). Using a combined dataset of six nuclear protein-coding genes and an array of morphological characters (n =80), Reidenbach *et al.* (2009) performed maximum parsimony and maximum likelihood analyses to understand phylogenetic radiation among mosquitoes representing 25 genera. Their analysis provided a 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) 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 showed 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 vectorial capacity of *Anopheles* to transmission of malaria. 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). Introgression, an important source of genetic variation in natural populations, is caused by stable integration of genetic material from one species into another through repeated back-crossing. Introgression is particularly a problem in phylogeny reconstruction of among closely related lineages partly due to the fact that 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, 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 is 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 resolution of holometabolous insects. Savard *et al.* (2006) 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, 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). 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 Mecoptera 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 monophyly of butterflies with Hesperidae (skippers) and Hedyidae (moth-butterflies), and provided support for placing butterflies sister to the obtectomeran Lepidoptera. Similarly, the aculeate Hymenoptera that are extensively investigated for eusocial behavior of many species such as ants, bees and stinging insects has 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 identification of 1:1 orthologous genes among the species. This has been a challenging step particularly when genes are present in duplicated copies within genome (Jensen 2001, Dalquen *et al.* 2013), and also for the reason that building orthology models relies upon all-against-all gene comparisons (Sonnhammer *et al.* 2014). Similarly, the presence of horizontally transferred genes further confound the problem of identification of orthologs (Dalquen *et al.* 2013). Furthermore, the number of single copy ortholog groups sharply decreases as we consider more diverse species (Figure 1). This is based on known orthologies predicted among different insect species where genome sequences are available (Waterhouse *et al.* 2013). Thus, the number of single copy orthologous genes that can be used in the phylogenomic analysis becomes a limiting factor to carry out phylogenomic investigations among distantly related species. Besides single copy orthologs, phylogenomics studies are also carried out with partitioning the data into predicted protein domains. This approach may also have a limitation in 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) may impact 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 is well known that 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), horizontal gene transfer (Doolittle 1999) and incomplete lineage sorting (Pamilo and Nei 1988) 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). 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 them (Salichos and Rokas 2013, Salichos *et al.* 2014, Dell'Ampio *et al.* 2014).

Long-branch attraction (where species with high evolutionary rates tend to group together) and heterotachy (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) 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), study

(Huelsenbeck and Lander 2003) has suggest 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 homotachous 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. Kolaczowski and Thornton (2008) 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) may also lead to conflicting phylogenetic inferences. By analyzing ~300,000 protein sequences across eukaryotes and prokaryotes, Driskell *et al.* (2004) 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). Similarly, bias in data matrix composition is known to mislead phylogenomic inferences. The phylogenomic investigation by Letsch *et al.* (2012), using expressed sequence tags of different Polyneoptera and Paraneoptera species, clearly demonstrated the effect of matrix composition on 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 inaccurate placement of the human louse along with Holometabola. This study suggested that when transcript sequences used in phylogenomic analyses are not comprehensive for each genome, the results may have potential pitfalls.

Future 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 vectorial ability of malaria mosquitoes (Fontaine *et al.* 2015) are recent examples of this progress.

In the 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). The emerging concept of ‘integrative phylogenomics’ is now gaining popularity in diverse domains of evolutionary biology (Bapteste and Burian 2010). Integration of anatomical data generated by the non-invasive three dimensional reconstruction techniques along with fossil and genomic data (Giribet and Edgecombe 2012) 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 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.

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 utilities in future phylogenomic studies of insects. The phylogenomic approaches to study functional adaption of a large number of genes may be beneficial compared to the phylogenetic approaches that rely on the correlation between single gene tree and the associated trait (gene function) (Pagel 1999). Besides, phylogenomic tools are also applicable in studying evolutionary histories of endosymbionts (Comas *et al.* 2007, Kembel *et al.* 2011). They may also find utility in cell biology to define 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) is 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 deeper understanding of biology and evolution in a holistic manner. 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 integration of phylogenomics tools to provide new insights into systematics and evolutionary relationships of several unresolved or poorly resolved insect lineages. With no doubt, insect phylogenomics holds huge promise in understanding insect biology in evolutionary terms; and sooner or later, it is possible that they may unravel some of the many secrets of why insects are so diverse and adaptive in nature.

Acknowledgments

My effort to prepare this article was partly supported by a grant (R21 AI101345-A1) from National Institutes of Health (NIH), USA. I am thankful to two anonymous reviewers for helpful suggestions and, also Joanne Cunningham for carefully reading the manuscript.

References

- 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]
- Baptiste E, Burian RM. On the need for integrative phylogenomics, and some steps towards its creation. *Biol Philos.* 2010; 25:711–736.
- Behura SK. Molecular marker systems in insects: current trends and future avenues. *Mol Ecol.* 2006; 15:3087–3113. [PubMed: 16968257]
- Behura SK. Insect microRNAs: Structure, function and evolution. *Insect Biochem Mol Biol.* 2007; 37:3–9. [PubMed: 17175441]

- 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]
- Blair C, Murphy RW. Recent trends in molecular phylogenetic analysis: where to next? *J Hered.* 2011; 102:130–138. [PubMed: 20696667]
- Boussau B, Walton Z, Delgado JA, Collantes F, Beani L, Stewart II, et al. Strepsiptera, phylogenomics and the long branch attraction problem. *PLoS One.* 2014; 9:e107709. [PubMed: 25272037]
- Brown D, Sjölander K. Functional classification using phylogenomic inference. *PLoS Comput Biol.* 2006; 2:e77. [PubMed: 16846248]
- Buljan M, Bateman A. The evolution of protein domain families. *Biochem Soc Trans.* 2009; 37:751–755. [PubMed: 19614588]
- Calvo E, Pham VM, Ribeiro JM. An insight into the sialotranscriptome of the non-blood feeding *Toxorhynchites amboinensis* mosquito. *Insect Biochem Mol Biol.* 2008; 38:499–507. [PubMed: 18405828]
- Campbell LI, Rota-Stabelli O, Edgecombe GD, Marchioro T, Longhorn SJ, Telford MJ, et al. MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. *Proc Natl Acad Sci U S A.* 2011; 108:15920–15924. [PubMed: 21896763]
- Chan CX, Ragan MA. Next-generation phylogenomics. *Biol Direct.* 2013; 8:3. [PubMed: 23339707]
- 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]
- 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. *bioRxiv.* 2014:009837.
- 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]
- Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference, and the multispecies coalescent. *Trends Ecol Evol.* 2009; 24:332–340. [PubMed: 19307040]
- Dell'Ampio E, Meusemann K, Szucsich NU, Peters RS, Meyer B, Borner J, et al. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol Biol Evol.* 2014; 31:239–349. [PubMed: 24140757]
- Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet.* 2005; 6:361–375. [PubMed: 15861208]
- Doolittle WF. Lateral genomics. *Trends Cell Biol.* 1999; 9:M5–M8. [PubMed: 10611671]
- 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]
- Eisen JA. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 1998; 8:163–167. [PubMed: 9521918]
- Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool.* 1978; 27:401–410.
- Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, et al. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science.* 2015; 347:1258524. [PubMed: 25431491]
- Footitt, RG.; Adler, PH. *Insect biodiversity: science and society.* Oxford (UK): Wiley-Blackwell; 2009.
- Friend WG, Smith JJ. Factors affecting feeding by bloodsucking insects. *Annu Rev Entomol.* 1977; 22:309–331. [PubMed: 319741]
- Giribet G, Edgecombe GD. Reevaluating the arthropod tree of life. *Annu Rev Entomol.* 2012; 57:167–186. [PubMed: 21910637]
- Hu F, Lin Y, Tang J. MLGO: phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinformatics.* 2014; 15:354. [PubMed: 25376663]

- 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]
- Huelsenbeck JP, Lander KM. Frequent inconsistency of parsimony under a simple model of cladogenesis. *Syst Biol.* 2003; 52:641–648. [PubMed: 14530131]
- i5K consortium. The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered.* 2013; 104:595–600. [PubMed: 23940263]
- Jensen RA. Orthologs and paralogs - we need to get it right. *Genome Biol.* 2001; 2:INTERACTIONS1002. [PubMed: 11532207]
- Johnson BR, Borowiec ML, Chiu JC, Lee EK, Atallah J, Ward PS. Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Curr Biol.* 2013; 23:2058–2062. [PubMed: 24094856]
- Johnson KP, Walden KK, Robertson HM. Next-generation phylogenomics using a target restricted assembly method. *Mol Phylogenet Evol.* 2013; 66:417–422. [PubMed: 23000819]
- Jones B. Building the insect tree-of-life. *Nat Rev Genet.* 2015; 16:2–3. [PubMed: 25404110]
- 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]
- Kamali M, Xia A, Tu Z, Sharakhov IV. A new chromosomal phylogeny supports the repeated origin of vectorial capacity in malaria mosquitoes of the *Anopheles gambiae* complex. *PLoS Pathog.* 2012; 8:e1002960. [PubMed: 23055932]
- Kawahara AY, Breinholt JW. Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proc Biol Sci.* 2014; 281:20140970. [PubMed: 24966318]
- Kembel SW, Eisen JA, Pollard KS, Green JL. The Phylogenetic diversity of metagenomes. *PLoS One.* 2011; 6:e23214. [PubMed: 21912589]
- Kolaczowski B, Thornton JW. A mixed branch length model of heterotachy improves phylogenetic accuracy. *Mol Biol Evol.* 2008; 25:1054–1066. [PubMed: 18319244]
- 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 edn. Vol. I. Melbourne University Press; Carlton: 1991. p. 125-140.
- 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]
- Legg DA, Sutton MD, Edgecombe GD. Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nat Commun.* 2013; 4:2485. [PubMed: 24077329]
- Letsch H, Simon S. Insect phylogenomics: new insights on the relationships of lower neopteran orders (Polyneoptera). *Sys Entomol.* 2013; 38:783–793.
- 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]
- Liang J, Cai W, Sun Z. Single-cell sequencing technologies: current and future. *J Genet Genomics.* 2014; 41:513–528. [PubMed: 25438696]
- Maddison WP. Gene trees in species trees. *Syst Biol.* 1997; 46:523–536.
- May RM. How many species are there on earth? *Science.* 1988; 241:1441–1449. [PubMed: 17790039]
- 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]
- 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]
- Murphy WJ, Pevzner PA, O'Brien SJ. Mammalian phylogenomics comes of age. *Trends Genet.* 2004; 20:631–639. [PubMed: 15522459]
- 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]

- Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science*. 2015; 347:1258522. [PubMed: 25554792]
- O'Loughlin SM, Magesa S, Mbogo C, Mosha F, Midega J, Lomas S, et al. Genomic analyses of three malaria vectors reveals extensive shared polymorphism but contrasting population histories. *Mol Biol Evol*. 2014; 31:889–902. [PubMed: 24408911]
- 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]
- Pagel M. Inferring the historical patterns of biological evolution. *Nature*. 1999; 401:877–884. [PubMed: 10553904]
- Pamilo P, Nei M. Relationships between gene trees and species trees. *Mol Biol Evol*. 1988; 5:568–583. [PubMed: 3193878]
- Parkinson J, Blaxter M. Expressed sequence tags: an overview. *Methods Mol Biol*. 2009; 533:1–12. [PubMed: 19277571]
- Peters RS, Meusemann K, Petersen M, Mayer C, Wilbrandt J, Ziesmann T, et al. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *BMC Evol Biol*. 2014; 14:52. [PubMed: 24646345]
- Philippe H, Snell EA, Baptiste E, Lopez P, Holland PW, Casane D. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol Biol Evol*. 2004; 21:1740–1752. [PubMed: 15175415]
- Qi J, Wang B, Hao BI. Whole proteome prokaryote phylogeny without sequence alignment: a K-string composition approach. *J Mol Evol*. 2004; 58:1–11. [PubMed: 14743310]
- Rannala B, Yang Z. Phylogenetic inference using whole genomes. *Annu Rev Genomics Hum Genet*. 2008; 9:217–231. [PubMed: 18767964]
- 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]
- Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, et al. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*. 2008; 452:949–955. [PubMed: 18362917]
- Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, Goldsmith MR, et al. Creating a buzz about insect genomes. *Science*. 2011; 331:1386. [PubMed: 21415334]
- Rota-Stabelli O, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, et al. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc Biol Sci*. 2011; 278:298–306. [PubMed: 20702459]
- Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*. 2013; 497:327–331. [PubMed: 23657258]
- 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]
- Salipante SJ, Horwitz MS. A phylogenetic approach to mapping cell fate. *Curr Top Dev Biol*. 2007; 79:157–184. [PubMed: 17498550]
- 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]
- Severson DW, Behura SK. Mosquito genomics: progress and challenges. *Annu Rev Entomol*. 2012; 57:143–166. [PubMed: 21942845]
- Siepel A. Phylogenomics of primates and their ancestral populations. *Genome Res*. 2009; 19:1929–1941. [PubMed: 19801602]
- Simon S, Narechania A, Desalle R, Hadrys H. Insect phylogenomics: exploring the source of incongruence using new transcriptomic data. *Genome Biol Evol*. 2012; 4:1295–1309. [PubMed: 23175716]
- Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet*. 2013; 14:618–630. [PubMed: 23897237]

- Sharakhov IV. Chromosome phylogenies of malaria mosquitoes. *Tsitologiya*. 2013; 55:238–240. [PubMed: 23875455]
- Sjölander K. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics*. 2004; 20:170–179. [PubMed: 14734307]
- 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]
- Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO, Rogers JS. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst Biol*. 2001; 50:525–539. [PubMed: 12116651]
- Theodorides K, De Riva A, Gómez-Zurita J, Foster PG, Vogler AP. Comparison of EST libraries from seven beetle species: towards a framework for phylogenomics of the Coleoptera. *Insect Mol Biol*. 2002; 11:467–475. [PubMed: 12230545]
- 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]
- 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]
- Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res*. 2013; 41(Database issue):D358–D365. [PubMed: 23180791]
- Wipfler B, Bai M, Schoville S, Dallai R, Uchifune T, Machida R, et al. Ice crawlers (Grylloblattodea) – the history of the investigation of a highly unusual group of insects. *J Insect Biodiversity*. 2014; 2:1–25.
- 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]
- Zhang H, Zhong Y, Hao B, Gu X. A simple method for phylogenomic inference using the information of gene content of genomes. *Gene*. 2009; 441:163–168. [PubMed: 18675889]

Figure 1.

Number of single copy orthologs among different insect species. The screenshot image shows 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 number of species on the top and number of single copy orthologous genes on the bottom.

Table 1

Recent developments in insect phylogenomics. A list of studies reported within the last 2 years only are provided*.

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

* The list may not be complete.