OXFORD

Briefings in Functional Genomics, 14(6), 2015, 424–431

doi: 10.1093/bfgp/elv010
Advance Access Publication Date: 30 March 2015
Review paper

Beyond fruit-flies: population genomic advances in non-*Drosophila* arthropods

Martin Hasselmann, Luca Ferretti and Amro Zayed

Corresponding author. M. Hasselmann, Livestock Population Genomics Group (460h), Institute of Animal Science, University of Hohenheim, Stuttgart, Germany. Tel.: +49 711 459 22481; E-mail: martin.hasselmann@uni-hohenheim.de

Abstract

Understanding the evolutionary processes driving the adaptive differentiation of populations is of broad interest in biology. Genome-wide nucleotide polymorphisms provide the basis for population genetic studies powered by advances in high-throughput sequencing technologies. These advances have led to an extension of genome projects to a variety of non-genetic model organisms, broadening our view on the evolution of gene families and taxonomic-restricted novelties. Here, we review the progress of genome projects in non-*Drosophila* arthropods, focusing on advances in the analysis of large-scale polymorphism data and functional genomics and examples of population genomic studies.

Key words: population genomics; adaptive evolution; neutrality tests; genetic differentiation; pool sequencing

Introduction

We are faced with a rapid progress of genome projects arising from non-model organisms. Starting with the release of the first insect genome of the fruit fly *Drosophila melanogaster* in the year 2000 [1], we are in the enviable position of having access to both inexpensive high-throughput sequencing technology (HTS) and advanced statistical methods for analyzing large-scale data. These two points have led to a nearly exponential increase in the number of genome projects from organisms broadly scattered across the animal kingdom. The growing availability of many reference genomes spurred on research in comparative evolutionary genomics, which aims to decipher the genetic composition and lineage-specific processes of, for example, gene gain and losses between species. Comparative genomic studies can provide insights into not only evolutionary processes shaping the differentiation between larger phylogenetic units [2–4] but also species-specific differentiation [4–6]. Further, the re-sequencing of genomes from a population sample of individuals (i.e. population genomics [7]) provides the platform to analyze genome-wide polymorphism to address a variety of interesting ecological, evolutionary and genetic questions in non-model organisms.

The decreasing sequencing costs over the past decade led to the recent emergence of new arthropod models that are suitable for addressing long-standing questions in ecology and evolution. The fascinating and seemingly endless diversity of arthropods provide fertile grounds for understanding how some ecological (e.g. diet and habitat specialization) and genetic traits influence patterns of genome evolution. For example, hymenopteran genomes have opened up the door to studying the effect of haplodiploidy and recombination (e.g. which is highly elevated in many social insects) and low effective population sizes on patterns of molecular evolution. Sociality represents a major transition in evolution and genomic research on social and solitary Hymenoptera (i.e. ants, bees and wasps) promises to help us understand how and why sociality evolved [8]. We now can broaden our view on evolution of obligate mutualism (e.g. extreme anatomical sexual dimorphism found in the fig wasp [9]), host–parasite interactions (e.g. the vectorial capacity for human malaria among *Anopheles* mosquito species [10]) or

Martin Hasselmann is professor and head of the Livestock Population Genomics Group at University of Hohenheim. His group focuses on studying the natural variation in populations of social insects and livestocks shaped by environmental interactions.
Luca Ferretti is postdoctoral researcher in Université Pierre et Marie Curie and Collège de France, Paris. He works in population genomics, statistical analysis of sequence data and fitness landscapes.
Amro Zayed is an associate professor at Department of Biology at York University, where he studies the evolutionary genomics of social insects.

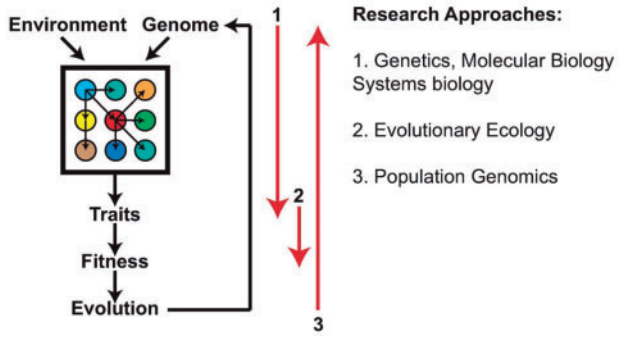

Figure 1. The next-generation synthesis: Population genomic studies act as a 'glue' connecting the genetics of complex traits and the relationship between organismal traits and fitness in nature. Population genomics explicitly link mutations in DNA sequences to patterns of selection that indicate past or ongoing changes in fitness. By integrating genetic, evolutionary ecology and population genomic studies, it is possible to understand how genes underlying complex traits evolved and how mutations in such genes influence fitness. (A colour version of this figure is available online at: http://bfg.oxfordjournals.org)

host–plant interaction in stick insects as example of how ecological sources affect phenotypic targets of selection [11]. Genomic work on Lepidoptera promises to transform our knowledge of the evolution of mimicry and migration [12], while studies on key planktonic crustaceans such as *Daphnia* will advance knowledge on the capacity of arthropods to adapt to toxins in their environment [13].

In addition to the large number of *de novo* genome sequencing projects, inexpensive sequencing has made way to powerful population genomic studies that characterize genome-wide patterns of within-species diversity and between-species divergence. Understanding the former is critical because, ultimately, differences between species first originate as differences among individuals [14, 15]. Population genomic studies allow researchers to map out genomic loci that underlie adaptive evolution [16–18] or introgression and chromosomal evolution signatures such as rearrangements in supergene systems [19]; previously an unfathomable goal, now an easier to achieve reality. Population genomics provide the glue between functional studies that link genes with phenotypes, and ecological and evolutionary studies that attempt to link differences in phenotype with differences in fitness (Figure 1).

In this review, we summarize progress of genome projects in non-*Drosophila* arthropods and advances in the statistical analysis of large-scale polymorphism data and provide examples of population genomic studies.

Progress in genome projects of arthropods: an overview

We found 135 completed or ongoing arthropod genome projects registered with the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov, circa June, 2014). Of these 135 genome projects, 57 belong to Dipteran species (flies and mosquitos). The remaining 77 genome projects are distributed among several insect orders, including the Hymenoptera (bees, wasp and ants), the Lepidoptera (butterflies and moths), the Coleoptera (beetles) and the Hemiptera (true bugs), although there is clear overrepresentation of hymenopteran genomes among non-dipteran insects. Beside insects, there are several genome projects of Chelicerata (spiders, scorpions and mites) and some Crustacean (e.g. the ostracod *Daphnia pulex*).

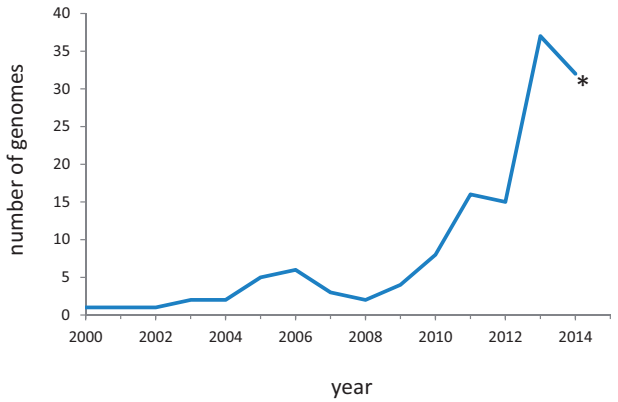

Figure 2. Number of arthropod genome projects starting from the year 2000 (first *Drosophila* genome) until now (2014, until June *).

Myriapods are the most basal of the four major extant arthropode clades, and the recently published Centipede genome will provide new insights on the key genetic and phenotypic innovations that facilitated subterranean life [20]. The number of new insect genomes sequenced per year appears to be exponentially increasing since 2000, with the most remarkable increase recorded in 2013 (Figure 2). This peak coincided with the release of the first insect genomes from the i5k initiative, a coordinated effort to sequence up to 5000 insect genomes in the next few years [21].

Multiple genome comparisons have allowed researchers to discover and better annotate genes and to identify, for example, lineage-specific gene (or gene family) gains or losses associated with the amazing phenotypic differences observed between different insect lineages or conserved noncoding sequences that are likely important for gene regulation in insect genomes [22–24]. The importance of DNA methylation and phenotypic evolution in social insects (e.g. caste determination) has been also emphasized by comparative genomics [25–27], as well as the important role of non-protein coding elements (e.g. micro RNAs [28]). Within the Hymenoptera, several genome projects of species representing different levels of social organization are currently under way, providing novel insights into the evolutionary processes leading to eusociality. Another initiative focuses on the evolution of parasitic life styles that independently evolved many times within insects.

Advances in methods for high-throughput genome analysis

HTS methods [29] have many useful applications for studying the genomics of non-model arthropods. Modern HTS technologies deliver a large amount of data; at present, for example, a single lane of Illumina HiSeq 2500 delivers about 300 million reads (average read length 160 bp) for a total length of 50 Gbp (billions of base pairs). For the smallest arthropod genomes (about 100 Mbp [30]), this corresponds to an impressive coverage of 500×. Arthropod genome sizes vary by more than an order of magnitude: consequently, the most effective *de novo* sequencing strategy will depend on genome size. For small genomes, it would be cost-effective to sequence a pool of individuals in a single lane, whereas for genomes larger than a few hundred Mb, sequencing each individual separately would be most effective.

HTS strategies for population genetics are based on re-sequencing several individuals from a population. The reads are either mapped to the reference genome or, if it is not available,

clustered into small scaffolds. Calling genetic variants, typically single-nucleotide polymorphisms (SNPs), from the aligned sequences are the starting point for subsequent population genetics analyses, and few of them (e.g. phylogenetic reconstruction, average nucleotide diversity, inference of demography and migration rates between populations) can be performed without SNP data mapping to a reference sequence. However, most population genetics methods aim at detecting genomic regions under selection (e.g. genomic scans [31]) and therefore require the knowledge of the location of SNPs along the genome, including their coding context; these methods require alignment to a reference genome. If the species of interest lacks an annotated reference genome, the alternatives are either to obtain a genome by *de novo* sequencing combined with automatic or semiautomatic annotation, or to use the annotated reference genome of the closest species available as used in the *Drosophila miranda* genome project [32].

Sequencing individuals separately and aligning them to the reference genome of a close species is the most immediate approach for population genetic studies. There are two different strategies for the choice of the appropriate read depth [33]. For studies that do not use haplotype information, a low coverage (2–4× per individual) is often sufficient for standard population genetics analyses, provided that the sample size is large enough (tens of individuals). For studies that focus on haplotype tests or specific mutations, a moderate to high coverage (8–20× per individual) is advised. Above this coverage, it is convenient to increase the number of individuals sequenced rather than the coverage per individual [33].

Pooled samples

For HTS approaches using the cost-efficient strategy of pooling a number of different individual DNA samples together (pool-seq)[34], two parameters are most relevant: (1) the sample size, i.e. the number of individuals combined in one pool, and (2) the total sequencing depth (coverage), which measures how often the pool has been sequenced. The quality of the analyses increases with both sample size and read depth. Although several analyses have been published with pools sequenced at a total read depth of 10×, simulations show that better results can be obtained with a total read depth of 20× or higher [35]. For experimental design, there are both theoretical and practical reasons to increase sample size instead of read depth. First, increasing the number of sampled individuals is usually much less expensive than increasing the amount of sequence. Second, pools with a few individuals could be prone to biases owing to unequal DNA contributions from each individual, while this effect is negligible for large samples [35]. Third, although pools with a large number of individuals relative to sequencing depth have reduced SNP quality (high error probability), they summarize more accurately the mutation frequency spectrum, as it is averaged over more evolutionary histories [36]. Most important for the analysis of pooled genome data is to use SNP callers specific for pooled samples or that have a specific option for pools [37].

Summary about barcoding (parallel tagged sequencing)

To increase cost-effectiveness, it is possible to sequence multiple individuals or multiple pools from different populations in a single lane by barcoding/tagging/multiplexing technologies. Here, samples are uniquely tagged with short sequences (barcodes), pooled, sequenced and subsequently sorted by barcode [38]. An increasingly popular strategy is the restriction-site-associated DNA sequencing (RADseq) [39], which combines restriction-enzyme recognition sites with HTS and barcoding and has been established as a powerful method for population genomic studies, initially applied by Hohenlohe *et al*. in sticklebacks [40]. Alternative RAD approaches are ddRAD [41], 2bRAD [42] and genotype by sequencing (GBS) variants, such as GBS [43], RESTseq [44], SLAFseq [45]. These methods reduce the complexity of the genome, providing a large number of dense SNP markers sequenced at much higher read depths compared with whole-genome sequencing. RADseq approaches provide excellent data for studying genome-wide variability within single or closely related populations, for association studies and breeding programs. However, whole-genome sequencing is more appropriate when the focus is on genetic differences between distant populations or close species, or on specific genomic regions.

RNA-seq

There are several advantages in doing RNA-seq [46] of neglected genomes. First, by definition, RNA-seq provides only transcribed sequences, including the exome as well as functional noncoding RNA, therefore focusing on the most interesting functional regions of the genome. As the true goal of many studies in non-model species is exome sequencing rather than whole-genome sequencing, RNA-seq appears as an effective alternative. Second, RNA-seq simultaneously produces sequence information and expression levels, which are correlated to the read depth of the sequences. Furthermore, in principle, the most expressed genes can be assembled *de novo* from the data without the need of a reference genome, even if the quality of the assembly is much lower than what would be obtained by mapping to a reference genome [46]. For individual RNA-seq data, it is possible to use the same methods as for individual DNA HTS samples. In this way, the recent study of McCoy *et al*. generated first genomic resources for Gillette's checkerspot butterfly, *Euphydryas gillettii*, by RNA-seq and then used the discovered SNPs to infer population demography of individuals from two geographic regions [47]. However, there are also several major drawbacks to population genomic studies conduced using RNA-Seq. The most relevant is related to the skewed distribution of expression levels, which translates into a skewed distribution of read depths among different genes. Only the most expressed genes reach a sufficient read depth for population analyses, therefore covering only a fraction of the exome [48]. This issue has been mitigated by the high read depths currently available. Other drawbacks are the increased experimental noise and the uncertainties related to allele-specific expression and related effects (e.g. tissue-specific expression, not covered by the sampling, might hamper the discovery of genomic variants), which are difficult to control [49], especially with pooled RNA-seq. In addition, the missing out of regulatory (noncoding) regions should be taken into account.

Software and genetic technologies

The processing of large-scale genomic data is the key issue of all population genomic approaches. Over the past years, several software suites have been developed to improve the data handling and to estimate several test statistics (e.g. VarScan [50], PoPoolation and PoPoolation2 [51, 52] and PopGenome [53]). These packages contain modules for completing all necessary tasks for basic and advanced population genomic data handling. This includes (i) reading a variety of input data files

resulting from re-sequencing projects, (ii) including annotation files and select regions of interest, and (iii) applying a battery of test statistics (Neutrality tests, linkage disequilibirum, recombination, diversity, selective sweeps, genetic differentiation—F_{ST}/BayeScanR, site frequency spectrum). The detection of selective sweeps in pool-seq data is statistically challenging mainly owing to sequencing errors and random sampling among chromosomes but has been initially solved by Boitard and colleagues [54, 55]. The software package NPStat [35] by Ferretti and colleagues is focused on the statistics of neutrality test for population genomic data and their improvement. Specifically, the authors generalized the most common allele-frequency-spectrum-based estimators and tests (e.g. variability, heterozygosity, Tajima's D, Fu and Li's D and F, Fay and Wu's H) for the application of pooled sequencing data, correcting for sequencing errors and ascertainment bias. There are also now methods for testing neutral evolution while incorporating demographic information (e.g. recent changes in population size), as recently applied to a human genome-wide SNP data set [56]. The demographic dynamics of the population can be also inferred from a single diploid genome by methods like the pairwise sequentially Markovian coalescent model by Li and Durbin [57]. These methods exploit the information stored in the distribution along the genome of the time to the most recent common ancestor of two sequences. Demographic methods like these are effective for genome-wide data, thanks to the large size of animal genomes.

Beside the ongoing progress in finding ways to exploit rapidly accumulating DNA sequence data, only few genetic technologies are available that enable us to experimentally manipulate the genomes of non-genetic model arthropods of interest. Promising initiatives to foster collaborations and promote genetic technologies are, for example, the NSF-funded Insect Genetic Technologies Research Coordination network (http://igtrcn.org/) and, supported by the German Research Foundation, the *ibeetle*-program (http://ibeetle-base.uni-goettingen.de/) that establish the Red flour beetle *Tribolium castaneum* as model for genome-wide RNA interference (RNAi) screens. This *iBeetle* screen aims to target diverse biological questions and to overcome the currently prevailing candidate gene approach in arthropods. Recently, significant advances in establishing genetic tools for the honey bees have been made by Schulte *et al.* [58, 59]. The authors identified honey bee promoter sequences that can drive constitutive and tissue-specific induced gene expression and established a stable piggyBac transposon mediated platform to manipulate gene function by microinjection, a technique that now extends the previously established RNAi technique in honey bees [60, 61].

A summary of population genomics in non-model arthropods

Population genomics of social insects

Recent studies in social insects using population genomics provided insights into how the honey bee genome is shaped by evolutionary processes related to the local environment and the dynamics of eusociality. A recent study by Harpur *et al.* [62] focused on the prevalence of positive selection on the honey bee genome, and to test some ideas about how sociality evolves in insects, using 39 sequenced workers from the European honey bee *Apis mellifera* and one worker from the closely related Asiatic honey bee *Apis cerana*. The *A. mellifera* workers belonged to four highly distinct geographic populations, which facilitated tests of selection over short and intermediate timescales, since divergence of the different *A. mellifera* populations less than 1 million years ago, and since the divergence of *A. mellifera* and *A. cerana* ~5 million years ago, respectively. The authors used McDonald–Kreitman(MK)-type tests and outlier tests to identify genomic regions with statistical signatures of positive selection (Figure 3).

This study offered some interesting insights about adaptive evolution in the honey bee, for example, that taxonomically restricted genes (TRGs) were associated with higher levels of positive selection relative to conserved genes. In bees, TRGs had significantly higher selection coefficients relative to genes that were taxonomically restricted to the Hymenoptera (ants, bees and wasps), which in turn, had higher selection coefficients relative to genes that were found in two or more insect orders. The importance of genes that lack similarity to genes of other species (also known as orphan genes) as driver of lineage-specific adaptation has been pointed out also by Wissler et al [63] in a comparative study of 30 arthropod genomes. The authors found that different arthropods' lineages accumulate orphan genes at different rate, and some social insects, including leaf-cutter ants, appear to have a dynamic model of frequent gene birth and death. These findings indicate that novel genes play a significant and previously underappreciated role in facilitating the adaptive evolution of social insects, which need to be functionally characterized (e.g. by RNAi-experiments) in the future.

Worker bees are effectively sterile members of honey bee colonies, and—under typical circumstances—have no direct fitness. Positive selection can still fix mutations that affect worker traits, so long as these traits benefit related reproductive individuals (i.e. mother queen and brother drones); this type of indirect selection is referred to as kin-selection [64, 65]. Harpur *et al.* [62] provided strong empirical evidence for kin-selection by showing that genes associated with worker phenotypes are enriched for signatures of positive selection. For example, proteins that are, on average, upregulated in workers relative to queens have significantly higher selection coefficients relative to proteins that are upregulated in queens relative to workers (Figure 3). Moreover, genes that are associated with several aspects of worker behavior, based on microarray studies [66], are often enriched for outlier loci with putative signs of positive selection. The population genomic data clearly portray worker traits as major vectors for adaptive evolution of honey bee colonies.

The evolution of *cis*-regulatory sequences in the honey bee genome was investigated by Molodtsova *et al.* 2014 [67], estimating patterns of natural selection at putative *cis*-regulatory regions for most genes using a modified MK-test on the data set of [62]. The authors studied how the structure of a brain transcriptional regulatory network that influences several aspects of worker behavior evolves. The transcriptional regulatory network contains a few hundred transcription factors that in turn regulate several thousand target genes. Molodtsova *et al.* 2014 discovered that the most connected 'hub' genes had the highest levels of negative selection on their coding sequences; genes with signs of adaptive protein evolution mostly resided at the network periphery. In contrast, connectedness appeared to have no influence of patterns of molecular evolution at putative *cis*-regulatory sequences. These results show how the structure of regulatory networks can facilitate adaptive evolution via both regulatory and coding mutations.

Another honey bee population genomic study used a comprehensive worldwide survey of sequence variation in 140 genomes of honey bees to understand the genetic basis of local

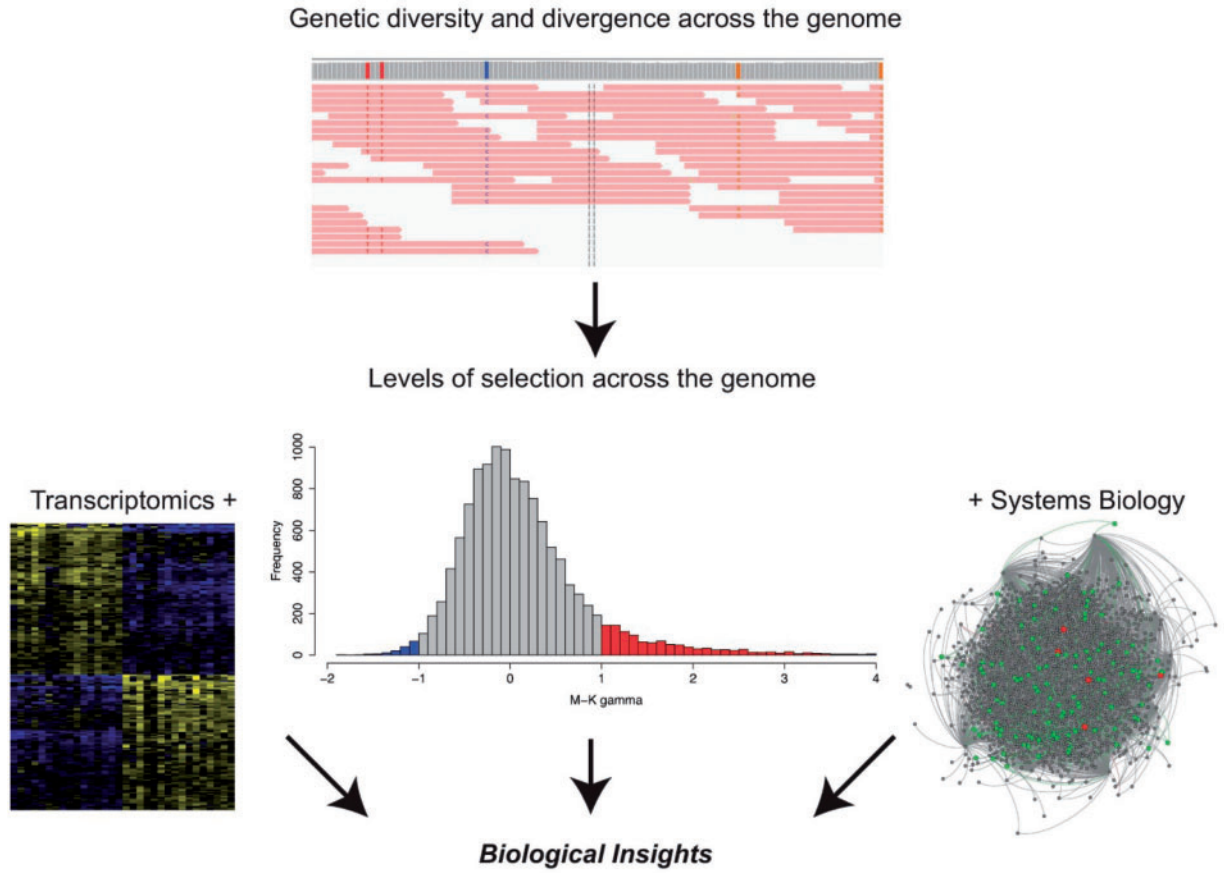

Figure 3. Population genomics leverages functional genomic data to generate novel biological insights. The honey bee is a model organism for the study of complex social behaviors. Functional genomic research conducted over the past decade identified many genes that are associated with worker behaviors using transcriptomic and proteomic approaches. Systems biology studies have also identified key transcriptional regulatory modules that control behavioral states in bees. Harpur *et al.* carried out a population genomic study by sequencing 40 honey bee genomes, and used this data set to map signatures of positive (M-K gamma >1) or negative (M-K gamma < −1) selection across the bee genome. By integrating the population genomics data set with the existing transcriptomic and proteomic data sets, the authors were able to address long-standing questions in the field of sociobiology—questions that were otherwise impossible to investigate (data taken from [45]). (A colour version of this figure is available online at: http://bfg.oxfordjournals.org)

adaptation of the honey bee *A. mellifera* [18]. The genetic variation at 8.3 million SNPs laid ground to a powerful evaluation of genetic differentiation and signatures of local adaptation. In the study of Wallberg *et al.*, measurement of F_{ST} at single SNPs presents evidence for high differentiation between African and European bees in intergenic, intronic and coding regions. Genomic signatures of local adaptation enriched in genes related to the immune system and sperm motility likely reflect geographic variation in reproduction and disease resistance, one of the novel insights when compared with the results from [62].

In conclusion, these studies reflect nicely the utility of population genomic approaches to elucidate complex evolutionary processes.

Population genomics suggests that social insects have reduced population sizes

Theories predict that small effective populations sizes (N_e) result in decreased amount of within-species polymorphism and decreased efficiency of natural selection owing to stronger impact of genetic drift when compared with species with larger N_e [68]. In social insects, N_e is expected to be small [69], as the number of reproductive individuals is usually relatively low (e.g. one mated queen per honey bee colony, but reproductive workers may occur). Romiguier *et al.* [70] used population genomics to test this hypothesis. Romiguer *et al.* used RNA-seq polymorphism and divergence data of eight eusocial and non-social insects and calculated low levels of genetic polymorphism, which are markedly lower in eusocial than in non-social insects. These results, along with other evidence presented by Romiguer *et al.*, support theoretical expectations of low N_e in eusocial insects comparable with that of mammals. However, in this study, 75% of the social insects belonged to the haplodiploid insect order Hymenoptera, while all the solitary insects belonged to diploid insect orders. More population genomics data on solitary hymenoptera species are needed to rule out the possibility that low N_e in social insects is simply an artifact of haplodiploidy rather than sociality. Interestingly, the study of Wallberg *et al.* [18] provided evidence that N_e of honey bees are much higher than previously expected (for European populations ~200 000, for African populations ~500 000), which cast some doubt over Romiguer *et al.*'s findings.

The genomics of adaptive traits—lessons from the Postman butterfly *Heliconius*

Population genomics of *Heliconius* species have provided insights into the genomic basis for complex adaptive traits. Butterfly wing patterning is a well-studied example of a complex adaptive trait driven by functional changes in natural populations (see e.g. [71–73]). The *Heliconius* Genome

Consortium [74] sequenced the genome of the postman butterfly *Heliconius melpomene* and used RAD sequencing of 84 *H. melpomene* butterflies and its relative to reconstruct a robust phylogeny of the genus and to test for introgression between a subset of individuals (the sympatric co-mimetic races *H. melpomene amaryllis* and *Heliconius timareta ssp. nov*. The *Heliconius* Genome consortium found strong evidence for adaptive introgression contributing to evolutionary radiation within this butterfly group and driving genomic divergence (F_{ST}) in color pattern region and non-color region between *H. timareta* and of *H. melpomene* populations. A 400 kb chromosomal region identified by Joron *et al*. [19] provides evidence for a mimicry supergene system in *Heliconius numata*, maintained by balanced polymorphism and highlight, how chromosomal rearrangements can contribute to the coexistence of adaptive phenotypes. A recent population genomics approach of Supple *et al*. [5] using RNA-seq short read data from pooled samples of *Heliconius erato* races and *H. melpomene* narrowed down a 65 kb region, showing high levels of differentiation between hybridizing races, low nucleotide diversity within races and increased linkage disequilibrium relative to other genomic regions. These findings strongly indicate a recent fixation of haplotypes driven by positive selection. The authors propose that tightly linked *cis*-regulatory regions of the transcription factor *optix* located within that region control differentiation of color patterning between *H. erato* races. Another tempting strategy to analyze candidate loci controlling wing pattern phenotypes in hybridizing *Heliconius* butterflies was used by Nadeau *et al*. [6]. The authors used targeted next-generation sequence capture to survey polymorphism across 3.5 Mb SNPs among divergent geographical races and species in *Helconius*. By this, several island of elevated divergence ($F_{ST} > 0.3$) even between closely related species were identified, harboring loci for wing color patterning. The combination of population genomic resequencing, association mapping, positional cloning and developmental data applied by Gallant *et al*. [3] shed light on the molecular basis of ancient phenotypic divergence of wing patterning in butterflies. In this study, comprehensive data from two butterfly lineages diverged >65 million years ago (*Limenitis* and *Heliconius*) provide evidence that the positional orthologous region of the WntA locus has independently driven the evolution of adaptive wing patterning.

Signatures of introgression and local adaptation among microcrustacean water flea *Daphnia*

Obligate asexual organisms may be prone to accumulate increased number of deleterious mutations, caused by the lack of recombination, which otherwise would break up linkage between selected sites and newly arisen deleterious ones [75], and promotes elevated heterozygosity. Tucker *et al*. [76] obtained whole-genome sequences of 11 sexual and 11 asexual genotypes from *D. pulex* to investigate pattern of variation across the genomes of both reproductive types. The study shows that genome-wide total heterozygosity per site (π_t) in asexual genotypes is 31% elevated when compared with sexual genotypes. The clonal reproduction of *D. pulex* is also manifested by the 33,575 SNPs that are fully associated with obligately asexual genotypes. Interestingly, the authors could determine by using divergence data from the sister species *Daphnia pulicaria*, that the asexual phenotype of *D. pulex* originated by introgression of chromosomal parts. Specifically, parts of chromosome VIII and IX show particularly high level of heterozygosity, caused by the historical introgression of chromosomal parts of the *D. pulicaria* genome. A first step to decipher local adaptation of *Daphina* population using RNA-seq has been recently performed by Schwarzenberger *et al*. [13]. The authors identified several candidate genes (transporter genes) that respond specifically in the presence of toxic microcystins, produced by cyanobacteria as serious threat to freshwater ecosystems.

These examples provide promising evidence, that the use of HTS technology opens the avenue to extend research projects in diverse ecosystems to better understand macro- and microevolutionary processes, driven by structural genomic and adaptive regulatory changes, taking advantage of the enormous richness of non-*Drosophila* arthropods with distinct life-history traits.

Conclusions and future directions

Population genomics has become a powerful tool for addressing fundamental questions in biology and evolution. By studying genome-wide polymorphism data sets, genomic regions with signs of positive selection can be narrowed down to the locus of adaptation. High-resolution genome-wide association mapping may even help to identify causative mutations associated with the interesting and charismatic phenotypes of arthropods. For instance, comparative population genomic studies across groups of solitary, primitively social and advanced social insects are needed to make full sense of the adaptive evolutionary changes that gave rise to sociality, and the subsequent evolutionary changes that established social societies from primitive to advanced forms. Speciation processes, driven by local environmental adaptation, reduced fitness of hybrids, and step-wise accumulation of genomic island of increased divergences can be surveyed now in more non-*Drosophila* arthropods by using genome-wide sequence data. Consequently, we are now in the comfortable situation to test population genetic theories with additional empirical data representing distinct phylogenetic lineages. By this, we can gain a broader picture of the evolutionary convergent or homologous trajectories shaping organismic diversity and avoid a biased view on animal evolution.

Although a powerful tool on its own, it is difficult to interpret population genomic data sets without knowledge of the genetics and molecular biology of complex traits. It is even more difficult to make sense of population genomic data without a good understanding of the ecology and biogeography of the organisms sequenced. For future research, it is of utmost importance to unravel and test the function of mutations in regulatory sequences and genes that may be linked to adaptive evolutionary processes in populations identified by genome-wide studies of polymorphism. By integrating population genomic data with functional genomics and ecological knowledge, we will unravel evolutionary processes driving organismic development.

Key points

- Population genomics may provide insight into adaptive evolutionary processes within and between species
- Ongoing progress in sequencing technology will facilitate the generation of high-throughput data and technologies to manipulated genomes now reach non-*Drosophila* arthropods
- Statistical advances to analyze genome-wide polymorphism increases the prediction for signatures of adaptive evolution
- High-throughput analysis of the divers non-*Drosophila* arthropods may foster an unbiased view on macro- and microevolutionary processes

Acknowledgements

We thank three anonymous reviewers for their constructive comments.

Funding

M.H. is supported by the German Research Foundation (DFG, HA 5499/3-2). A.Z. is funded by a discovery grant from the Natural Sciences and Engineering Research Council of Canada, and an Early Researcher Award from the Ontario Ministry of Research and Innovation.

References

1. Adams MD, Celniker SE, Holt RA, *et al*. The genome sequence of *Drosophila melanogaster*. *Science* 2000;**287**:2185–95.
2. Gallant JR, Traeger LL, Volkening JD, *et al*. Nonhuman genetics. Genomic basis for the convergent evolution of electric organs. *Science* 2014;**344**:1522–5.
3. Gallant JR, Imhoff VE, Martin A, *et al*. Ancient homology underlies adaptive mimetic diversity across butterflies. *Nat Commun* 2014;**5**:4817.
4. Woodard SH, Fischman BJ, Venkat A, *et al*. Genes involved in convergent evolution of eusociality in bees. *Proc Natl Acad Sci USA* 2011;**108**:7472–7.
5. Supple MA, Hines HM, Dasmahapatra KK, *et al*. Genomic architecture of adaptive color pattern divergence and convergence in Heliconius butterflies. *Genome Res* 2013;**23**:1248–57.
6. Nadeau NJ, Whibley A, Jones RT, *et al*. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Phil Trans R Soc Biol Sci* 2012;**367**:343–53.
7. Black WC, Baer CF, Antolin MF, *et al*. Population genomics: Genome-wide sampling of insect populations. *Annu Rev Entom* 2001;**46**:441–69.
8. Smith M, Szathmary E. *The Major Transition in Evolution*. Oxford, England: Oxford University Press, 1995.
9. Xiao JH, Yue Z, Jia LY, *et al*. Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biol* 2013;**14**:R141.
10. Neafsey DE, Waterhouse RM, Abai MR, *et al*. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science* 2015;**347**:1258522.
11. Soria-Carrasco V, Gompert Z, Comeault AA, *et al*. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 2014;**344**:738–742.
12. Nadeau NJ, Martin SH, Kozak KM, *et al*. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Mol Ecol* 2013;**22**:814–26.
13. Schwarzenberger A, Sadler T, Motameny S, *et al*. Deciphering the genetic basis of microcystin tolerance. *BMC Genomics* 2014;**15**:776.
14. Harr B. Genomic islands of differentiation between house mouse subspecies. *Genome Res* 2006;**16**:730–7.
15. Hutter S, Saminadin-Peter SS, Stephan W, *et al*. Gene expression variation in African and European populations of *Drosophila melanogaster*. *Genome Biol* 2008;**9**:R12.
16. Bradic M, Teotonio H, Borowsky RL. The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus*. *Mol Biol Evol* 2013;**30**:2383–400.
17. Zayed A, Whitfield CW. A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. *Proc Natl Acad Sci USA* 2008;**105**:3421–6.
18. Wallberg A, Han F, Wellhagen G, *et al*. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat Genet* 2014;**46**:1081–8.
19. Joron M, Frezal L, Jones RT, *et al*. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 2011;**477**:203–6.
20. Chipman AD, Ferrier DE, Brena C, *et al*. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biol* 2014;**12**:e1002005.
21. Consortium Tik. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered* 2013;**104**:595–600.
22. Weinstock GM, Robinson GE, Gibbs RA, *et al*. Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* 2006;**443**:931–49.
23. Elsik CG, Worley KC, Bennett AK, *et al*. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* 2014;**15**:86.
24. Werren JH, Richards S, Desjardins CA, *et al*. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* 2010;**327**:343–8.
25. Kucharski R, Maleszka J, Foret S, *et al*. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 2008;**319**:1827–30.
26. Lyko F, Foret S, Kucharski R, *et al*. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol* 2010;**8**:e100506.
27. Drapeau MD, Albert S, Kucharski R, *et al*. Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res* 2006;**16**:1385–94.
28. Hausser J, Zavolan M. Identification and consequences of miRNA-target interactions—beyond repression of gene expression. *Nat Rev Genet* 2014;**15**:599–612.
29. Glenn TC. Field guide to next-generation DNA sequencers. *Mol Ecol Res* 2011;**11**:759–69.
30. Kelley JL, Peyton JT, Fiston-Lavier AS, *et al*. Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nat Commun* 2014;**5**:4611.
31. Nielsen R, Williamson S, Kim Y, *et al*. Genomic scans for selective sweeps using SNP data. *Genome Res* 2005;**15**:1566–75.
32. Kaiser VB, Bachtrog D. De novo transcriptome assembly reveals sex-specific selection acting on evolving neo-sex chromosomes in *Drosophila miranda*. *BMC Genomics* 2014;**15**:241.
33. Nevado B, Ramos-Onsins SE, Perez-Enciso M. Resequencing studies of nonmodel organisms using closely related reference genomes: optimal experimental designs and bioinformatics approaches for population genomics. *Mol Ecol* 2014;**23**:1764–79.
34. Futschik A, Schlötterer C. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 2010;**186**:207–18.
35. Ferretti L, Ramos-Onsins SE, Perez-Enciso M. Population genomics from pool sequencing. *Mol Ecol* 2013;**22**:5561–76.
36. Ferretti L, Raineri E, Ramos-Onsins S. Neutrality tests for sequences with missing data. *Genetics* 2012;**191**:1397–401.
37. Raineri E, Ferretti L, Esteve-Codina A, *et al*. SNP calling by sequencing pooled samples. *BMC Bioinformatics* 2012;**13**:239.

38. Wong KH, Jin Y, Moqtaderi, Z. Multiplex illumina sequencing using DNA barcoding. *Curr Prot Mol Biol* 2013;**Chapter 7**:Unit 7.11
39. Baird NA, Etter PD, Atwood TS, *et al*. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 2008;**3**:e3376.
40. Hohenlohe PA, Bassham S, Etter PD, *et al*. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 2010;**6**:e1000862.
41. Peterson BK, Weber JN, Kay EH, *et al*. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 2012;**7**:e37135.
42. Wang S, Meyer E, McKay JK, *et al*. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods* 2012; **9**:808–10.
43. Elshire RJ, Glaubitz JC, Sun Q, *et al*. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 2011;**6**:e19379.
44. Stolle E, Moritz RFA. RESTseq - efficient benchtop population genomics with restriction fragment sequencing. *PLoS One* 2013;**8**.
45. Sun X, Liu D, Zhang X, *et al*. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 2013;**8**:e58700.
46. Gayral P, Melo-Ferreira J, Glemin S, *et al*. Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *Plos Genet* 2013;**9**: e1003457.
47. McCoy RC, Garud NR, Kelley JL, *et al*. Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Mol Ecol* 2014;**23**:136–50.
48. Sims D, Sudbery I, Ilott NE, *et al*. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014;**15**:121–32.
49. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet* 2013;**93**: 641–51.
50. Koboldt DC, Chen K, Wylie T, *et al*. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 2009;**25**:2283–5.
51. Kofler R, Orozco-terWengel P, De MN, *et al*. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* 2011;**6**:e15925.
52. Kofler R, Pandey RV, Schlotterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 2011;**27**:3435–6.
53. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, *et al*. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol Biol Evol* 2014;**31**:1929–36.
54. Boitard S, Schlotterer C, Futschik A. Detecting selective sweeps: a new approach based on hidden markov models. *Genetics* 2009;**181**:1567–78.
55. Boitard S, Schlotterer C, Nolte V, *et al*. Detecting selective sweeps from pooled next-generation sequencing samples. *Mol Biol Evol* 2012;**29**:2177–86.
56. Rafajlovic M, Klassmann A, Eriksson A, *et al*. Demography-adjusted tests of neutrality based on genome-wide SNP data. *Theor Popul Biol* 2014;**95**:1–12.
57. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature* 2011;**475**:493–6.
58. Schulte C, Theilenberg E, Muller-Borg M, *et al*. Highly efficient integration and expression of piggyBac-derived cassettes in the honeybee (*Apis mellifera*). *Proc Natl Acad Sci USA* 2014;**111**: 9003–8.
59. Schulte C, Leboulle G, Otte M, *et al*. Honey bee promoter sequences for targeted gene expression. *Insect Mol Biol* 2013;**22**: 399–410.
60. Beye M, Hasselmann M, Fondrk MK, *et al*. The gene csd is the primary signal for sexual development in the honey bee and encodes a SR-type protein. *Cell* 2003;**114**:419–29.
61. Beye M, Hartel S, Hagen A, *et al*. Specific developmental gene silencing in the honey bee using a homeobox motif. *Insect Mol Biol* 2002;**11**:527–32.
62. Harpur BA, Kent CF, Molodtsova D, *et al*. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci USA* 2014;**111**: 2614–19.
63. Wissler L, Gadau J, Simola DF, *et al*. Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome Biol Evol* 2013;**5**:439–55.
64. Hamilton WD. The genetical evolution of social behavior I. *J Theor Biol* 1964;**7**:1–16.
65. Hamilton WD. The genetical evolution of social behavior II. *J Theor Biol* 1964;**7**:17–52.
66. Zayed A, Robinson GE. Understanding the relationship between brain gene expression and social behavior: Lessons from the honey bee. *Annu Rev Gen* 2012;**46**:591–615.
67. Molodtsova D, Harpur BA, Kent CF, *et al*. Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. *Front Genet* 2014;**5**:431.
68. Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 2009;**10**: 195–205.
69. Pamilo P, Crozier RH. Population biology of social insect conservation. *Mem Mus Vic* 1997;**56**:411–19.
70. Romiguier J, Lourenco J, Gayral P, *et al*. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J Evol Biol* 2014;**27**:593–603.
71. Brakefield PM, Gates J, Keys D, *et al*. Development, plasticity and evolution of butterfly eyespot patterns. *Nature* 1996;**384**: 236–42.
72. Beldade P, Brakefield PM. The genetics and evo-devo of butterfly wing patterns. *Nat Rev Genet* 2002;**3**:442–52.
73. Beldade P, Brakefield PM, Long AD. Contribution of Distal-less to quantitative variation in butterfly eyespots. *Nature* 2002; **415**:315–18.
74. Heliconius Genome C. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 2012;**487**:94–8.
75. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res* 1966;**8**:269–94.
76. Tucker AE, Ackerman MS, Eads BD, *et al*. Population-genomic insights into the evolutionary origin and fate of obligately asexual Daphnia pulex. *Proc Natl Acad Sci USA* 2013;**110**: 15740–5.