Application of next-generation sequencing to the study of non-model insects

Wachi, Nakatada

Matsubayashi, Kei W.

Maeto, Kaoru

(Citation)
Entomological Science,21(1):3-11

(Issue Date)
2018-03

(Resource Type)
journal article

(Version)
Accepted Manuscript

Application of next-generation sequencing to the study of non-model insects

Nakatada Wachi[1, 4], Kei W. Matsubayashi[2], Kaoru Maeto[3]

[1]Biohistory Research Hall, Takatsuki, Japan

[2]Faculty of Arts and Science, Kyushu University, Fukuoka, Japan.

[3]Graduate School of Agricultural Science, Kobe University, Kobe, Japan

[4]Present address: Department of Environmental Research, Research Institute of Environment, Agriculture, and Fisheries, Osaka Prefecture, Habikino, Japan

[1]nwachi@kyudai.jp

[2]matsubayashi.kei@artsci.kyushu-u.ac.jp

[3]maeto@kobe-u.ac.jp

Keywords: amplicon sequencing, high-throughput sequencing, MIG-seq, non-model organism, RAD-seq, RNA-seq, target capture

ABSTRACT

Recent developments in and widespread availability of platforms, technologies, and computer software and hardware have enhanced the use of genetic markers to address major scientific questions in non-model organisms, even by researchers with limited expertise in genomics. However, there are few studies investigating the genomes of non-model insects using these approaches. This review discusses the application of next-generation sequencing (NGS) technologies to the study of genomes of wild organisms. We first introduce currently available NGS-based methods—including restriction site-associated DNA sequencing; multiplexed inter-simple sequence repeat genotyping by sequencing; target capture; and amplicon, transcriptome, and whole genome sequencing—as useful tools for studies of non-model insects. We also provide guidelines for first-time users of NGS systems. In addition to the massive amount of information that it provides, a major advantage of NGS data is the scalability to future research projects. Studies using NGS technology can answer questions related to basic entomology by focusing on the unique nature of non-model insects in wild environments in a way that is not possible for model organisms.

35 **INTRODUCTION**

36 Genomic studies of non-model species (i.e., those for which genomic information is
37 non-existent or limited) have multiplied with the emergence of next generation sequencing
38 (NGS). There are many advantages to using NGS-based methods for investigations of
39 non-model organisms (Ekblom & Galindo 2011; McCormack *et al.* 2013; Lemmon &
40 Lemmon 2013). Indeed, the levels and patterns of genetic diversity, phylogenetic
41 relationships, and genetic regions responsible for adaptation in non-model organisms have
42 been investigated by NGS (Nadeau & Jiggins 2010; Stapley *et al.* 2010).

43 Despite the availability of technology and knowledge, there have been relatively
44 few studies of non-model insects especially in Japan in the 10 years since the development of
45 NGS. The reasons for this relate to four aspects of phylogenetic and phylogeographic studies
46 (McCormack *et al*. 2013): (1) a lack of focus on non-model species in general; (2) the need
47 for a large number of samples; (3) a lack of standardized protocols; and (4) the transitional
48 state of techniques. NGS-based methods require sophisticated bioinformatics analyses, and
49 the lack of expertise in this area has constrained the application of this technology to studies
50 in entomology.

51 In this review, we introduce studies that have used NGS technologies to a wide range of

52 entomologists who are especially interested in insect taxonomy, ecology, and natural history
53 but not necessarily genomic information. Since the general methods (Davey *et al*. 2011;
54 Ekblom & Galindo 2011; Lemmon & Lemmon 2013; Jones & Good 2016) and their
55 application to phylogenetic studies (McCormack *et al*. 2013) and genomics (Rice *et al*. 2010;
56 Stapley *et al*. 2010) have already been reviewed in detail elsewhere, we focus here on
57 describing methods that can be applied to studies of non-model insects. After a brief review
58 of these methods, we provide practical examples and address the feasibility of using these
59 techniques. The reviews and references herein (Lemmon & Lemmon 2013; McCormack *et al*.
60 2013; Jones & Good 2016) provide a more detailed analysis of the advantages and caveats of
61 each approach.
62
63 *Restriction site-associated DNA sequencing (RAD-seq)*
64 RAD-seq can detect genetic markers in limited amounts of sequencing data from individuals
65 rather than from whole genome sequences (Davey *et al*. 2011, Andrews *et al*. 2016). There
66 are several variations of RAD-seq (Andrews *et al*. 2016) that differ in terms of the type,
67 number, and restriction enzyme combinations that are used. Numerous short sequence reads
68 (< 300 bp) are clustered to 'loci' based on their similarities and analyzed using bioinformatics

69 approaches (e.g., STACKS, Catchen *et al*. 2011; PYRAD, Eaton & Ree 2013). These
70 methods were first applied to linkage mapping in fish (Baird *et al*. 2008), and later used to
71 reconstruct the phylogeography of the pitcher plant mosquitos *Wyeomyia smithii* (Emerson *et*
72 *al*. 2008). Major benefits of RAD-seq are its applicability to insects for which there is no
73 reference genome information available and adaptability to different types of investigation;
74 indeed, it has been used for studies on insect population genomics (*Heliconius* butterflies,
75 Nadeau *et al*. 2013, 2014; a stick insect of the genus *Timema*, Soria-Carrasco *et al*. 2014; a
76 cricket of the genus *Teleogryllus*, Pascoal *et al*. 2014; and *Blastophaga* fig wasps, Wachi *et al*.
77 2016) and phylogenetics (*Carabus* ground beetles, Cruaud *et al*. 2014, Takahashi *et al*. 2014;
78 *Adelpha* butterflies, Ebel *et al*. 2015; and *Magicicada* periodical cicadas, Koyama *et al*.
79 2016). However, one disadvantage is the issue of missing data, which can affect the results of
80 population genetic analyses (Gautier *et al*. 2013). When genomic information is available,
81 RAD-seq can provide insight into the genetic basis of adaptive differentiation (e.g., as
82 observed for *Heliconius* butterflies, Nadeau *et al*. 2014; a stick insect of the genus *Timema*,
83 Soria-Carrasco *et al*. 2014; and a fruit fly of the genus *Rhagoletis*, Egan *et al*. 2015).
84 RAD-seq and related methods require a large amount of high-quality genomic DNA (20–150
85 ng/μl concentration, 120–900 ng in total); therefore, the use of a DNA extraction kit on

86 frozen samples preserved in absolute ethanol is recommended. The requirements in terms of
87 the quality and quantity of genomic DNA are similar to those for amplified fragment length
88 polymorphism (AFLP).
89
90 *Multiplexed inter-simple sequence repeats (ISSR) genotyping by sequencing (MIG-seq)*
91 MIG-seq is a PCR-based method used to identify large numbers of genetic markers
92 throughout the genome (Suyama & Matsuki 2015). RAD-seq and its iterations require large
93 amounts of genomic DNA; MIG-seq is an alternative approach that can yield high-resolution
94 NGS data from small amounts of low-quality genomic DNA using multiplex PCR primers
95 targeting ISSRs (Suyama & Matsuki 2015). One drawback of this method is that it identifies
96 fewer single-nucleotide polymorphisms (SNPs) than RAD-seq methods; although the number
97 of loci obtained is sufficient for population genetic studies (Suyama & Matsuki 2015;
98 Takahashi *et al*. 2016). In fact, the genetic structure of *Ischnura senegalensis* and *I. asiatica*
99 damselfly (Odonata) populations have been investigated using this method, leading to the
100 identification of 231 and 270 SNPs, respectively (Takahashi *et al*. 2016). Given its successful
101 application to studies of various taxonomic groups (e.g., mushrooms, *Pholiota microspore*;
102 copepods, *Eodiaptomus japonicas*; sea cucumbers, *Apostichopus japonicus*; sea snails,

103 *Laguncula pulchella*; Carolina anoles, *Anolis carolinensis*; and orchids, *Cypripedium*
104 *macranthos* var. *rebunense*) (Suyama & Matsuki 2015), MIG-seq—like RAD-seq—can be
105 useful for genetic studies of non-model insects.
106
107 *Target capture*
108 Target capture is a solution-based hybridization method in which genomic regions of
109 interest—such as those that are phylogenetically significant or encode functional genes—are
110 enriched using specific probes. In phylogenetic analyses, target capture uses probes that
111 hybridize to and enrich regions flanking sequences that are conserved among samples
112 (Lemmon & Lemmon 2013; Jones & Good 2016) for NGS. However, genomic information
113 must be available for target capture probe design. PCR capture, de novo assembly capture,
114 and divergent reference capture provide solutions for species for which genomic information
115 is lacking (Jones & Good 2015). These strategies differ in the number of target loci, cost
116 (mainly associated with probe synthesis), applicable taxonomic range, and requirement for
117 reference genome sequences (Jones & Good 2015). Currently, divergent reference capture
118 dominates the phylogenetic study of non-model insects. In this approach, a probe set is
119 designed for genomic sequences of interest based on the reference genome assembly of a

120 closely related species (Jones & Good 2015). An average of 863 and 721 regions from 12
121 genome-available and 30 newly investigated species of Hymenoptera, respectively, were thus
122 captured (Faircloth *et al*. 2015) using 1,510 ultra-conserved elements (UCEs; Faircloth *et al*.
123 2012) from *Apis mellifera* and *Nasonia vitripennis* genomes, yielding phylogenies that were
124 consistent with those obtained in previous studies. Based on these 1,510 UCEs in
125 Hymenoptera, 959 loci (30% missing data) were obtained across Formicinae species
126 (Blaimer *et al*. 2015) along with 944 loci (30% missing data) for 23 species of *Acropyga* ants
127 (Blaimer *et al*. 2016). Using anchored hybrid enrichment (Lemmon *et al*. 2012) similar to the
128 UCE approach, a set of 962 probes was developed based on 4,485 protein-coding genes from
129 13 insect species (Diptera, Hymenoptera, Lepidoptera, Strepsiptera, and Coleoptera, with
130 Hemiptera and Anoplura as outgroups) (Young *et al*. 2016). To investigate phylogenetic
131 relationships among syrphid species (Diptera), these investigators obtained alignments of 343
132 loci (6% missing data) from 30 Syrphidae and five dipteran species. Target capture is
133 applicable to small insect species as it enables library construction from relatively small
134 amounts of DNA (70 ng; Faircloth *et al*. 2014). Thus, even museum collections with genomic
135 DNA of limited quality and quantity can be used for phylogenomic studies (Faircloth *et al*.
136 2015).

137 Exon capture has been used to assess sequence variations at various functional loci
138 (Smadja *et al*. 2012). Using probes targeting 1,021 exon sequences from 203 loci of both
139 candidate (chemosensory) and non-candidate (randomly selected) genes, sequences from
140 samples of *Acyrthosiphon pisum* pea aphids on three different host plants were captured,
141 revealing a signature of divergent selection consisting of 128 SNPs (95% of which were in
142 chemosensory genes). As an alternative approach to target capture, organelle enrichment is a
143 feasible method for ordinary entomology using mitochondrial DNA fragments, and has been
144 reviewed in detail elsewhere (Lemmon & Lemmon 2013).
145
146 *Amplicon sequencing (amplicon-seq)*
147 NGS of PCR amplicons is a feasible method for ordinary entomology, for instance in
148 phylogeography and phylogenetics (McCormack *et al*. 2013). There are two general
149 approaches to amplicon-seq:
150 1. Uniplex PCR amplification of target regions within a sample, with tagging by several
151 base indices as in other methods and pooling of amplicons from each loci (parallel
152 tagged sequencing, Meyer *et al*. 2008, targeted amplicon-seq: Bybee *et al*. 2011).
153 Based on six loci from 44 samples, phylogenies of Pancrastacea (including Diptera,

two Coleoptera, Hymenoptera, Lepidoptera, Blattodea, Plecoptera, and Hemiptera) with well-supported nodes and few missing data were reconstructed (Bybee *et al*. 2011). The drawback of this method is that PCR is required for each locus and sample, which is time-consuming and labor-intensive when analyzing multiple samples.

2. Multiplex PCR using multiple primer sets for different loci in the same reaction. Limitations of multiplex PCR include biased amplification of specific amplicons and DNA sequence chimerism (Mamanova *et al*. 2009).

Amplicon-seq has the advantage that DNA fragments can be sequenced from multiple loci or samples without cloning; as such, it is more time- and cost-effective than Sanger sequencing. An early application of amplicon-seq to non-model insects used sequenced amplicons covering the whole mitochondrial genomes of 30 Coleoptera species (Timmermans *et al*. 2010). Instead of labeling individuals, raw reads were assembled to reconstruct contigs, which were assigned to each species using 'baits' (*cox1*, *cytb*, and *nad5*) as queries for homology searches. In this manner, near-complete mitochondrial genomes of 21 species and partial sequences from the remainder were obtained. These sequences were used to reconstruct phylogeny at the family level (Timmermans *et al*. 2010). Amplicon-seq of DNA barcode sequences provides ordinary entomologists with novel opportunities for

171 taxonomic identification and analyses of host-parasite and predator-prey interactions. General
172 approaches to DNA barcoding using NGS for ecological network analyses have been
173 previously reviewed (Toju 2015; Kishimoto & Ito 2015). For the identification of 1,010
174 individual specimens caught by Malaise trap, at least one sequence from 983 samples
175 (97.3%) was obtained (Shokralla *et al*. 2015); the order-level identification based on
176 sequence homology from 757 individuals (75.0%) was matched to morphological
177 characteristics. These proportions were greater than for sequences obtained by the Sanger
178 method (53.2% and 50.4%, respectively). To evaluate the utility of molecular data from a
179 quantitative analysis of food web structure in host-parasitoid interactions, a set of
180 lepidopteran hosts and their parasitoids (Ichneumonidae, Braconidae, and Eulophidae:
181 Hymenoptera; Tachinidae: Diptera) were examined by Sanger sequencing-based DNA
182 barcoding (Wirta *et al*. 2014); 14 of 21 host-parasitoid relationships were revealed by
183 ordinary rearing methods, and an additional 40 were newly detected interactions. Thus, DNA
184 barcoding with NGS can reveal interactions among insect species with high resolution.
185
186 *Transcriptome sequencing (RNA-seq)*
187 RNA-seq is more suitable for detecting functional genes since all obtained sequences are

188 from coding genes. The advantage of this method is that it directly connects genomic
189 information and function, unlike other genome-based methods. However, its requires
190 high-quality RNA, which poses a challenge when working with non-model insects, especially
191 field samples and from museum collections. There are various methods for obtaining
192 high-quality RNA from wild insect specimens—e.g., direct extraction from living samples;
193 flash-freezing in liquid nitrogen followed by storage at −80°C; and use of RNA-stabilizing
194 solutions such as RNAlater (Thermo Fisher Scientific) that enable extraction of high-quality
195 RNA even from samples that have been stored at 0°C for several months after collection.
196 However, even with the availability of such solutions, specimens for RNA extraction should
197 not be keep at room temperature for more than one night.
198 A database of transcriptome data from thousands of insects has been constructed
199 (1K Insect Transcriptome Evolution; http://www.1kite.org). Based on analyses of orthologous
200 loci in transcriptome sequences, several satisfactory phylogenies have been reconstructed
201 (Hymenoptera: Johnson *et al*. 2013; Insecta: Misof *et al*. 2014). The first robust phylogenetic
202 tree of Lepidoptera was reconstructed using newly obtained and available transcriptome data
203 (13 and 33 species, respectively, for a total of 2,696 genes) (Kawahara & Breinhost 2014)
204 and revealed that butterflies are a monophyletic group nested within macromoths.

205 Additionally, sequences obtained by RNA-seq can be used as references to design probes and
206 PCR primers for target capture and amplicon-seq, respectively.
207
208 *Whole genome sequencing (WGS)*
209 WGS yields very long (sometimes nearly complete) genomic DNA sequences of an organism
210 through the assembly of numerous shotgun reads that cover the genome multiple times. WGS
211 is always useful as a reference for any of the NGS techniques described above, and its
212 application to non-model organisms is garnering considerable interest, with at least four
213 suppliers in Japan currently distributing WGS platforms for non-model insects
214 The PacBio RS II sequencing system (Pacific Bioscience of California) has been
215 applied to various materials, including non-model insects (e.g., a *Belgica* antarctic midge,
216 which has one of the smallest insect genomes at 99 Mbp; Kelley *et al*. 2014). This system
217 enables the construction of adequately long scaffolds without a reference genome; however,
218 one drawback is the requirement for large amounts of genomic DNA (concentration > 50
219 ng/μl and over several micrograms in total). WGS depends on comparison to closely related
220 individuals (i.e., within the same family or ideally, genetic clones), but it is difficult to obtain
221 the required amounts of genomic DNA from small insects. Although novel platforms such as

222 with PacBio Sequel (Pacific Bioscience of California) have been developed, their efficiency
223 has not yet been validated in various non-model organisms.
224 Dovetail's Chicago Method (Dovetail Genomics LLC) consists of chromosome-scale
225 shotgun assembly based on chromatin interactions (Putnam *et al*. 2016). This system
226 provides extraordinarily long (possibly chromosome-level) scaffolds. The system is
227 completely serviced by the manufacturers from DNA extraction to genome assembly. It
228 requires relatively small amount of genomic DNA (concentration > 100 ng/μl and < 2 μg in
229 total), which can be obtained from a single individual. WGS based on the Chicago Method
230 has been applied to various non-model animals (the vampire bat *Desmodus rotundus*; the
231 prairie chicken *Tympanuchus cupido*; and the cichlid *Amphilophus citrinellus*) and plants (the
232 common onion *Allium cepa*; and the dwarf cashew tree *Anacardium occidentale*)
233 (https://dovetailgenomics.com/services/case-studies/, cited 30 Jan 2017). Although only a few
234 non-model insects have been analyzed with the system, high-resolution WGS (N50 > 3.8
235 Mbp) has been performed for Japanese *Henosepilachna* ladybird beetle (K. Matsubayashi,
236 unpublished data)..
237 The GemCode system (10X Genomics), which also yields chromosome-length
238 scaffolds based on linked reads by emulsion PCR has the advantage of requiring a small

239 quantity of genomic DNA for library preparation (< 5 ng in total). This system was optimized
240 for the human genome, but a novel Chromium platform from the same manufacturers is
241 available for general use and can be useful even for non-model insects. WGS of the Sitka
242 Spruce chloroplast genome (Coombe *et al*. 2016) has been achieved with this system.
243 The Irys system (Bionano Genomics) yields 1 Mbp-long scaffolds based on
244 single-molecule fluorescence detection by high-resolution imaging with a charge-coupled
245 device camera. The system allows whole genome mapping data to be rapidly obtained from a
246 small amount of DNA (less than several ng), and has been used for genome assembly of the
247 common bed bug *Cimex lectularius* (Rosenfeld *et al*. 2016). Additional details for each
248 system can be found on the manufacturers' web pages.
249
250 *Advantages of NGS for ordinary entomology*
251 Although it may be of interest to researchers in ordinary entomology, NGS is costly in terms
252 of equipment and time required for data analysis. Nonetheless, NGS has some advantages
253 such as rapid data acquisition, lower cost and time investment than Sanger sequencing, and
254 the large amount of genomic information obtained from the source material. Among the
255 papers published in *Entomological Science* between 2004 to 2016, 66 contained DNA

256 sequence data obtained by Sanger sequencing for various applications including taxonomy
257 (18), molecular identifications (15), population genetics (14), phylogeny (10), marker
258 development (4), phylogeography (2), molecular biology (2), and genotyping (1). Many of
259 these (44/68) used a small number of loci and samples and partial mitochondrial DNA
260 sequences. For such cases, NGS and Sanger sequencing may be similar in terms of total cost
261 (Nelson *et al*. 2012; Hodel *et al.* 2016), although less time is required to obtain sequence data
262 by NGS. Amplicon-seq or PCR capture may be more suitable than Sanger sequencing for a
263 small number of loci and/or samples, whereas RAD-seq, MIG-seq, or target capture may be
264 more appropriate for a large number of loci and/or samples (~100 and more). RNA-seq and
265 even low-coverage WGS has been useful for identifying genetic markers (e.g., a wasp of the
266 genus *Sirex*, Santana *et al*. 2009; a scuttle fly of the genus *Megaselia*, Rasmussen & Noor
267 2009; a water strider of the genus *Gerris*, Perry & Rowe 2011; and *Melitaea* butterflies,
268 Nakahama *et al*. 2015). Sequences obtained by these NGS methods can also be used to
269 design probes for target capture or primers for amplicon-seq. NGS data for specific insect
270 species acquired in current research projects can also be used as a basis for future research
271 topics (Table 1).
272 Ordinary entomology often relies on specimens in museum collections. NGS can

potentially be applied to genomic studies based on these relatively old specimens, which then become useful repositories of DNA sequences for future research (i.e., museomics; Buerki & Baker 2016 and references therein).

Preparations required for NGS

We suggest the following preparations for projects involving NGS of non-model insects (Fig. 1).

1. Consider the benefits and disadvantages of using genomic approaches for the species of interest. The greatest merit of using non-model insects is the unique nature of their presence in wild environments, which is never achievable for model organisms. Since NGS and data analysis are costly and time consuming, researchers should maximize the scientific advantages of valuable materials. The study of material-specific ecology, physiology, behavior, and natural history is a good basis for applying NGS. Similarly, there may be challenges associated with the material such as small body size, limited amounts of fresh organs, rarity of specimens, polyploidy, and large genome size. These factors may negatively affect DNA extraction, sequencing, and de novo sequence assembly. Specifically, target genome

290 size is important information to consider prior to initiating any NGS method;
291 sequencing and sequence assembly of a large genome is currently prohibitively
292 expensive. One large database of animal genome size (http://www.genomesize.com/)
293 is available (Gregory *et al*. 2007); in addition, the genome size of a specific material
294 can be easily measured by flow cytometry (Sota *et al*. 2013; Matsubayashi &
295 Ohshima 2015).
296 2. Choice of NGS method. There is no method capable of capturing data for all
297 purposes of a genomic study. Therefore, the first step is to identify an appropriate
298 NGS method and platform to address the research question. The platforms used in
299 the studies cited herein are shown in Table 2. Early case studies that applied NGS to
300 non-model insects (Ekblom & Galindo 2011) primarily used 454 (Roche FLX)
301 sequencers, but more recently, Illumina sequencers (Hiseq and Miseq) have been
302 favored (Table 2). There are several types of sequencers varying in terms of length
303 and number of reads, cost, and running time (briefly reviewed in Toju 2015).
304 3. Consider the collection and storage of tissues to be used for DNA extraction. In
305 general, DNA should be extracted from the freshest possible tissue. Although
306 extraction from living or flash-frozen specimens is the best approach, specimens

fixed with absolute ethanol are also suitable for RAD-seq, target capture, amplicon-seq, and MIG-seq. Specimens should be stored between −20°C and −80°C before extraction. Genomic DNA extraction should be performed with muscle or other clean tissue after degutting the specimen to avoid potential contamination from symbiotic microorganisms and food contents. Bioinformatics techniques are available for the removal of contaminating sequences. For examples, symbiont or contaminant sequences were identified in the genomes of *Dirofilaria* and *Caenorhabditis* nematodes (Kumar & Blaxter 2011; Kumar *et al*. 2013; Percudani 2013), a heartworm of the genus *Dirofilaria* (Godel *et al*. 2012), and a tardigrade of genus *Hypsibius* (Koutsovoulos *et al*. 2016; Delmont & Eren 2016). Storage and extraction of RNA and genomic DNA for WGS require special attention to sample quality and quantity; RAD-seq and target capture require careful handling, as do AFLP and other methods used for genomic DNA analysis; and amplicon-seq and MIG-seq require normal handling for Sanger sequencing of mitochondrial DNA.

ACKNOWLEDGMENTS

We thank Dr. M. J. Toda for giving us the opportunity to write this review. This research was

324 supported by JSPS KAKENHI (grant nos. JP15K21705 to N.W. and JP25292034 to K.M.)
325 from the Japan Society for the Promotion of Science.

326

327 **REFERENCES**

328 Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power
329 of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81–92.

330

331 Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP Discovery and Genetic Mapping
332 Using Sequenced RAD Markers. *PLoS ONE* **3**, Article ID e3376, DOI:
333 10.1371/journal.pone.0003376.

334

335 Blaimer BB, Brady SG, Schultz TR, Lloyd MW, Fisher BL, Ward PS (2015) Phylogenomic
336 methods outperform traditional multi-locus approaches in resolving deep evolutionary
337 history: a case study of formicine ants. *BMC Evolutionary Biology* **15**, 1–14.

338

339 Blaimer BB, LaPolla JS, Branstetter MG, Lloyd MW, Brady SG (2016) Phylogenomics,
340 biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*.

341 *Molecular Phylogenetics and Evolution* **102**, 20–29.

342

343 Buerki S, Baker WJ (2016) Collections□based research in the genomic era. *Biological*

344 *Journal of the Linnean Society* **117**, 5–10.

345

346 Bybee SM, Bracken-Grissom H, Haynes BD *et al.* (2011) Targeted amplicon sequencing

347 (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology*

348 *and Evolution* **3**, 1312–1323.

349

350 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) *Stacks*: building and

351 genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171–

352 182.

353

354 Coombe L, Warren RL, Jackman SD *et al*. (2016) Assembly of the complete Sitka spruce

355 chloroplast genome using 10X Genomics' GemCode sequencing data. *PLoS ONE* 11, Article

356 ID e0163059, DOI:10.1371/journal.pone.0163059.

357

358 Cruaud A, Gautier M, Galan M et al. (2014) Empirical assessment of RAD sequencing for
359 interspecific phylogeny. *Molecular Biology and Evolution* **31**, 1272–1274.
360
361 Davey JW, Hohenlohe P, Etter P, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide
362 genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews*
363 *Genetics* **12**, 499–510.
364
365 Delmont TO, Eren AM (2016) Identifying contamination with advanced visualization and
366 analysis practices: metagenomic approaches for eukaryotic genome assemblies. *PeerJ*, **4**,
367 Article ID e1839, DOI: 10.7717/peerj.1839.
368
369 Eaton DA, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an
370 example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology* **62**, 689–
371 706.
372
373 Ebel ER, DaCosta JM, Sorenson MD *et al.* (2015) Rapid diversification associated with
374 ecological specialization in Neotropical *Adelpha* butterflies. *Molecular Ecology* **24**, 2392–

375 2405.

376

377 Egan SP, Ragland GJ, Assour L *et al*. (2015) Experimental evidence of genome□wide impact

378 of ecological selection during early stages of speciation□with□gene□flow. *Ecology Letters*

379 **18**, 817–825.

380

381 Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular

382 ecology of non-model organisms. *Heredity* **107**, 1–15.

383

384 Emerson KJ, Merz CR, Catchen JM *et al*. (2010) Resolving postglacial phylogeography

385 using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the*

386 *United States of America* **107**, 16196–16200.

387

388 Faircloth BC, Branstetter MG, White ND, Brady SG (2015) Target enrichment of

389 ultraconserved elements from arthropods provides a genomic perspective on relationships

390 among Hymenoptera. *Molecular Ecology Resources* **15**, 489–501.

391

392 Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012)
393 Ultraconserved elements anchor thousands of genetic markers spanning Multiple
394 Evolutionary Timescales. *Systematic Biology* **61**, 717–726.
395
396 Gautier M, Gharbi K, Cezard T *et al*. (2013) The effect of RAD allele dropout on the
397 estimation of genetic variation within and between populations. *Molecular Ecology* **22**,
398 3165–3178.
399
400 Godel C, Kumar S, Koutsovoulos G *et al.* (2012) The genome of the heartworm, *Dirofilaria*
401 *immitis*, reveals drug and vaccine targets. *The FASEB Journal* **26**, 4650–4661.
402
403 Gregory TR, Nicol JA, Tamm H *et al.* (2007) Eukaryotic genome size databases. *Nucleic*
404 *Acids Research* **35**, 332–338.
405
406 Hodel RG, Segovia-Salcedo CM, Landis JB *et al*. (2016) The Report of my death was an
407 exaggeration: a review for researchers using microsatellites in the 21st century. *Applications*
408 *in plant sciences* **4**, Article ID 1600025. DOI: 10.3732/apps.1600025.

409

410 Johnson BR, Borowiec ML, Chiu JC, Lee EK, Atallah J, Ward PS (2013) Phylogenomics

411 resolves evolutionary relationships among ants, bees, and wasps. *Current Biology* **23**, 2058–

412 2062.

413

414 Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics.

415 *Molecular Ecology* **25**, 185–202.

416

417 Kawahara AY, Breinholt JW (2014) Phylogenomics provides strong evidence for

418 relationships of butterflies and moths. *Proceedings of the Royal Society of London B:*

419 *Biological Sciences* **281**, Article ID 20140970, DOI: 10.1098/rspb.2014.0970.

420

421 Kelley JL, Peyton JT, Fiston-Lavier A *et al*. (2014) Compact genome of the Antarctic midge

422 is likely an adaptation to an extreme environment. *Nature communications* **5**, Article ID 4611,

423 DOI: 10.1038/ncomms5611.

424

425 Kishimoto K, Ito M (2015) Studies of interaction networks among insect species: Insights

426 from DNA barcoding. *Japanese Journal of Ecology* **65**, 255–266. (In Japanese, with English
427 abstract).
428
429 Koutsovoulos G, Kumar S, Laetsch D *et al.* (2016) No evidence for extensive horizontal gene
430 transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proceedings of the National*
431 *Academy of Sciences of the United States of America* **113**, 5053–5058.
432
433 Koyama T, Ito H, Fujisawa T *et al*. (2016) Genomic divergence and lack of introgressive
434 hybridization between two 13□year periodical cicadas support life cycle switching in the face
435 of climate change. *Molecular Ecology* **25**, 5543–5556.
436
437 Kumar S, Blaxter ML (2011) Simultaneous genome sequencing of symbionts and their hosts.
438 *Symbiosis* **55**, 119–126.
439
440 Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M (2013) Blobology: exploring raw
441 genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage
442 plots. *Frontiers in Genetics* **29**, Article ID 237, DOI: 10.3389/fgene.2013.00237.

443

444 Lemmon AR, Emme SA, Lemmon EM (2012) Anchored hybrid enrichment for massively

445 high-throughput phylogenomics. *Systematic Biology* **61**, 727–744.

446

447 Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and

448 phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **44**, 99–121.

449

450 Mamanova L, Coffey A, Scott C *et al*. (2010) Target-enrichment strategies for

451 next-generation sequencing. *Nature Methods* **7**, 111–118.

452

453 Matsubayashi KW, Ohshima I (2015) Genome size increase in the phytophagous ladybird

454 beetle *Henosepilachna vigintioctomaculata* species complex (Coleoptera: Coccinellidae).

455 *Entomological Science* **18**, 134-137.

456

457 McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of

458 next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics*

459 *and Evolution* **66**, 526–538.

460

461 Meyer M, Stenzel U, Hofreiter M (2008) Parallel tagged sequencing on the 454 platform.

462 *Nature Protocols* **3**, 267–278.

463

464 Misof B, Liu S, Meusemann K *et al*. (2014) Phylogenomics resolves the timing and pattern

465 of insect evolution. *Science* **346**, 763–767.

466

467 Nadeau N, Jiggins C (2010) A golden age for evolutionary genetics? Genomic studies of

468 adaptation in natural populations. *Trends in Genetics* **26**, 484–492.

469

470 Nadeau N, Martin S, Kozak K *et al*. (2013) Genome□wide patterns of divergence and gene

471 flow across a butterfly radiation. *Molecular Ecology* **22**, 814–826.

472

473 Nadeau N, Ruiz M, Salazar P *et al*. (2014) Population genomics of parallel hybrid zones in

474 the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research* **24**, 1316–1333.

475

476 Nakahama N, Izuno A, Arima K, Isagi Y (2015) Development of microsatellite markers for

two endangered grassland butterflies, *Melitaea ambigua* and *M. protomedia* (Nymphalidae), using Ion Torrent next-generation sequencing. *Conservation Genetics Resources* **7**, 525–527.

Nelson L, Lambkin C, Batterham P *et al*. (2012) Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). *Gene* **511**, 131–142.

Pascoal S, Cezard T, Eik-Nes A *et al*. (2014) Rapid convergent evolution in wild crickets. *Current Biology* **24**, 1369–1374.

Percudani R (2013) A microbial metagenome (*Leucobacter* sp.) in *Caenorhabditis* whole genome sequences. *Bioinformatics and Biology Insights* **7**, 55–72.

Perry JC, Rowe L (2010) Rapid microsatellite development for water striders by next-generation sequencing. *Journal of Heredity* **102**, 125–129.

Putnam NH, O'Connell BL, Stites JC *et al*. (2016) Chromosome-scale shotgun assembly

494 using an in vitro method for long-range linkage. *Genome Research* **26**, 342-350.

495

496 Rasmussen DA, Noor MAF (2009) What can you do with 0.1× genome coverage? A case

497 study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). *BMC*

498 *Genomics* **10** Article ID 382, DOI: 10.1186/1471-2164-10-382.

499

500 Rosenfeld JA, Reeves D, Brugler MR *et al*. (2016). Genome assembly and geospatial

501 phylogenomics of the bed bug *Cimex lectularius*. *Nature Communications* **7** Article ID 10164,

502 DOI: 10.1038/ncomms10164.

503

504 Rice A, Rudh A, Ellegren H, Qvarnström A (2011) A guide to the genomics of ecological

505 speciation in natural animal populations. *Ecology Letters* **14**, 9–18.

506

507 Santana QC, Coetzee MPA, Steenkamp ET *et al.* (2009) Microsatellite discovery by deep

508 sequencing of enriched genomic libraries. *BioTechniques* **46**, 217–223.

509

510 Shokralla S, Porter T, Gibson J *et al*. (2015) Massively parallel multiplex DNA sequencing

511 for specimen identification using an Illumina MiSeq platform. *Scientific reports* **5**, Article ID
512 9687, DOI:10.1038/srep09687.
513
514 Smadja C, Canbäck B, Vitalis R *et al*. (2012) Large-scale candidate gene scan reveals the role
515 of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*
516 **66**, 2723–2738.
517
518 Soria-Carrasco V, Gompert Z, Comeault A *et al*. (2014) Stick insect genomes reveal natural
519 selection's role in parallel speciation. *Science* **344**, 738–742.
520
521 Sota T, Konuma J, Fujiwara M, Shoguchi E (2013) Genome sizes of three species in the
522 subtribe Carabina (Coleoptera: Carabidae). *Entomological Science* **16**, 122-124.
523
524 Stapley J, Reger J, Feulner P *et al*. (2010) Adaptation genomics: the next generation. *Trends*
525 *in Ecology & Evolution* **25**, 705–712.
526
527 Suyama Y, Matsuki Y (2015) MIG-seq: an effective PCR-based method for genome-wide

528 single-nucleotide polymorphism genotyping using the next-generation sequencing platform.
529 *Scientific Reports* **5**, Article ID 16963, DOI: 10.1038/srep16963.
530
531 Takahashi T, Nagata N, Sota T (2014) Application of RAD-based phylogenetics to complex
532 relationships among variously related taxa in a species flock. *Molecular Phylogenetics and*
533 *Evolution* **80**, 137–144.
534
535 Takahashi Y, Suyama Y, Matsuki Y, Funayama R, Nakayama K, Kawata M (2016) Lack of
536 genetic variation prevents adaptation at the geographic range margin in a damselfly.
537 *Molecular Ecology* **25**, 4450–4460.
538
539 Timmermans MJTN, Dodsworth S, Culverwell CL *et al*. (2010) Why barcode?
540 High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics.
541 *Nucleic Acids Research* **38**, Article ID e197, DOI: 10.1093/nar/gkq807.
542
543 Toju H (2015) High-throughput DNA barcoding for ecological network studies. *Population*
544 *Ecology* **57**, 37–51.

545

546 Wachi N, Kusumi J, Tzeng HY, Su ZH (2016) Genome□wide sequence data suggest the

547 possibility of pollinator sharing by host shift in dioecious figs (Moraceae, *Ficus*). *Molecular*

548 *Ecology* **25**, 5732–5746.

549

550 Wirta HK, Hebert PD, Kaartinen R, Prosser SW, Várkonyi G, Roslin T (2014)

551 Complementary molecular information changes our perception of food web structure.

552 *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1885–

553 90.

554

555 Young AD, Lemmon AR, Skevington JH *et al*. (2016) Anchored enrichment dataset for true

556 flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae).

557 *BMC Evolutionary Biology* **16**, Article ID 143, DOI: 10.1186/s12862-016-0714-0.

558

559

560 **Figure legend**

561 **Figure 1.** Work flows of NGS methods. PCR-based methods (MIG-seq and amplicon-seq)

562 can be carried out with small amounts of low-quality genomic DNA, whereas

563 fragmentation-based methods (RAD-seq, target capture, RNA-seq, and WGS) require large

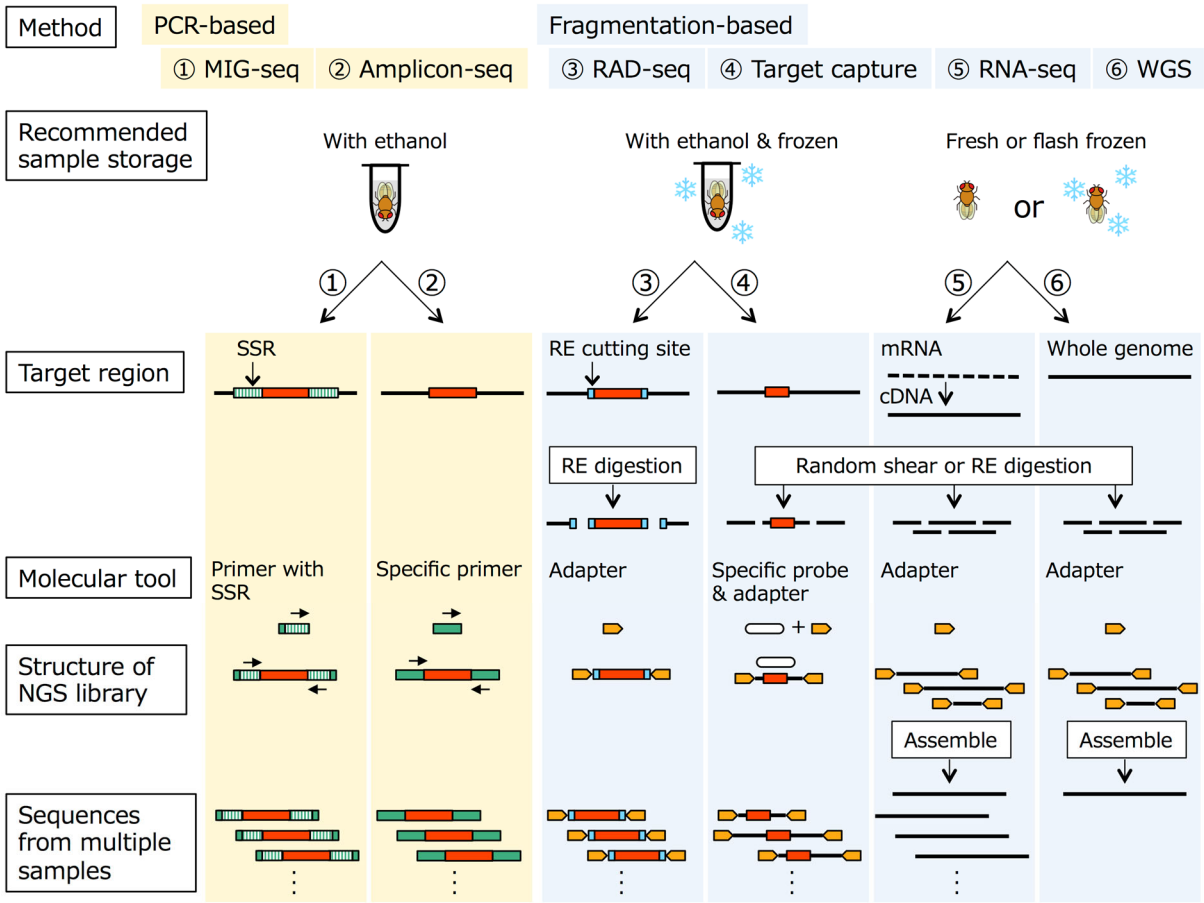
564 amounts of high-quality genomic DNA. “RAD-seq” represents double-digest RAD-seq. SSR:

565 simple sequence repeat; mRNA: messenger RNA; cDNA: complementary DNA; RE:

566 restriction enzyme.

567

568


Method
PCR-based
Fragmentation-based
① MIG-seq
② Amplicon-seq
③ RAD-seq
④ Target capture
⑤ RNA-seq
⑥ WGS
Recommended sample storage
With ethanol
With ethanol & frozen
Fresh or flash frozen
or
①
②
③
④
⑤
⑥
Target region
SSR
RE cutting site
mRNA
cDNA
Whole genome
RE digestion
Random shear or RE digestion
Molecular tool
Primer with SSR
Specific primer
Adapter
Specific probe & adapter
Adapter
Adapter
Structure of NGS library
Assemble
Assemble
Sequences from multiple samples

569

570 Table 1. The applicability of next generation sequencing methods for each research topic. Relative suitability of each application is denoted as

571 +++ (very high), ++ (high), + (moderate), and - (low), respectively. [*1] WGS can be useful for all NGS methods as a reference.

Research field	NGS method					
	RAD-seq	MIG-seq	Target capture	Amplicon-seq	RNA-seq	WGS[*1]
Phylogenomics	++	++	+++	++	+++	++
Phylogeography	+++	+++	+	+	+	++
Population genomics	+++	+++	++	+	+	++
Functional genomics	++	++	+++	++	+++	++
Museomics	+	+	++	+++	-	++
DNA barcoding	-	-	++	+++	-	++

572

573 Table 2. Case studies for application of NGS in non-model insects. RAD: Restriction site associated DNA sequencing; MIG: Multiplexed ISSR

574 genotyping by sequencing; TC: Target capture; AMP: Amplicon sequencing; RNA: RNA sequencing; LWGS: Low coverage whole genome

575 sequencing; WGS: whole genome sequencing.

Insect samples	Method	Platform	Author	Year
Wyeomyia mosquito; Diptera	RAD	GAIIx	Emerson *et al*.	2010
Heliconius butterfly; Lepidoptera	RAD	GAIIx & Hiseq	Nadeau *et al*.	2013
Heliconius butterfly; Lepidoptera	RAD	Hiseq	Nadeau *et al*.	2014
Teleogryllus cricket; Orthoptera	RAD	Hiseq	Pascoal *et al*.	2014
Timema stick insect; Phasmatodea	RAD	Hiseq	Soria-Carrasco *et al*.	2014
Carabus ground beetle; Coleoptera	RAD	Hiseq	Takahashi *et al*.	2014
Carabus ground beetle; Coleoptera	RAD	Hiseq	Cruaud *et al*.	2014
Adelpha butterfly; Lepidoptera	RAD	Hiseq	Ebel *et al*.	2015
Rhagoletis fruit fly; Diptera	RAD	GAII & Hiseq	Egan *et al*.	2015
Blastophaga fig wasp; Hymenoptera	RAD	Miseq	Wachi *et al*.	2016
Magicicada periodical cicada; Hemiptera	RAD	Hiseq	Koyama *et al*.	2016
Ischnura damselfly; Odonata	MIG	Hiseq	Takahashi *et al*.	2016
Acyrthosiphon pea aphid; Hemiptera	TC	454	Smadja *et al*.	2012
Hymenoptera	TC	Miseq	Faircloth *et al.*	2015
formicine ant; Hymenoptera	TC	Hiseq	Blaimer *et al.*	2015

Acropyga ant; Hymenoptera	TC	Hiseq	Blaimer *et al.*	2016
syrphid flower fly; Diptera	TC	Hiseq	Young *et al.*	2016
Coleoptera	AMP	454	Timmermans *et al*.	2010
Pancrastacea (including Diptera, two Coleoptera, Hymenoptera, Lepidoptera, Blattodea, Plecoptera, Hemiptera)	AMP	454	Bybee *et al*.	2011
Coleoptera, Diptera, Hemiotera, Hymenoptera, Lepidoptera, Psocoptera etc.	AMP	Miseq	Shokralla *et al*.	2015
Hymenoptera	RNA	Hiseq	Johnson *et al.*	2013
Lepidoptera	RNA	Hiseq	Kawahara & Breinhost	2013
Insecta	RNA	Hiseq	Misof *et al.*	2014
Sirex wasp; Hymenoptera	LWGS	454	Santana *et al*.	2009
Megaselia scuttle fly; Diptera	LWGS	454	Rasmussen & Noor	2009
Gerris water strider; Hemiptera	LWGS	454	Perry & Rowe	2011
Melitaea butterfly; Lepidoptera	LWGS	Ion PGM	Nakahama *et al*.	2015
Belgica antarctic midge; Diptera	WGS	Hiseq & PacBio RSII	Kelly *et al*.	2014
Cimex bed bug; Hemiptera	WGS	Hiseq, Moleculo & Irys	Rosenfeld *et al*.	2016

576

577