How consistent is RAD-seq divergence with DNA-barcode based clustering in insects?

Marie Cariou, Hélène Henri, Sonia Martinez, Laurent Duret, Sylvain Charlat

How consistent is RAD-seq divergence with DNA-barcode based clustering in insects?

Marie Cariou[1,2], Hélène Henri[1], Sonia Martinez[1], Laurent Duret[1] & Sylvain Charlat[1*].

[1] Université de Lyon, Université Claude Bernard Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne France

[2] Current address: Université de Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR5023 LEHNA, F-69622, Villeurbanne, France

*Corresponding author: sylvain.charlat@univ-lyon1.fr

Running tittle: Assessing DNA barcoding in insects

Key words: RAD-seq, divergence, barcode, mitochondria

Abstract

Promoted by the barcoding approach, mitochondrial DNA is more than ever used as a molecular marker to distinguish and identify species. Yet, it has been repeatedly argued that it may be poorly suited for this purpose, especially in insects where mitochondria are often associated with invasive intracellular bacteria that may promote their introgression. Here we inform this debate by assessing how divergent nuclear genomes can be when mitochondrial barcodes indicate very high proximity. To this end, we obtained RAD-seq data from 92 barcode-based species-like units (that is Operational Taxonomic Units, OTU), spanning 4 insect orders. In 100% of the cases, the observed median nuclear divergence was lower than 2%, a value that was recently estimated as one below which nuclear gene flow is not uncommon. These results suggest that although mitochondria may occasionally leak between species, this process is rare enough in insects to make DNA barcoding a reliable tool for clustering specimens into species-like units.

Mitochondrial DNA is widely used to track population histories and identify species, with a number of pros and cons making this issue controversial (Galtier, Nabholz, Glémin, & Hurst, 2009; Hebert, Hollingsworth, & Hajibabaei, 2016; Hebert, Stoeckle, Zemlak, & Francis, 2004; Hurst & Jiggins, 2005; Moritz & Cicero, 2004; Ratnasingham & Hebert, 2007; Sloan, Havird, & Sharbrough, 2017). Beyond their technical ease of use, mitochondrial sequences possess characteristics that make them particularly suitable for short evolutionary time scale inferences. Thanks to a high, albeit variable, mutation rate (Allio, Donega, Galtier, & Nabholz, 2017), they tend to efficiently discriminate closely related lineages. Their short expected coalescence time, due to maternal inheritance, should also make them less susceptible than nuclear sequences to the persistence of ancestral polymorphism in diverging populations, that is, incomplete lineage sorting. On the other hand, because they lack recombination and are genetically linked with invasive cytoplasmic elements such as *Wolbachia* (Werren, Baldo, & Clark, 2008), mitochondria might be particularly prone to selective sweeps (Cariou, Duret, & Charlat, 2017). Such events may also facilitate introgression between incipient species following hybridizations, a process that introduces disagreements between demographic and genetic genealogies (Hurst & Jiggins 2005; Galtier et al. 2009). Several reports have indeed demonstrated cases of *Wolbachia*-driven mitochondrial introgressions (Charlat et al., 2009; Jiggins, 2003; Narita, Nomura, Kato, & Fukatsu, 2006), and *Wolbachia* infected species generally harbour a reduced mitochondrial polymorphism (Cariou et al., 2017). Mitochondrial DNA may also be subject to much wider variations in mutation rates and evolve much less neutrally, than has been traditionally acknowledged (Allio et al., 2017; Galtier et al., 2009; Sloan et al., 2017). Because of these problematic issues, cautions have been repeatedly raised regarding the use of mitochondrial markers to infer evolutionary histories (Chan & Levin, 2005; Galtier et al.,

2009; Hurst & Jiggins, 2005; Schmidt & Sperling, 2008; Sloan et al., 2017; Taylor & Harris, 2012; Towes & Brelsford, 2012). Should these valid criticisms be taken as good reasons for not relying on mtDNA to cluster unknown specimens into species-like units?

Whatever the markers considered, the frequency of gene flow is known to decrease with increasing divergence time. Roux et al. (2016) analysed this effect by exploring the prevalence of gene flow along the continuum of divergence leading to complete isolation. This study was based on transcriptome data from 61 pairs of populations (or closely related species), showing variable levels of divergence, and selected in order to sample evenly the phylogenetic and ecological diversity of animals. The analysis revealed a "grey zone" between 0.5% and 2% of net synonymous divergence (defined as the synonymous divergence measured from non-polymorphic sites), within which species definition is often controversial and where alleles can be exchanged at some but not all loci. However, above a threshold of about 2%, they observe that gene flow is suppressed and that species are indeed isolated. Are mitochondrial markers specific to this regard, often moving between lineages beyond the "grey zone"? Or is their bad reputation a caricature, as much as would be the idyllic picture of a "perfect" molecular marker ?

To address this question, we used RAD-seq, recognized as a robust "genome reduction approach" (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016), to estimate of the genome-wide nuclear divergences in 92 pairs of insect specimens, each composed of two individuals that presumably belonged to the same species, or "Operational Taxonomic Units" (OTUs) as indicated by previously obtained DNA barcodes (figure 1). These specimens, spanning 4 insects orders (figure 2) represent a phylogenetically diverse subset of a larger sample of Arthropods collected in Tahiti and surrounding islands (Ramage et al., 2017). In seven OTUs,

82 the RAD-seq experiment was repeated for one of the two specimens, in order to estimate the
83 reliability of the data. In all cases, the global nuclear divergence, as estimated by the median
84 RAD distance, were indeed very similar among replicates (S2 table).

85 The observed median RAD divergence ranged from 0% to 1.1% (Figure 3, S3 Table),
86 that is, remained within the standard range of within species polymorphism. In 75% of the
87 presumed species (69 cases) the median nuclear divergence was below 0.5%, that is, showed
88 a typical value for specimens collected in randomly mixed populations (Roux et al., 2016). In
89 the remaining 25% (23 cases), the median RAD distance remained well below 2%, that is, within
90 the "grey zone of speciation", suggesting the two specimens sampled likely belong to slightly
91 isolated or large populations (Roux et al., 2016). Notably, this general picture of small nuclear
92 distances is further reinforced by noting that, without polymorphism data in hands, we used
93 here raw divergence, which is inflated in comparison with the "net divergence" used by Roux
94 et al. (2016).

95 Arguably, mitochondrial DNA is not the ideal molecular marker. Even though it has a
96 small effective population size, a high mutation rate, and usually doesn't recombine, it is
97 certainly not evolving as neutrally as would be required to make it suitable for any kind of
98 evolutionary or demographic inferences. Some mitochondrial sites are likely under direct
99 selection, and perhaps more importantly, mitochondria are genetically linked with maternally
100 inherited elements, such as *Wolbachia* bacteria that tend to rapidly invade populations thanks
101 to selfish drive strategies. For these reasons, some have argued that "mtDNA is perhaps
102 intrinsically the worst population genetic and phylogenetic molecular marker we can think of"
103 (Galtier et al., 2009). Yet, the Barcoding Of Life Database, mostly based on the mitochondrial
104 locus CO1, hosts over 7 million records at the time of writing, from an estimated 600 000

105 species, suggesting it still has a future in the field, at least as a species assignment and
106 delimitation tool.

107 How can we reconcile the above mentioned critical views with this massive usage of
108 mitochondrial DNA? First, we may object that many of the criticisms would apply to any single
109 locus, be it sampled from the nuclear or the cytoplasmic genome. Our data says that a very
110 small mitochondrial distance is indicative of a very small global nuclear divergence in 92 out of
111 92 cases, suggesting that in insects, mitochondrial DNA does not so often depart from the
112 common genomic history. Of course, a deeper survey would likely reveal exceptions, but this
113 would be true for any randomly picked nuclear locus that may sometimes fall at the edge of
114 the overall distance distribution.

115 Second, it seems important to distinguish, among the many questions relevant to
116 evolutionary biology and systematics, those that can be addressed at low risk with mtDNA from
117 those that cannot. Here we answered one specific question: are very small mtDNA distances
118 indicative of very small overall nuclear distances? In other words, how can we trust barcode-
119 based clustering of specimens into species? “Very much” is the answer suggested by our
120 random sample of 92 barcode-based OTUs. However, we did not assess all possible sources of
121 cyto-nuclear discordance (Toews & Brelsford, 2012). For example, our data is not designed for
122 testing if specimens harbouring distant mtDNA are always similarly distant at the nuclear level,
123 which may not be the case in species subject to incomplete introgressions making them
124 paraphyletic at the mtDNA level (Charlat et al., 2009; Funk & Omland, 2003; Hurst & Jiggins,
125 2005). We did not either assess how much barcode-based genealogies are reliable, which is
126 well known to depend on the timescale considered, giving more or less weight to signal
127 saturation or introgression. Neither did we assess how often mtDNA may be subject to selective

sweeps, making them unreliable indicators of demographic histories. It may also be argued that our conclusion holds only for the taxonomic groups surveyed, that is, four orders of insects. However, insects are notoriously prone to *Wolbachia* infections, which may make them more prone than others to cytonuclear discordance, but likely not less.

In brief, there is no doubt that mtDNA is far from a perfect marker (Galtier et al., 2009). Yet, some specific objectives, such as the clustering of specimens into species-like units, can be attained at low cost and low risk with mitochondrial DNA barcodes. These markers are less appropriate for other categories of questions, that should generally be addressed not with a single locus, be it cytoplasmic or nuclear, but with genome wide data.

Material and Methods

Samples and RAD-seq library preparation

To compare mitochondrial and nuclear estimates of divergence, we selected 544 specimens distributed in 261 mitochondrial OTUs from the SymbioCode system, a large sample of Arthropods collected in French Polynesia, previously subject to DNA barcoding (Ramage et al., 2017). We obtained reliable data from 92 OTUs falling in 4 insect orders (Diptera, Hemiptera, Hymenoptera and Lepidoptera), as detailed in table S1. Importantly, the choice of OTUs and specimens was blind to the *Wolbachia* infection status, although it had been previously determined (Bailly-Bechet et al., 2017). In order to obtain RAD-seq data, DNA extracts were digested using Sbf1 restriction enzyme and distributed in 3 libraries (table S1). Seven samples

were analysed twice, in order to assess the accuracy of the average nuclear divergence estimates. Molecular Identifiers (MIDs) were designed using the barcrawl program (Frank, 2009) and prepared according to Henri et al (2015). A mixture of 8 and 10 bp-MIDs was used to avoid sequence homogeneity at the restriction site that disturbs base calling (Krueger, Andrews, & Osborne, 2011) and 15% of phiX DNA were also added to mitigate low complexity issues. Sequence data was acquired in three experiments. The first library was sequenced using the HiSeq 2000 system at the « Génomique & Microgénomique » platform of Lyon1 University (ProfileXpert), producing 100 bp-long reads. The second and third libraries were sequenced using the HiSeq 3000 system at the Genotoul platform (Toulouse), producing 150 bp-long reads, with additional PCR amplification cycles to reach more suitable DNA quantities. We aimed at obtaining paired-end reads in all experiments, but only forward reads were obtained in one experiment, which was thus excluded from our analysis.

RAD-seq data analysis

Reads were demultiplexed using the process_radtag program from the Stacks software pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013), allowing for one mismatch in the molecular identifier or restriction site sequence. We used paired-end reads to identify and remove PCR duplicates, using a filterPCRdupl.pl script from the ConDeTri program (Smeds & Künstner, 2011). Forward reads were then clustered into loci using pyRAD (Eaton, 2014). Minimal identity (the threshold for inferring that two reads belong to the same locus) was set to 95%. Minimal coverage (the minimal number of reads in a cluster for it to be defined as a valid locus) was set to five, following preliminary analyses that showed lower thresholds

173 introduced errors that inflated the estimated genetic distances. Additional filters were used to
174 identify spurious reads containing adaptors and MID sequences, as well as those showing
175 significant blast hits with bacterial and archaeal genomes (as present in Ensembl Genome in
176 October 2014) that represent potential contaminants. Notably, these additional filters removed
177 only 4% of the data. Homologous loci shared between two specimens within an OTU were
178 identified using SiLiX (Miele, Penel, & Duret, 2011), with clustering thresholds of 35% minimum
179 identity and 80% minimum overlap. Clusters including more than one locus per species, that is,
180 paralogous loci, were excluded.
181

182 Data accessibility

183

184 The mtDNA data used in this paper have been made accessible through earlier publications.
185 The RAD-seq will be made available upon on the pbil repository hosted by University of Lyon 1
186 at the following address: ftp://pbil.univ-lyon1.fr/pub/datasets/
187

188 Authors' contributions

189

190 MC designed and conducted the experiments, analyzed de the data and wrote the
191 manuscript; HH conducted the experimental setup and contributed to analyses and writing;
192 SM contributed to the experiments and writing; LD designed the experiments and analyses

and contributed to writing; SC designed the experiments and analyses and contributed to writing.

Competing interest

We have no competing interest.

Funding

This work was supported by the Centre National de la Recherche Scientifique (CNRS) (ATIP grant SymbioCode to S.C.).

Acknowledgments

This work benefitted from the computing facilities of the CC LBBE/PRABI. Sequencing reactions were performed in collaboration with the ProfileXpert (University Claude Bernard Lyon 1) and GeT core facilities (Toulouse, France, http://get.genotoul.fr) which is supported by the "France Génomique" National infrastructure and the Agence Nationale pour la Recherche (contract ANR-10-INBS-09).

References

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, *msx197*. doi:10.1093/molbev/msx197

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

Bailly-Bechet, M., Simoes, P., Szöllősi, G., Mialdea, G., Sagot, M.-F., & Charlat, S. (2017). How long does Wolbachia remain on board? *Molecular Biology and Evolution*, *34*, 1183–1193. doi:10.1093/molbev/msx073

Cariou, M., Duret, L., & Charlat, S. (2017). The global impact of Wolbachia on mitochondrial diversity and evolution. *Journal of Evolutionary Biology*, *30*, 2204–2210.

Catchen, J., Hohenlohe, P. a, Bassham, S., Amores, A., & Cresko, W. a. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–40. doi:10.1111/mec.12354

Chan, K., & Levin, S. (2005). Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, *59*(4), 720–729.

Charlat, S., Duplouy, A., Hornett, E. A. A., Dyson, E. A. A., Davies, N., Roderick, G. K. K., … Hurst, G. D. D. (2009). The joint evolutionary histories of Wolbachia and mitochondria in Hypolimnas bolina. *BMC Evol Biol*, *9*(1), 64. doi:1471-2148-9-64 [pii]10.1186/1471-2148-9-64

237 Eaton, D. A. R. (2014). PyRAD : assembly of de novo RADseq loci for phylogenetic analyses,
238 *30*(13), 1844–1849. doi:10.1093/bioinformatics/btu121
239 Frank, D. N. (2009). BARCRAWL and BARTAB: software tools for the design and
240 implementation of barcoded primers for highly multiplexed DNA sequencing. *BMC*
241 *Bioinformatics*, *10*, 362. doi:10.1186/1471-2105-10-362
242 Funk, D. J., & Omland, K. E. (2003). Species-Level Paraphyly and Polyphyly: Frequency, Causes,
243 and Consequences, with Insights from Animal Mitochondrial DNA. *Annual Review of*
244 *Ecology, Evolution, and Systematics*, *34*(1), 397–423.
245 doi:10.1146/annurev.ecolsys.34.011802.132421
246 Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of
247 molecular diversity : a reappraisal. *Molecular Ecology*, *18*, 4541–4550.
248 doi:10.1111/j.1365-294X.2009.04380.x
249 Hebert, P. D. N., Hollingsworth, P. M., & Hajibabaei, M. (2016). From writing to reading the
250 encyclopedia of life. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
251 *371*(1702). doi:10.1098/rstb.2015.0321
252 Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds
253 through DNA barcodes. *PLoS Biology*, *2*(10). doi:10.1371/journal.pbio.0020312
254 Henri, H., Cariou, M., Terraz, G., Martinez, S., el Filali, A., Veyssiere, M., … Charlat, S. (2015).
255 Optimization of multiplexed RADseq libraries using low-cost adaptors. *Genetica*, *143*(2),
256 139–143. doi:10.1007/s10709-015-9828-3
257 Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in
258 population, phylogeographic and phylogenetic studies: the effects of inherited
259 symbionts. *Proc Biol Sci*, *272*(1572), 1525–1534. doi:10.1098/rspb.2005.3056

260 Jiggins, F. M. (2003). Male-Killing Wolbachia and Mitochondrial DNA. Selective sweeps, hybrid
261 introgression and parasite population dynamics. *Genetics*, *164*(1), 5–12.
262 Krueger, F., Andrews, S. R., & Osborne, C. S. (2011). Large scale loss of data in low-diversity
263 illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS ONE*,
264 *6*(1), 4–10. doi:10.1371/journal.pone.0016607
265 Miele, V., Penel, S., & Duret, L. (2011). Ultra-fast sequence clustering from similarity networks
266 with SiLiX. *BMC Bioinformatics*, *12*(1), 116. doi:10.1186/1471-2105-12-116
267 Moritz, C., & Cicero, C. (2004). DNA Barcoding: Promise and Pitfalls. *PLoS Biology*, *2*(10), e354.
268 doi:10.1371/journal.pbio.0020354
269 Narita, S., Nomura, M., Kato, Y., & Fukatsu, T. (2006). Genetic structure of sibling butterfly
270 species affected by Wolbachia infection sweep : evolutionary and biogeographical
271 implications. *Molecular Ecology*, *15*, 1095–1108. doi:10.1111/j.1365-294X.2006.02857.x
272 Ramage, T., Martins-Simoes, P., Mialdea, G., Allemand, R., Duplouy, A. M. R., Rousse, P., …
273 Charlat, S. (2017). A DNA barcode-based survey of terrestrial arthropods in the Society
274 Islands of French Polynesia: host diversity within the SymbioCode project. *European*
275 *Journal of Taxonomy*, *272*(272), 1–13. doi:10.5852/ejt.2017.272
276 Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD : The Barcode of Life Data System
277 (www.barcodinglife.org). *Molecular Ecology Notes*, *7*, 355–364. doi:10.1111/j.1471-
278 8286.2006.01678.x
279 Roux, C., Fra, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding Light on
280 the Grey Zone of Speciation along a Continuum of Genomic Divergence, 1–22.
281 doi:10.1371/journal.pbio.2000234
282 Schmidt, B. C., & Sperling, F. A. H. (2008). Widespread decoupling of mtDNA variation and

283 species integrity in Grammia tiger moths (Lepidoptera: Noctuidae). *Systematic*
284 *Entomology*, *33*(4), 613–634. doi:10.1111/j.1365-3113.2008.00433.x
285 Sloan, D., Havird, J., & Sharbrough, J. (2017). The on-again , off-again relationship between
286 mitochondrial genomes and species boundaries. *Molecular Ecology*, *26*, 2212–2236.
287 doi:10.1111/mec.13959
288 Smeds, L., & Künstner, A. (2011). CONDETRI - A Content Dependent Read Trimmer for
289 Illumina Data. *PloS One*, *6*(10), e26314.
290 Taylor, H. R., & Harris, W. E. (2012). An emergent science on the brink of irrelevance: A review
291 of the past 8years of DNA barcoding. *Molecular Ecology Resources*, *12*(3), 377–388.
292 doi:10.1111/j.1755-0998.2012.03119.x
293 Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear
294 discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. doi:10.1111/j.1365-
295 294X.2012.05664.x
296 Towes, D., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance
297 in animals. *Molecular Ecology*, *21*, 3907–3930. doi:10.1111/j.1365-294X.2012.05664.x
298 Werren, J. H., Baldo, L., & Clark, M. E. (2008). Wolbachia: master manipulators of invertebrate
299 biology. *Nature Reviews. Microbiology*, *6*(10), 741–51. doi:10.1038/nrmicro1969
300
301 Bailly-Bechet, M., Simoes, P., Szöllősi, G., Mialdea, G., Sagot, M.-F., & Charlat, S. (2017). How
302 long does *Wolbachia* remain on board? *Molecular Biology and Evolution*, *34*, 1183–1193.
303 doi:10.1093/molbev/msx073
304 Cariou, M., Duret, L., & Charlat, S. (2017). The global impact of *Wolbachia* on mitochondrial
305 diversity and evolution. *Journal of Evolutionary Biology*, *30*, 2204–2210.

306 Catchen, J., Hohenlohe, P. a, Bassham, S., Amores, A., & Cresko, W. a. (2013). Stacks: an
307 analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–40.
308 doi:10.1111/mec.12354
309 Chan, K., & Levin, S. (2005). Leaky prezygotic isolation and porous genomes: rapid introgrssion
310 of maternally inherited DNA. *Evolution*, *59*(4), 720–729.
311 Charlat, S., Duplouy, A., Hornett, E. A. A., Dyson, E. A. A., Davies, N., Roderick, G. K. K., …
312 Hurst, G. D. D. (2009). The joint evolutionary histories of *Wolbachia* and mitochondria in
313 Hypolimnas bolina. *BMC Evol Biol*, *9*(1), 64. doi:1471-2148-9-64 [pii]10.1186/1471-2148-
314 9-64
315 Eaton, D. A. R. (2014). PyRAD : assembly of de novo RADseq loci for phylogenetic analyses,
316 *30*(13), 1844–1849. doi:10.1093/bioinformatics/btu121
317 Frank, D. N. (2009). BARCRAWL and BARTAB: software tools for the design and
318 implementation of barcoded primers for highly multiplexed DNA sequencing. *BMC*
319 *Bioinformatics*, *10*, 362. doi:10.1186/1471-2105-10-362
320 Funk, D. J., & Omland, K. E. (2003). Species-Level Paraphyly and Polyphyly: Frequency, Causes,
321 and Consequences, with Insights from Animal Mitochondrial DNA. *Annual Review of*
322 *Ecology, Evolution, and Systematics*, *34*(1), 397–423.
323 doi:10.1146/annurev.ecolsys.34.011802.132421
324 Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of
325 molecular diversity : a reappraisal. *Molecular Ecology*, *18*, 4541–4550.
326 doi:10.1111/j.1365-294X.2009.04380.x
327 Hebert, P. D. N., Hollingsworth, P. M., & Hajibabaei, M. (2016). From writing to reading the
328 encyclopedia of life. *Philosophical Transactions of the Royal Society B: Biological Sciences*,

329 *371*(1702). doi:10.1098/rstb.2015.0321
330 Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds
331 through DNA barcodes. *PLoS Biology*, *2*(10). doi:10.1371/journal.pbio.0020312
332 Henri, H., Cariou, M., Terraz, G., Martinez, S., el Filali, A., Veyssiere, M., … Charlat, S. (2015).
333 Optimization of multiplexed RADseq libraries using low-cost adaptors. *Genetica*, *143*(2),
334 139–143. doi:10.1007/s10709-015-9828-3
335 Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in
336 population, phylogeographic and phylogenetic studies: the effects of inherited
337 symbionts. *Proc Biol Sci*, *272*(1572), 1525–1534. doi:10.1098/rspb.2005.3056
338 Jiggins, F. M. (2003). Male-Killing *Wolbachia* and Mitochondrial DNA. Selective sweeps, hybrid
339 introgression and parasite population dynamics. *Genetics*, *164*(1), 5–12.
340 Krueger, F., Andrews, S. R., & Osborne, C. S. (2011). Large scale loss of data in low-diversity
341 illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS ONE*,
342 *6*(1), 4–10. doi:10.1371/journal.pone.0016607
343 Miele, V., Penel, S., & Duret, L. (2011). Ultra-fast sequence clustering from similarity networks
344 with SiLiX. *BMC Bioinformatics*, *12*(1), 116. doi:10.1186/1471-2105-12-116
345 Moritz, C., & Cicero, C. (2004). DNA Barcoding: Promise and Pitfalls. *PLoS Biology*, *2*(10), e354.
346 doi:10.1371/journal.pbio.0020354
347 Narita, S., Nomura, M., Kato, Y., & Fukatsu, T. (2006). Genetic structure of sibling butterfly
348 species affected by *Wolbachia* infection sweep : evolutionary and biogeographical
349 implications. *Molecular Ecology*, *15*, 1095–1108. doi:10.1111/j.1365-294X.2006.02857.x
350 Ramage, T., Martins-Simoes, P., Mialdea, G., Allemand, R., Duplouy, A. M. R., Rousse, P., …
351 Charlat, S. (2017). A DNA barcode-based survey of terrestrial arthropods in the Society

352 Islands of French Polynesia: host diversity within the SymbioCode project. *European*
353 *Journal of Taxonomy*, *272*(272), 1–13. doi:10.5852/ejt.2017.272
354 Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD : The Barcode of Life Data System
355 (www.barcodinglife.org). *Molecular Ecology Notes*, *7*, 355–364. doi:10.1111/j.1471-
356 8286.2006.01678.x
357 Roux, C., Fra, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding Light on
358 the Grey Zone of Speciation along a Continuum of Genomic Divergence, 1–22.
359 doi:10.1371/journal.pbio.2000234
360 Schmidt, B. C., & Sperling, F. A. H. (2008). Widespread decoupling of mtDNA variation and
361 species integrity in Grammia tiger moths (Lepidoptera: Noctuidae). *Systematic*
362 *Entomology*, *33*(4), 613–634. doi:10.1111/j.1365-3113.2008.00433.x
363 Sloan, D., Havird, J., & Sharbrough, J. (2017). The on-again , off-again relationship between
364 mitochondrial genomes and species boundaries. *Molecular Ecology*, *26*, 2212–2236.
365 doi:10.1111/mec.13959
366 Smeds, L., & Künstner, A. (2011). CONDETRI - A Content Dependent Read Trimmer for
367 Illumina Data. *PloS One*, *6*(10), e26314.
368 Taylor, H. R., & Harris, W. E. (2012). An emergent science on the brink of irrelevance: A review
369 of the past 8years of DNA barcoding. *Molecular Ecology Resources*, *12*(3), 377–388.
370 doi:10.1111/j.1755-0998.2012.03119.x
371 Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear
372 discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930. doi:10.1111/j.1365-
373 294X.2012.05664.x
374 Towes, D., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance

375 in animals. *Molecular Ecology*, *21*, 3907–3930. doi:10.1111/j.1365-294X.2012.05664.x

376 Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate

377 biology. *Nature Reviews. Microbiology*, *6*(10), 741–51. doi:10.1038/nrmicro1969

378

379

Figures

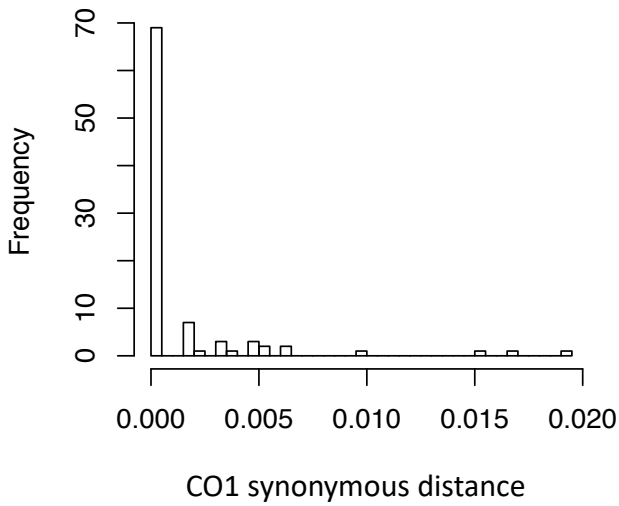


Figure 1. Distribution of the CO1 distance within each of the 92 OTUs under study, computed from 3rd codon positions using data from Ramage et al (2017).

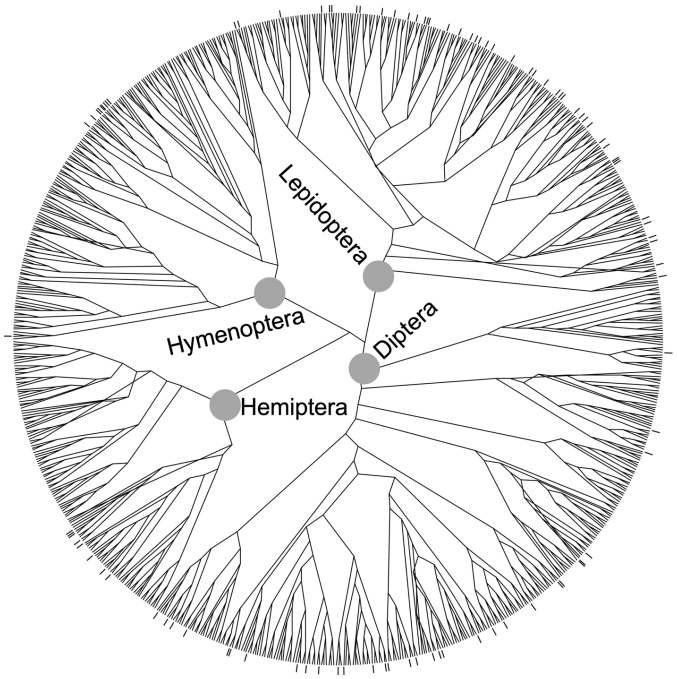
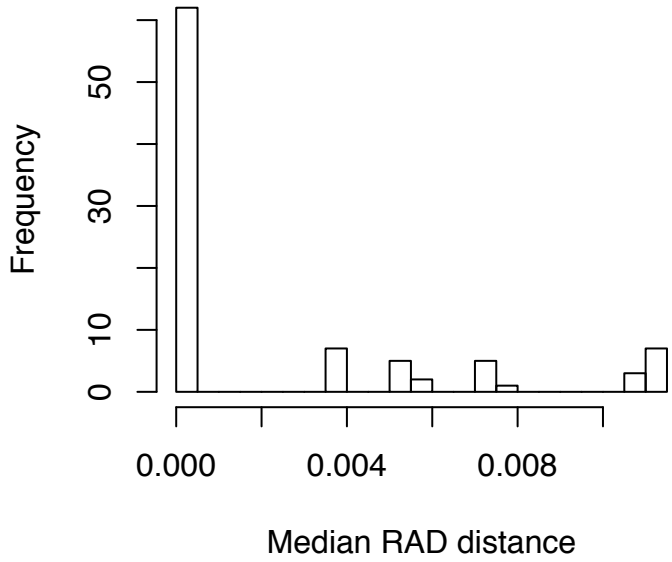


Figure 2. Phylogenetic diversity of our sample (92 OTUs) in comparison with the local diversity. The full tree includes one representative of all Lepidoptera, Diptera, Hemiptera and Hymenoptera OTUs from a previously reported extensive survey of the Arthropods from Tahiti and surrounding islands (Ramage et al., 2017). Finer scale taxonomic details can be found in table S1. Vertical bars on the surface indicate the species included in the present analysis.

396



397

398 **Figure 3.** Distribution of median RAD distances within the 92 OTUs under study.