Pushing the limits of whole genome amplification: successful sequencing of RADseq libraries from single micro-hymenoptera (Chalcidoidea, Trichogramma)

Astrid Cruaud, Geraldine Groussier, Guénaëlle Genson, Laure Saune, Jean Yves Rasplus

Pushing the limits of whole genome amplification: Successful sequencing of RADseq libraries from single micro-hymenoptera (Chalcidoidea, *Trichogramma*)

Astrid Cruaud [Corresp., 1], **Géraldine Groussier** [2], **Guenaëlle Genson** [1], **Laure Sauné** [1], **Jean-Yves Rasplus** [1]

[1] CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France

[2] Institut Sophia Agrobiotech, INRA, CNRS, Université Côte d'Azur, Sophia Antipolis, France

Corresponding Author: Astrid Cruaud
Email address: astrid.cruaud@inra.fr

A major obstacle to high-throughput genotyping of micro-hymenoptera is their small size. As species are difficult to discriminate and because complexes may exist, the sequencing of a pool of specimens is hazardous. Thus, one should be able to sequence pangenomic markers (e.g. RADtags) from a single specimen. To date, whole genome amplification (WGA) prior to library construction is still a necessity as only *ca* 10ng of DNA can be obtained from single specimens. However this amount of DNA is not compatible with manufacturer's requirements for commercialised kits. Here we tested the accuracy of the GenomiPhi kit V2 on *Trichogramma* wasps by comparing RAD libraries obtained from the WGA of single specimens (generation F0 and F1, *ca* 1 ng input DNA for the WGA) and a biological amplification of genomic material (the pool of the progeny of the F1 generation). Globally, we found that *ca* 99% of the examined loci (up to 48,189; 109 bp each) were compatible with the mode of reproduction of the studied model (haplodiploidy) or a Mendelian inheritance of alleles. The remaining 1% (ca 0.01% of the analysed nucleotides) could represent WGA bias or other experimental / analytical bias. This study shows that the multiple displacement amplification method on which the GenomiPhi kit relies, could also be of great help for the high-throughput genotyping of micro-hymenoptera used for biological control or other organisms from which only a very low amount of DNA can be extracted such as human disease vectors (e.g. sand flies, fleas, ticks etc.).

Pushing the limits of whole genome amplification: successful sequencing of RADseq libraries from single micro-hymenoptera (Chalcidoidea, *Trichogramma*)

Astrid Cruaud[1]*, Géraldine Groussier[2], Guenaëlle Genson[1], Laure Sauné[1], and Jean-Yves Rasplus[1]

[1] CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France

[2] Institut Sophia Agrobiotech, INRA, CNRS, Université Côte d'Azur - 400, route des chappes – BP 167 - 06903 Sophia Antipolis Cedex – France

*corresponding author : astrid.cruaud@inra.fr

Abstract

A major obstacle to high-throughput genotyping of micro-hymenoptera is their small size. As species are difficult to discriminate and because complexes may exist, the sequencing of a pool of specimens is hazardous. Thus, one should be able to sequence pangenomic markers (e.g. RADtags) from a single specimen. To date, whole genome amplification (WGA) prior to library construction is still a necessity as only *ca* 10ng of DNA can be obtained from single specimens. However this amount of DNA is not compatible with manufacturer's requirements for commercialised kits. Here we tested the accuracy of the GenomiPhi kit V2 on *Trichogramma* wasps by comparing RAD libraries obtained from the WGA of single specimens (generation F0 and F1, *ca* 1 ng input DNA for the WGA) and a biological amplification of genomic material (the pool of the progeny of the F1 generation). Globally, we found that *ca* 99% of the examined loci (up to 48,189; 109 bp each) were compatible with the mode of reproduction of the studied

model (haplodiploidy) or a Mendelian inheritance of alleles. The remaining 1% (ca 0.01% of the analysed nucleotides) could represent WGA bias or other experimental / analytical bias. This study shows that the multiple displacement amplification method on which the GenomiPhi kit relies, could also be of great help for the high-throughput genotyping of micro-hymenoptera used for biological control or other organisms from which only a very low amount of DNA can be extracted such as human disease vectors (e.g. sand flies, fleas, ticks etc.).

Running Head : RADseq from low DNA amount through WGA

Keywords : DNA quantity, GenomiPhi, high-throughput genotyping, microarthropods, outside manufacturer recommendations, RAD.

INTRODUCTION

Parasitoid wasps (especially Chalcidoidea; Heraty et al. 2013) are increasingly used as biocontrol agents of many crop pests to reduce pesticide use (Austin et al. 2000). Among them, minute wasps of the genus *Trichogramma* (210 species worldwide, 40 in Europe), which develop within the eggs of ca. 200 species of moths damaging crops (e.g. corn, grapes, apple, pines; Consoli et al. 2010) are the most commercialized worldwide.

It is acknowledged that successful and safe biological control depends on accurate genetic and phenotypic characterization of the strains released. Furthermore, host preferences and the potential of strains to hybridize with each other or with native species should be carefully studied. This is critical to avoid non-target effects such as gene introgression with indigenous species (Van Driesche & Hoddle 2016). However, probably because most species of chalcids are minute wasps (less than a few millimetres long) and are difficult to identify to species by non-specialists, strains are often released without in depth characterization.

RADseq, the sequencing of hundreds of thousands of DNA fragments flanking restriction sites (Miller et al. 2007) has been successfully used for population genetics or phylogeography (Emerson et al. 2010), to infer relationships between closely (Jones et al. 2013; Nadeau et al. 2013; Wagner et al. 2013) or more distantly (Cruaud et al. 2014; Hipp et al. 2014) related species, to detect hybridization processes (Eaton & Ree 2013; Hohenlohe et al. 2011), to identify markers under selection and detect genes that are candidates for phenotype evolution (Hohenlohe et al. 2010), or to better understand the genomic architecture of reproductive isolation (Gagnaire et al. 2013). Thus, sequencing RAD markers appear relevant for in depth characterisation of *Trichogramma* strains used in biocontrol.

A major obstacle to RAD sequencing of oophagous parasitoids is their small size. Ideally, one should be able to sequence RAD markers from a single specimen. Indeed, species complexes may exist that are difficult to identify based on morphology only (Al Khatib et al. 2014; Kenyon et al. 2015; Mottern & Heraty 2014), which makes sequencing of a pool of specimens risky. However, to date, the DNA amount obtained from single specimens is not sufficient enough to build a RADseq library. Usually, for minute specimens, *ca* 10 ng are obtained while *ca* 150ng input DNA are required to build a RADseq library. Performing whole genome amplification (WGA) prior to library construction is thus a necessity. So far a few studies have formally examined the accuracy of WGA methods, mostly on human DNA and either a few loci (Hosono et al. 2003; Lovmar et al. 2003; Sun et al. 2005) or a higher number of SNPs and loci but always with 10ng of more input DNA (Abulencia et al. 2006; Barker et al. 2004; Blair et al. 2015; ElSharawy et al. 2012; Paez et al. 2004; Pinard et al. 2006). All studies have concluded that the multiple displacement amplification method (MDA; Dean et al. 2002; Lasken 2009), which relies on isothermal DNA amplification using a high-fidelity polymerase (bacteriophage phi29; Paez et al. 2004) and random hexamer primers to decrease amplification bias and increase product size, is among the most accurate.

So far, only one study has quantified sequence bias that might result from WGA prior to double-digest RAD sequencing (ddRADseq; Peterson et al. 2012); a variant of RADseq that uses two restriction enzymes to cut DNA instead of one enzyme and a DNA shearing system. In their study, Blair et al. (2015) use the Qiagen REPLI-g Mini Kit and 100 ng of input DNA (as requested by the kit) extracted from liver samples of specimens of the grey mouse lemur (*Microcebus murinus*). They conclude that the kit does not introduce bias for i) SNP calling as compared to what is obtained from native DNA of the same samples or ii) genome coverage as

compared to the published genome of *M. murinus*. Here we test the accuracy of the GE Healthcare Life Sciences™ illustra™ GenomiPhi V2 for the WGA of single *Trichogramma* wasps prior to RADseq library construction. As for the REPLI-g Mini Kit, WGA is performed using the MDA. However, the GenomiPhi kit requires 10 times less DNA (1 µl of input DNA at 10ng/ µl) but still more than what can be extracted from single *Trichogramma* wasps. As a consequence, we had to push the limits of the kit, increasing the risk of inconsistent or not representative amplification of the genome. To test the accuracy of the GenomiPhi kit in these challenging conditions we compared RADtags obtained from the WGA of single individuals and RADtags obtained from the pool of their progeny (Fig 1). Thus, we compared RAD libraries obtained from a technical / artificial amplification (WGA) and a biological / natural amplification (pool of specimens).

MATERIALS AND METHODS

Sampling and experimental design

Males and females (F0) were taken from the strain collection hosted by the Biological Resource Centre "Egg Parasitoid Collection" (EP-Coll, Sophia-Antipolis, France) (Marchand et al. 2017). All specimens belong to the species *Trichogramma brassicae* Bezdenko, 1968. Ten crossing experiments were attempted (Fig. 1). Ten pairs (F0) of male (haploid) / female (diploid) were sorted out from the rest of the specimens. Each pair was then reared in individual glass tubes and left free to mate. Droplets of honey were provided as food and eggs of *Ephestia kuehniella* (Pyralidae) were used as hosts. Females and males were killed in ethanol 70% before emergence of the F1 generation. Emerging females of the F1 generation were kept separated from males (no

mating) and reared in new glass tubes (1 female per tube). As for F0, droplets of honey were provided as food and eggs of *Ephestia kuehniella* were used as hosts. Females F1 were killed before the emergence of the F2 generation (males only, arrhenotokous parthenogenesis). For each cross, all males F2 were pooled prior to DNA extraction.

DNA extraction and whole genome amplification

DNA extraction was performed with the Qiagen DNeasy 96 Blood & Tissue Kit, following manufacturer protocol with the following modifications to increase DNA yield: two successive elutions (50 µL each) were performed with heated buffer AE (55°C) and an incubation step of 15 minutes followed by plate centrifugation (6000 rpm for 2 minutes).

DNA was quantified with a Qubit® 2.0 Fluorometer (Invitrogen). To fit as much as possible the recommendations of the GenomiPhi protocol (1 µl DNA input at 10 ng/µl), ethanol precipitation of DNA was performed prior to WGA. 1/10 volume of sodium acetate 3M pH 5.2 was added to the extract. Then, 2 volumes of cooled absolute ethanol were added to the mix. The mix was incubated at -20°C overnight. The mix was then centrifuged (30 min, 13 000 rpm, 4°C) and the pellet was washed with 500 µl of cooled ethanol 70%. After another centrifugation (15 min, 13 000 rpm, 4°C), the pellet was dried at room temperature and resuspended in 4 µl of sterile molecular biology ultrapure water, as a total resuspension of the pellet would not have been obtained in a smaller volume. Concentrate DNA was quantified with a Qubit® 2.0 Fluorometer (Invitrogen). DNA extracts were then subjected to Whole Genome Amplification using the GenomiPhi™ V2 DNA Amplification kit (GE Healthcare) with 1ul of concentrate DNA used as input. The resulting DNA was quantified with a Qubit® 2.0 Fluorometer (Invitrogen).

RADseq library construction

Library construction followed Baird et al. (2008) and Etter et al. (2011) with modifications detailed in Cruaud et al. (2014). The *PstI* enzyme was chosen as cutter. The number of expected cut sites was estimated with an *in silico* digestion of the genome of *T. pretiosum* (assembly Tpre_1.0, 196Mb) using a custom script. The experiment to test the accuracy of WGA for RADseq of micro-hymenoptera is part of a larger project that aims at resolving the phylogenetic relationships of European *Trichogramma* wasps. Thus more samples (N=40) than what was used to answer our technical question were included in the library. About 250ng of input DNA was used for each sample. The quantity of P1 adapters (100nM) to be added to saturate restriction sites (result=3uL) as well as the optimal time for DNA sonication on a Covaris S220 ultrasonicator to obtain fragments of 300 – 600 bp (results = duty cycle 10%, intensity 5, cycles/burst 200, duration 70s) that are both specific to the studied group were evaluated in a preliminary experiment. After tagging with barcoded P1 adapters and prior to sonication, samples were pooled eight by eight. Five pools were thus obtained, and each pool was sheared and then tagged with a different barcoded P2 adapter. 2*125nt paired-end sequencing of the library was performed at MGX-Montpellier GenomiX on one lane of an Illumina HiSeq 2500 flow cell.

Data analysis

Cleaning of raw data was performed with the wrapper RADIS (Cruaud et al. 2016) that relies on Stacks (Catchen et al. 2013; Catchen et al. 2011) for demultiplexing of data and removing PCR duplicates. Data analysis was performed with Stacks v1.46. Individual loci were built using *ustacks* [m=3 ; M=1 ; N=2 ; with the removal (r) and deleveraging (d) algorithms enabled]. Catalogs of loci were built with *cstacks* (n=2) for each of the four crosses. Females F1 and their

progeny were analysed first (4 data sets), other catalogs grouping all specimens involved in the cross were built (4 other data sets). *sstacks* was used to map individual loci to the catalog. *rxstacks* was then used to correct genotype and haplotype calls : i) Loci for which at least 50% of the samples (when a pair composed of one female F1 and a pool of males F2 was analysed) or 25% of the samples (when F0, F1 and F2 were analysed together) had a confounded match to the catalog were removed; ii) excess haplotypes were pruned; iii) SNPs were recalled after removal of possible sequencing errors using the bounded SNP model (--bound_high 0.1), and iv) loci with an average log likelihood less than -10.0 were discarded. After this filtering step, *cstacks* and *sstacks* were rerun. The program *populations* was then used to compare the RADtags obtained with or without WGA (parsing of the *haplotypes.tsv* and *populations.log* files). Loci were kept only if i) they had a minimum stack depth of 10 and ii) all samples had a sequence. Analyses were performed on the Genotoul Cluster (INRA, Toulouse).

RESULTS

On the ten attempted crosses, only three leaded to enough F2 males (N > 100) to get a sufficient amount of DNA for RADseq library construction without WGA. Consequently RADseq libraries were constructed only on these crosses. DNA extraction of one third of the tested specimens provided an amount of DNA that stand below the detection limit of the Qubit (Table 1). WGA was not attempted on these specimens. For other specimens, the average amount of DNA obtained with the Qiagen kit was 10.4 ng (min = 6.2 – max = 13.9) (Table1). After DNA re-concentration, the average DNA quantity used as input for the WGA was *ca* 1.0 ng (0.17 – 2.9). In average, 947.5 ng of DNA was obtained with the WGA (226 - 2393).

172 *In silico* digestion of the genome of *T. pretiosum* revealed 59,433 *PstI* cut sites (i.e. 118,866
173 tags). An average of 2*3,757,867 reads (109 bp) was obtained for the different samples after
174 quality filtering, demultiplexing and removal of PCR clones (Table 1). Two females F1
175 (TRIC00027_1103 and TRIC00027_3103) were represented by much less reads than other
176 samples (595,204 and 1,991,305 respectively). The number of tags recovered by *ustacks* and
177 *cstacks* varied but was comparable among the samples and in line with the predictions made on
178 the genome of *T. pretiosum* when these two females were excluded from calculation (average
179 number of ustacks tags = 132,787; average number of cstacks tags = 128,293, Table1).
180 The comparison of the loci obtained after filtering steps with *rxstacks* and *populations* revealed
181 that, in average, 97.6% of the loci were homozygous and identical for females F1 and the pool of
182 males F2 (min=96.8% - max=98.2%, Table 2). In average, 0.7% (0.3%-1.3%) of the loci were
183 heterozygous and identical in both samples. Thus there was a *ca* 98.3% (97.4%-99.0%) exact
184 match between the loci of the females F1 included in the library (and whose DNA was amplified
185 with WGA) and the whole progeny of the F1 generation (pool of males whose DNA was not
186 amplified).
187 Between 1.0 and 2.6% of the loci were not identical between analysed pairs (Table 2). A careful
188 inspection of the haplotypes revealed that ca 60% of these differences could be explained by the
189 experimental setup, *i.e.* the sequencing of a single female of the F1 generation versus the
190 sequencing of the whole progeny of the F1 generation (pool of males heterozygous, female F1
191 homozygous with an allele present in the pool of males; pool of males with three alleles, female
192 F1 homozygous with an allele present in the pool of males; pool of males with three alleles,
193 female F1 heterozygous with two alleles present in the pool of males). About 40% of the
194 observed SNPs were not compatible with either the experimental setup (pool of males and

195 female F1 homozygous but with different alleles; female F1 heterozygous, pool of males
196 homozygous with an allele present in the female F1; female F1 heterozygous, pool of males
197 homozygous with an allele not present in the female F1; pool of males heterozygous, female F1
198 homozygous with an allele not present in the pool of males; females F1 and pool of males
199 heterozygous with only one of the two alleles in common. Globally, 99.3% (98.9 - 99.6%) of the
200 shared loci were either identical or displayed differences that could be explained by the
201 experimental setup. The first cross was used to check in details the overall coherence of the
202 haplotypes from the parental generation to the whole progeny of the F1 generation (Table 3).
203 98.8% of the 32,913 loci shared by the four samples displayed haplotypes consistent with
204 experimental the setup and a Mendelian inheritance of alleles (97.3% being homozygous and
205 identical between samples). SNPs observed in 385 loci (which represent 1.2% of the loci and
206 0.01% of the analysed nucleotides) were not compatible with the mode of reproduction of the
207 studied model (haplodiploidy) or a Mendelian inheritance of alleles. Considering the haplotype
208 observed in the pool of males F2 as a reference, questionable SNPs could be categorized into
209 five categories as listed in Table 4. In *ca* 90% of the situations, SNPs found either in the male F0
210 (21%), the female F0 (32.5%) or the female F1 (36.6%) were incompatible with haplodiploidy or
211 with a Mendelian inheritance of alleles. 96 cases (ca 25.0%) represented situations where one
212 allele was missing for the female F0 or the female F1 to fit with a Mendelian inheritance of
213 alleles (possible cases of allele drop-out).
214

215 **DISCUSSION**

216 Here we compare RAD libraries obtained from a technical / artificial amplification of DNA
217 (WGA of single specimen of micro-hymenoptera, F0 and F1 generations) and a biological /

218 natural amplification (pool of the progeny of the F1 generation). We push the limits of the kit
219 used for the WGA (GenomiPhi) by using ca 90% less DNA (ca. 1.0ng) than the required amount
220 specified on the manufacturer's protocol (10ng). Globally, we show that 99% of the examined
221 loci (up to 48,189; 109 bp each) were compatible with haplodiploidy and either identical among
222 specimens or compatible with a Mendelian inheritance of alleles. These results are consistent
223 with observations by Blair et al. (2015) who used the Qiagen REPLI-g Mini Kit and 100 ng of
224 input DNA and showed that SNP calling between ddRAD libraries from native and amplified
225 DNA presented a > 98% match (up to 11,309 loci examined). They are also in agreement with
226 older studies that attempted to quantify bias induced by multiple displacement amplification
227 method (MDA) on which the GenomiPhi kit relies (> 99% match; Barker et al. 2004; ElSharawy
228 et al. 2012; Paez et al. 2004), though with more input DNA (10ng).
229 To the exception of two samples, for which the construction of the library seems to have failed
230 (much less reads were obtained), comparable numbers of tags were obtained. This indicates that
231 the coverage of the genome is the same regardless if native or amplified DNA is used as
232 suggested by previous studies on the potential bias induced by MDA (Abulencia et al. 2006;
233 Blair et al. 2015; Paez et al. 2004). Studies have suggested that WGA may induce allele dropout
234 especially when the starting amount of DNA is low (<1ng) (Handyside et al. 2004; Lovmar et al.
235 2003; Lovmar & Syvänen 2006; Sun et al. 2005). ElSharawy et al. (2012) and Blair et al. (2015)
236 concluded that MDA had no significant effect on levels of homozygosity. Here about 1% of the
237 loci retained by our analytical pipeline (ie ca 0.01% of the examined nucleotides) presented
238 problematic SNPs that were not compatible with the biology of *Trichogramma* wasps or a
239 Mendelian inheritance of alleles. 0.3% of the SNPs were possible cases of allele drop-out (one
240 allele was missing for the female F0 or the female F1 to fit with a Mendelian inheritance of

alleles). A larger sampling would be required to examine these few problematic SNPs in more details. Here, a correlation may exist between the number of problematic SNPs and the quantity of input DNA used for the WGA [less bias in haploid male F0 (1.5 ng; 81 problematic SNPs; 0.002% of the examined nucleotides) as compared to female F0 (0.39 ng; 125; 0.003%) and female F1 (0.35 ng; 141; 0.004%)] but no definite conclusion can be drawn. It is noteworthy that if these problematic SNPs can indeed result from bias caused by WGA, other explanations are possible (competition between fragments for ligation of P1 or adapters, mutation during enrichment PCR, sequencing error). Indeed, although they are less frequent, bias are also observed in the pool of males F2 (29 problematic SNPs; 0.0002% of the examined nucleotides). Regarding the possible improvements of our protocol. Extraction failed for a third of our specimens (especially single males that are much smaller than females). Here we used the Qiagen kit 96-well-plate format in order to facilitate the processing of many specimens at a time. However, especially for precious specimens, DNA yield could be increased with the spin-column format, as higher centrifuge speed could be used. Furthermore, for projects that aim to target a high number of specimens, re-concentration on SPRI beads may be used instead of using ethanol precipitation of DNA. Indeed, such methods are compatible with robotic sample preparation. However, while DNA yield could be better, working with very low amount of buffer to resuspend DNA could be troublesome.

CONCLUSION

In this study we pushed the limits of the GenomiPhi kit V2 and successfully built RADseq libraries from single micro-wasps (*Trichogramma*). Globally, we found that *ca* 99% of the examined loci (up to 48,189; 109 bp each) were compatible with the mode of reproduction of the

studied model (haplodiploidy) and/or a Mendelian inheritance of alleles. The remaining 1% (ca 0.01% of the analysed nucleotides) could represent WGA bias or other experimental / analytical bias. It is noteworthy that the GenomiPhi kit V2 (and the new GenomiPhi kit V3) are affordable and easy to use by the vast majority of laboratories, which is an important point to consider given the increasing demand for the genomic characterisation of parasitoids used in biocontrol programs or other disease-transmitting micro-arthropods (e.g. sand flies, fleas, ticks etc.).

ACKNOWLEDGEMENTS

We are grateful to Montpellier GenomiX (Montpellier, France) for sequencing of the library and to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing computing resources. This work was funded by the ANR project TriPTIC (ANR-14-CE18-0002).

REFERENCES

Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie EL, Hazen TC, and Keller M. 2006. Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Applied and Environmental Microbiology*:3291-3301.

Al Khatib F, Fusu L, Cruaud A, Gibson G, Borowiec N, Ris N, Rasplus J-Y, and Delvare G. 2014. An integrative approach for species discrimination in the *Eupelmus urozonus* complex (Hymenoptera, Eupelmidae), with the descriptions of eleven new species from the West Palaearctic. *Systematic Entomology* 39:806-862.

Austin AD, Dowton M, and (Eds.). 2000. *Hymenoptera: Evolution, Biodiversity and Biological Control*: CSIRO PUBLISHING.

287 Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA,
288 and Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD
289 markers. *PLoS ONE* 3:e3376.
290 Barker DL, Hansen MS, Faruqi AF, Giannola D, Irsula OR, Lasken RS, Latterich M, Makarov
291 V, Oliphant A, Pinter JH, Shen R, Sleptsova I, Ziehler W, and Lai E. 2004. Two methods
292 of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage
293 panel. *Genome Research* 14:901-907.
294 Blair C, Campbell CR, and Yoder AD. 2015. Assessing the utility of whole genome amplified
295 DNA for next-generation molecular ecology. *Molecular Ecology Resources* 15:1079-
296 1090.
297 Catchen J, Hohenlohe PA, Bassham S, Amores A, and Cresko WA. 2013. Stacks: an analysis
298 tool set for population genomics. *Molecular Ecology* 22:3124-3140.
299 Catchen JM, Amores A, Hohenlohe P, Cresko W, and Postlethwait JH. 2011. Stacks: Building
300 and Genotyping Loci De Novo From Short-Read Sequences. *G3-Genes Genomes*
301 *Genetics* 1:171-182. 10.1534/g3.111.000240
302 Consoli FL, Parra JRP, Zucchi RA, and (Eds.). 2010. *Egg parasitoids in agroecosystems with*
303 *emphasis on Trichogramma*: Springer, Dordrecht.
304 Cruaud A, Gautier M, Galan M, Foucaud J, Sauné L, Genson G, Dubois E, Nidelet S, Deuve T,
305 and Rasplus J-Y. 2014. Empirical assessment of RAD sequencing for interspecific
306 phylogeny. *Molecular Biology and Evolution* 31:1272-1274.
307 Cruaud A, Gautier M, Rossi J-P, Rasplus J-Y, and Gouzy J. 2016. RADIS: Analysis of RAD-seq
308 data for InterSpecific phylogeny *Bioinformatics* 32:3027-3028.

309 Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J,
310 Driscoll M, Song W, Kingsmore SF, Egholm M, and Lasken RS. 2002. Comprehensive
311 human genome amplification using multiple displacement amplification. *Proceedings of*
312 *the National Academy of Sciences of the United States of America* 99:5261-5266.
313 Eaton DAR, and Ree RH. 2013. Inferring Phylogeny and Introgression using RADseq Data: An
314 Example from Flowering Plants (Pedicularis: Orobanchaceae). *Systematic Biology*
315 62:689-706. 10.1093/sysbio/syt032
316 ElSharawy A, Warner J, Olson J, Forster M, Schilhabel MB, Link DR, Rose-John S, Schreiber S,
317 Rosenstiel P, Brayer J, and Franke A. 2012. Accurate variant detection across non-
318 amplified and whole genome amplified DNA using targeted next generation sequencing.
319 *BMC Genomics* 13:500.
320 Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, and Holzapfel
321 CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing.
322 *Proceedings of the National Academy of Sciences of the United States of America*
323 107:16196-16200. 10.1073/pnas.1006538107
324 Etter PD, Bassham S, Hohenlohe PA, Johnson EA, and Cresko WA. 2011. SNP discovery and
325 genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo V, and
326 Rockman MV, eds. *Molecular methods for evolutionary genetics*: Humana Press, New
327 York., 157-178.
328 Gagnaire PA, Normandeau E, Pavey SA, and Bernatchez L. 2013. Mapping phenotypic,
329 expression and transmission ratio distortion QTL using RAD markers in the Lake
330 Whitefish (*Coregonus clupeaformis*). *Molecular Ecology* 22:3036-3048.

331 Handyside AH, Robinson MD, Simpson RJ, Omar MB, Shaw MA, Grudzinskas JG, and
332 Rutherford A. 2004. Isothermal whole genome amplification from single and small
333 numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease.
334 *Molecular Human Reproduction* 10:767-772.
335 Heraty JM, Burks RA, Cruaud A, Gibson GAP, Liljeblad J, Munro J, Rasplus J-Y, Delvare G,
336 Janšta P, Gumovsky A, Huber J, Woolley JB, Krogmann L, Heydon S, Polaszek A,
337 Schmidt S, Darling DC, Gates MW, Mottern J, Murray E, Dal Molin A, Triapitsyn S,
338 Baur H, Pinto JD, van Noort S, George J, and Yoder M. 2013. A phylogenetic analysis of
339 the megadiverse Chalcidoidea (Hymenoptera). *Cladistics* 29:466-542.
340 Hipp AL, Eaton DAR, Cavender-Bares J, Fitzek E, Nipper R, and Manos PS. 2014. A
341 framework phylogeny of the american oak clade based on sequenced RAD data. *PLoS*
342 *ONE* 9:e93975. 10.1371/journal.pone.0093975
343 Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, and Luikart G. 2011. Next-generation
344 RAD sequencing identifies thousands of SNPs for assessing hybridization between
345 rainbow and westslope cutthroat trout. *Molecular Ecology Resources* 11 (Suppl. 1):117-
346 122.
347 Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, and Cresko WA. 2010. Population
348 genomics of parallel adaptation in threespine stickleback using sequenced RAD tags.
349 *PloS Genetics* 6:e1000862.
350 Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, and
351 Lasken RS. 2003. Unbiased whole-genome amplification directly from clinical samples.
352 *Genome Research* 13:954-964.

353 Jones JC, Fan SH, Franchini P, Schartl M, and Meyer A. 2013. The evolutionary history of
354 *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using
355 restriction site-associated DNA sequencing. *Molecular Ecology* 22:2986-3001.
356 10.1111/mec.12269
357 Kenyon SG, Buerki S, Hansson C, Alvarez N, and Benrey B. 2015. Uncovering cryptic
358 parasitoid diversity in *Horismenus* (Chalcidoidea, Eulophidae). *PLoS ONE* 10:e0136063.
359 Lasken RS. 2009. Genomic DNA amplification by the multiple displacement amplification
360 (MDA) method. *Biochemical Society Transactions* 37:450.
361 Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, and Syvänen AC. 2003. Quantitative
362 evaluation by minisequencing and microarrays reveals accurate multiplexed SNP
363 genotyping of whole genome amplified DNA. *Nucleic Acids Research* 31.
364 Lovmar L, and Syvänen AC. 2006. Multiple displacement amplification to create a long-lasting
365 source of DNA for genetic studies. *Human Mutation* 27:603-614.
366 Marchand M, Sellier N, Warot S, Ion Scotta M, Ris N, and Groussier-Bout G. 2017.
367 Formalisation d'un Centre de Ressources Biologiques dédié aux parasitoïdes oophages :
368 CRB Ep-Coll. *Cahier des Techniques de l'INRA* Special Volume "Innovations
369 entomologiques : du laboratoire au champ":49-58.
370 Miller MR, Dunham JP, Amores A, Cresko WA, and Johnson EA. 2007. Rapid and cost-
371 effective polymorphism identification and genotyping using restriction site associated
372 DNA (RAD) markers. *Genome Research* 17:240-248.
373 Mottern JL, and Heraty JM. 2014. Revision of the *Cales noacki* species complex (Hymenoptera,
374 Chalcidoidea, Aphelinidae). *Systematic Entomology* 39:354-379.

375 Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, Baxter SW,
376 Blaxter ML, Mallet J, and Jiggins CD. 2013. Genome-wide patterns of divergence and
377 gene flow across a butterfly radiation. *Molecular Ecology* 22:814-826. 10.1111/j.1365-
378 294X.2012.05730.x
379 Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, Gabriel S, Herman P, Sasaki H,
380 Altshuler D, Li C, Meyerson M, and Sellers WR. 2004. Genome coverage and sequence
381 fidelity of φ29 polymerase-based multiple strand displacement whole genome
382 amplification. *Nucleic Acids Research* 32:e71. doi.org/10.1093/nar/gnh069
383 Peterson BK, Weber JN, Kay EH, Fisher HS, and Hoekstra HE. 2012. Double digest RADseq:
384 an inexpensive method for de novo SNP discovery and genotyping in model and non-
385 model species. *PLoS ONE* 7:e37135.
386 Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg
387 JM, and Leamon JH. 2006. Assessment of whole genome amplification-induced bias
388 through high-throughput, massively parallel whole genome sequencing. *BMC Genomics*
389 7:216.
390 Sun G, Kaushal R, Pal P, Wolujewicz M, Smelser D, Cheng H, Lu M, Chakraborty R, Jin L, and
391 Deka R. 2005. Whole-genome amplification: relative efficiencies of the current methods.
392 *Legal Medicine* 7:279-286.
393 Van Driesche R, and Hoddle MS. 2016. Non-target effects of insect biocontrol agents and trends
394 in host specificity since 1985. *CAB Reviews* 44.
395 Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, and Seehausen
396 O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species

boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22:787-798.

DATA ACCESSIBILITY

Cleaned reads are available as a NCBI Sequence Read Archive (#SRP136713).

AUTHOR CONTRIBUTIONS

A.C. and J.-Y.R. designed the study, analysed the data and wrote the manuscript. G. Groussier reared *Trichogramma* strains. G. Genson and L.S. performed molecular work. All authors commented on the manuscript.

TABLES AND FIGURES

Figure 1. Experimental setup

Photo *Trichogramma brassicae* © J.-Y. Rasplus

Table 1. Extraction, whole genome amplification and sequencing results

* reads obtained after demultiplexing and quality filtering with *process_radtags*; ** reads obtained after removal of PCR clones (input reads for the *ustacks* step); *** one catalog was built for each cross

Table 2. Pairwise comparison of loci obtained for females of the F1 generation and pools of males of the F2 generation. Analysed loci have been first corrected by *rxstacks* for genotype

and haplotype calls and filtered with *populations*. Only loci that were present in the two samples with a stack depth of 10 were kept.

Table 3. Comparison of loci obtained for the first crossing experiment. Analysed loci have been first corrected by *rxstacks* for genotype and haplotype calls and filtered with *populations*. Only loci that were present in the four samples with a stack depth of 10 were kept.

Table 4. Categories of SNPs not compatible with the mode of reproduction of the studied model (haplodiploidy) or a Mendelian inheritance of alleles and number of occurrences of each case. The different situations are illustrated by examples taken from the analysis of the 385 questionable SNPs.

Table 1(on next page)

Extraction, whole genome amplification and sequencing results

* reads obtained after demultiplexing and quality filtering with *process_radtags*; ** reads obtained after removal of PCR clones (input reads for the *ustacks* step); *** one catalog was built for each cross

Table 1

Cross #	Sample code	Description	qDNA (ng)	Input DNA for WGA (ng)	Output DNA from WGA (ng)	Input DNA for RAD library (ng)	Demultiplexed Reads* (forward only)	Cleaned reads** (forward only)	*ustacks* : Nb of loci	*ustacks* : Nb of loci***
1	TRIC00027_2101	Male F0, haploid, WGA	11.5	1.5	500	169.0	6,774,680	5,173,711	136,623	130,060
1	TRIC00027_2102	Female F0, diploid, WGA	10.6	0.39	1048	203.3	4,073,370	3,179,891	128,212	122,860
1	TRIC00027_2103	Female F1, diploid, WGA	6.20	0.35	2393	281.2	4,597,505	3,566,986	130,565	124,845
1	TRIC00027_2199	Pool of haploid males F2 (n=933), no WGA	735.4	N.A.	N.A.	269.0	4,818,385	3,745,752	127,709	125,047
2	TRIC00027_1101	Male F0, haploid, WGA	Too low	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
2	TRIC00027_1102	Female F0, diploid, WGA	Too low	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
2	TRIC00027_1103	Female F1, diploid, WGA	9.4	0.17	1128	164.6	774,450	595,204	43,763	42,062
2	TRIC00027_1199	Pool of haploid males F2 (n=229), no WGA	359.6	N.A.	N.A.	270.6	5,878,301	4,437,984	127,380	125,302
3	TRIC00027_3101	Male F0, haploid, WGA	Too low	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
3	TRIC00027_3102	Female F0, diploid, WGA	13.9	2.9	390	247.7	7,006,980	5,365,499	147,718	140,690
3	TRIC00027_3103	Female F1, diploid, WGA	10.7	0.43	226	137.3	2,616,330	1,991,305	93,606	89,543
3	TRIC00027_3199	Pool of haploid males F2 (n=1415), no WGA	670.7	N.A.	N.A.	228.4	7,510,904	5,764,475	131,267	129,249

Table 2(on next page)

Pairwise comparison of loci obtained for females of the F1 generation and pools of males of the F2 generation

Analysed loci have been first corrected by *rxstacks* for genotype and haplotype calls and filtered with *populations*. Only loci that were present in the two samples with a stack depth of 10 were kept.

Table 2

Pair of samples	Nb of shared loci	Percentage of identical loci (homozygous)	Percentage of identical loci (heterozygous)	Percentage of loci with differences possibly explained by the experimental setup	Percentage of loci with differences not explained by the experimental setup
Cross #1 Female F1 x pool of males F2	48,189	97.7	1.3	0.6	0.4
		Total percentage of identical loci 99.0		Total percentage of loci with differences 1.0	
Cross #2 Female F1 x pool of males F2	5,184	96.8	0.6	1.5	1.1
		Total percentage of identical loci 97.4		Total percentage of loci with differences 2.6	
Cross #3 Female F1 x pool of males F2	20,095	98.2	0.3	0.9	0.6
		Total percentage of identical loci 98.5		Total percentage of loci with differences 1.5	

Table 3(on next page)

Comparison of loci obtained for the first crossing experiment.

Analysed loci have been first corrected by *rxstacks* for genotype and haplotype calls and filtered with *populations*. Only loci that were present in the four samples with a stack depth of 10 were kept.

1 **Table 3**

Studied samples	Nb of shared loci	Percentage of identical loci (homozygous)	Percentage of loci consistent with the experimental setup and Mendelian inheritance of alleles	Percentage of loci not consistent with the experimental setup and Mendelian inheritance of alleles
Cross #1 - female F0 & male F0 - one female F1 - progeny of the F1 generation	32,913	97.3	98.8	1.2

2

Table 4(on next page)

Categories of SNPs not compatible with the mode of reproduction of the studied model (haplodiploidy) or a Mendelian inheritance of alleles and number of occurrences of each case.

The different situations are illustrated by examples taken from the analysis of the 385 questionable SNPs.

1 **Table 4**

Description	Male F0	Female F0	Female F1	Pool of males F2	Occurrences
Male F0 incompatible	**A/G**	G	G	G	81 (21.04 % of the problematic SNPs ; 0.002% of the analysed nt)
	TG	GA	GA	GA	
Female F0 incompatible	C	C/**T**	C	C	125 (32.47 % of the problematic SNPs ; 0.003% of the analysed nt)
	A	**A**	A/G	A/G	
	G	**A**	G	G	
Female F1 incompatible	C	C	**A**/C	C	141 (36.62 % of the problematic SNPs ; 0.004% of the analysed nt)
	T	C/T	**C**	C/T	
	GG	AG/GG	GG/**GT**	AG/GG	
	T	A	**A**	A/T	
Pool of males F2 incompatible	T	T	T	**C**/T	29 (7.53% ; of the problematic SNPs ; 0.0008% of the analysed nt)
	A	A/G	A/G	A	
	AA	AA	AA	**CC**	
	C	T	C/T	**C**	
Combination of the different situations	**C/G**	C/G	C/G	**C**	9 (2.34%; of the problematic SNPs ; 0.0002% of the analysed nt)

2

3

Figure 1

Experimental setup

Photo *Trichogramma brassicae* J.-Y. Rasplus