A peer-reviewed version of this preprint was published in PeerJ on 16 October 2018.

View the peer-reviewed version (peerj.com/articles/5640), which is the preferred citable publication unless you specifically need to cite this preprint.

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

Astrid Cruaud Coresp. 1, Géraldine Groussier 2, Guenaëlle Genson 1, Laure Sauné 1, Jean-Yves Rasplus 1

1 CBGP, INRA, CIRAD, IRD, 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 Cruaud1*, Géraldine Groussier2, Guenaëlle Genson1, Laure Sauné1, and Jean-Yves Rasplus1

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 150 ng 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 10 ng 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 ; \text{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).
In silico digestion of the genome of *T. pretiosum* revealed 59,433 *PstI* cut sites (i.e. 118,866 tags). An average of 2*3,757,867 reads (109 bp) was obtained for the different samples after quality filtering, demultiplexing and removal of PCR clones (Table 1). Two females F1 (TRIC00027_1103 and TRIC00027_3103) were represented by much less reads than other samples (595,204 and 1,991,305 respectively). The number of tags recovered by *ustacks* and *cstacks* varied but was comparable among the samples and in line with the predictions made on the genome of *T. pretiosum* when these two females were excluded from calculation (average number of *ustacks* tags = 132,787; average number of *cstacks* tags = 128,293, Table 1).

The comparison of the loci obtained after filtering steps with *rxstacks* and *populations* revealed that, in average, 97.6% of the loci were homozygous and identical for females F1 and the pool of males F2 (min=96.8% - max=98.2%, Table 2). In average, 0.7% (0.3%-1.3%) of the loci were heterozygous and identical in both samples. Thus there was a ca 98.3% (97.4%-99.0%) exact match between the loci of the females F1 included in the library (and whose DNA was amplified with WGA) and the whole progeny of the F1 generation (pool of males whose DNA was not amplified).

Between 1.0 and 2.6% of the loci were not identical between analysed pairs (Table 2). A careful inspection of the haplotypes revealed that ca 60% of these differences could be explained by the experimental setup, *i.e.* the sequencing of a single female of the F1 generation versus the sequencing of the whole progeny of the F1 generation (pool of males heterozygous, female F1 homozygous with an allele present in the pool of males; pool of males with three alleles, female F1 homozygous with an allele present in the pool of males; pool of males with three alleles, female F1 heterozygous with two alleles present in the pool of males). About 40% of the observed SNPs were not compatible with either the experimental setup (pool of males and
female F1 homozygous but with different alleles; female F1 heterozygous, pool of males

homozygous with an allele present in the female F1; female F1 heterozygous, pool of males

homozygous with an allele not present in the female F1; pool of males heterozygous, female F1

homozygous with an allele not present in the pool of males; females F1 and pool of males

heterozygous with only one of the two alleles in common. Globally, 99.3% (98.9 - 99.6%) of the
shared loci were either identical or displayed differences that could be explained by the
experimental setup. The first cross was used to check in details the overall coherence of the
haplotypes from the parental generation to the whole progeny of the F1 generation (Table 3).

98.8% of the 32,913 loci shared by the four samples displayed haplotypes consistent with
experimental the setup and a Mendelian inheritance of alleles (97.3% being homozygous and
identical between samples). SNPs observed in 385 loci (which represent 1.2% of the loci and
0.01% of the analysed nucleotides) were not compatible with the mode of reproduction of the
studied model (haplodiploidy) or a Mendelian inheritance of alleles. Considering the haplotype
observed in the pool of males F2 as a reference, questionable SNPs could be categorized into
five categories as listed in Table 4. In ca 90% of the situations, SNPs found either in the male F0
(21%), the female F0 (32.5%) or the female F1 (36.6%) were incompatible with haplodiploidy or
with a Mendelian inheritance of alleles. 96 cases (ca 25.0%) represented situations where one
allele was missing for the female F0 or the female F1 to fit with a Mendelian inheritance of
alleles (possible cases of allele drop-out).

DISCUSSION

Here we compare RAD libraries obtained from a technical / artificial amplification of DNA
(WGA of single specimen of micro-hymenoptera, F0 and F1 generations) and a biological /
natural amplification (pool of the progeny of the F1 generation). We push the limits of the kit used for the WGA (GenomiPhi) by using ca 90% less DNA (ca. 1.0ng) than the required amount specified on the manufacturer’s protocol (10ng). Globally, we show that 99% of the examined loci (up to 48,189; 109 bp each) were compatible with haplodiploidy and either identical among specimens or compatible with a Mendelian inheritance of alleles. These results are consistent with observations by Blair et al. (2015) who used the Qiagen REPLI-g Mini Kit and 100 ng of input DNA and showed that SNP calling between ddRAD libraries from native and amplified DNA presented a > 98% match (up to 11,309 loci examined). They are also in agreement with older studies that attempted to quantify bias induced by multiple displacement amplification method (MDA) on which the GenomiPhi kit relays (> 99% match; Barker et al. 2004; ElSharawy et al. 2012; Paez et al. 2004), though with more input DNA (10ng).

To the exception of two samples, for which the construction of the library seems to have failed (much less reads were obtained), comparable numbers of tags were obtained. This indicates that the coverage of the genome is the same regardless if native or amplified DNA is used as suggested by previous studies on the potential bias induced by MDA (Abulencia et al. 2006; Blair et al. 2015; Paez et al. 2004). Studies have suggested that WGA may induce allele dropout especially when the starting amount of DNA is low (<1ng) (Handyside et al. 2004; Lovmar et al. 2003; Lovmar & Syvänen 2006; Sun et al. 2005). ElSharawy et al. (2012) and Blair et al. (2015) concluded that MDA had no significant effect on levels of homozygosity. Here about 1% of the loci retained by our analytical pipeline (ie ca 0.01% of the examined nucleotides) presented problematic SNPs that were not compatible with the biology of Trichogramma wasps or a Mendelian inheritance of alleles. 0.3% of the SNPs were possible cases of allele drop-out (one allele was missing for the female F0 or the female F1 to fit with a Mendelian inheritance of
241 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.

260 CONCLUSION

261 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


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


10.1111/mec.12269


**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

<table>
<thead>
<tr>
<th>Cross #</th>
<th>Sample code</th>
<th>Description</th>
<th>qDNA (ng)</th>
<th>Input DNA for WGA (ng)</th>
<th>Output DNA from WGA (ng)</th>
<th>Input DNA for RAD library (ng)</th>
<th>Demultiple xed Reads* (forward only)</th>
<th>Cleaned reads** (forward only)</th>
<th>ustacks : Nb of loci</th>
<th>ustacks : Nb of loci***</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRIC00027_2101</td>
<td>Male F0, haploid, WGA</td>
<td>11.5</td>
<td>1.5</td>
<td>500</td>
<td>169.0</td>
<td>6,774,680</td>
<td>5,173,711</td>
<td>136,623</td>
<td>130,060</td>
</tr>
<tr>
<td>1</td>
<td>TRIC00027_2102</td>
<td>Female F0, diploid, WGA</td>
<td>10.6</td>
<td>0.39</td>
<td>1048</td>
<td>203.3</td>
<td>4,073,370</td>
<td>3,179,891</td>
<td>128,212</td>
<td>122,860</td>
</tr>
<tr>
<td>1</td>
<td>TRIC00027_2103</td>
<td>Female F1, diploid, WGA</td>
<td>6.20</td>
<td>0.35</td>
<td>2393</td>
<td>281.2</td>
<td>4,597,505</td>
<td>3,566,986</td>
<td>130,565</td>
<td>124,845</td>
</tr>
<tr>
<td>1</td>
<td>TRIC00027_2199</td>
<td>Pool of haploid males F2 (n=933), no WGA</td>
<td>735.4</td>
<td>N.A.</td>
<td>N.A.</td>
<td>269.0</td>
<td>4,818,385</td>
<td>3,745,752</td>
<td>127,709</td>
<td>125,047</td>
</tr>
<tr>
<td>2</td>
<td>TRIC00027 1101</td>
<td>Male F0, haploid, WGA</td>
<td>Too low</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>2</td>
<td>TRIC00027 1102</td>
<td>Female F0, diploid, WGA</td>
<td>Too low</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>2</td>
<td>TRIC00027 1103</td>
<td>Female F1, diploid, WGA</td>
<td>9.4</td>
<td>0.17</td>
<td>1128</td>
<td>164.6</td>
<td>774,450</td>
<td>595,204</td>
<td>43,763</td>
<td>42,062</td>
</tr>
<tr>
<td>2</td>
<td>TRIC00027 1199</td>
<td>Pool of haploid males F2 (n=229), no WGA</td>
<td>359.6</td>
<td>N.A.</td>
<td>N.A.</td>
<td>270.6</td>
<td>5,878,301</td>
<td>4,437,984</td>
<td>127,380</td>
<td>125,302</td>
</tr>
<tr>
<td>3</td>
<td>TRIC00027 3101</td>
<td>Male F0, haploid, WGA</td>
<td>Too low</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>3</td>
<td>TRIC00027 3102</td>
<td>Female F0, diploid, WGA</td>
<td>13.9</td>
<td>2.9</td>
<td>390</td>
<td>247.7</td>
<td>7,006,980</td>
<td>5,365,499</td>
<td>147,718</td>
<td>140,690</td>
</tr>
<tr>
<td>3</td>
<td>TRIC00027 3103</td>
<td>Female F1, diploid, WGA</td>
<td>10.7</td>
<td>0.43</td>
<td>226</td>
<td>137.3</td>
<td>2,616,330</td>
<td>1,991,305</td>
<td>93,606</td>
<td>89,543</td>
</tr>
<tr>
<td>3</td>
<td>TRIC00027 3199</td>
<td>Pool of haploid males F2 (n=1415), no WGA</td>
<td>670.7</td>
<td>N.A.</td>
<td>N.A.</td>
<td>228.4</td>
<td>7,510,904</td>
<td>5,764,475</td>
<td>131,267</td>
<td>129,249</td>
</tr>
</tbody>
</table>
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

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

<table>
<thead>
<tr>
<th>Studied samples</th>
<th>Nb of shared loci</th>
<th>Percentage of identical loci (homozygous)</th>
<th>Percentage of loci consistent with the experimental setup and Mendelian inheritance of alleles</th>
<th>Percentage of loci not consistent with the experimental setup and Mendelian inheritance of alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross #1</td>
<td>32,913</td>
<td>97.3</td>
<td>98.8</td>
<td>1.2</td>
</tr>
<tr>
<td>- female F0 &amp; male F0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- one female F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- progeny of the F1 generation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.
### Table 4

<table>
<thead>
<tr>
<th>Description</th>
<th>Male F0</th>
<th>Female F0</th>
<th>Female F1</th>
<th>Pool of males F2</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male F0 incompatible</td>
<td>A/G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>TG</td>
<td>GA</td>
<td>GA</td>
<td>GA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(21.04 % of the problematic SNPs ; 0.002% of the analysed nt)</td>
</tr>
<tr>
<td>Female F0 incompatible</td>
<td>C</td>
<td>C/T</td>
<td>C</td>
<td>C</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>A</td>
<td>A/G</td>
<td>A/G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(32.47 % of the problematic SNPs ; 0.003% of the analysed nt)</td>
</tr>
<tr>
<td>Female F1 incompatible</td>
<td>C</td>
<td>C</td>
<td>A/C</td>
<td>C</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>C/T</td>
<td>C</td>
<td>C/T</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>AG/GG</td>
<td>GG/GT</td>
<td>AG/GG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A/T</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(36.62 % of the problematic SNPs ; 0.004% of the analysed nt)</td>
</tr>
<tr>
<td>Pool of males F2 incompatible</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C/T</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>A/G</td>
<td>A/G</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>AA</td>
<td>AA</td>
<td>CC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>T</td>
<td>C/T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(7.53% ; of the problematic SNPs ; 0.0008% of the analysed nt)</td>
</tr>
<tr>
<td>Combination of the different situations</td>
<td>C/G</td>
<td>C/G</td>
<td>C/G</td>
<td>C</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2.34%; of the problematic SNPs ; 0.0002% of the analysed nt)</td>
</tr>
</tbody>
</table>
Figure 1

Experimental setup

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