**Supplementary material**

Reisser et al. 2017

1

**S1:** Supplementary_Material_S1_S2_S7_S8.xls

Excel sheet S1 of the Supplementary_Material-S1_S2_S7_S8.xls file, giving the list of the 81 microsatellite markers tested in this study.

2

**S2:** Supplementary_Material_S1_S2_S7_S8.xls

Excel sheet S2 of the Supplementary_Material-S1_S2_S7_S8.xls file, containing the list of SNPs obtained from the RAD-sequencing panel for all individuals. Information listed: Linkage Group; order of SNP on the genetic map; CentiMorgan position on the genetic map; basepair position on the physical map; MegaBase position; scaffold where the SNP maps to; position of the SNP on the scaffold (in bp); major allele; minor allele. Additional information: Chi square value; associated P-value; mutation location and type (intergenomic, intron, 5' / 3' UTR; amino acid substitution); synonymous or non synonymous mutation; gene impacted.

3

**S3:** Relative heterozygosity of the sample set.

Distribution of the percentage of relative heterozygosity in MP (black) and NMP (grey) individuals. Calculations were performed without LG3, as this chromosome shows a higher heterozygosity in NMP individuals.
S4: RAD-sequencing and SNP calling protocol

We used the RAD-sequencing protocol developed by Etter et al. (2011) with a few modifications. The 72 individuals were divided in 2 libraries. Prior to DNA extraction, individuals were treated for 72 hours with three antibiotics (Streptomycin, Tetracyclin, Ampicilin) at a concentration of 50 mg/L of each antibiotic and fed with microscopic glass beads (Sephadex “Small” by Sigma Aldrich: 50 µm diameter) at a concentration to 0.5g/100mL. The aim of this treatment was to minimize contaminant DNA (i.e., bacterial DNA or algal DNA) in the gut and on the surface of the carapace. Genomic DNA was extracted using the Qiagen Blood and Tissue kit following manufacturer’s instructions and digested with PstI (New England Biolabs). Digested DNA was barcoded with individual-specific P1 adapters and pooled to create a library containing 2100ng DNA. The pooled library was sheared on a Bioruptor using 2 times 3
cycles (1 cycle 30 seconds ON, 1 minute OFF), and fragments between 300 and 500bp were selected through agarose gel electrophoresis. DNA fragments were blunted and a P2 adapter was ligated. The library was amplified through PCR (30 seconds at 98°C, followed by 18 cycles of 10 sec. at 98°C, 30 sec. at 65°C and 30 sec. at 72°C; a final elongation step was performed at 72°C for 5 min.). A final electrophoresis was performed to select and purify fragments between 350 and 600bp. Each library was sequenced on a single lane of an Illumina HiSeq 2000, using single-end 100 cycle sequencing by the Quantitative Genomics Facility service of the Department of Biosystem Science and Engineering (D-BSSE, ETH), Basel, Switzerland.

The quality of the raw sequencing reads (library-wide and per-base) was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and reads were checked for barcode integrity, absence of adapter sequences within the reads, and integrity of the PstI cut site. The reads were sorted individually by barcode and filtered to remove reads with uncalled bases and an overall base quality score of less than 24. Reads were subsequently aligned to the *Daphnia magna* genome (V2.4; *Daphnia* Genomic Consortium, BioProject reference PRJNA298946, on the NCBI repository: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA298946) using BWA v.0.7.10 (Li and Durbin 2009). Reads that did not map to the reference genome or that mapped to more than one place were discarded. The successfully mapped reads were filtered according to mapping quality (end-to-end mapping with a mapping quality score of at least 25, no more than eight high quality substitutions).

Assignment of reads to RAD loci (defined by unique 95 bp locations on the reference genome) and genotype calling was performed in Stacks V1.19 with a bounded SNP model in pstacks (--bound_high of 0.04, according to the base call error rate provided by the sequencing facility) and allowing a maximum of two high frequency haplotypes (i.e. alleles) per locus per individual. Loci with more than two high frequency alleles were excluded because of a too high
risk of falsely mapping paralogous reads to a single locus. Cstacks and sstacks were operated with default settings and with the -g option to use genomic location as method to group reads. The distribution of the minor allele frequency indicated that heterozygous loci usually had a minor allele frequency ranging between 0.2 and 0.5 within an individual. We thus fixed the max_het_seq parameter to 0.2 in the program genotypes. As such, potentially heterozygous genotypes with a minor allele frequency of between 0.05 (default homozygote cut-off) and 0.2 were considered ambiguous and were scored as missing in the results. Loci were also filtered according to sequencing depth: Loci with less than 20 reads were discarded (to reduce uncertainty in genotype calls) as were reads with a more than five times higher depth than the average depth across individuals (to reduce the risk of including repetitive elements).

After final genotype calling, loci were mapped to the *Daphnia magna* genetic map v.3.0 (Dukič et al, 2016, referenced in article). This was done by extracting for each RAD locus the linkage group and cM position of the nearest map-markers on the same scaffold and, if needed, by extrapolating the cM position of the RAD locus by linear extrapolation between the two nearest map-markers.

**References**


S5 : association.R

Fully annotated R script providing input and output format information, as well as the function we created to perform the association analysis at the genome wide level.

S6 : Protocol for the physical ordering of scaffold in the NMP associated non-recombining region of LG3.

The region controlling the NMP phenotype maps to a region with a low recombination rate in the reference genetic map. This results in the fact that numerous scaffolds are mapped to the exact same cM position, and their relative position and orientation amongst each other are not resolved. Hence, no physical order of the scaffolds in the region can be obtained from the genetic map. This is problematic for genome-wide association studies and fine mapping of the NMP locus, especially for determining whether the NMP phenotype maps to one or multiple specific sub-regions. In an attempt to physically order and orientate the scaffold in the NMP region, we performed linkage disequilibrium (LD) mapping, which uses data on LD from a single population and therefore can make use of historical recombination events present in the data. LD mapping relies on the expectations that two physically close loci should show a high correlation in their segregation patterns in a population (and thus high LD), since recombination events between the two loci should be rare. It is a population based method, so that individual discrepancies with the global population pattern cannot be tested.

Because the NMP region on the incipient W chromosome might carry phenotype specific rearrangements, we based the physical ordering using LD mapping only on the 54 MP individuals sampled from the MOS population. We performed LD mapping on a region of LG3 between
85cM and 95cM, in order to use SNPs just outside the NMP linked region as anchor. The MOS dataset contains SNPs on 30 mapped scaffolds in this region. Among these, there are three groups of scaffolds for which the relative position and orientation could not be resolved with the genetic map: two scaffolds at position 88.8 cM, 17 scaffolds at 90.8 cM and 4 scaffolds at 93.7 cM. For physical ordering of the scaffolds within each of these groups, we first calculated $r^2$ values for each pair of SNPs with MCLD (Zaykin 2008, referenced in article), which is based on the correlation of segregation of genotypes in natural populations, avoiding the need to phase the data, but removing any individual particularities in segregation pattern. We then averaged the $r^2$ values of the three terminal SNPs on each side of each scaffold and created a matrix of pairwise average $r^2$ values between each pair of scaffold extremities for each of the cM groups separately. When more than two scaffolds had to be ordered in a group, we perform a hierarchical clustering analysis to identify “starting clusters” (highly linked scaffold extremities), using the hclust function of the R core package stats. Scaffolds were then added one by one to the starting clusters following the hierarchical order obtained from the hclust dendogram, and oriented in a way that maximized the average $r^2$ values between adjacent scaffold extremities.

S7 : Supplementary Material_S1_S2_S7_S8.xls
Excel sheet S7 of the Supplementary Material-S1_S2_S7_S8.xls file, containing calculated $r^2$ values for the SNPs present in the NMP non recombining region, in order to order and orientate scaffolds in the region not recombining in the reference MPxMP cross.
Excel sheet S8 of the Supplementary_Material-S1_S2_S7_S8.xls file, containing the list of (a) the raw and (b) the corrected phased haplotypes used in this study, along with the SNP coordinates for each position.

FASTA formatted document listing the 283 genes used in the analysis.