Genomics of Sexual and Asexual Reproduction in *Daphnia magna*

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Marinela Dukić

aus Split, Kroatien

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Dieter Ebert und Dr. Christoph R. Haag

Basel, den 23. Februar 2016

Prof. Dr. Jörg Schibler, Dekan

Genomics of Sexual and Asexual Reproduction in *Daphnia magna*

Marinela Dukić – PhD thesis

Table of Contents

SUMMARY

During my PhD I used the next generation sequencing technology to investigate patterns of recombination and the genetic consequences of different reproductive modes of *Daphnia magna*. More precisely, I have used Restriction site Associated (RAD) sequencing to construct a high-density genetic map that can be coupled with the draft genome assembly of *D. magna*, thus, providing an essential tool for genome investigations in this widely used model organism (**Chapter I**). Such a map has enabled me to characterize variation in meiotic recombination rates across the genome of *D. magna* for the first time. Since recombination rates are an important parameter in almost any type of genetic research, this newly gained insight into recombination landscape of *D. magna* offers fundamental information for future studies of genome evolution, identification of genes underlying phenotypic traits and population genetic analyses.

In addition to sexual reproduction, *D. magna* can also reproduce asexually to generate clutches of clonal offspring (ameiotic parthenogenesis). This feature of *Daphnia* biology is extremely useful for scientific experimentation where the genetic variation among tested individuals has to be minimized. However, over the last decade, reports of genome homogenization (loss of heteroygosity - LOH) in asexual lineages of *D. pulex* have indicated that asexual genomes are not static as it was previously assumed, and that some levels of ameiotic recombination, in addition to mutation, may induce genetic variation among putative clones. However, comparing parthenogenetic offspring with their mothers at several thousand genetic markers generated by RAD-sequencing, I was not able to detect any LOH events in *D. magna* (**Chapter II**). I cannot exclude the possibility that ameiotic recombination indeed occurs in *D. magna*, however, my results indicate that such phenomenon is extremely rare or restricted to the very short genomic regions that I was unable to investigate, despite a high-density of markers used in this study.

Nevertheless, the implementation of RAD-sequencing protocol for genome studies of *D. magna* still enables interrogation of the transmission of genetic information from parents to offspring at unprecedented resolution. For an example, a RAD-sequencing based analysis of reduction in parental heterozygosity among rare ephippial hatchlings (typically produced by sexual reproduction) found in non-male producing populations of *D. magna*, has enabled differentiation between self-fertilization and automixis (meiotic parthenogenesis), by uncovering the subtle differences in genetic consequences of these reproductive strategies (**Chapter III**). Harnessing the ability of high resolution genetic analysis it was demonstrated that, in the absence of males, *D. magna* can produce diapause eggs by automixis, and additional type of asexual reproduction that was not previously reported for this species.

Finally, RAD-sequencing European populations of *D. magna* revealed an association of genetic variation with the geographic location of individual samples (**Chapter IV**), a task which was not previously amenable using mitochondrial or microsatellite markers. This study provided a better insight into population genetic structure of *D. magna* and suggested that genetic differentiation is mainly driven by geographic distance. These results set a foundation for forthcoming studies aiming to disentangle past and future evolutionary processes shaping populations of this intriguing model organism.

Taken together, research presented in my thesis illustrates the practicality of reduced representation genome sequencing for tackling diverse topics in evolutionary biology. By increasing awareness of non-randomness of meiotic recombination across the genome of *D. magna*, the diversity of reproductive mechanisms it can employ, and its large-scale population structure, I hope this work will contribute to further understanding of the remarkable adaptive capacity this crustacean is famous for.

THESIS

INTRODUCTION

HOMOLOGOUS RECOMBINATION – GENERAL OVERVIEW

In eukaryotes, homologous recombination is an essential and the most truthful mechanism for the repair of DNA double strand breaks (DSBs) which are the most common and the most cytotoxic types of DNA damage (Huertas 2010). DSBs can be induced by exogenous agents like radiation or mutagenic chemicals, but often they are a consequence of intrinsic cellular process such as oxidative DNA damage (reactive oxygen species) and DNA replication failures (Aguilera and Gómez-González 2008; Woodbine et al. 2011). Such damages can affect both mitotic and meiotic cells; however, in meiotic cells during the meiotic prophase I, DSBs are also induced in a highly controlled manner by the enzyme Spo11 (Keeney 2008). Finally, if left unrepaired, DSBs will lead to genomic instability with detrimental fitness consequences (Aguilera and Gómez-González 2008; Huertas 2010).

The most accurate way to repair DSBs is homology-directed repair, i.e. using a homologous DNA sequence (more than 97 % of sequence identity) as a template. This type of repair is called homologous recombination and it may involve several pathways with different genetic outcomes. Mechanisms of homologous recombination have been reviewed extensively (Sung and Klein 2006; San Filippo et al. 2008; Huertas 2010; Krejci et al. 2012). One of the most studied models is the DSB repair (DSBR) model, proposed by Szostak et al. (1983) and it is often used synonymously for homologous recombination. Briefly, after the DSB formation, one strand from both sides of the break will be enzymatically degraded, creating a single-stranded DNA stretch (ssDNA) in a process known as DNA-end resection. This ssDNA then invades the nearby homologous sequence (homologous chromosome, sister-chromatid or non-allelic homologous region, depending upon the cell cycle phase at which DSB repair occurs) that will serve as a template (donor sequence) for the DNA synthesis. The invading strand gets captured between the strands of a donor molecule and causes the displacement of one of the donor strands into the D-loop. In the DSBR model, the non-invading resected end of the break anneals with the D-loop (second strand capture) and further synthesis leads to the formation of a cruciform structure known as double Holliday junction (dHJ). The genetic consequences of the DSBR model will depend on the resolution of dHJs. When dHJs are nicked in a non-symmetric fashion (horizontal and vertical cleavage) homologous recombination results in cross-overs (COs), i.e., reciprocal exchange of the genetic information between two DNA molecules. If this exchange happens between sister-chromatids (two DNA molecules resulting from the replication prior to the cell division, carrying the same alleles, and attached to each other by a centromere) no genetic trace will be left. If CO involves non-sister chromatids (DNA molecules from homologous chromosomes having different parental origin and carrying different alleles) it will yield the new combination of parental alleles on the recombinant chromosomes. When dHJs are nicked in a symmetric way the final outcome of DSBR model is a non-crossover resolution. However, even in the case of non-crossover resolution, highly localized genetic exchanges can take place due to the mismatch repair of intermediate heteroduplex DNA (pairing of non-complementary bases). Mismatch repair can lead to non-reciprocal exchange of genetic material, where one of the participating chromatids is modified to have the same genetic information as its homologue while the original information is lost. This exchange usually spans over several hundred bases and it is called a gene conversion. Contrary to that, CO recombination results in the exchange of genetic information across whole chromosomal

arms from the site of CO formation distal to the centromere, unless another CO structure is encountered which would reverse the effect of the previous CO. Hence, the genetic exchange due to COs can be measured in the order of megabases (Mb), whereas gene conversions usually affect less than a kilobase (kb) of genetic sequence. For this reason, the frequency of CO recombination is more easily measured and serves for linkage mapping and as a proxy for the estimation of recombination rates (see below).

Unlike DSBR model, other homologous recombination pathways do not include the formation of HJs and result exclusively in non-crossovers with possibility of non-reciprocal gene conversions. The synthesis-dependant strand annealing (SDSA) pathway also involves the formation of D-loop; however, it does not lead to the second end capture but instead the invading strand gets displaced and anneals with the second resected end (San Filippo et al. 2008). In the break-induced replication (BIR) pathway, single strand invasion is simply followed by replication until the end of the chromosome (Lydeard et al. 2010). Same as the SDSA pathway, BIR results in non-reciprocal transfer of genetic information from the template to a damaged strand; however, with BIR, this transfer may involve substantially longer tracts (Lydeard et al. 2010).

RECOMBINATION DURING SEXUAL REPRODUCTION

Meiosis is a central feature of sexually reproducing eukaryotes. Meiosis involves two cellular divisions (meiosis I and meiosis II) coupled with a single round of DNA replication resulting in four potential daughter cells (haploid gametes) which have only half of the genetic material of the parental cell. During the first meiotic (reductional) division, homologous chromosomes, each made of two sister-chromatids, segregate into separate daughter cells and the number of chromosomes is halved. Sister-chromatids then separate during the second meiotic (equational) division in a mitosis-like process. Even though its benefits are still under discussion (see below), one of the most prominent attributes of meiosis is the generation of genetic variation among resulting cells (gametes). The independent orientation of homologous chromosomes in metaphase I and their subsequent distribution to daughter cells will create new combinations of chromosomes, as one portion of the final chromosome number will be of paternal and the other of maternal origin. Moreover, during prophase I, CO recombination creates novel combinations of alleles along chromosomes by reciprocal exchange of genetic information between homologous chromosomes in gamete precursor cells.

As it was mentioned earlier, homologous recombination during meiosis is actively promoted by programmed induction of DSBs and it is 100-1000-fold more frequent than recombination during mitosis (Sung and Klein 2006). In addition, recombination resulting in COs is very common during meiosis. Meiotic COs generate genetic diversity among the progeny of same parents but also, in majority of species, COs serve as a physical link between homologous chromosomes during prophase I, ensuring their proper alignment and succeeding segregation during anaphase I (see also Page and Hawley 2003; Jones and Franklin 2006).

It is now clear that meiotic COs are not distributed randomly across chromosomes, but instead, tend to occur only in a small portion of genome (Nachman 2002; de Massy 2013). On the other hand, large chromosomal regions where COs are suppressed, have been identified. These include heterochromatic regions around centromeres (Talbert and Henikoff 2010) or chromosomal inversions that enable formation of gene clusters, i.e., supergenes (Thompson and Jiggins 2014).

The determinants of CO distribution are not fully understood, however, latest research indicates that the CO patterning is regulated on a structural basis, with the synaptonemal complex (proteinaceous structure between aligned homologous chromosomes) playing a major role (Libuda et al. 2013; Zhu and Keeney 2014). Moreover, available experimental data largely fit the model of CO patterning governed by structural properties of chromosomes and redistribution of mechanical stress elicited by chromosome dynamics in meiosis (Zhang et al. 2014; Wang et al. 2015). Besides mechanical constraints on CO distribution due to its role in meiosis, the frequency and distribution of COs should also be considered from an evolutionary perspective, since it determines the pace of formation of new haplotypes upon which selection can act. To understand this adaptive role of meiosis, scientist are more and more interested in how exactly variation in recombination rates correlates with ecology (Hoffmann and Hercus 2000), life-history traits (Wilfert et al. 2007) or genome architecture (Pál and Hurst 2003). Moreover, meiotic recombination lies at the heart of almost any genetic analysis and better understanding of meiotic processes can largely empower further advancements in genomics era.

Linkage Maps and Recombination Rates

Genetic linkage maps of a species or of an experimental cross display the linear ordering of genetic elements (genes or genetic markers) along chromosomes. Unlike physical maps, the position of studied genetic elements is determined based on their recombination frequency. Recombination frequency is defined as the rate at which a single CO will occur between two genetic elements. The mapping unit is the centimorgan (cM) which describes a recombination frequency of 1 %, i.e., "a portion of the chromosome of such length that, on the average, one CO will occur in it out of every 100 gametes formed" (Sturtevant 1913).

After Sturtevant's seminal work (Sturtevant 1913), when the concept of linkage mapping was introduced for the first time, genetic linkage maps were constructed following the inheritance of just a few phenotypic markers (heritable polymorphism that can be measured in different populations of individuals) and measuring the frequency of meiotic crossovers by the number of recombinant individuals obtained in genetic crosses. Subsequently, DNA-based markers such as microsatellites, restriction fragment length polymorphisms (RFLPs) or amplified fragment length polymorphisms (AFLPs) replaced the phenotypic markers and made genetic maps far denser (up to thousand markers) and, consequently, more informative. More recently, the ease of studying genome-wide single nucleotide polymorphisms (SNPs), due to advancements in sequencing technologies, has enabled the construction of high-density genetic maps that include several thousand markers and provide unprecedented resolution in identifying genomic regions that affect the phenotype of interest. Besides mapping quantitative trait loci (QTL), currently produced high-density genetic linkage maps are an essential resource in studying almost every aspect of genome

biology. Such maps provide a framework for the improvements of genome assemblies generated by the Next Generation Sequencing (NGS) technologies which result in stretches of genomic sequences of different length (contigs and scaffolds), that have to be connected using additional mapping approaches. Thus, high-density linkage maps that can be amalgamated with pre-existing scaffolds enable anchoring and orienting of genomic parts along chromosomes, resulting in the construction or improvement of physical maps for the species or population under study.

Finally, comparing genetic and physical distances between markers is a foundation for studying recombination rates, which are defined as the observed frequency of COs per unit of physical distance (e.g. cM/Mb). Even though the process of meiotic recombination is mainly conserved among eukaryotes, the resulting recombination rates vary substantially between species, populations, individuals and different genomic regions. The resulting recombination landscape can have a profound influence on the structure of genetic variation within and between species, but also it can influence the design of quantitative trait mapping, population genetics studies and the interpretation of their inferences.

For example, Noor et al.(2001) have shown that the gene density per centimorgan can have a strong influence on biases in inferred effect sizes of QTLs as well as biases on which chromosome a QTL is likely to be detected (autosomes vs. sex chromosomes). Their simulations showed that the strongest QTLs are likely to be found in the regions of low recombination since they usually represent the combined effect of several contributing loci. On the other hand QTLs located in regions of high recombination will be more isolated from others, and for that reason, it will be easier to pinpoint the loci underlying the trait of interest. Thus, QTL mapping can be largely improved with the availability of detailed genetic and physical maps, and the design of mapping protocols should be corrected by taking into account variation in recombination rate along the chromosomes in order to get more accurate estimates of number of QTLs and their effect sizes.

Besides affecting the interpretation of mapping results in experimental crosses, variation in recombination rate can also have an influence on population genetic studies and understanding the effects of selection on genetic variation within and between populations. Genomic regions with low recombination rates are expected to have lower levels of neutral polymorphism than genomic regions with high recombination rates because of positive (hitchhiking) or negative (background) selection on sites in their physical neighbourhood (Cutter and Payseur 2013). In population genomics studies, this will be manifested as loci-specific reduction in effective population size (N_e). In addition to the effect on neutral diversity, the absence of CO can also reduce the efficiency of selection at the loci that are linked to others under selection as described by the process of Hill-Robertson interference (Hill and Robertson 1966; Cutter and Payseur 2013). This could also have an impact on genome organization since it enables fixation of deleterious mutations in the regions of low recombination, whereas beneficial mutations can be expected to approach fixation much faster in genomic regions where CO occurs. Hence, the density of sites under selection and the variation in CO rate across the genome should be considered as important factors in population studies of diversity and divergence.

RECOMBINATION DURING ASEXUAL REPRODUCTION

In eukaryotes, asexual reproduction shows a bewildering diversity of mechanisms and the terminology used for their description is not always used consistently. Moreover, there is a disparity in terminology describing asexual reproduction in plants and animals; however, in this section I will mainly be dealing with asexual reproduction in animals. Readers interested in an exhaustive description of the different modes of asexual reproduction and the history of their classification should refer to Bell (1982), Suomalainen et al. (1987) and Schön et al. (2009). Also, for the purpose of clarity, here I will focus on the classification of asexual reproduction based on the mechanisms of how the parental (somatic) number of chromosomes is maintained without syngamy (i.e. fertilization). More precisely, for the discussion of the expected levels of homologous recombination in asexual reproduction, the most relevant classification is based on the presence or absence of meiosis. Throughout my thesis, I use the term apomixis for types of asexual reproduction where meiosis is completely absent (mitotic reproduction) or partially suppressed (suppression of the first or the second meiotic division) so that the reduction of chromosome number does not take place. The term automixis is used for asexual reproduction where meiosis is normal, reduction of chromosome number takes place and several different mechanisms can be employed to restore parental ploidy.

Apomixis

The simplest type of apomixis is mitotic reproduction which is mainly found in unicellular eukaryotes. Special types of mitosis-based reproduction, such as fission in planaria, budding in hydra or vegetative reproduction in plants, are not considered as apomixis here. During mitosis, homologous recombination is used for the repair of endogenously or exogenously induced DSBs. CO resolution is very rare compared to meiosis, and the offspring produced by mitotic reproduction are considered to be clones of their parent with the only exception of rare mutations. Often, this expectation is oversimplified since DSBs are inevitable and significant levels of spontaneous mitotic COs have been reported to occur in *Saccharomyces cerevisiae, Aspergillus niger* and *Candida albicans* (Debets et al. 1993; Mandegar and Otto 2007; Forche et al. 2011). Nevertheless, there is accumulating evidence that DSB repair during mitosis is highly biased towards the non-CO resolution (reviewed in Andersen and Sekelsky 2010) leaving no genetic trace or only short stretches of gene conversions.

After the CO event, during mitosis, homologous chromosomes will line up one after another, with the metaphase plane "passing" between sister-chromatids. If the chromatids containing different alleles (two non-recombinant or two recombinant ones) align on the same side of the metaphase plate and segregate towards the same pole, the resulting daughter cells will retain parental heterozygosity and the occurrence of recombination will not be detectable. However, if the homologous chromatids, containing the same allele (one recombinant and one non-recombinant), align on the same side of the metaphase plate, the resulting daughter cells will become homozygous, i.e., carrying the same copy of the gene on both homologous chromosomes. Thus, recombination in the latter case will result in loss of heterozygosity (LOH) in both daughter cells. When genetic material between homologous chromatids is exchanged in a non-reciprocal way, as it is the case of gene-conversions, only one of the recombining chromatids receives genetic information from its homologue, leading

to a 3:1 allelic ratio (instead of 1:1 as it is the case for a reciprocal exchange). Thus, mitotic gene conversions will always result in LOH in one daughter cell. Again, it is important to note that the occurrence of COs during mitosis might lead to LOH in whole chromosomal arms, while gene conversions usually result in stretches of LOH shorter than 1 kb (but see Lee et al. 2009).

Even very rare occurrences of CO recombination can have pivotal consequences for the fitness of apomictic organisms due to LOH and unmasking of mutations that had little or no effect in a heterozygous state. If unmasked mutations are beneficial, LOH will hasten the spread of those mutations and their fixation in a population (Mandegar and Otto 2007). However, if unmasked mutations are deleterious, as the majority of mutations are (Lynch and Gabriel 1990), LOH can have detrimental fitness consequences. Unmasking of recessive deleterious mutations is also known as loss of complementation (LOC) which might play an important role in understanding the advantages of sexual reproduction (Archetti 2010) or the "two-hit" hypothesis for tumour development (Knudson 1971).

Besides mitosis, apomictic reproduction can represent meiosis with the suppression of one division. A very common type of asexual reproduction in animals is meiosis with the suppression of the first division. Genetic consequences will be similar to mitotic reproduction since only the second meiotic (mitosis-like) division takes place. However, if meiosis is suppressed after the prophase I, there is a possibility that DSBs will be induced in a meiotic like fashion, largely increasing probability for the occurrence of CO recombination and large scale LOH.

The opposite situation can be found if we consider meiosis with the suppression of the second meiotic division. In that case, sister-chromatids are segregating together and will represent homologous chromosomes in the offspring individual. Thus, if those two chromatids are the same, i.e., without any CO, offspring individuals should exhibit LOH across whole chromosomes. However, since the first meiotic division is supposed to be normal, high levels of CO may occur, leading to partial retention of heterozygosity in offspring individuals.

Automixis

In automixis, meiosis is normal and meiotic levels of homologous recombination are expected to occur. The genetic consequences of automixis will depend on the mechanisms for the recovery of parental ploidy. In the majority of cases this is achieved by a fusion of meiotic products similar to fertilization; however, in automixis, fusing cells are the products of a single meiosis (see Chapter III).

Gamete duplication is a type of automixis where the final haploid products duplicate through mitosis. This might involve diakinesis (cell division) and their subsequent fusion or chromosomes might duplicate without cell division. In both scenarios, gamete duplication will result in a complete genomic LOH since homologous chromosomes will be copies of each other.

The term terminal fusion is used to describe the fusion of products of the second meiotic division (second polar body and an egg cell). Since this type of fusion represents the reunion

of sister chromatids, all centromeric regions where recombination is suppressed will be homozygous while in regions where COs can be found, heterozygosity gradually increases with an increasing distance from the centromere (see Chapter III).

Central fusion denotes the fusion of haploid cells derived from two separate products of the first meiotic division. In this case, centromeric regions will retain heterozygosity as each chromatid (i.e., chromosome) will be derived from homologous chromosomes separated during the first meiotic division. In regions where recombination occurs, parental heterozygosity will decrease from 100 % to 67 % of the parental heterozygosity, depending on the frequency of COs between a locus and the centromere (see Chapter III).

In the best studied cases of endomitosis, chromosomes undergo two rounds of replication before the onset of meiosis. Each chromosome pairs and recombines with its sister copy rather than its homologue and, therefore, genetic information remains the same regardless of the recombination pathway utilized (Lutes et al. 2010). Asexual offspring will fully retain parental heterozygosity and can be considered clones of their mother.

SEXUAL *versus* ASEXUAL REPRODUCTION

Sex is a ubiquitous mode of reproduction among eukaryotes even though, compared to asexual reproduction, sex seems to be quite costly both from genetic and ecological point of view. There are many hypothesised costs of sex (Lehtonen et al. 2012; Meirmans et al. 2012); however, the "two-fold cost of sex" is conceptually the most straightforward and it is often used in models aiming to explain the persistence of sex when sexually reproducing individuals are faced with asexual competitors (Kondrashov 1993; Hartfield and Keightley 2012). Asexual reproduction is a state derived from sexuality in multicellular eukaryotes and such a transition has occurred many times independently throughout the metazoan evolution. When an asexual mutant occurs in a sexual population, asexual females would on average produce twice as many asexual daughters than sexual females; thus, asexuals should at least initially double in each generation, driving the sexual population extinct over few dozen generations (Maynard Smith 1978). The fact that asexual taxa appear only at the tips of phylogenetic trees within the sexual clades, indicates that the emergence of asexuality usually represents an evolutionary dead end(Schwander and Crespi 2009). This "paradox of sex" poses a fundamental problem to the theory of evolution by natural selection and considerable number of theoretical models have been proposed in order explain the evolutionary advantage of sex which can counterbalance its hypothesised costs and thwart the invasion of asexual mutants. Historically, models that have caught the great deal of attention can be classified in two groups (but see West et al. 1999; Schwander and Crespi 2009). Ecological (environmental) models argue that new combinations of genes produced by sexual reproduction (and recombination) provide a higher evolutionary potential in changing environments while mutation based models promote the efficiency of purging out deleterious mutations as the main advantage of sex. The general conclusion of these models is that the absence of recombination in asexual reproduction severely reduces the efficiency of natural selection in asexual compared to sexual organisms (Barton and Charlesworth 1998; Otto and Lenormand 2002; Haag et al. 2009). As a consequence, asexual populations should

adapt more slowly to changing environments and suffer from genetic load (Haag et al. 2009), while sex and recombination produce new genetic variability on which natural selection can act (reviewed in Otto 2009). However, sexual reproduction does not need to increase genetic variation and if it does, generated variation often results in reduced fitness (Otto and Lenormand 2002; Otto 2009). These models also lack the explanations for the immediate advantage of sex and most of them do not consider the diversity of asexual mechanisms that occur in nature. Even though proposed models could be proved valid under very restricted conditions, finding a consensus explanation for the evolutionary success of sexual reproduction still remains one of the main challenges in evolutionary biology.

Recently, Archetti (2004a,b, 2010) proposed an alternative model to explain the maintenance of sex in nature. His "Loss of complementation - LOC" hypothesis takes into account the frequency of recombination in different types of asexual reproduction (see above), and their varying genetic consequences, to encounter all major problems in the existing hypotheses concerning the evolution of sex. One of the most prominent ideas arising from the LOC hypothesis is that recombination takes place during asexual reproduction. As it was explained earlier, this should not be surprising since DSBs are an inevitable type of DNA damages and can only be repaired truthfully by one of the homologous recombination pathways. Details of LOC will depend on the portion of genome that becomes homozygous (LOH) within a single asexual generation, as well as the number of recessive deleterious mutations (lethal equivalents) present. If the right combination of these two parameters existed, the cost of asexual reproduction due to LOC would be higher than two-fold cost of sex, thus providing an explanation for an immediate advantage and the evolutionary success of sexual reproduction (Archetti 2004b, 2010).

Daphnia AS A MODEL ORGANISM

The species of the planktonic crustacean genus *Daphnia* (common name, water fleas) have been in the focus of biological research for centuries, with early studies dating back to 17th century (Korovchinsky 1997). The outstanding versatility of this system was instrumental for many studies yielding historical contributions in elucidating fundamental concepts in biology (Ebert 2011). There are many features of *Daphnia* biology that make them extremely valuable for scientific experimentation. These include a cosmopolitan distribution, short generation times, possibility of sexual and asexual (nearly clonal) reproduction, body transparency, the production of diapausing eggs and the ease of laboratory maintenance (Ebert 2005; Miner et al. 2012). The extensive literature on *Daphnia* research exceeds 10,000 scientific papers published up to date, profiling them as one of the oldest and most extensively used model organisms in biology.

Daphnia spp. can be found all over the world and are considered to be keystone species in many fresh water habitats. Multiple *Daphnia* lineages independently colonized water bodies that may differ in size, seasonal stability or composition of biotic and abiotic factors (Colbourne et al. 1997), indicating the remarkable ability of the genus for rapid adaptation to a wide range of environmental conditions. *Daphnia* spp. are well known for exhibiting a variety of environmentally induced phenotypes which include environmental sex

determination and switching from asexual to sexual reproduction. Females of the genus usually reproduce asexually (ameiotic parthenogenesis) through several clutches of 5-30 directly-developing females. However, in response to environmental conditions, such as crowding or seasonal change, male offspring may be produced parthenogenetically (genetically identical to their mothers). Unfavourable environments also induce meiosis in females and the production of two haploid eggs that have to be fertilized. Sexually produced embryos go through the resting stage encased in ephippia that enable survival of harsh conditions (e.g. desiccation or freezing) and migration mainly assisted by birds, terrestrial mammals or humans (Havel and Shurin 2004). Diapausing eggs can also sink to the bottom of lakes and remain viable in sediment for decades (Hairston 1996). Only when conditions become favourable again, a female offspring will hatch out of ephippia and through asexual reproduction, population has fast, exponential growth until limiting conditions are reached (e.g. overcrowding or food scarcity).

Extensive research on *Daphnia* has resulted in a superb knowledge of its ecology, life-history traits and evolutionary adaptations to different environments. However, due to scarcity of genomic and molecular resources, less progress has been made in characterizing the genetic bases of phenotypic traits in the context of changing environments (Ebert 2011). Nevertheless, this situation is gradually changing and *Daphnia* is becoming acknowledged as one of the most prominent models for the emerging field of evolutionary genomics (Eads et al. 2008; Shaw et al. 2008; Colbourne et al. 2011). This is largely due to the joint efforts of scientists gathered in *Daphnia* Genomic Consortium (DGC). The genome sequences and high-density genetic maps of *D. pulex* and *D. magna* can be considered as the key foundation for the further development of genomic resources.

The only draft genome published up to date is the one of *D. pulex* since it is pertinent to a larger number of researchers (Colbourne et al. 2011). The genome draft assembly of *D. magna* (v2.4) is also available to the members of DGC. Already from the preliminary draft sequences it is becoming obvious that these genomes harbour some interesting features that might help to explain the large adaptive potential of the genus as well as the high level of phenotypic plasticity *Daphnia* is famous for. A large number of genes is reported (at least 30,907 genes) probably arising through elevated rates of gene duplications and lower rates of gene loss compared to other lineages (Colbourne et al. 2011). Interestingly, Colbourne et al. (2011) have shown that numerous paralogs have divergent expression patterns in different environments, suggesting their novel roles in responsiveness to ecological challenges. Besides the *D. pulex* genome, recently, a high-density linkage map became available for this species (Xu et al. 2015).

The draft genome assembly of *D. magna* (v.2.4) consists of 40,356 scaffolds and contigs, summing up to 131,266,987 bp of genomic sequence. The size of *D. magna* genome, as estimated by flow-cytometry is 238 Mb (Routtu et al. 2014), indicates substantial gaps in the currently available reference genome. On-going endeavours to fill the gaps include different types of sequencing methods (PacBio and Hi-C) with high number of repetitive sequences representing a major challenge (P. Fields and D. Ebert, personal communication). The first linkage map for *D. magna* was constructed using 109 microsatellite markers and it was unsaturated, i.e., number of linkage groups exceeded the number of chromosomes (Routtu et

al. 2010). The second-generation linkage map was based on a highly error-prone SNP array (Routtu et al. 2014) and, therefore, I suspect that the ascertained linkages between markers are not highly reliable. The third-generation linkage map presented in Chapter I of this thesis was based on RAD-sequencing (see below) and it has the highest density and the highest resolution when compared with previous genetic maps of *D. magna*. Moreover, this map, coupled with the draft genome assembly, enabled the characterization of recombination landscape in *D. magna* for the first time.

All of this, and many other genomic resources that are already available for *Daphnia* species (reviewed in Shaw et al. 2008) can now be coupled with well the documented phenotypic diversity, within and between species, and the good understanding of the impact these traits have on fitness in different environments. Finally, integrating genomics and ecology will largely empower quantitative trait and association genetics studies to identify heritable genetic basis of molecular modifications, individual phenotypes and population-level responses that may govern the process of adaptation to changing environments.

Reproduction in *Daphnia*

The majority of *Daphnia* species are cyclic parthenogenetic, that is, they can switch between sexual and asexual reproduction mainly in response to environmental conditions, as it was explained earlier. The term used to describe asexual reproduction in *Daphnia* is apomixis (Zaffagnini 1987; but see Svendsen et al. 2015). As for other apomictic animals it was believed for a long time that the mechanism of asexual reproduction is mitosis without chromosome pairing or possibility for recombination (Zaffagnini 1987). However, the most detailed cytological study of asexual reproduction in *D. pulex* demonstrated that diploidy is maintained by meiotic arrest at early anaphase of the first meiotic division (Hiruta et al. 2010; Hiruta and Tochinai 2012). Thus, the reduction in chromosome number does not take place, but instead, all homologous chromosomes align to form a metaphase plate II and sister-chromatids separate to opposite poles as in meiosis II (mitosis-like process). Importantly, reported mechanisms of apomixis includes chromosome pairing during the prophase I, implying the possibility for CO recombination, although no chiasmata (cytological representation of COs) were observed (Hiruta et al. 2010). For *D. magna*, which is the focal species of the research presented in this thesis, such detailed cytological studies are not yet available. However, the mechanism of asexual reproduction in *D. magna* was described as a mitotic-like division with the extrusion of the polar body (Zaffagnini 1987), what is suggestive of a meiotic process as it was depicted for *D. pulex*. Studies on the genetic consequences of apomictic reproduction in *Daphnia* were also restricted to *D. pulex* until recently (see Chapter II). Interestingly, high levels of genome homogenization were reported using microsatellite markers in asexual mutation accumulation lines of *D. pulex* and were attributed to the occurrence of ameiotic (i.e., mitotic) recombination (Omilian et al. 2006; Xu et al. 2011). More recently, using a whole-genome sequencing approach, Keith et al. (2015) showed that the incidents of LOH are localized to short genomic tracts and most probably caused by gene conversions. In addition, majority of LOH tracts were associated with large scale duplications and deletions (Keith et al. 2015), demonstrating a dynamic nature of asexual genomes which were previously assumed to be clonal.

Obligate asexual lineages have also been reported among the *D. pulex* complex (Zaffagnini and Sabelli 1972; Paland et al. 2005). Some of these lineages still have the ability to produce functional males; however, meiosis in females is lost and even diapause eggs are produced asexually. On the other hand, both among *D. pulex* and *D. magna*, non-male producing strains were reported (Innes and Dunbrack 1993; Galimov et al. 2011). These females can still reproduce sexually but only through the production of diapause eggs that have to be fertilized by males from "regular" male-producing strains. For non-male producing strains of *D. magna*, it was reported that even in isolation, only very rarely, a few offspring will hatch from these ephippia (Galimov et al. 2011). Interestingly, the analyses of genome-wide patterns of heterozygosity reduction in these rare hatchlings lead to the discovery of automictic reproduction in *D. magna* (Svendsen et al. 2015, Chapter III). This diversity in reproductive modes of *Daphnia* makes them excellent model organisms for studying transitions to asexuality and finally elucidating the evolutionary advantage of sex.

REDUCED REPRESENTATION GENOME SEQUENCING

Over the last decade we have witnessed dramatic advances in sequencing technologies that have revolutionised the ways genomes can be interrogated. Combining the Next (2nd) Generation Sequencing (NGS) technologies, also known as massively parallel sequencing, with appropriate experimental selection of sequencing targets conceded expansion of methods for high-throughput, genome-wide sequence studies or transcriptional and regulatory profiling, at exceptional speed and resolution (Mardis 2011). Moreover, since the majority of these methods do not rely on previous knowledge of genetics or molecular tools available for the study system, high resolution genome research is possible for almost any organism or population. Despite constantly dropping costs of DNA sequencing, sequencing of entire genomes remains rather expensive for analysing a large number of organisms. Moreover, the amount of data that are generated in that way represent a challenge for computational storage and the downstream analysis, thus requiring highly expensive computational power and bioinformatics expertise. An alternative approach is to use reduced representation genome sequencing by targeted subsampling of the genome.

Restriction site associated DNA (RAD) sequencing is a method that combines digestion of genomic DNA with restriction enzyme(s) and sequencing the ends of those fragments using NGS technology, such as Illumina (Davey et al. 2011). Modified sequencing adapters, containing unique barcode sequence are used, thus it is possible to pool fragments from many individuals into the library that can be sequenced on a single sequencing lane. This allows simultaneous marker discovery and genotype estimation (homo- and heterozygotes) at several thousand random locations in a genome. Since RAD-sequencing targets only a subset of the genome (regions adjacent to a restriction site), when compared to whole genome sequencing it provides greater sequencing depth (the number of obtained sequencing reads for a given locus) per locus and the ability to analyse higher number of samples for a given budget (Andrews et al. 2016). All of this makes RAD-sequencing a suitable tool for tackling many biological questions with high accuracy. Just to name a few, RAD-sequencing has been successfully used for the detection of recombination breakpoints for linkage analysis (see Chapter I) and QTL mapping (Miller et al. 2007; Laporte et al. 2015), studying genomics of

adaptation (Hohenlohe et al. 2010), inbreeding and genome-wide heterozygosity (Hoffman et al. 2014; see Chapter III) or shading light onto species evolutionary histories through phylogenomcs and phylogeography (Wagner et al. 2013; see Chapter IV).

THESIS OUTLINE

To enrich the available genomic resources for *D. magna*, I have used RAD-sequencing for the construction of a high-density genetic map (**Chapter I**). This third-generation linkage map of *D. magna* includes more than four thousand markers, encompassing 77 % of the genomic sequence currently available (genome assembly v.2.4; *Daphnia* Genomics Consortium) and 55 % of the estimated genome size (238 Mb, Routtu et al. 2014). Since all major scaffolds were covered with multiple markers, I was able to determine their orientation within the chromosomes and their linkage with other genome segments that was previously unknown. Therefore, such a high-density map can considerably assist the on-going improvements of *D. magna* genome assembly. Including a large number of genetic markers also allowed the comparison between genetic and physical distances. This newly gained ability has provided a basis for estimating variation in recombination rate along the chromosomes of *D. magna* for the first time. The results presented in Chapter I of this thesis clearly indicate that meiotic COs are more likely to occur in chromosomal peripheries while appear to be very rare or absent around chromosomal centres which, in *D. magna*, coincide with centromeres (see Chapter III). These novel insights into the recombination landscape of *D. magna* can provide a valuable assistance in future studies of genome architecture, mapping of quantitative traits, and in population genetic studies.

One of the important features of *Daphnia* as a model system is the ability to reproduce asexually (ameiotic parthenogenesis). This enables laboratory maintenance of clonal lines over many generations and "genotype replication" for experimental purposes. However, several studies using asexual mutation accumulation lines of *D. pulex* showed unexpectedly high rates of LOH accompanied by segmental deletions (hemizygosity) and duplications (Xu et al. 2011; Keith et al. 2015). These phenomena indicate the dynamic nature of asexual genomes and might have a strong influence on the evolutionary potential of asexual lineages. One of the possibilities is that ameiotic recombination, resulting in LOH, leads to the unmasking of recessive deleterious mutations. This would impose a strong selective pressure against LOH in asexual lineages, suggesting that the rates of LOH reported for mutation accumulation lines of *D. pulex* might be underestimated. For the assessment of LOH rate in *D. magna*, while minimizing the effect of selection, I have used RAD-sequencing to compare heterozygosity patterns between asexual daughters and their mothers, hence within a single generation of asexual reproduction (**Chapter II**). Even though a substantial number of LOH events were detected using RAD-sequencing, a subsequent validation proved that these LOH events were false positives (i.e., heterozygotes appearing as homozygotes) and most probably reflecting the sequencing errors (allele dropouts) that could not be recognised and rejected through bioinformatics analysis only. I cannot exclude the possibility that ameiotic recombination indeed occurs in *D.magna*, however, results presented in Chapter II indicate that such a phenomenom is extremely rare or restricted to the very short genomic regions which I was not able to investigate, despite a high density of markers utilized.

While RAD-sequencing turned out to be error-prone for detecting LOH in apomixis, a rare phenomenon that is probably restricted to small genomic regions, it is still a highly valuable method for the analysis of genome-wide heterozygosity patterns. Exactly this was the key feature, combined with the availability of a high-density genetic map (Chapter I), that enabled the differentiation between self-fertilization and automixis (meiotic parthenogenesis) in rare hatchlings from isolated non-male producing strains of *D. magna* (**Chapter III**). Since self-fertilization and automixis result in offspring with a similar pattern of overall parental heterozygosity reduction, using small numbers of markers, such as microsatellites, might yield misleading results. However, during automixis, zygosity of centromeric regions should be the same within a single individual (i.e., all centromeric regions should be heterozygous or all homozygous) while in self-fertilization each centromeric region is expected to retain parental heterozygosity or become homozygous independently. Therefore, distinct analysis of inter-chromosomal and intra-chromosomal patterns of genome-wide heterozygosity provided an unequivocal evidence of automixis in *D. magna*. Also, the inter-chromosomal retention (or loss) of heterozygosity during apomixis confirmed centromere locations that were presumed from the variation in recombination rate along chromosomes of *D. magna* (see Chapter I). This study increased the knowledge of the versatility of the possible reproductive routes in *Daphnia*, but also provided a demonstration of the utility of genomic approaches in elucidating breeding systems that are difficult to examine with purely observational data.

Finally, RAD-sequencing European populations of *D. magna* ensured a high resolution analysis that enabled to uncover gene-geography correspondence patterns (**Chapter IV**), a task which was not previously possible using mitochondrial or microsatellite markers. More precisely, identifying tens of thousands SNPs has largely increased the power of detection of genetic differentiation among *D. magna* populations. A principle component analysis (PCA) of genetic variation and Procrustres analytical approach were used to quantify spatial genetic structure across Europe revealing remarkable consistency between the first two PCA axes and the geographic location of individual samples. This study showed that, contrary to earlier studies, *D. magna* indeed has a population genetic structure that is consistent with isolation by distance (IBD), i.e., genetic differentiation of *D. magna* populations across Europe can largely be explained by geographical distance and unimpeded migration at large spatial scales. Furthermore, new insights on the type and magnitude of population genetic structure of *D. magna* will allow a better understanding of the evolutionary history of this intriguing model species.

REFERENCES

Aguilera, A., and B. Gómez-González. 2008. Genome instability: a mechanistic view of its causes and consequences. Nat. Rev. Genet. 9:204–17.

Andersen, S. L., and J. Sekelsky. 2010. Meiotic *versus* mitotic recombination: Two different routes for double-strand break repair: The different functions of meiotic versus mitotic DSB repair are reflected in different pathway usage and different outcomes. BioEssays 32:1058–1066.

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet advance on:81–92.

Archetti, M. 2010. Complementation, genetic conflict, and the evolution of sex and recombination. J. Hered. 101 Suppl :S21–33.

Archetti, M. 2004a. Loss of complementation and the logic of two-step meiosis. J. Evol. Biol. 17:1098–105.

Archetti, M. 2004b. Recombination and loss of complementation: a more than two-fold cost for parthenogenesis. J. Evol. Biol. 17:1084–97.

Barton, N. H., and B. Charlesworth. 1998. Why sex and recombination? Science (80-.). 281:1986–1990.

Bell, G. 1982. The Masterpiece of Nature: The Evolution and Genetics of Sexuality. Croom Helm, London.

Colbourne, J. K., P. D. N. Hebert, and D. J. Taylor. 1997. Evolutionary origins of phenotypic diversity in *Daphia*. Pp. 163–188 *in* T. J. Givnish and K. J. Sytsma, eds. Molecular Evolution and Adaptive Radiation. Cambridge: Cambridge University Press

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Cáceres, L. Carmel, C. Casola, J.-H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Fröhlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V Kriventseva, D. Kültz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V Koonin, E. M. Zdobnov, I. V Grigoriev, M. Lynch, and J. L. Boore. 2011. The ecoresponsive genome of *Daphnia pulex*. Science 331:555–561.

Cutter, A. D., and B. A. Payseur. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14:262–74.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510.

de Massy, B. 2013. Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. Annu. Rev. Genet. 47:563–99.

Debets, F., K. Swart, R. F. Hoekstra, and C. J. Bos. 1993. Genetic maps of eight linkage groups of Aspergillus niger based on mitotic mapping. Curr. Genet. 23:47–53.

Eads, B. D., J. Andrews, and J. K. Colbourne. 2008. Ecological genomics in *Daphnia*: stress responses and environmental sex determination. Heredity 100:184–90.

Ebert, D. 2011. A genome for the environment. Science 331:539.

Ebert, D. 2005. Ecology, Epidemiology and Evolution of Parasitism in *Daphnia*. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology.

Forche, A., D. Abbey, T. Pisithkul, M. A. Weinzierl, T. Ringstrom, D. Bruck, K. Petersen, and J. Berman. 2011. Stress alters rates and types of loss of heterozygosity in *Candida albicans*. MBio 2:e00129–11.

Galimov, Y., B. Walser, and C. R. Haag. 2011. Frequency and inheritance of non-male producing clones in *Daphnia magna*: Evolution towards sex specialization in a cyclical parthenogen? J. Evol. Biol. 24:1572–1583.

Haag, C. R., S. J. McTaggart, A. Didier, T. J. Little, and D. Charlesworth. 2009. Nucleotide polymorphism and within-gene recombination in *Daphnia magna* and *D. pulex*, two cyclical parthenogens. Genetics 182:313–23.

Hairston, N. G. 1996. Zooplankton egg banks as biotic reservoirs in changing environments. Limnol. Oceanogr. 41:1087–1092.

Hartfield, M., and P. D. Keightley. 2012. Current hypotheses for the evolution of sex and recombination. Integr. Zool. 7:192–209.

Havel, J. E., and J. B. Shurin. 2004. Mechanisms, effects, and scales of dispersal in freshwater zooplankton. Limnol. Oceanogr. 49:1229–1238.

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.

Hiruta, C., C. Nishida, and S. Tochinai. 2010. Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. Chromosome Res. 18:833–40.

Hiruta, C., and S. Tochinai. 2012. Spindle assembly and spatial distribution of γ-tubulin during abortive meiosis and cleavage division in the parthenogenetic water flea *Daphnia pulex*. Zoolog. Sci. 29:733–7.

Hoffman, J. I., F. Simpson, P. David, J. M. Rijks, T. Kuiken, M. a S. Thorne, R. C. Lacy, and K. K. Dasmahapatra. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. Proc. Natl. Acad. Sci. U. S. A. 111:3775–80.

Hoffmann, A. A., and M. J. Hercus. 2000. Environmental stress as an evolutionary force. Bioscience 50:217–226.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6.

Huertas, P. 2010. DNA resection in eukaryotes: deciding how to fix the break. Nat. Struct. Mol. Biol. 17:11–16.

Innes, D. J., and L. Dunbrack. 1993. Sex allocation variation in *Daphnia pulex*. J. Anim. Ecol. 6:559–575.

Jones, G. H., and F. C. H. Franklin. 2006. Meiotic Crossing-over: Obligation and Interference. Cell 126:246–248.

Keeney, S. 2008. Spo11 and the formation of DNA double-strand breaks in meiosis. Genome Dyn. Stab. 2:81–123.

Keith, N., A. E. Tucker, C. E. Jackson, W. Sung, J. I. Lucas, D. R. Schrider, and A. J. Younge. 2015. High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. Genome Res., doi:

10.1101/gr.191338.115.

Knudson, A. G. J. 1971. Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. U. S. A. 68:820–823.

Kondrashov, A. S. 1993. Classification of hypotheses on the advantage of amphimixis. J. Hered. 84:372–387.

Korovchinsky, N. M. 1997. On the history of studies on cladoceran taxonomy and morphology, with emphasis on early work and causes of insufficient knowledge of the diversity of the group. Pp. 1–11 *in* Hydrobiologia.

Krejci, L., V. Altmannova, M. Spirek, and X. Zhao. 2012. Homologous recombination and its regulation. Nucleic Acids Res. 40:5795–5818.

Laporte, M., S. M. Rogers, A.-M. Dion-Côté, E. Normandeau, P.-A. Gagnaire, A. C. Dalziel, J. Chebib, and L. Bernatchez. 2015. RAD-QTL Mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in Lake Whitefish (*Coregonus clupeaformis*) species pairs. G3 (Bethesda). 5:1481–91.

Lee, P. S., P. W. Greenwell, M. Dominska, M. Gawel, M. Hamilton, and T. D. Petes. 2009. A fine-structure map of spontaneous mitotic crossovers in the yeast *Saccharomyces cerevisiae*. PLoS Genet. 5.

Lehtonen, J., M. D. Jennions, and H. Kokko. 2012. The many costs of sex. Trends Ecol. Evol. 27:172–178.

Libuda, D. E., S. Uzawa, B. J. Meyer, and A. M. Villeneuve. 2013. Meiotic chromosome structures constrain and respond to designation of crossover sites. Nature 502:703–706.

Lutes, A. A., W. B. Neaves, D. P. Baumann, W. Wiegraebe, and P. Baumann. 2010. Sister chromosome pairing maintains heterozygosity in parthenogenetic lizards. Nature 464:283–286.

Lydeard, J. R., Z. Lipkin-Moore, Y. J. Sheu, B. Stillman, P. M. Burgers, and J. E. Haber. 2010. Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. Genes Dev. 24:1133–1144.

Lynch, M., and W. Gabriel. 1990. Mutation load and the survival of small populations. Evolution 44:1725–1737.

Mandegar, M. A., and S. P. Otto. 2007. Mitotic recombination counteracts the benefits of genetic segregation. Proc. R. Soc. B Biol. Sci. 274:1301–1307.

Mardis, E. R. 2011. A decade's perspective on DNA sequencing technology. Nature 470:198–203.

Maynard Smith, J. 1978. The evolution of sex. Cambridge: Cambridge University Press.

Meirmans, S., P. G. Meirmans, and L. R. Kirkendall. 2012. The costs of sex: Facing real-world complexities. Q. Rev. Biol. 87:19–40.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240–248.

Miner, B. E., L. De Meester, M. E. Pfrender, W. Lampert, and N. G. Hairston. 2012. Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. Proc. R. Soc. B Biol. Sci. 279:1873–1882.

Nachman, M. W. 2002. Variation in recombination rate across the genome: Evidence and implications. Curr. Opin. Genet. Dev. 12:657–663.

Noor, M. A. F., A. L. Cunningham, and J. C. Larkin. 2001. Consequences of recombination rate variation on quantitative trait locus mapping studies: Simulations based on the *Drosophila melanogaster* genome. Genetics 159:581–588.

Omilian, A. R., M. E. A. Cristescu, J. L. Dudycha, and M. Lynch. 2006. Ameiotic recombination in asexual lineages of *Daphnia*. Proc. Natl. Acad. Sci. U. S. A. 103:18638–43.

Otto, S. P. 2009. The evolutionary enigma of sex. Am. Nat. 174 Suppl :S1–S14.

Otto, S. P., and T. Lenormand. 2002. Resolving the paradox of sex and recombination. Nat. Rev. Genet. 3:252–261.

Page, S. L., and R. S. Hawley. 2003. Chromosome choreography: the meiotic ballet. Science 301:785–789.

Pál, C., and L. D. Hurst. 2003. Evidence for co-evolution of gene order and recombination rate. Nat. Genet. 33:392–395.

Paland, S., J. K. Colbourne, and M. Lynch. 2005. Evolutionary history of contagious asexuality in *Daphnia pulex*. Evolution 59:800–813.

Routtu, J., M. D. Hall, B. Albere, C. Beisel, R. D. Bergeron, A. Chaturvedi, J.-H. Choi, J. Colbourne, L. De Meester, M. T. Stephens, C.-P. Stelzer, E. Solorzano, W. K. Thomas, M. E. Pfrender, and D. Ebert. 2014. An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. BMC Genomics 15:1033.

Routtu, J., B. Jansen, I. Colson, L. De Meester, and D. Ebert. 2010. The first-generation *Daphnia magna* linkage map. BMC Genomics 11:508.

San Filippo, J., P. Sung, and H. Klein. 2008. Mechanism of eukaryotic homologous recombination. Annu. Rev. Biochem. 77:229–257.

Schön, I., K. Martens, and P. J. van Dijk. 2009. Lost sex: the evolutionary biology of parthenogenesis. Springer Dordrecht Heidelberg London New York, The Netherlands.

Schwander, T., and B. J. Crespi. 2009. Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. Mol. Ecol. 18:28–42.

Shaw, J. R., M. E. Pfrender, B. D. Eads, R. Klaper, A. Callaghan, R. M. Sibly, I. Colson, B. Jansen, D. Gilbert, and J. K. Colbourne. 2008. *Daphnia* as an emerging model for toxicological genomics.

Sturtevant, A. H. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. J. Exp. Zool. 14:43–59.

Sung, P., and H. Klein. 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat. Rev. Mol. Cell Biol. 7:739–750.

Suomalainen, E., A. Saura, and J. Lokki. 1987. Cytology and Evolution in Parthenogenesis. Boca Raton (FL): CRC Press.

Svendsen, N., C. M. O. Reisser, M. Dukic, V. Thuillier, C. Liautard-haag, D. Fasel, E. Hürlimann, T. Lenormand, Y. Galimov, and C. R. Haag. 2015. Uncovering cryptic asexuality in *Daphnia magna* by RAD-sequencing. Genetics 201:1143–1155.

Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl. 1983. The double-strand-break repair model for recombination. Cell 33:25–35.

Talbert, P. B., and S. Henikoff. 2010. Centromeres convert but don't cross. PLoS Biol. 8:1–5.

Thompson, M. J., and C. D. Jiggins. 2014. Supergenes and their role in evolution. Heredity 113:1–8.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22:787–798.

Wang, S., D. Zickler, N. Kleckner, and L. Zhang. 2015. Meiotic crossover patterns: Obligatory crossover, interference and homeostasis in a single process. Cell Cycle 14:305–314.

West, S. A., C. M. Lively, and A. F. Read. 1999. A pluralist approach to sex and recombination. J. Evol. Biol. 12:1003–1012.

Wilfert, L., J. Gadau, and P. Schmid-Hempel. 2007. Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity 98:189–97.

Woodbine, L., H. Brunton, A. A. Goodarzi, A. Shibata, and P. A. Jeggo. 2011. Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair. Nucleic Acids Res. 39:6986–6997.

Xu, S., M. S. Ackerman, H. Long, L. Bright, K. Spitze, J. S. Ramsdell, W. K. Thomas, and M. Lynch. 2015. A male-specific genetic map of the microcrustacean *Daphnia pulex* based on single sperm whole-genome sequencing. Genetics 201:31–38.

Xu, S., A. R. Omilian, and M. E. Cristescu. 2011. High rate of large-scale hemizygous deletions in asexually propagating *Daphnia*: implications for the evolution of sex. Mol. Biol. Evol. 28:335–42.

Zaffagnini, F. 1987. Reproduction in *Daphnia*. Pp. 245–284 *in* Memorie dell'Istituto Italiano di Idrobiologia.

Zaffagnini, F., and B. Sabelli. 1972. Karyologic observations on the maturation of the summer and winter eggs of *Daphnia pulex* and *Daphnia middendorffiana*. Chromosoma 36:193–203.

Zhang, L., Z. Liang, J. Hutchinson, and N. Kleckner. 2014. Crossover patterning by the beam-film model: analysis and implications. PLoS Genet. 10:e1004042.

Zhu, X., and S. Keeney. 2014. Zip it up to shut it down. Cell Cycle 13:2157–2158.

Chapter I

HIGH-DENSITY GENETIC MAP REVEALS VARIATION IN RECOMBINATION RATE ACROSS THE GENOME OF *Daphnia magna*

Manuscript in submission:
Dukić, M., D. Berner, M. Roesti, C. R. Haag and D. Ebert. 2016. High-density genetic map reveals variation in recombination rate across the genome of *Daphnia magna*.

ABSTRACT

Recombination rate is an essential parameter for many genetic analyses. It is now clear that recombination rates are highly variable across species, populations, individuals and different genomic regions. Due to the profound influence that recombination can have on intraspecific diversity and interspecific divergence, characterization of recombination rate variation emerges as a key resource for population genomic studies and emphasises the importance of high-density genetic maps as tools for studying genome biology. Here we present such a high-density genetic map for *Daphnia magna*, and analyse patterns of recombination rate across the genome. A F2 intercross panel was genotyped by Restriction-site Associated DNA sequencing to construct the third-generation linkage map of *D. magna*. The resulting high-density map included 4037 markers covering 813 scaffolds and contigs that sum up to 77 % of the currently available genome draft sequence (v2.4) and 55 % of the estimated genome size (238 Mb). Total genetic length of the map presented here is 1614.5 cM and the genome-wide recombination rate is estimated to 6.78 cM/Mb. Merging genetic and physical information we found that recombination rate estimates are consistently high towards the peripheral parts of the chromosomes, while chromosome centres, harbouring centromeres in *D. magna*, show very low recombination rate estimates. The third-generation linkage map for *D. magna,* constructed in this study provides an essential tool for genome investigation in this model organism. Due to its high-density, our linkage map can be used for the on-going improvements of the genome assembly, but more importantly, it has enabled us to characterize variation in recombination rate across the genome of *D. magna* for the first time. These new insights can provide a valuable assistance in future studies of the genome evolution, mapping of quantitative traits and population genetic studies.

INTRODUCTION

Meiotic recombination is an essential process in sexually reproducing eukaryotes since it is involved in the maintenance of genome stability, in proper segregation of chromosomes into haploid gametes, and in shaping patterns of genetic variation among offspring individuals (Baudat et al. 2013). Mechanistically, recombination between homologous chromosomes is crucial for accurate repair of DNA double strand breaks that are induced in a highly controlled manner during early meiotic prophase I (reviewed in Gerton and Hawley 2005; Baudat et al. 2013). Such homology-based repair ensures the maintenance of genome integrity, but also often represents a physical bond between homologous chromosomes, critical for their positioning and proper segregation into the gamete cells (Gerton and Hawley 2005). In proceeding meiotic processes, physical connection between homologs will lead to reciprocal (crossover; CO) or unidirectional (gene conversion) exchange of DNA between paternal and maternal chromosomes. It has been shown in several organisms that more abundant and uniformly distributed gene conversions have a limited influence on inherited genetic variation as they affect small genomic regions (350-2000 bp; Padhukasahasram and Rannala 2013). On the other hand, CO involves reciprocal allelic exchange across longer chromosomal segments resulting in recombination of genetic variation that can be readily detected following the inheritance of genetic markers in large pedigrees or experimental crosses. Consequently, recombination rate is traditionally approximated as the observed frequency of COs (i.e. neglecting gene conversions) per unit of physical distance (e.g. cM/Mb).

Advancements in sequencing techniques have revolutionized genome research over the last decade, including valuable insight concerning recombination rate in many different species (Jensen-Seaman et al. 2004; McVean et al. 2004; Chan et al. 2012; Roesti et al. 2013). Importantly, there is accumulating evidence for large amounts of variation in recombination rate across species, populations, individuals, and different genomic regions (McVean et al. 2004; Smukowski and Noor 2011; Chan et al. 2012). This is especially interesting from an evolutionary perspective since the distribution of recombination events across the genome defines the size of genomic fragments that will be incorporated into haplotypes exposed to selection. When recombination is rare, selection wields its influence across long genomic tracts that may contain multiple loci with differing fitness effects. Theory predicts that genetic linkage between multiple sites under selection leads to a reduction of the overall efficiency of selection (Hill and Robertson 1966; Comeron et al. 2008) and high levels of CO recombination are considered favourable for breaking up association between loci subjected to contrasting selective pressures (Barton 2010). In addition, genomic regions with low recombination are expected to have lower levels of neutral polymorphism than genomic regions with high recombination rates because of positive (hitch-hiking) or negative (background) selection on sites in their physical neighbourhood (Cutter and Payseur 2013). Considering the profound influence that the recombination landscape can have on genome-wide genetic variation and diversity, analysis of the recombination rate emerge as a key resource for population and evolutionary genomics studies, emphasising the importance of high-density genetic maps as essential tools for studying many features of genome biology.

Waterfleas of the genus *Daphnia* have emerged as a well-suited model system for studying genetics of fitness related traits in environmental contexts, due to the extensive knowledge of their ecology, a life-cycle including clonal and sexual reproduction, and the development of genomic resources (Eads et al. 2008; Colbourne et al. 2011; Ebert 2011). However, to take full advantage of this model-system, a better understanding of the genome architecture of *Daphnia* is needed, as well as of the mechanisms that are shaping it. In the present study, we use a standard F2 intercross panel and Restriction-site Associated DNA (RAD) sequencing for the construction of a high-density genetic map of *D. magna,* one of the best known and most widely used study species of the genus. Including more than four thousand markers across the 238 Mb genome (Routtu et al. 2014), we provide the first characterization of recombination landscape along the chromosomes of *D. magna*. Our data clearly show high levels of recombination towards chromosomal peripheries with chromosomal centres being almost deprived of CO. We discuss these findings in comparison with other organisms and address possible mechanisms underlying the observed patterns of recombination rate variation across the genome of this species.

MATERIALS & METHODS

Design of genetic crosses and DNA isolation

D. magna individuals used in this study were obtained by asexually (clonally) propagating lines selected from an F2 intercross panel that had already been used for the construction of microsatellite and SNP-array based genetic maps (Routtu et al. 2010, 2014) as well as for QTL mapping (Roulin et al. 2013; Routtu and Ebert 2014). Details about the crossing design can be found in these papers. Briefly, the F2 panel was established by crossing individuals from two inbred clonal lines of *D. magna* (Xinb3 and Iinb1, hereafter referred as parental lines). One of the parental lines (Xinb3) was a third-generation inbred offspring (three rounds of within-clone mating, each round being genetically equivalent to self-fertilization) of a clone from Southern Finland, the other (Iinb1) was a first-generation inbred offspring of a German clone. A female from the Xinb3 and a male from the Iinb1 parental line were crossed to obtain IXF1 hybrid line. By mating genetically identical brothers and sisters of the IXF1 hybrid, F2 offspring were generated, with each initial offspring individual (hatchling from a sexually produced egg) being a founder of a clonal F2 line that was maintained via asexual reproduction as a part of our F2 intercross panel. The Xinb3 line is also the clone on which the *D. magna* reference genome is based (Daphnia Genomics Consortium). The draft genome sequence version 2.4 was used in the present study. Finally, 2-3 females from each parental line, the IXF1 line, and 66 randomly chosen F2 lines were used to establish asexually propagated sub-lines that were used for DNA extractions (pooling nine individuals for each line).

Prior to DNA extractions, all individuals were cleaned by an antibiotic and starvation treatment to minimize algal and bacterial contamination in the sample of genomic DNA. Animals were kept without food during three days in a medium containing Ampicillin (Sigma), Streptomycin (Sigma) and Tetracycline (Sigma) at a concentration of 50 mg/L each, and transferred daily to fresh antibiotic medium. To enforce the cleansing of gut contents, a

small amount of superfine Sephadex ® G-25 (Sigma-Aldrich) was added frequently to the antibiotic medium, making dextran beads accessible to *Daphnia* for ingestion and gut evacuation. Animals with clear intestine were sampled and used for DNA extractions. In the majority of cases, DNA was isolated immediately after sampling, but in some instances, animals were stored in 70 % ethanol at -20 ˚C until extraction. DNA extraction was done using the DNeasy Blood and Tissue kit (Qiagen) including the RNaseA (100 mg/ml; Sigma) digestion step.

RAD library preparation and sequencing

We prepared libraries for RAD-sequencing (Baird et al. 2008) adopting the protocol of Etter et al. (2011) with modifications according to Roesti et al. (2013). Specifically, 1 μg of genomic DNA from each sample was digested with the *PstI* HF restriction enzyme (NEB) in 50 μl reaction volume, for 90 min. at 37 ˚C and then heat-inactivated following the manufacturer's manual. A P1 sequencing adapter (5 μl of 100 nM stock solution), containing a unique 5-bp barcode, was ligated to each sample using T4 DNA-ligase (NEB, 0.5 μl of 2,000,000 units/mL stock solution) in 60 μl reaction volume for 45 min at room temperature followed by heat-inactivation for 20 min at 65 ˚C. The total of 70 samples (Xinb3, Iinb1, IXF1 and 66 F2 lines, with one F2 individual replicated twice) were then combined into 2 pools (one with 30 and one with 40 samples) and sheared using a Bioruptor (Diagenode). The rationale of combining fewer individuals into the first pool, which included the parental, IXF1 and 26 F2 lines ("parental" library), was to ensure higher sequencing depth and genotyping quality for the founder individuals of the F2 panel, thus facilitating the robust identification of informative SNPs for genetic mapping. The second library contained F2 lines exclusively.

DNA fragments in a range of 250-500 bp were selected using agarose gel electrophoresis (1.25 %, 0.5X TBE), purified and blunt-ended (Quick Blunting Kit, NEB). Klenow fragment exo$^-$ (NEB) was used to add dA-overhangs, followed by P2 adapter ligation (1 μl of 10 mM stock solution). Products were purified and PCR amplification was done using Phusion High-Fidelity DNA polymerase (NEB). To minimize the probability of PCR error, master mixes for each library were divided into six separate 12.5 μl reactions for amplification (30 s at 98 ˚C, 17 cycles of 98 ˚C 10 s, 65 ˚C 30 s, 72 ˚C 30 s, then a final extension for 5 min at 72 ˚C).

The enriched RAD libraries were sequenced on separate Illumina HiSeq2000 lanes using 100 bp single-end sequencing (Quantitative Genomics Facility service platform, Deep Sequencing Unit Department of Biosystems Science and Engineering, ETH-Zurich in Basel, Switzerland).

Defining genetic markers for linkage mapping

In total 259,580,561 raw 100 bp reads were generated by sequencing (120,336,323 and 139,244,238 reads in the first and the second library, respectively). Overall read quality was inspected using FastQC (Babraham Bioinformatics, The Babraham Institute). Custom scripts coded in R (Gentleman et al. 2004) were used to sort raw reads according to unique barcodes into individual samples. Low-quality reads (containing ambiguous bases) and reads that did not feature valid barcode or restriction-site sequence were discarded from further analysis (23 % of the total raw reads). Cleaned and sorted reads were aligned to the reference draft

genome assembly v2.4 of *D. magna* using Novoalign v2.07 (http://novocraft.com). We allowed on average one high-quality mismatch per 14 bases and accepted only reads that aligned to unique location within the reference. Eight F2 samples were discarded because they were sequenced at substantially lower depths compared to the other samples within the same library. In summary, we achieved a mean coverage of 68-fold among the individuals from the parental library (including 22 F2 lines), and 40-fold among the final 37 F2 lines from the second library.

Stacks v1.08 (Catchen and Hohenlohe 2013) was used for identification of putative marker loci and for genotyping. The samples from the two libraries were analysed separately, taking the differences in sequencing depth into account. In both cases, SAM files were analysed with the ref_map.pl pipeline and parental lines were used to construct a "catalog" of loci (3 mismatches allowed between reads mapping to the same locus). The minimum coverage depth was set to 25 (parental library) and 15 (lower-coverage library) to call a stack (group of identical reads). Custom MySQL scripts were used for merging the results from both libraries. Deleveraged loci (see Catchen and Hohenlohe 2013) and loci with more than 2 alleles were excluded from the analysis as we were interested only in loci that were homozygous for alternative alleles (aa, bb) in the parental lines and heterozygous (ab) in the F1 hybrids. In total, haplotype and genotype data for 7183 putative markers were retrieved from the Stacks analysis.

We inspected the obtained genotypes for biases, which are potential source of errors during linkage map construction. This resulted in the removal of six F2 lines from further analysis because they had more than 30 % of missing genotypes. Furthermore, we removed markers exhibiting more than 20 % missing data across the F2 lines, as suggested by Catchen and Hohenlohe (2013) and Davey et al. (2013). The resulting dataset comprised 52 F2 lines and 4849 genetic markers in total.

Linkage analysis

JoinMap 4.0 (Van Ooijen 2006) was used as the main software for genetic map construction. However, several additional steps were taken to maximize the number of markers that could be mapped and to avoid the reduction in mapping accuracy that is expected in large datasets (> 1000 markers; Cheema and Dicks 2009). First we selected a subset of 253 "anchor" markers representing large scaffolds of the *D. magna* draft genome assembly (v2.4). Using the regression mapping algorithm with default parameters in JoinMap, these markers were grouped into 10 preliminary linkage groups (LGs) at LOD = 3. Assuming no assembly error at this point of the analysis, all other markers on the same scaffolds were attributed to the same preliminary LG as the respective anchor marker. We then continued to expand the preliminary LGs by performing contingency table analyses of segregation patterns. More precisely, we compared terminal markers of scaffolds that were attributed to one of the preliminary LGs, against the dataset of so-far un-mapped markers. Only markers with very similar segregation patterns (< 3 different genotypes among the 52 F2 lines) were assigned to the same preliminary LG (Pearson's $\chi 2$ test, cut-off threshold $P<0.0001$), whereas markers showing ambiguous association to two or more LGs were discarded at this point of the analysis. Markers with the extreme segregation ratio distortion (SRD) that contradicted surrounding markers within the same scaffold were removed. Following this procedure, 4045

markers were assigned to one of the ten preliminary LGs, 75 markers were removed while 729 markers remained unattributed.

Many markers included in preliminary map showed identical segregation patterns across all F2 individuals (i.e. they did not show any evidence of CO recombination). In total, 756 segregation patterns could be distinguished within our preliminary dataset and the groups of co-segregating markers are hereafter referred to as "bins" (1 to 384 markers per bin). One of the markers exhibiting the lowest number of missing genotypes from each bin was denoted as "frame marker" and was used for creating a framework map, a non-redundant representation of all detected segregation patterns suitable for further analysis with JoinMap. The grouping of frame markers into 10 LGs was confirmed at LOD = 7 (maximum likelihood, ML, option and otherwise default parameter values of the program). We then continued by iteratively adding sets of the remaining, unattributed markers to the preliminary map. In each round, newly grouped markers were designated as frame or non-frame markers, depending on whether their genotypes matched one of the previously defined bins. Non-frame markers were continuously omitted from the framework map and kept separately for later construction of a composite map. After several iterations, we managed to include a total of 4761 markers in the composite map while 13 markers did not map to any of the ten LGs and consequently, were omitted from the final dataset. Once all markers were included, the composite map was inspected visually, and the ordering of the markers within the LGs was corrected, based on the available information of physical position within the scaffolds (mostly applying to markers within the same bin, the position of which could not be determined based on segregation patterns). Dubious genotypes were corrected, making the assumption that the vast majority of singletons reflect genotyping errors rather than double CO within short physical distance (i.e. between the focal marker and the adjacent markers on both sides). Thus, if the genotype was not observed in at least 3 adjacent markers within the same scaffold, it was replaced with a missing value (Isidore et al. 2003). We also checked marker pairs obtained from sister RAD-tags (i.e. markers obtained from RAD loci flanking the same *PstI* restriction site, hence with a distance of <200 bp) and removed one marker of the pair as redundant. If both sister RAD-tags were highly reliable markers (up to three missing values), the consensus segregation pattern was kept (thus reducing the number of missing genotypes in the data set). If the RAD-tag pair showed inconsistent genotypes within the same F2 individual, these instances were replaced with missing values as it is highly improbable that a recombination event happened within such a short distance.

The final composite map comprised 4037 markers, out of which 952 were defined as frame markers (952 bins with 1 to 354 markers). Grouping and ordering of markers within the framework map was confirmed using the ML algorithm at the LOD = 7. Afterwards, each LG was analysed individually with markers in fixed order and genetic distances were calculated using the Kosambi mapping function. Furthermore, the mapping quality of the framework map was validated through an independent approach using the CheckMatrix program (http://www.atgc.org/XLinkage/).

Estimating physical distances between markers

The current version (v2.4) of the *D. magna* genome is a still unfinished draft version. Hence, we used the following procedures to estimate the physical distances between markers and the

cumulative physical length of each LG: (i) Mapped scaffolds were considered oriented if they had two or more markers separated by at least one recombination event (so the orientation of the scaffold ends could be estimated). Within oriented scaffolds, the distances between markers were known from their alignment position while distances between two terminal markers of adjacent scaffolds were calculated based on the position of markers within their scaffolds and the number of remaining base pairs up to the scaffold's ends. Note that this assumes no gaps between adjacent scaffolds (see below). (ii) Scaffolds and contigs with only one marker or without detected recombination events were designated as un-oriented. Physical lengths of un-oriented regions were estimated based on the sum of the total lengths of scaffolds and contigs included in those regions. Distances between markers within the non-recombining region were attributed an average value (estimated physical length divided by the number of segments defined by markers). This was done because it was unknown which end segment was adjacent to the next oriented scaffold. (iii) When small contigs mapped inside a longer, oriented scaffolds, their size was not considered, as it was assumed that these contigs mapped to the region of uncertain nucleotides (Ns) inside the scaffold. Such regions are present on all scaffolds due to paired-end sequencing with long, un-sequenced inserts.

Analysis of recombination rates

R/qtl (countXO function, Broman and Sen 2009) was used to count the recombination breakpoints observed in each F2 for each LG. Recombination breakpoints were detected as a change in a genotype along the LGs. More precisely, observed genotype transitions A → H, H → A, B → H or H → B were counted as a single recombination breakpoint, while double breakpoints between successive markers would appear as A → B or B → A genotype transitions ("A" being homozygote for the alleles from the German father clone, "B" is homozygote for the alleles from the Finnish mother clone while "H" annotates heterozygote genotype). The mean number of recombination breakpoints observed in F2 offspring corresponds to the expected mean number of COs during meiosis, averaged across males and females.

Genome-wide recombination rate (GWRR) was calculated by summing cumulative genetic distances of all LGs and dividing it by the most recent estimate of the total length of the *D. magna* genome (238Mb; Routtu et al. 2014). An average recombination rate for each LG (chromosomal recombination rate) was estimated in the same way but we used the physical length that was based only on scaffolds included in our map (see above). We calculated the intra-chromosomal (local) recombination rate between each pair of adjacent markers as the ratio of genetic distance and estimated physical distance between those markers (cM/Mb). Marey maps were used to plot genetic distance (in cM) against physical distance (in Mb) and to visualise variation in recombination rates along LGs (Rezvoy et al. 2007). In addition, local recombination rates were plotted against the physical midpoints of marker intervals, and LOESS (locally weighted scatterplot smoothing) was used for smoothing the estimated values (polynomial degree = 1, α value was adjusted to the density of markers in each linkage group to cover approximately 2 Mb windows). It is important to note here that the chromosomal and also some of the intra-chromosomal recombination rates are probably overestimates because the mapped scaffolds of the reference genome assembly do not cover

the full genomic sequence of *D. magna* (only 131 Mb in total). Therefore, the physical distances used here have to be considered as minimum estimates.

GC content analysis

To test whether the sequence composition is associated with the recombination landscape in *D. magna*, we investigated how GC content correlates with differences in recombination rate. All analyses of the GC content were done using the available reference genome sequence (v.2.4). At the chromosomal scale, we tested for differences in sequence composition of scaffolds found in recombining vs. non-recombining regions (see results): We compared the average GC content of all scaffolds mapping to regions of low recombination with the ones mapping to regions with high recombination, omitting scaffolds found at the borders of these regions. Furthermore, to assess whether the magnitude of recombination rate correlates with GC content in more discrete intervals, we restricted our analysis to recombining regions only. For this, the two longest scaffolds of each LG were selected and the GC content was extracted for each interval between two markers for which local recombination rate was estimated (interval size between 5 and 100 kb, depending on the spacing between markers).

RESULTS

Linkage map

The genetic map of *D. magna* constructed in this study includes 4037 markers, assigned to ten LGs that correspond to the ten chromosomes of *D. magna* (n = 10). 952 clusters of co-segregating markers (bins) were identified, and only one marker from each cluster was used to assemble a framework map ("Frame" markers; Table 1.). The cumulative genetic lengths (Kosambi corrected) estimated for each LG ranged from 205.4 cM for LG1 to 131.4 cM for LG10, with the total map spanning 1614.5 cM (Table 1). LGs were numbered according to their genetic length estimated in this study (from largest to smallest); not exactly corresponding to the previously published *D. magna* linkage maps (Routtu et al. 2010; Routtu et al. 2014). The average genetic distance between frame markers was 1.7 cM with 78 % of the distances being under 2 cM and the largest gap being 14.5 cM (LG3, Figure 1a), possibly corresponding to a region with a large assembly gap. The independent validation of the framework map is shown as a heatplot (Figure 1b) with clearly visible LG borders and a red diagonal area, which is generally considered as a sign of high mapping quality (http://www.atgc.org/XLinkage/).

Three regions showing significant segregation ratio distortion (SRD) were identified. A region spanning 0.77 Mb within LG5 has been described previously, and is due to an allele responsible for the "Unviable Eggs" phenotype (Routtu et al. 2010). Homozygotes for the alleles from the Finnish mother individual (hereafter B alleles) are highly underrepresented in this region, with complete deficiency located at 80.01 cM (within scaffold00084). Another region, carrying the infertility allele responsible for the "Red Dwarf" phenotype (Routtu et al. 2010) also displayed SRD in our analysis. This region spans approximately 0.69 Mb within LG10 and shows complete deficiency of homozygotes for the alleles originating from the German father individual (A alleles) at 72.29 cM (within scaffold01036). In addition to these

previously described regions, we also found a relatively small region with SRD, spanning 0.15 Mb on LG7 (at 81.92 cM). However SRD in this region was weaker than in the two above regions as none of the two homozygotes was completely absent. Nevertheless, the strong deficiency of BB homozygotes in this region (4 % genotype frequency among F2 offspring) suggests the presence of a strongly deleterious, recessive allele in the Finnish mother clone.

Genome coverage and scaffold mapping

The total size of the *D. magna* genome is estimated at 238 Mb (Routtu et al. 2014). The draft genome assembly used in this study (v2.4) comprises 40,356 scaffolds and contigs summing up to 131,266,987 bp of genomic sequence (55 % of the estimated genome size). 813 scaffolds and contigs were incorporated in this map (Table 2); this fraction, however, represents 77 % (100,609,459 bp) of the sequence currently assembled and 42 % of the estimated genome size. The high density of markers enabled us to determine the orientation of 97 scaffolds (representing 63,321,641 bp, i.e. 48% of the reference genome; Table 2). We found only five scaffolds exhibiting inconsistency between the physical position of markers in the current assembly and their segregation pattern. In all instances, these scaffolds comprised two fragments mapping to separate LGs or to different regions of the same LG (Table 3), while the ordering of markers within these fragments remained consistent. These few discrepancies likely indicate errors in the reference genome assembly. Nevertheless, the small portion of scaffolds displaying putative assembly mismatches indicates an overall high quality of the draft genome assembly used here. In addition, scaffold01409 and scaffold01036, spanning parts of the SRD region on LG10 (see above), showed partial overlap, probably due to our inability to precisely map the markers within the region showing SRD.

Recombination Rate Estimates

A total of 1564 recombination breakpoints were detected across all F2 individuals and across all LGs. The number of detected recombination breakpoints per individual and LG mainly lies between zero and six, with an average of three and a maximum of 14 (Figure 2). These are counts of the number of recombination breakpoints observed in F2 offspring, the average of which also estimates number of CO events that occurred during meiosis in F1, averaged across male and female meiosis. The variance in F2 recombination breakpoints and CO numbers during F1 meiosis is, however not equal as can be seen from the following consideration: If each chromosome pair undergoes exactly 1 CO per meiosis, 50 % of the resulting gametes will have one recombination breakpoint and the other 50 % will have zero. If these gametes are randomly combined to form F2 individuals, the number of recombination breakpoints in F2 individuals is the sum of those on the two gametes. Hence 25 % of the F2 individuals would have two recombination breakpoints (if each of the two gametes has one), 50 % would have one and 25 % would have zero. Hence the observation that no recombination breakpoint was observed on some LGs in some individuals (see Figure 2) does not mean that zero CO occurred in F1 meiosis during gamete formation that gave rise to these individuals.

A genome-wide recombination rate (GWRR) of 6.78 cM/Mb was calculated based on the ratio of the total cumulative genetic map length (1614.48 cM) and the estimated genome size of *D. magna* (238 Mb; Routtu et al. 2014). For an estimation of the GWRR based on the

genome length that was effectively covered by our markers, we used the total length of the current genome assembly (131 Mb) and accordingly obtained a substantially higher estimate of 12.32 cM/Mb. Due to the gaps within the genome assembly, the later GWRR value has to be regarded as an overestimate.

However, assuming that the missing genomic sequence is uniformly distributed among chromosomes (largely in heterochromatic regions), we can make comparisons between recombination rates estimated for each LG. Genetic length increases linearly with the estimated physical length of each LG (Figure 3a) with an intercept larger than zero, indicating that even the smallest chromosomes harbour at least one CO. Consequently, smaller chromosomes display more recombination per unit of physical distance resulting in strong negative correlation between recombination rate and the estimated physical length of LGs (Figure 3b; Pearson's correlation; R= -0.839; n=10; P<0.002).

Recombination rate varied extensively within LGs (Figure 4). In each of the 10 LGs, we detected one large region (two in the case of LG3) where recombination was rare or apparently absent. These low-recombination regions are situated mainly in the chromosomal centres and comprise up to 40 % of the mapped genomic sequence. In all cases (except only one of the two regions of LG3), these regions span the map position of the centromere (Svendsen et al. 2015). In each LG the low-recombination regions are flanked by regions of high recombination. Furthermore, we observed a drop in recombination rates towards the very ends of the LGs. However, due to the current state of the genome assembly and the generation of markers sensitive to sequence motifs (RAD), these terminal regions were difficult to study in more detail.

GC content analysis

We found no difference in the mean GC content between the scaffolds mapping to low-recombination regions and the ones located in regions with high recombination (Paired *t*-test; n = 10; P = 0.97). Focusing only on scaffolds in highly recombining regions, we found a weak positive correlation between GC content and recombination rate (Pearson's correlation; r = 0.184; n = 907 marker intervals; $P<10^{-8}$).

DISCUSSION

We present a high-density genetic map for *D. magna* that can be coupled with the draft genome assembly, thus providing a valuable resource for genomic investigation and QTL mapping. In contrast to the previously published maps for *D. magna*, all large scaffolds in our map are covered by multiple markers, enabling us to determine their orientation within the chromosome (unless situated in a non-recombining region) and the linkage to other genome segments, which were not previously known. Thus, the linkage map constructed in this way can be used for the on-going *D. magna* genome assembly. Co-segregating markers were used to confirm that the observed patterns of segregation are true biological events rather than methodological artefacts. Hence, although a relatively small number of F2 lines was included in our study, the accuracy of final ordering of markers within and between scaffolds is likely high, much higher compared to previous maps, which were based on few microsatellites (Routtu et al. 2010) or an error-prone SNP-array (Routtu et al. 2014). In addition to

increased reliably, the here presented third-generation linkage map enables merging of genetic and physical information, and therefore addressing the variation in recombination rate across the genome of *D. magna* for the first time. This is also the most comprehensive study of recombination landscape for any crustacean species reported so far.

Genome-Wide and Chromosomal Recombination Rate

The genome-wide recombination rate (GWRR) of *D. magna* as estimated in the present study is 6.8 cM/Mb, which is slightly higher than the value of 6.2 cM/Mb assessed from the previously published SNP-based map (Routtu et al. 2014). Similarly, the GWRR of the related species *D. pulex* is estimated at 7.2 cM/Mb (Xu et al. 2015), suggesting conserved levels of recombination among *Daphnia* species. Much lower GWRRs were reported for a handful of crustacean species for which genetic maps and genome size estimates are available (mean = 1.2 cM/Mb; Du et al. 2010; You et al. 2010; Foley et al. 2011; De Vos et al. 2013). Also compared to other animal taxa, GWRR of *Daphnia* is high, similar to some Hymenoptera and Lepidopteran species (Wilfert et al. 2007). It has been hypothesized that the elevated GWRRs are favoured in systems with reduced opportunity for sex and recombination including haplodiploidy, cyclic parthenogenesis or species where recombination is restricted to one sex (Wilfert et al. 2007). However, many exceptions from this pattern (Wilfert et al. 2007; Niehuis et al. 2010; Foley et al. 2011) indicate that peculiar life-cycles *per se* are likely not the only explanation for high recombination rates.

More consistently, it has been shown that recombination rate scales negatively with genome size in many organisms, mainly due to the fact that majority of species have at least one COs per chromosome, even on the smallest chromosomes (Lynch 2006). Consistent with this, we found that the positive linear relationship between genetic distance and physical distance of chromosomes in *D. magna* has a positive y-intercept, and, hence, smaller chromosomes experience more recombination per physical distance when compared to larger ones.

The mean numbers of observed recombination breakpoints in F2 individuals, as well as the estimated genetic map length per chromosome indicate that the different chromosomes of *D. magna* undergo on average between 2.6 and 4.1 CO per meiosis (one expected CO corresponds to 50 cM of genetic map length). It is also interesting to notice that the number of detected recombination breakpoints varies considerably between F2 offspring and individual chromosomes. In 4.8 % of all cases, no recombination breakpoints were detected along an entire LG within a given individual. These instances likely represent the chance union of two gametes that were non-recombinant for this linkage group. Such gametes occur even in meioses with one or several COs and therefore do not represent evidence for meioses without CO. Furthermore, we may have missed some breakpoints when they were too closely spaced or when they occurred in the terminal chromosome regions, i.e. peripheral to the last marker. However, we believe that this would only explain part of the cases without any detected breakpoints. On the other extreme, a few individuals had very high number of recombination breakpoints along a given LG (up to fourteen). These may suggest a rather high variance in the number of COs per meiosis. We cannot fully rule out that they are the result of genotyping errors. Overall, these instances (in both directions) are, however, rare and hence it is unlikely that they significantly influence the summary statistics on the overall genetic map length presented here.

Local recombination rates

The genetic map of *D. magna* described in this study revealed major intra-chromosomal variation in recombination rates. The determinants of non-random CO patterning are not yet clear, though several lines of evidence indicate that the hierarchical combination of multiple factors plays a role in shaping the recombination landscape across genomes. These factors include chromosomal size and structural properties, large subchromosomal domains, chromatin structure and the local nucleotide composition (Libuda et al. 2013; Zhu and Keeney 2014). In *D. magna* we found that, CO recombination is more likely to occur in the peripheral parts of the chromosome, while large regions of low or no recombination occur near the central parts of all chromosomes. As for many species that were studied earlier (Jensen-Seaman et al. 2004; Wong et al. 2010; Davey et al. 2011; Salomé et al. 2012; Farré et al. 2013), these regions of extremely reduced recombination coincide with centromeres of *D. magna* (Svendsen et al. 2015). LG3 is an exception because two regions without recombination were detected, though only one of these two regions (the one at 96 cM, also containing a centromere) was also found by Svendsen et al. (2015). The second non-recombining region on this LG might be the result of an inversion or a large indel suppressing recombination specifically in the inter-population cross used for the present study. It is also interesting to note that regions without recombination probably extend to the pericentromeric heterochromatin regions because centromeric regions are usually not included in genetic maps due to the repetitive nature of their sequence (tandem sequence repeats).

Along with the structural confines on recombination landscape, in the majority of animal species that were studied hitherto it has been shown that recombination rates covary with the local nucleotide composition (Nachman 2002; Beye et al. 2006; Backström et al. 2010; Niehuis et al. 2010; Tortereau et al. 2012; Roesti et al. 2013). High GC content is considered as a predictor of regions with high recombination rate due to involvement of GC-rich elements in the process of recombination (recognition sites of DNA binding proteins) or, conversely, high recombination rates can lead to high GC content due to GC-biased gene conversions that accompany CO events. In *D. magna* there is no difference in GC content between recombining and non-recombining regions. Within the recombining regions, we found that GC content indeed correlates positively with recombination rate, although the detected correlation is weak. These findings are not surprising considering that the correlation between nucleotide composition and recombination rate occurs at very small physical scales, so testing for this association is strongly dependent on the interval size used and on the precision at which recombination hotspots can be identified.

Future perspective for genomics studies in *Daphnia*

Due to the high density of markers included, the genetic map presented here has enabled us to investigate how CO varies in frequency and distribution along the chromosomes of *D. magna*. We have identified large regions of low or no recombination in the chromosomal centres covering approximately 40% of the mapped genome. These regions also contain the centromeres, but likely extend much beyond the actual centromeric regions. In contrast, CO recombination occurs mainly towards the chromosomal peripheries. These insights into the recombination landscape of *D. magna* can provide a valuable assistance in future studies of

the genome architecture, mapping of quantitative traits and population genetic studies. Following improvements in genome annotation, it will be important to understand how gene density correlates with variation in recombination rate. Both the density of sites potentially under selection and the variation in CO rate across the genome can bias genomic analyses (Noor et al. 2001; Cutter and Payseur 2013) and should be considered as important factors in QTL mapping protocols or population genetic studies aiming to understand the effects of selection on genetic variation within and between populations of *D. magna*.

REFERENCES

Backström, N., W. Forstmeier, H. Schielzeth, H. Mellenius, K. Nam, E. Bolund, M. T. Webster, T. Ost, M. Schneider, B. Kempenaers, and H. Ellegren. 2010. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. Genome Res. 20:485–95.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376.

Barton, N. H. 2010. Mutation and the evolution of recombination. Philos. Trans. R. Soc. B Biol. Sci. 365:1281–1294.

Baudat, F., Y. Imai, and B. de Massy. 2013. Meiotic recombination in mammals: localization and regulation. Nat. Rev. Genet. 14:794–806.

Beye, M., I. Gattermeier, M. Hasselmann, T. Gempe, M. Schioett, J. F. Baines, D. Schlipalius, F. Mougel, C. Emore, O. Rueppell, A. Sirviö, E. Guzmán-Novoa, G. Hunt, M. Solignac, and R. E. Page. 2006. Exceptionally high levels of recombination across the honey bee genome. Genome Res. 16:1339–44.

Broman, K. W., and S. Sen. 2009. A Guide to QTL Mapping with R / qtl. Springer, New York City, New York.

Catchen, J., and P. Hohenlohe. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

Chan, A. H., P. A. Jenkins, and Y. S. Song. 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. PLoS Genet. 8:e1003090.

Cheema, J., and J. Dicks. 2009. Computational approaches and software tools for genetic linkage map estimation in plants. Brief. Bioinform. 10:595–608.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Cáceres, L. Carmel, C. Casola, J.-H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Fröhlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V Kriventseva, D. Kültz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V Koonin, E. M. Zdobnov, I. V Grigoriev, M. Lynch, and J. L. Boore. 2011. The ecoresponsive genome of *Daphnia pulex*. Science 331:555–561.

Comeron, J. M., A. Williford, and R. M. Kliman. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity 100:19–31.

Cutter, A. D., and B. A. Payseur. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14:262–74.

Davey, J. W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi, and M. L. Blaxter. 2013. Special features of RAD Sequencing data: implications for genotyping. Mol. Ecol. 22:3151–64.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510.

De Vos, S., P. Bossier, G. Van Stappen, I. Vercauteren, P. Sorgeloos, and M. Vuylsteke. 2013. A first AFLP-based genetic linkage map for brine shrimp *Artemia franciscana* and its application in mapping the sex locus. PLoS One 8:e57585.

Du, Z.-Q., D. C. Ciobanu, S. K. Onteru, D. Gorbach, A. J. Mileham, G. Jaramillo, and M. F. Rothschild. 2010. A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*. Anim. Genet. 41:286–294.

Eads, B. D., J. Andrews, and J. K. Colbourne. 2008. Ecological genomics in *Daphnia*: stress responses and environmental sex determination. Heredity 100:184–90.

Ebert, D. 2011. A genome for the environment. Science 331:539.

Etter, P. D., S. Bassham, P. A. Hohenlohe, E. A. Johnson, and W. A. Cresko. 2011. SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. Pp. 157–178 *in* V. Orgogozo and M. V. Rockman, eds. Molecular Methods for Evolutionary Genetics.

Farré, M., D. Micheletti, and A. Ruiz-Herrera. 2013. Recombination rates and genomic shuffling in human and chimpanzee - A new twist in the chromosomal speciation theory. Mol. Biol. Evol. 30:853–864.

Foley, B. R., C. G. Rose, D. E. Rundle, W. Leong, G. W. Moy, R. S. Burton, and S. Edmands. 2011. A gene-based SNP resource and linkage map for the copepod *Tigriopus californicus*. BMC Genomics 12:568. BioMed Central Ltd.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5:R80.

Gerton, J. L., and R. S. Hawley. 2005. Homologous chromosome interactions in meiosis: diversity amidst conservation. Nat. Rev. Genet. 6:477–487.

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.

Isidore, E., H. Van Os, S. Andrzejewski, J. Bakker, I. Barrena, G. J. Bryan, B. Caromel, H. Van Eck, B. Ghareeb, W. De Jong, P. Van Koert, D. Milbourne, E. Ritter, J. R. Van Der Voort, and R. Waugh. 2003. Toward a marker-dense meiotic map of the potato genome : Lessons from linkage group I. Genetics 165:2107–2116.

Jensen-Seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin, C.-F. Chen, M. A.

Thomas, D. Haussler, and H. J. Jacob. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 14:528–38.

Libuda, D. E., S. Uzawa, B. J. Meyer, and A. M. Villeneuve. 2013. Meiotic chromosome structures constrain and respond to designation of crossover sites. Nature 502:703–706.

Lynch, M. 2006. The origins of eukaryotic gene structure. Mol. Biol. Evol. 23:450–468.

McVean, G. A. T., S. R. Myers, S. Hunt, P. Deloukas, D. R. Bentley, and P. Donnelly. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304:581–584.

Nachman, M. W. 2002. Variation in recombination rate across the genome: Evidence and implications. Curr. Opin. Genet. Dev. 12:657–663.

Niehuis, O., J. D. Gibson, M. S. Rosenberg, B. A. Pannebakker, T. Koevoets, A. K. Judson, C. A. Desjardins, K. Kennedy, D. Duggan, L. W. Beukeboom, L. van de Zande, D. M. Shuker, J. H. Werren, and J. Gadau. 2010. Recombination and its impact on the genome of the haplodiploid parasitoid wasp *Nasonia*. PLoS One 5:e8597.

Noor, M., A. Cunningham, and J. Larkin. 2001. Consequences of recombination rate variation on quantitative trait locus mapping studies: simulations based on the *Drosophila melanogaster* genome. Genetics 159: 581-588

Padhukasahasram, B., and B. Rannala. 2013. Meiotic gene-conversion rate and tract length variation in the human genome. Eur. J. Hum. Genet. 1–8.

Rezvoy, C., D. Charif, L. Guéguen, and G. A. B. Marais. 2007. MareyMap: An R-based tool with graphical interface for estimating recombination rates. Bioinformatics 23:2188–2189.

Roesti, M., D. Moser, and D. Berner. 2013. Recombination in the threespine stickleback genome--patterns and consequences. Mol. Ecol. 22:3014–27.

Roulin, A. C., J. Routtu, M. D. Hall, T. Janicke, I. Colson, C. R. Haag, and D. Ebert. 2013. Local adaptation of sex induction in a facultative sexual crustacean: Insights from QTL mapping and natural populations of *Daphnia magna*. Mol. Ecol. 22:3567–3579.

Routtu, J., and D. Ebert. 2014. Genetic architecture of resistance in *Daphnia* hosts against two species of host-specific parasites. Heredity 1–8.

Routtu, J., M. D. Hall, B. Albere, C. Beisel, R. D. Bergeron, A. Chaturvedi, J.-H. Choi, J. Colbourne, L. De Meester, M. T. Stephens, C.-P. Stelzer, E. Solorzano, W. K. Thomas, M. E. Pfrender, and D. Ebert. 2014. An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. BMC Genomics 15:1033.

Routtu, J., B. Jansen, I. Colson, L. De Meester, and D. Ebert. 2010. The first-generation *Daphnia magna* linkage map. BMC Genomics 11:508.

Salomé, P. A., K. Bomblies, J. Fitz, R. A. E. Laitinen, N. Warthmann, L. Yant, and D. Weigel. 2012. The recombination landscape in *Arabidopsis thaliana* F2 populations. Heredity 108:447–455.

Smukowski, C. S., and M. A. F. Noor. 2011. Recombination rate variation in closely related species. Heredity 107:496-508

Svendsen, N., C. M. O. Reisser, M. Dukić, V. Thuillier, C. Liautard-Haag, D. Fasel, E. Hürlimann, T. Lenormand, Y. Galimov, and C. R. Haag. 2015. Uncovering cryptic asexuality in *Daphnia magna* by RAD-sequencing. Genetics 201:1143–1155.

Tortereau, F., B. Servin, L. Frantz, H. Megens, D. Milan, G. Rohrer, R. Wiedmann, J. Beever, A. L. Archibald, L. B. Schook, and M. A. M. Groenen. 2012. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics 13:586–598.

Wilfert, L., J. Gadau, and P. Schmid-Hempel. 2007. Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity 98:189–97.

Wong, A. K., A. L. Ruhe, B. L. Dumont, K. R. Robertson, G. Guerrero, S. M. Shull, J. S. Ziegle, L. V. Millon, K. W. Broman, B. A. Payseur, and M. W. Neff. 2010. A comprehensive linkage map of the dog genome. Genetics 184:595–605.

Xu, S., M. S. Ackerman, H. Long, L. Bright, K. Spitze, J. S. Ramsdell, W. K. Thomas, and M. Lynch. 2015. A male-specific genetic map of the microcrustacean *Daphnia pulex* based on single sperm whole-genome sequencing. Genetics 201:31–38.

You, E.-M., K.-F. Liu, S.-W. Huang, M. Chen, M. L. Groumellec, S.-J. Fann, and H.-T. Yu. 2010. Construction of integrated genetic linkage maps of the tiger shrimp (*Penaeus monodon*) using microsatellite and AFLP markers. Anim. Genet. 41:365–376.

Zhu, X., and S. Keeney. 2014. Zip it up to shut it down. Cell Cycle 13:2157–2158.

Table 1. Linkage map summary. The physical lengths refer to the cumulative length of the scaffolds mapped in each linkage group.

Linkage group	Number of markers	Number of "Frame" markers	Genetic length (cM)	Physical length (Mb)	Recombination rate (cM/Mb)
1	441	124	205.38	13.57	**15.14**
2	706	112	177.51	15.97	**11.12**
3	312	62	175.77	10.15	**17.32**
4	449	97	170.68	9.84	**17.35**
5	426	104	168.02	8.77	**19.16**
6	407	100	165.91	9.04	**18.35**
7	377	85	139.98	9.23	**15.16**
8	362	98	139.96	9.41	**14.87**
9	319	91	139.89	7.41	**18.88**
10	238	79	131.38	7.23	**18.18**
total/ average	**4037**	**952**	**1614.48**	**100.61**	16.55

Table 2. Summary of scaffolds and contigs included in the linkage map.

Linkage group	No. of mapped scaffolds & contigs	No. of scaffolds	Scaffold bases	No. of contigs	Contig bases	Oriented scaffolds	Oriented scaffold bases
1	84	63	13938534	21	24389	11	10419401
2	112	84	16116820	28	30759	7	9015316
3	81	61	11176544	20	16793	11	2994136
4	73	60	9026485	13	14395	5	6621437
5	102	77	8909170	25	28609	13	5360882
6	81	62	8721446	19	21200	10	6421448
7	72	55	9304851	17	13654	5	6019697
8	72	52	8742755	20	20271	11	5913714
9	77	54	7544014	23	28315	12	5497291
10	59	47	7119613	12	10754	12	5058319
TOTAL	813	615	100600232	198	209139	97	63321641

Table 3. Sequence scaffolds of *Daphnia magna* genome assembly v2.4 whose markers map to different linkage groups. An exception is scaffold02227 which is divided into two fragments mapped to different parts of the LG8.

Misassembled scaffolds	Total length (bp)	No. markers; position within scaffold	Linkage Group
scaffold00093	237880	1 marker; 28344 bp	**8**
scaffold00093		4 markers; 135580 - 210844 bp	**2**
scaffold03387	219786	6 markers; 11029 - 100849 bp	**7**
scaffold03387		1 marker; 133916 bp	**5**
scaffold02486	541490	20 markers; 35514 - 304827 bp	**2**
scaffold02486		7 markers; 400767 - 528088 bp	**4**
scaffold00233	263417	3 markers; 10570 - 50454 bp	**1**
scaffold00233		3 markers; 82320 - 131292 bp	**4**
scaffold02227	412371	7 markers; 311258 - 396970 bp	**8**
scaffold02227		9 markers; 11932 - 261034 bp	**8**

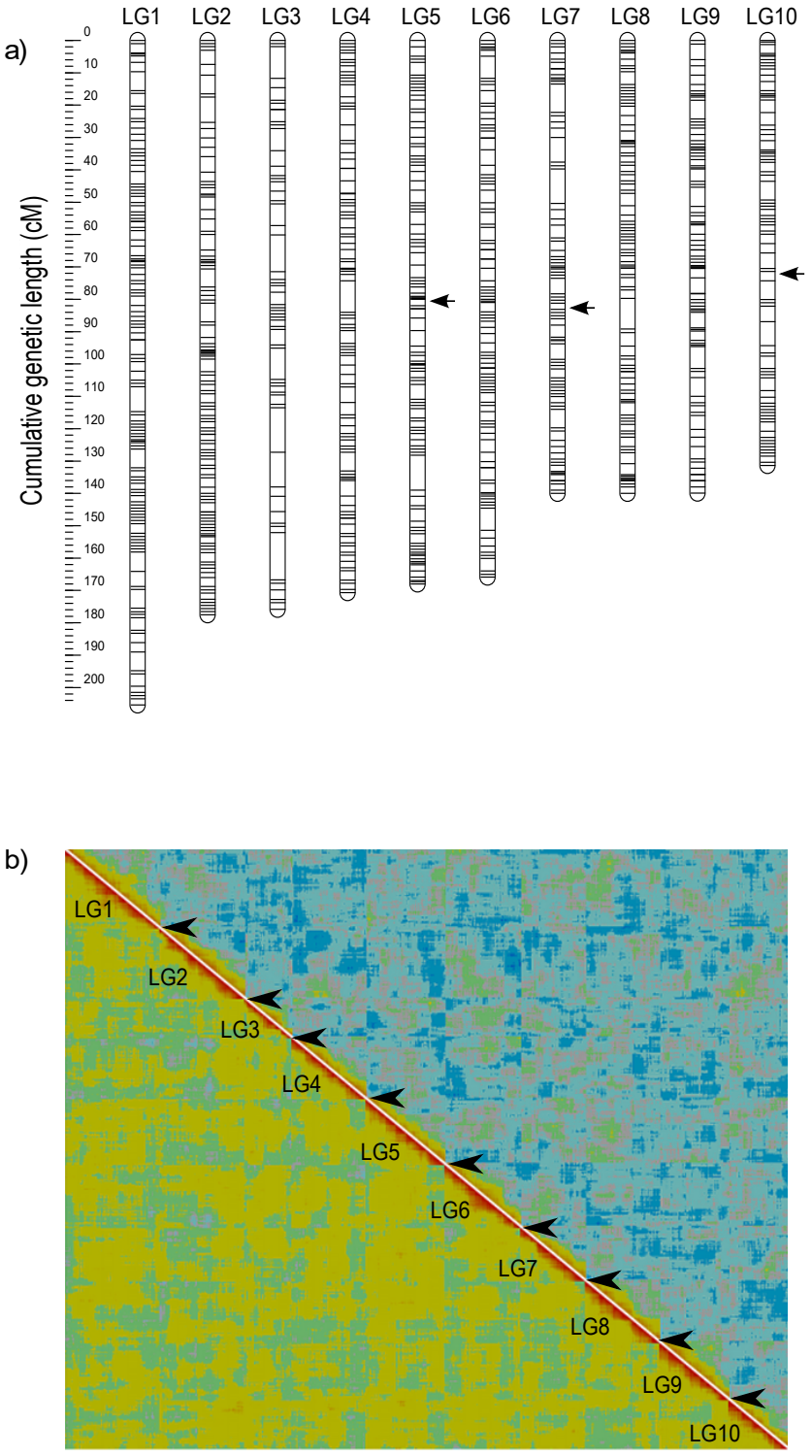


Figure 1. **a) Linkage length and marker distribution of the framework linkage map**. The linkage groups (LGs) are ordered from LG1 to LG10 by decreasing genetic length. Only "Frame" markers are shown with grey lines. Black arrows indicate regions with segregation ratio distortion (see text). **b) Heatplot as graphical representation of the quality of the linkage map.** The image is produced with CheckMatrix (http://www.atgc.org/XLinkage/) to validate the quality of mapping using REC score (low-left diagonal) and BIT score (top-right diagonal). Red colour represents tight linkage and green to blue colour indicates no linkage. Borders of the LGs are indicated by interruptions of the red diagonal (arrows) which confirms the quality of ordering markers within the LG.

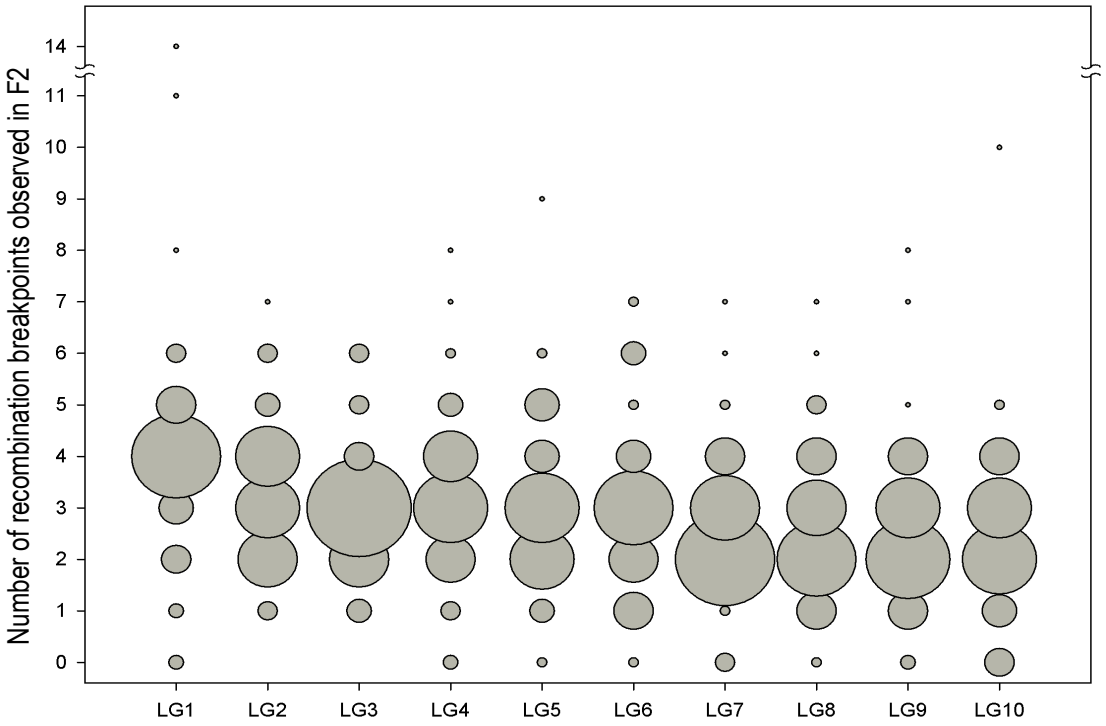


Figure 2. Bubble plot of recombination breakpoints count in each F2 line and LG. Size of circles corresponds to the number of F2 lines with a specific count

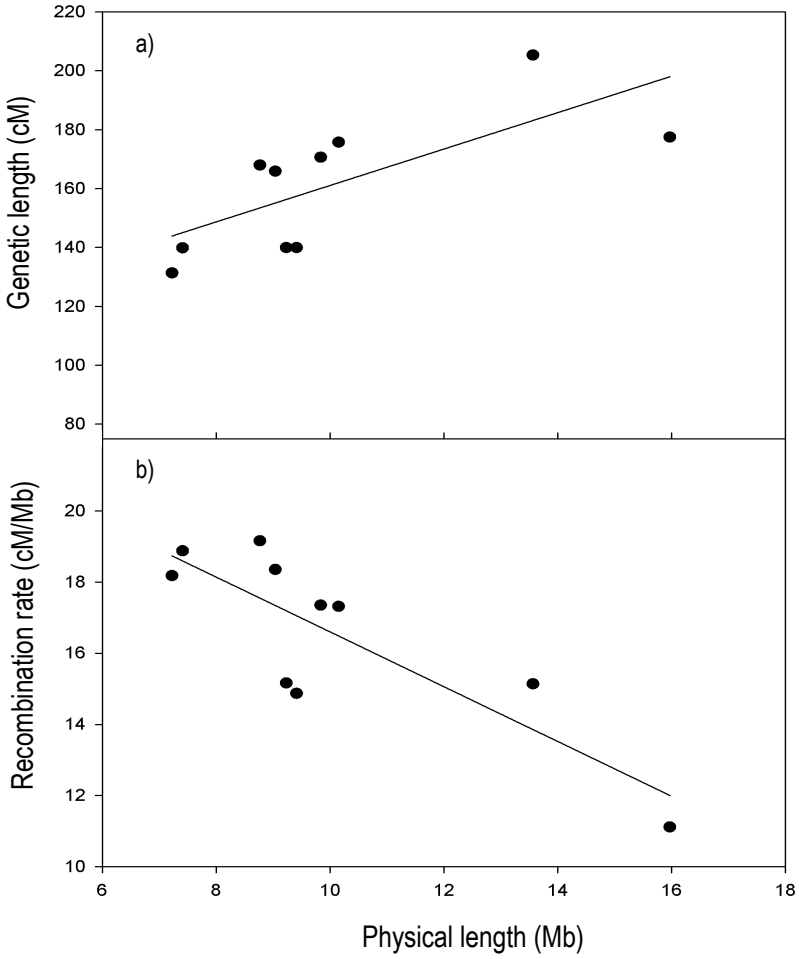


Figure 3. Relationship between estimated physical length of chromosomes and a) genetic length and b) chromosomal recombination rate.

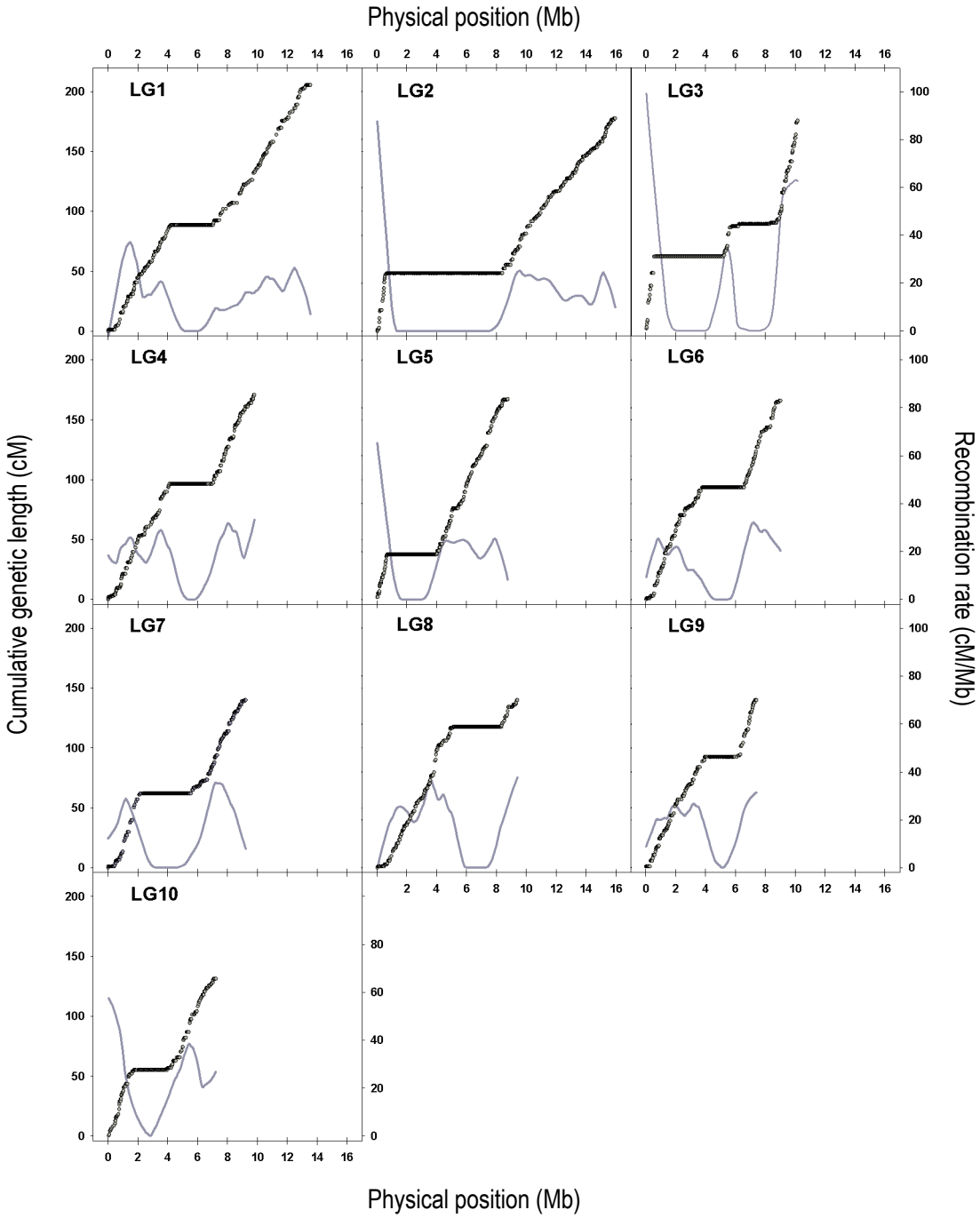


Figure 4. Recombination along the 10 chromosomes of *Daphnia magna*. Dots indicate genetic position of markers in centimorgans (referring to the left axis), plotted against their estimated physical position in the genome (in megabases). An average recombination rate (cM/Mb) was calculated for intervals between adjacent markers and plotted against their physical midpoint (Mb). Data points are not shown but a grey curves indicating data smoothed by LOESS, with the polynomial degree of one and the sampling portion adjusted for each LG according to the density of markers to obtain a constant smoothing resolution across the panels.

Chapter II

HOW CLONAL ARE CLONES?
A QUEST FOR LOSS OF HETEROZYGOSITY DURING ASEXUAL REPRODUCTION IN *Daphnia magna*

Manuscript in preparation:
Dukić, M., C. R. Haag, D. Berner and D. Ebert: How clonal are clones? – A quest for Loss of Heterozygosity during Asexual Reproduction in *Daphnia magna*

ABSTRACT

Due to the lack of recombination, asexual organisms are expected to accumulate mutations and show high levels of within-individual allelic divergence (heterozygosity) when compared to their sexual counterparts. However, there is little empirical support for this prediction, and, in contrast, there is accumulating evidence for genome homogenization during asexual reproduction. In particular, ameiotic crossover recombination is a mechanism that can lead to long stretches of loss of heterozygosity (LOH) and unmasking of mutations that have little or no effect in heterozygous state. Thus, LOH might be an important force for inducing variation among asexual offspring on which natural selection could act. To illuminate the genetic consequences of asexual reproduction in more detail, here we used high-throughput sequencing of *Daphnia magna* for assessing the rate of LOH in a single generations of asexual reproduction, thus minimizing the effect of selection. Comparing parthenogenetic offspring with their mothers at several thousand genetic markers generated by Restriction site Associated DNA (RAD) sequencing, we estimated an average rate of LOH to be 1.97×10^{-3} per locus per generation. Furthermore, we investigated if the detected differences are due to biological events or technical artefacts by re-sequencing the RAD-loci that apparently showed LOH. However, in all eighteen cases putative LOH events proved to be false positives, probably reflecting sequencing biases that could not be detected by bioinformatics analysis alone. We cannot exclude the possibility that short stretches of LOH indeed occur in genomic regions not covered with our markers, although the high number of loci screened, should have enabled us to detect long stretches of LOH if there were any. This suggests that in *D. magna* ameiotic crossovers are very rare or absent.

INTRODUCTION

In eukaryotes asexuality is a state derived from sexual reproduction and this transition has occurred many times independently throughout metazoan evolution (Simon et al. 2003). Parthenogenesis (asexual reproduction in animals where asexual offspring develops from unfertilized eggs) is an umbrella term for diverse mechanisms in which parental ploidy can be maintained without fertilization (see Suomalainen et al. 1987). Based on the presence or absence of meiosis, there are 2 main types of parthenogenesis: automixis and apomixis. In automictic parthenogenesis (automixis), normal meiosis with chromosomes reduction takes place and the somatic ploidy is restored by duplication or fusion of products of the same meiosis. In apomictic parthenogenesis (apomixis), meiosis and recombination are considered to be absent, and such ameiotic conditions are assumed to result in clonal offspring. Since apomixis is the most common type of asexual reproduction in nature (Suomalainen et al. 1987; Archetti 2010) automictic organisms are often ignored, asexuality is habitually used synonymously for apomixis. It became generally accepted that asexual reproduction leads to the production of perfectly clonal offspring with rare mutations being the only source of genetic variation, and this assumption has played an important role in evolutionary biology, especially in models aiming to explain the evolutionary advantage of sex. Yet, despite its important consequences, little is known about the genetic consequences of asexual reproduction in animals.

Recombination in Ameiotic Reproduction (Apomixis)

In the complete absence of recombination during ameiotic reproduction, two alleles are expected to accumulate mutations independently over time, generating high levels of allelic divergence among asexual lineages (Mark Welch and Meselson 2000). However, closely studied asexual lineages do not show such an effect, but rather the opposite (Birky 2004; Hartfield 2015). For example, Darwinuloid ostracodes and oribatid mites show lower levels of allelic divergence than their sexual counterparts (Schon and Martens 2003; Schaefer et al. 2006), suggesting the presence of genome homogenization processes within asexual lineages.

Several mechanisms for allelic convergence were proposed (Birky 2004; Schaefer et al. 2006; Flot et al. 2013) and homology-based DNA double strand break (DSB) repair is a common factor among all of them (i.e. homologous recombination [HR]). Depending on the exact mechanisms employed, recombination in apomixis can result in long or short tracts of loss of heterozygosity (LOH). Reciprocal crossover (CO) recombination will lead to long stretches of LOH (assuming frequent co-segregation of recombined and non-recombined homologous chromatids) while non-reciprocal exchange results in the homozygosis of short DNA tracts (gene conversions), restricting LOH to few hundreds base pairs (bp).

Even though CO recombination is usually associated with the meiotic prophase I, evidence of COs under ameiotic conditions come from studies of mitotic recombination in fungi or somatic cells. In *Saccharomyces cerevisiae, Aspergillus niger* and *Candida albicans* rates of LOH caused by mitotic COs range from 10^{-2} to 10^{-7} per locus per generation (Debets et al. 1993; Mandegar and Otto 2007; Forche et al. 2011). However, in multicellular eukaryotes the situation is more complex. Since asexuality is derived from sexual reproduction, mechanistically, apomixis is more likely to be modified meiosis rather than mitosis. The most

common modification is the suppression of the first (reductional) meiotic division which includes homologue pairing and recombination (Archetti 2010). Since the first meiotic division is suppressed, reduction in the number of chromosomes is not taking place and only sister chromatids separate in a mitotic-like process (second meiotic division). However, reductional division does not have to be suppressed completely and the remnants of meiosis I are described in apomictic organisms (Suomalainen et al. 1987) .

Karyological studies on apomictic dandelions (van Baarlen et al. 2000) and weevils (Rozek et al. 2009) demonstrated pairing of chromosomes and the formation of structures resembling chiasmata (cytological indication of COs). Another compelling evidence of meiotic vestiges during apomixis comes from the study of parthenogentic mechanisms in *Daphnia pulex*. Hiruta et al. (2010) showed that the diploidy in *D. pulex* is maintained by meiotic arrest at an early anaphase I. The observed mechanism includes formation of bivalents in prophase I, implying the opportunity for CO recombination to occur. Chiasmata were not observed though, what is not surprising given the small size of *Daphnia* chromosomes (Hiruta et al. 2010).

However, unexpectedly high rates of LOH were reported to occur in mutation-accumulation lines of *D. pulex* and *D. obtusa* (Omilian et al. 2006; Xu et al. 2011). Analysing microsatellite markers after more than hundred parthenogenetic generations, Xu et al. (2011) estimated the rate of ameiotic recombination in *D. pulex* to be 3.3×10^{-5} per locus per generation. Besides LOH, a large number of segmental deletions resulting in hemizygosity were also reported in the same study (6.7×10^{-5} per locus per generation; Xu et al. 2011), thus making the mechanisms of genome homogenization difficult to disentangle. An important issue that arises from using mutation-accumulation lines is potential selection against LOH or hemizygosity because both may unmask recessive deleterious mutations. This would prevent or retard the propagation of asexual lines, leading to underestimated values of LOH rates. Indeed, mutation-accumulation lines from afore mentioned studies experienced high levels of mortality, enforcing authors to use back-ups in 15-20 % of transfers, thus increasing the opportunity for selection (Omilian et al. 2006; Xu et al. 2011).

In the current study we investigated whether LOH can be detected in a related species, *Daphnia magna*. The asexual reproduction of *D. magna* is described as apomixis with the extrusion of the polar body (Zaffagnini 1987), suggesting a process derived from meiosis. However the detailed mechanisms of apomixis in *D. magna* are not well studied and we assume diploidy is maintained by the suppression of the first meiotic division as it was described for *D. pulex* (Hiruta et al. 2010). To assess the rate of LOH with minimal selection, we sought to address this question by examining a single generation of asexual reproduction using several thousand markers generated by Restriction site Associated (RAD) sequencing (Baird et al. 2008). Moreover, since the nuclear genomes become inherently unstable during an organism senescence (McMurray and Gottschling 2003) we have analysed asexual daughters from young and old animals (Figure 1) to assess if the mother's age has any impact on the fidelity of apomictic process.

MATERIALS & METHODS

Experimental Design

D. magna is a cyclical parthenogen, that is, it can switch between sexual and asexual reproduction mainly in response to environmental conditions. Laboratory iso-female cultures are hereafter referred as lines, since they were initiated by a single female and propagated under conditions of continuous asexual reproduction. For our experiment, we haphazardly selected four individual females ("stem mothers", Figure 1) originating from four different lines. The IXF1 line is an inter-population F1 hybrid, maintained in the Ebert laboratory (University of Basel, Switzerland) since 2006. The lines RM1-02 and RM1-30 are two distinct clones obtained from a natural population in Moscow Zoo (Russia) and were maintained in the Haag laboratory (University of Fribourg, Switzerland) for six months prior to the start of the experiment. The fourth stem mother was hatched from a sexually produced resting egg that was collected from a pond near Frauenfeld (Switzerland) and it served as a founder female of the CH-H-876 line, also maintained in the Ebert laboratory. Each stem mother was cultured to produce thirteen consecutive clutches of all-female offspring. Five randomly chosen daughters from the second and from the twelfth clutch were designated as "asexual daughters" (Figure 1) and subsequently screened for LOH. Throughout the experiment, animals were kept individually in 100-ml beakers filled with artificial *Daphnia* medium (ADaM, Klüttgen et al. 1994) at 20 °C with 16 h light/ 8 h dark cycle and fed with fresh, chemostat-grown, unicellular algae *Scenedesmus* sp. (five million cells per individual per day).

Since we were not able to obtain a sufficient amount of DNA from single individuals for RAD-sequencing, we used genomic DNA extracted from a pool of nine offspring of a particular individual (see Figure 1). That is, the asexual F1 offspring of the stem mothers were grown to adulthood and the offspring of these individuals (i.e., the asexual F2) were pooled to reconstruct the genotype of the F1. Likewise, the genotypes of the stem mothers were inferred from pooled F1 offspring (from clutches not otherwise used in the experiment). Although the sequenced individuals might suffer from LOH with respect to their mother, we expect that each individual would show LOH at different locations in the genome. Thus, by pooling nine daughters, the individual LOH would not show up. However, if their mother (animal for which genotypes are inferred) had LOH, it would be visible in all her daughters and we would be able to detect it. Accordingly, the stem mother from each line was represented by two DNA samples extracted from their 5^{th}, 6^{th} and the 7^{th} clutch offspring. For genotyping asexual daughters we sampled and extracted DNA of individuals from their third clutch (Figure 1).

Prior to sampling, all individuals were cleaned by an antibiotic-starvation treatment to minimize algal and bacterial contamination of genomic DNA. More precisely, animals were kept for three days in a medium containing Ampicillin (Sigma), Streptomycin (Sigma) and Tetracycline (Sigma) at a concentration of 50 mg/L each, and transferred daily to fresh antibiotic medium. To enforce the evacuation of gut content, a small amount of superfine Sephadex ® G-25 (Sigma-Aldrich) was added frequently to the antibiotic medium. Animals with clear intestines were sampled and genomic DNA was extracted using the DNeasy Blood

and Tissue kit (Qiagen) with minor modifications and an inclusion of RNaseA (100 mg/ml; Sigma) digestion step.

RAD library preparation and sequencing

We prepared libraries for RAD-sequencing (Baird et al. 2008) adopting the protocol of Etter et al. (2011) with modifications according to Roesti et al. (2013). Specifically, 1 µg of genomic DNA from each sample (pooled genomic DNA of nine individuals) was digested with the *Pst1* HF restriction enzyme (NEB) in 50 µl reaction volume for 90 min at 37 °C and then heat-inactivated following the manufacturer's manual. P1 sequencing adapters (5 µl of 100 nM stock) containing a unique 5-bp barcode were ligated to each sample using T4 DNA-ligase (NEB, 0.5 µl of 2,000,000 units/mL stock) in a 60 µl volume for 45 min at room temperature. The reaction was then heat-inactivated for 20 min at 65 °C. The total of 48 samples (four *Daphnia* lines, each represented with two mother samples and ten daughter samples) were combined into two pools (each containing two *Daphnia* lines, i.e. 24 samples) and sheared using a Bioruptor (Diagenode). DNA fragments in a range of 250-500 bp were selected using agarose gel electrophoresis (1.25 %, 0.5X TBE), purified and blunt-ended (Quick Blunting Kit, NEB). dA-overhangs were added to the DNA fragments using the Klenow fragment exo$^-$ (NEB), followed by P2 adapter ligation (1 µl from 10 mM stock). Products were purified and PCR amplification was done using the Phusion High-Fidelity DNA polymerase. To minimize the probability of PCR errors, master mixes for each library were divided into eight separate 12.5 µl reactions for amplification (30 sec at 98 °C, 17 cycles of 98 °C 10 sec, 65 °C 30 sec, 72 °C 30 sec, then a final extension for 5 min at 72 °C). Prepared libraries were sequenced on separate Illumina HiSeq2000 lanes using 100 bp single-ends sequencing (Quantitative Genomics Facility service platform, Deep Sequencing Unit Department of Biosystems Science and Engineering, ETH Zurich in Basel, Switzerland).

Bioinformatics analysis

The quality of each sequenced library was assessed using FastQC (Babraham Bioinformatics, The Babraham Institute). A custom R script (available upon request) was used to sort reads into individual samples according to the unique barcodes. We discarded sequences containing ambiguous bases and reads that did not feature a valid *Pst1* restriction site. Reads were aligned to the currently available *D. magna* genome (v2.4; Daphnia Genomic Consortium, WFleaBase) using Novoalign v2.07 (http://novocraft.com) tolerating on average one high-quality mismatch per 14 bases. Only loci aligning to unique genomic locations were considered in further analyses. The average coverage obtained per individual per RAD locus was 115X for IXF1, 134X for RM1-30, 146X for RM1-02, and 136X for the CH-H-876 line, respectively. SAM files were converted into BAM format using SAMtools (Li et al. 2009). On average 24392 unique RAD-loci were obtained for each line.

SNP calling and genotyping was done using a custom R scripts, benefiting from Bioconductor packages Biostrings and Rsamtools (Gentleman et al. 2004). Even though only uniquely-aligned loci were considered in our analysis, we also excluded loci with excessive coverage (three times higher coverage than the overall mean for a given individual) in order to avoid repetitive sequences that are not represented in the *D. magna* reference genome v2.4. The minimum coverage required for a locus to be assigned as a diploid in each individual was three times lower than the estimated mean. We called homozygous genotype when a locus

showed only one haplotype with a read count greater than the threshold for calling a diploid locus, or when the second haplotype occurred in less than three copies (an *ad hoc* criterion to allow for sequencing error in Illumina generated data). A heterozygous locus was called when the total coverage exceeded the threshold for calling a diploid locus and the second most frequent haplotype occurred in not more than threefold less copies compared to the dominant haplotype. Highly polymorphic loci (more than three SNPs) were excluded from downstream analyses.

An error rate was estimated by comparing RAD loci from the two samples representing the stem mother of a given line. These two samples served as replicates and should be identical except for errors due to library preparation (e.g., PCR errors), sequencing, SNP-calling, and genotype-calling. The error rate was calculated by dividing the number of detected differences by the number of comparable loci (successfully sequenced in both stem mother samples).

Analysis of LOH was based on assessing whether the heterozygosity detected in stem mothers is retained in five asexual daughters from the second and the twelfth clutches. Thus, only RAD-loci that were heterozygous in stem mothers were informative for this analysis (Total number of informative loci, Table 1). In order to minimize reported biases of RAD-sequencing (Catchen et al. 2011) only loci that were informative in stem mothers and successfully genotyped in at least 8 out of 10 asexual daughters were considered as markers for estimating the rate of LOH (Number of markers, Table 1).

The rate of LOH λ (per locus per generation, Table 1) for each line was calculated following Omilian et al. (2006) $\lambda = h/(L \times i \times T)$, where h is the total number of LOH events observed, L is the number of parthenogenetic events analysed (for each line, we inspected 10 daughters, $L = 10$) and i is the number of markers. In all cases the number of generations (T) was one.

Fourteen loci showing LOH in eighteen daughters were selected and re-sequenced using Sanger sequencing to provide an independent test of LOH at these loci. Primers were designed using Primer3Plus (Untergasser et al. 2012) based on the *D. magna* reference genome sequence, capturing approximately 100 base pairs flanking the RAD locus on either sides. Re-sequenced loci and the primers used are listed in Supplementary Table 1. Each locus was re-sequenced in the daughter showing LOH, one of the daughters that retained maternal heterozygosity and both stem mother samples. Obtained DNA sequence electropherograms were analysed using CodonCode Aligner software v3.7.1.

RESULTS

We first estimated the error rates based on the two replicated samples of each of the stem mothers. Stem mother samples of the IXF1 and the RM1-30 line, which were sequenced as a part of the same sequencing library, showed genotype (homozygote/heterozygote) inconsistencies in 0.44 % and 0.22 % of loci, respectively. No inconsistencies were observed between stem mother samples of the RM1-02, while stem mother samples of the CH-H-876 line, which were part of the same library as the RM1-02 samples, showed inconsistencies in

0.02 % of loci that were successfully sequenced in both samples. Thus, error rate estimates for *D. magna* lines pooled into the first sequencing library were at least one order of magnitude higher than the error rates for lines that were part of the second sequencing library, indicating the presence of a "library effect" in our data (Table 1).

The IXF1 line had the highest level of heterozygosity as it was expected due to the fact that it is an inter-population hybrid (total number of informative loci, Table 1). More precisely, 33 % of RAD-loci were heterozygous in stem mothers of the IXF1 line, 21 % in the RM1-30 line, 24 % in the RM1-02 line, and 23 % in the CH-H-876 line. Among heterozygous loci, only those that were successfully genotyped in at least eight of the ten of daughter samples were used as markers for the assessment of LOH. In total, LOH was assessed at 4303 marker loci were obtained for the IXF1 line, 2930 markers for the RM1-30 line, 5409 markers for the RM1-02 line and 4785 markers for the CH-H-876 line (number of markers, Table 1).

For each marker locus we searched instances where asexual daughters became homozygous for maternal single nucleotide polymorphisms (SNPs) and those are referred to as LOH events. A total of 204 LOH events were detected for 192 markers in IXF1 line. Hundred and fifty-six LOH events were detected in five daughters from the second clutch while 48 markers showed LOH in five daughters from the twelfth clutch. The number of LOH events for each asexual daughter ranged from 1 to 99. In RM1-30 line, 53 LOH events were detected in the second clutch daughters and 37 LOH events were found in the twelfth clutch daughters, summing up to the total of 90 LOH events for 86 markers. One marker showed LOH in one of the daughters from the IXF1 line and one of the daughters from the RM1-30 line. Two LOH events were detected for two markers in the RM1-02 line, whereas only one LOH event was detected among asexual daughters of the CH-H-876 line. Detected instances of LOH encompassed all SNPs within markers (up to three SNPs). Genome-wide rates of LOH based on RAD-sequencing data were calculated per locus per generation, yielding the estimates of 4.7×10^{-3} for IXF1, 3×10^{-3} for RM1-30, 3.7×10^{-5} for RM1-02 and 2×10^{-5} for the CH-H-876 line, respectively (Table 1).

Per locus LOH rates were very similar to per locus error rates (Table 1). Hence it is possible that the observed LOH events actually were caused by erroneous genotype calls. Moreover, instances showing LOH mainly fall into a lower quartile of the coverage distribution (average 49X, sd = 5.5) indicating that those were poorly covered loci or possible deletions causing hemizygosity. To test between these two scenarios, we used Sanger sequencing for examination of fourteen most promising loci (showing LOH in different lines, having high coverage or showing LOH in adjacent markers flanking the same restriction site) that showed LOH in eighteen asexual daughters. In all eighteen instances putative LOH events proved to be false positives, i.e., presence of maternal polymorphism was confirmed in re-sequenced daughters.

DISCUSSION

Ameiotic recombination resulting in LOH may have important consequences for the evolutionary potential of asexual lineages. In this study we performed reduced representation genome sequencing using the RAD-sequencing protocol to address the occurrence of genome homogenization events within a single asexual generation of *D. magna*, thus minimizing possible selection against LOH. We have assayed in total 174270 RAD-markers (on average 4357 loci in 40 asexual daughters) covering 16.38 Mb of genomic sequence and including 310932 SNPs. Still we were not able to attest any LOH events in four lines of *D. magna*. Though, a substantial number of LOH events were detected in two out of four asexual lines with RAD–sequencing (see Results, Table 1), subsequent validation of putative LOH incidents, proved these were false positives (heterozygotes appearing as homozygotes). More precisely, using Sanger sequencing, maternal heterozygosity was confirmed in loci that showed up as homozygous in asexual daughters. This revealed that the LOH rate estimated from RAD-sequencing data most probably reflects sequencing biases that we were not able to detect using the bioinformatics analysis only.

Several possible sources of error in RAD-sequencing could have caused allele drop-outs that would appear as LOH in our data. These include restriction fragment length bias, stochastic events related to sequencing or PCR, and sequencing errors (Davey et al. 2013; Gautier et al. 2013). To minimise the systematic bias due to variation in restriction fragment length (Davey et al. 2013), we only considered informative loci (heterozygous in stem mothers) that were successfully genotyped in at least 80 % of daughters, as markers for our analysis. Allele dropout due to mutations in a restriction enzyme recognition site is not very likely since this would result in a failure to cut the DNA at that location and the given allele would not be sequenced at all. However, in majority of cases the second variant of a heterozygous locus was detected but at a very low coverage (less than 3-fold) and it was indistinguishable from a sequencing error. Taken together, this indicates that false homozygotes are primarily caused by preferential PCR amplification of one allele over the other (PCR duplicates), however, more detailed analysis of the error source for RAD-sequencing remains to be determined (Dukic et al. unpubl.) and it is out of the scope of this paper. Considering the chemistry of RAD-sequencing false negatives (homozygote appearing as heterozygote) are remote possibility as it is very unlikely that the PCR error will be identical to the ancestral SNP and in addition to this, variants appearing due to a sequencing error would be represented with only few reads.

Previous studies using microsatellite markers in mutation – accumulation lines of *D. pulex* and *D. obtusa* have estimated LOH to occur at a frequency of $\approx 10^{-5}$ to 10^{-4} per locus per generation (Omilian et al. 2006; Xu et al. 2011). If this would hold for our animals we should have detected between 1.7 and 17 LOH events across the four lines assayed. The number of putative LOH events detected by RAD - sequencing in IXF1 and RM1-30 lines, largely exceeded this expectation due to a high error rate inherent for the sequencing library containing these two lines. Among asexual daughters of CH-H-876 and RM1-02 lines, that had substantially lower error rates, we have detected one and two instances of LOH, respectively, and that was consistent with the expectations based on previous studies on *Daphnia*. However, these instances also proved to be erroneously typed as homozygotes.

Hence, despite the large number of markers that were assayed in this study, we were not able to detect LOH within a single asexual generation of *D. magna*.

Our inability to detect LOH could indicate that there is variation between *D. magna* and its related species *D. pulex* and *D. obtusa*, concerning the mechanism of ameiotic reproduction. But perhaps, more plausible explanation would be that this difference originates from the dissimilarity of methodologies employed in search for LOH. For example, bi-allelic SNP markers used in this study have much lower mutation rates than microsatellites. They are also far denser and relatively uniformly distributed across the genome. While using microsatellite markers would enable sampling of fairly unstable genomic regions, SNP markers are more suitable tool for the discovery of LOH across long genomic tracts caused by CO recombination. Moreover, genomic regions containing simple sequence repeats, such as microsatellites, are filtered out during the detection of SNPs with RAD-sequencing(Sharma et al. 2012) and clearly, this difference in methodology, might have influenced our ability to detect short genomic conversions or rearrangements that are usually associated with repetitive regions.

We cannot exclude the possibility that non-reciprocal gene-conversions occurred in genomic regions that were not covered by our markers; however, owing to a high- density of RAD-markers, we are quite confident that the exchange of long genomic tracts due to reciprocal CO recombination did not occur in forty parthenogenetic events screened in the present study, suggesting that its frequency is not exceeding 0.0025 per chromosome. Even mapping of RAD-markers to the third-generation linkage map of *D. magna* (Dukić et al. in prep.) indicated that none of the putative LOH events occurred in adjacent markers (data not shown). Latest study on *D. pulex* also demonstrates how the choice of methodology can influence the interpretation of results concerning LOH. In contrast to the previous assumption that LOH in adjacent microsatellite markers was caused by CO recombination (Omilian et al. 2006; Xu et al. 2011), using whole genome sequencing, Keith et al. (2015) reported the mean length of LOH tracts in *D. pulex* to be short (> 250 bp) and therefore, more likely caused by gene conversions. In addition, many LOH incidents were associated with large scale duplications typical for non-allelic homologous recombination (or ectopic gene conversions) that is invariably associated with repetitive (paralogous) sequences within the genome (Sasaki et al. 2010; Parks et al. 2015). Both *D. pulex* and *D. magna* genomes are extraordinarily rich in duplicated regions (Colbourne et al. 2011; *Daphnia* Genome Consortium), and these new findings open questions about the co-evolution of the genome architecture and the fidelity of mechanisms employed for DSB repair in *Daphnia* parthenogenesis.

We have also tested whether the mother's age has any impact on generation of LOH due to genomic instability that is expected to occur with organismal senescence (Burhans and Weinberger 2007). However, daughters produced at older age (the twelfth clutch) did not show LOH in our study, same like the daughters from the earlier clutches (the second clutch). High number of putative LOH incidents indicated by RAD-sequencing in the second clutch IXF1 daughters (hundred and fifty-six) can mainly be attributed to two samples in which 99 and 38 cases of LOH were detected. As explained earlier, these probably reflect highly erroneous samples. Interestingly, in yeast it was shown that LOH in

the offspring from young cells was caused by rare reciprocal CO recombination, while the rate of LOH was 40 – 200 fold higher in the cells produced by old mothers and it was mainly caused by non-reciprocal gene conversions (McMurray and Gottschling 2003). Thus our inability to detect elevation in LOH rates in older mothers could also be due to inherent limitations of the methodology employed in this study.

Implications for Evolutionary Biology

Clonality (with the exception of rare mutations) of asexual reproduction is a bedrock assumption in a great majority of models aiming to explain prevalence of sexual reproduction despite its high costs (Kondrashov 1993; West et al. 1999; Hartfield and Keightley 2012). However, since DSBs are inevitable by-product of cellular divisions (reviewed in Aguilera and Gómez-González 2008; Huertas 2010), it is difficult to anticipate that perfect clonality can be achieved within an asexual lineage. As we argue above, clonality is merely a concept "dependent upon the resolving power of molecular markers" (Loxdale and Lushai 2003) and some levels of homology-based DSB repair leading to LOH are expected to occur. The only model, at least to our knowledge, that takes into account the possibility of LOH in asexual reproduction is the "Loss of Complementation – LOC hypothesis" proposed by Archetti (2004a,b, 2010). The basic logic of LOC hypothesis is that the recombination processes causing LOH in asexual reproduction will lead to unmasking of recessive deleterious mutations (LOC). Consequently, the details of LOC will depend on the number of recessive deleterious mutations (lethal equivalents) and the portion of the genome that becomes homozygous within a single asexual generation. As Archetti proposed, with the right combination of parameters, asexual reproduction will have more than two – fold cost compared to sexual reproduction (Archetti 2004b, 2010). However, our study together with the latest findings in *D. pulex* (Keith et al. 2015) indicate that long reciprocal allelic exchanges (COs) during parthenogenesis are less frequent than it was previously assumed (Omilian et al. 2006; Xu et al. 2011) and LOH is probably restricted only to short genomic regions. This implies that the portion of the genome experiencing LOH in a single asexual generation predicted by LOC hypothesis (Archetti 2004b, 2010) is largely overestimated. Thus, for apomixis to bear high cost as predicted by the same hypothesis, asexual lineages should harbour a large number of lethal equivalents. Even though this possibility cannot be excluded, it is not very plausible since it would make asexual lineages difficult to maintain in the laboratory, which is not the case. Nevertheless, possibility for LOH during asexual reproduction should not be ignored when discussing the evolutionary potential and the age of asexual lineages (Hartfield et al. 2016) and more models accounting for this rare phenomenon are needed.

REFERENCES

Aguilera, A., and B. Gómez-González. 2008. Genome instability: a mechanistic view of its causes and consequences. Nat. Rev. Genet. 9:204–17.

Archetti, M. 2010. Complementation, genetic conflict, and the evolution of sex and recombination. J. Hered. 101 Suppl :S21–33.

Archetti, M. 2004a. Loss of complementation and the logic of two-step meiosis. J. Evol. Biol. 17:1098–105.

Archetti, M. 2004b. Recombination and loss of complementation: a more than two-fold cost for parthenogenesis. J. Evol. Biol. 17:1084–97.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376.

Birky, C. W. 2004. Bdelloid rotifers revisited. Proc. Natl. Acad. Sci. U. S. A. 101:2651–2.

Burhans, W. C., and M. Weinberger. 2007. DNA replication stress, genome instability and aging. Nucleic Acids Res. 35:7545–7556.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. G3 (Bethesda). 1:171–82.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Cáceres, L. Carmel, C. Casola, J.-H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Fröhlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V Kriventseva, D. Kültz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V Koonin, E. M. Zdobnov, I. V Grigoriev, M. Lynch, and J. L. Boore. 2011. The ecoresponsive genome of *Daphnia pulex*. Science 331:555–561.

Davey, J. W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi, and M. L. Blaxter. 2013. Special features of RAD Sequencing data: implications for genotyping. Mol. Ecol. 22:3151–64.

Debets, F., K. Swart, R. F. Hoekstra, and C. J. Bos. 1993. Genetic maps of eight linkage groups of *Aspergillus niger* based on mitotic mapping. Curr. Genet. 23:47–53.

Dukic, M., C. R. Haag, D. Berner, M. Roesti, and D. Ebert. in prep. Marker-dense genetic map reveals variation in recombination rate across the genome of *Daphnia magna*.

Etter, P. D., S. Bassham, P. A. Hohenlohe, E. A. Johnson, and W. A. Cresko. 2011. SNP Discovery and Genotyping for Evolutionary Genetics Using RAD Sequencing. Pp. 157–178 *in* V. Orgogozo and M. V. Rockman, eds. Molecular Methods for Evolutionary Genetics.

Flot, J.-F., B. Hespeels, X. Li, B. Noel, I. Arkhipova, E. G. J. Danchin, A. Hejnol, B.

Henrissat, R. Koszul, J.-M. Aury, V. Barbe, R.-M. Barthélémy, J. Bast, G. A. Bazykin, O. Chabrol, A. Couloux, M. Da Rocha, C. Da Silva, E. Gladyshev, P. Gouret, O. Hallatschek, B. Hecox-Lea, K. Labadie, B. Lejeune, O. Piskurek, J. Poulain, F. Rodriguez, J. F. Ryan, O. A. Vakhrusheva, E. Wajnberg, B. Wirth, I. Yushenova, M. Kellis, A. S. Kondrashov, D. B. Mark Welch, P. Pontarotti, J. Weissenbach, P. Wincker, O. Jaillon, and K. Van Doninck. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. Nature 500:453–457.

Forche, A., D. Abbey, T. Pisithkul, M. a. Weinzierl, T. Ringstrom, D. Bruck, K. Petersen, and J. Berman. 2011. Stress alters rates and types of loss of heterozygosity in *Candida albicans*. MBio 2:e00129–11.

Gautier, M., K. Gharbi, T. Cezard, J. Foucaud, C. Kerdelhué, P. Pudlo, J.-M. Cornuet, and A. Estoup. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol. Ecol. 22:3165–3178.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5:R80.

Hartfield, M. 2015. Evolutionary genetic consequences of facultative sex and outcrossing. J. Evol. Biol. 29:5–22.

Hartfield, M., and P. D. Keightley. 2012. Current hypotheses for the evolution of sex and recombination. Integr. Zool. 7:192–209.

Hartfield, M., S. I. Wright, and A. F. Agrawal. 2016. Coalescent times and patterns of genetic diversity in species with facultative sex: effects of gene conversion, population structure, and heterogeneity. Genetics 202:297–312.

Hiruta, C., C. Nishida, and S. Tochinai. 2010. Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. Chromosome Res. 18:833–40.

Huertas, P. 2010. DNA resection in eukaryotes: deciding how to fix the break. Nat. Struct. Mol. Biol. 17:11–6.

Keith, N., A. E. Tucker, C. E. Jackson, W. Sung, J. I. Lucas, D. R. Schrider, and A. J. Younge. 2015. High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. Genome Res., doi: 10.1101/gr.191338.115.

Klüttgen, B., U. Dülmer, M. Engels, and H. Ratte. 1994. ADaM, an artificial freshwater for the culture of zooplankton. Water Res. 28:743–746.

Kondrashov, A. S. 1993. Classification of hypotheses on the advantage of amphimixis. J. Hered. 84:372–387.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Loxdale, H. D., and G. Lushai. 2003. Rapid changes in clonal lines : the death of a " sacred cow ." Biol. J. Linn. Soc. 79:3–16.

Mandegar, M. A., and S. P. Otto. 2007. Mitotic recombination counteracts the benefits of genetic segregation. Proc. R. Soc. B Biol. Sci. 274:1301–1307.

Mark Welch, D., and M. Meselson. 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science 288:1211–1215.

McMurray, M. A., and D. E. Gottschling. 2003. An age-induced switch to a hyper-recombinational state. Science 301:1908–11.

Omilian, A. R., M. E. A. Cristescu, J. L. Dudycha, and M. Lynch. 2006. Ameiotic recombination in asexual lineages of *Daphnia*. Proc. Natl. Acad. Sci. U. S. A. 103:18638–43.

Parks, M. M., C. E. Lawrence, and B. J. Raphael. 2015. Detecting non-allelic homologous recombination from high-throughput sequencing data. Genome Biol. 16:72.

Roesti, M., D. Moser, and D. Berner. 2013. Recombination in the threespine stickleback genome-patterns and consequences. Mol. Ecol. 22:3014–27.

Rozek, M., D. Lachowska, M. Holecovà, and Ł. Kajtoch. 2009. Karyology of parthenogenetic weevils (Coleoptera, Curculionidae): Do meiotic prophase stages occur? Micron 40:881–885.

Sasaki, M., J. Lange, and S. Keeney. 2010. Genome destabilization by homologous recombination in the germ line. Nat. Rev. Mol. Cell Biol. 11:182–195.

Schaefer, I., K. Domes, M. Heethoff, K. Schneider, I. Schon, R. A. Norton, S. Scheu, and M. Maraun. 2006. No evidence for the "Meselson effect" in parthenogenetic oribatid mites (Oribatida, Acari). J. Evol. Biol. 19:184–193.

Schon, I., and K. Martens. 2003. No slave to sex. Proc. R. Soc. B Biol. Sci. 270:827–833.

Sharma, R., B. Goossens, C. Kun-Rodrigues, T. Teixeira, N. Othman, J. Q. Boone, N. K. Jue, C. Obergfell, R. J. O'Neill, and L. Chikhi. 2012. Two different high throughput sequencing approaches identify thousands of *de novo* genomic markers for the genetically depleted Bornean elephant. PLoS One 7:e49533.

Simon, J. C., F. Delmotte, C. Rispe, and T. Crease. 2003. Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. Biol. J. Linn. Soc. 79:151–163.

Suomalainen, E., A. Saura, and J. Lokki. 1987. Cytology and Evolution in Parthenogenesis. Boca Raton (FL): CRC Press.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res. 40:1–12.

van Baarlen, P., P. J. van Dijk, R. F. Hoekstra, and J. H. de Jong. 2000. Meiotic recombination in sexual diploid and apomictic triploid dandelions (*Taraxacum officinale* L.). Genome 43:827–835.

West, S. A., C. M. Lively, and A. F. Read. 1999. A pluralist approach to sex and recombination. J. Evol. Biol. 12:1003–1012.

Xu, S., A. R. Omilian, and M. E. Cristescu. 2011. High rate of large-scale hemizygous deletions in asexually propagating *Daphnia*: implications for the evolution of sex. Mol. Biol. Evol. 28:335–42.

Zaffagnini, F. 1987. Reproduction in *Daphnia*. Pp. 245–284 *in* Memorie dell'Istituto Italiano di Idrobiologia.

Table 1. Summary of LOH information in *Daphnia magna* estimated from RAD sequencing data.

Line	**IXF1**	**RM1-30**	**RM1-02**	**CH-H-876**
Number of asexual daughters	10	10	10	10
Total number of RAD-loci	22814	23060	26738	24957
Error rate (per RAD-locus)	0.00447	0.00221	0	0.00020
Number of informative loci (heterozygous in stem mothers)	7684	4840	6526	5644
Number of markers	4303	2930	5409	4785
Total number of LOH events detected	204	90	2	1
Rate of LOH ($locus^{-1}$ $generation^{-1}$)	0.004741	0.003072	0.000037	0.000021
Number of LOH events tested by Sanger sequencing	6	8	2	1

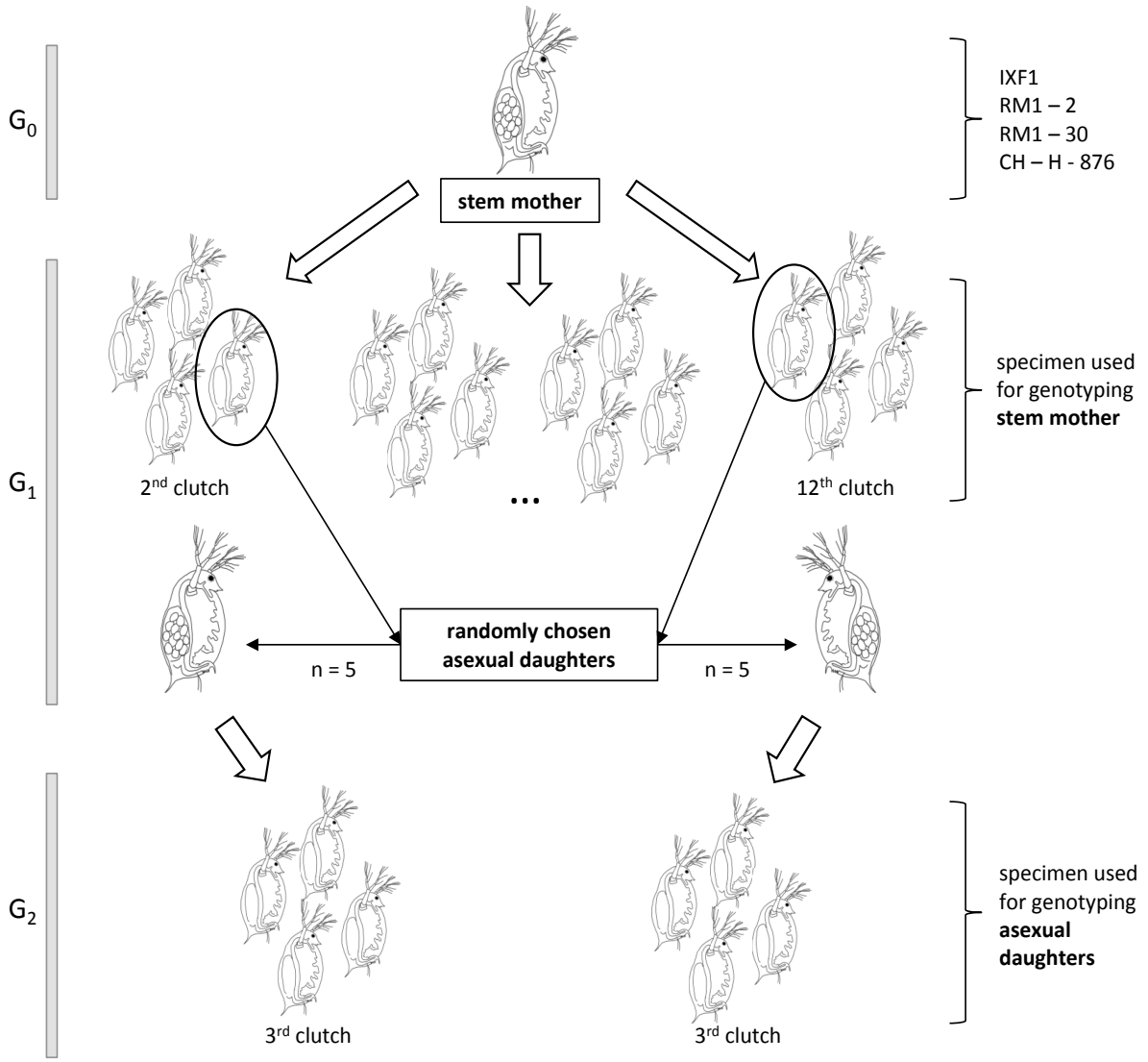


Figure 1. **Schematic diagram of the experimental design and sampling procedure**. The experiment was replicated for four different lines of *Daphnia magna* and encompassed three generations of asexual reproduction ($G_0 – G_2$). Empty arrows indicate generational transition and the production of all-female clutches. Circles with black arrows indicate five randomly chosen asexual daughters from the first asexual generation (G_1) that were subsequently screened for loss of heterozygosity (LOH). G_1 females from the 5th, 6th and the 7th clutch were used for genotyping the stem mother (G_0). G_2 females from the 3rd clutch were used for genotyping each asexual daughter (G_1).

Supplementary Table S1. PCR primers for re-sequenced markers that were selected for verification of loss of heterozygosity events detected using RAD-sequencing. RAD-markers are named based on the alignment position of RAD-reads to the genome draft of *Daphnia magna* (v4.2)

RAD_marker	Forward primer	Reverse primer
scaffold00084_181078	TTTTTGTGTGTGTGAAAGAGACC	CCTGGCAAGAAGAAAGAAGC
scaffold00024_965997	TGGTGCCCTGACTGAGTGTA	TTTCCCATGTAACGACGACA
scaffold01005_1090145	TTGAGGAAAGAGCGGGAATA	CACGGCCACAAAAATCTCAT
scaffold00687_215401	CCCCAGATACCCGTACACAC	AGCTATCCAACGCGATCATT
scaffold01005_615383	TTTTTGCTACCCCATGCAAT	CAAAGCCCCACAGCTATGAT
scaffold03258_520152	TTTGTCCACTTTTCCGGTTC	CGTTATGAAGTGGACGCTGT
scaffold01654_178568	GGCGGGTGTATAGCCAAGTA	AAAGAGACCGCGACTTTTGA
scaffold02723_2112	TGAAGCGTGTTGCTTCTGTT	ATTGACATAGCCGCCAGATT
scaffold02581_769135	AGGCCCTGATAGCATTACGA	GGAGACTCTGGAGCTGTTGG
scaffold02581_777742	CAGGGCTTCCAGAATTACCA	CTGGAGCAAAAGGAGATGCT
scaffold01115_5223	CCATTCTGATTGCGGTCTTT	CGTCCTTAATCCGACCACAT
scaffold00640_105924	TCATCCTTGCTCTGGTCCTC	AATATGGCGACACAACATGC
scaffold03102_238072	GGATTGCGTTAGGCAACAAA	GGCAACTGTGCCCTTGTATT
scaffold00512_163074	GAATATCTGGACGCCATGCT	ACAATCAACAAATGCCGAAA

Chapter III

UNCOVERING CRYPTIC ASEXUALITY IN *Daphnia magna* BY RAD-SEQUENCING

Manuscript published in Genetics as:
Svendsen, N., C. M. O. Reisser, M. Dukić, V. Thuillier, C. Liautard-Haag, D. Fasel, E. Hürlimann, T. Lenormand, Y. Galimov, and C. R. Haag. 2015. Uncovering cryptic asexuality in *Daphnia magna* by RAD-sequencing. Genetics 201:1143–1155.

CONTRIBUTION: I was involved in preparation of RAD-sequencing libraries, data analysis and writing the manuscript

GENETICS | INVESTIGATION

Uncovering Cryptic Asexuality in *Daphnia magna* by RAD Sequencing

Nils Svendsen,*,[1] Celine M. O. Reisser,*,†,[1] Marinela Dukić,‡ Virginie Thuillier,† Adeline Ségard,* Cathy Liautard-Haag,† Dominique Fasel,† Evelin Hürlimann,† Thomas Lenormand,* Yan Galimov,§ and Christoph R. Haag*,†,[2]

*Centre d'Ecologie Fonctionnelle et Evolutive (CEFE)–Unité Mixte de Recherche 5175, Centre National de la Recherche Scientifique (CNRS)–Université de Montpellier–Université Paul-Valéry Montpellier–Ecole Pratique des Hautes Etudes (EPHE), campus CNRS, 19, 34293 Montpellier Cedex 5, France, †Ecology and Evolution, University of Fribourg, 1700 Fribourg, Switzerland, ‡Zoology Institute, Evolutionary Biology, University of Basel, 4051 Basel, Switzerland, and §Koltsov Institute of Developmental Biology, Russian Academy of Sciences, 119334 Moscow, Russia

ABSTRACT The breeding systems of many organisms are cryptic and difficult to investigate with observational data, yet they have profound effects on a species' ecology, evolution, and genome organization. Genomic approaches offer a novel, indirect way to investigate breeding systems, specifically by studying the transmission of genetic information from parents to offspring. Here we exemplify this method through an assessment of self-fertilization *vs.* automictic parthenogenesis in *Daphnia magna*. Self-fertilization reduces heterozygosity by 50% compared to the parents, but under automixis, whereby two haploid products from a single meiosis fuse, the expected heterozygosity reduction depends on whether the two meiotic products are separated during meiosis I or II (*i.e.*, central *vs.* terminal fusion). Reviewing the existing literature and incorporating recombination interference, we derive an interchromosomal and an intrachromosomal prediction of how to distinguish various forms of automixis from self-fertilization using offspring heterozygosity data. We then test these predictions using RAD-sequencing data on presumed automictic diapause offspring of so-called nonmale producing strains and compare them with "self-fertilized" offspring produced by within-clone mating. The results unequivocally show that these offspring were produced by automixis, mostly, but not exclusively, through terminal fusion. However, the results also show that this conclusion was only possible owing to genome-wide heterozygosity data, with phenotypic data as well as data from microsatellite markers yielding inconclusive or even misleading results. Our study thus demonstrates how to use the power of genomic approaches for elucidating breeding systems, and it provides the first demonstration of automictic parthenogenesis in *Daphnia*.

KEYWORDS genome-wide heterozygosity; breeding system; inbreeding; automixis; tychoparthenogenesis; *Daphnia magna*; nonmale producers

WHILE humans and most other mammals reproduce exclusively by sexual reproduction with sexes being determined by the well-known XY sex-chromosome system, the breeding systems of many other organisms, including many pests and parasites, remain unknown (Bell 1982; Normark 2003). The breeding system *sensu lato*, (including details of meiosis, *e.g.*, recombination patterns and syngamy, *e.g.*, levels of inbreeding, as well as their variants, *e.g.*, modified meiosis in parthenogens) represents a key for understanding the biology of a species and has profound effects on its ecology, evolution, and genomics. Yet investigating breeding systems is often far from straightforward: Many species cannot easily be cultured and bred in the laboratory and observations of breeding behavior in nature are difficult. Even in species than can be bred in the laboratory, parts of the breeding system may be cryptic and not directly observable.

The advent of high-throughput genotyping methods opens an alternative possibility that can be used on a much larger array of species: indirect inference of the breeding system using genetic methods, which are based on differences among breeding systems in the transmission of genetic information from one generation to the next. In some cases, genome-wide information may not be needed. For instance, a few genetic

doi: 10.1534/genetics.115.179879
Manuscript received June 24, 2015; accepted for publication August 29, 2015; published Early Online September 3, 2015.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.179879/-/DC1.
[1]These authors contributed equally to this work.
[2]Corresponding author: CEFE UMR 5175, 1919, Route de Mende, 34293 Montpellier Cedex 5, France. E-mail: christoph.haag@cefe.cnrs.fr

markers such as microsatellites are sufficient to distinguish self-fertilization from outcrossing in hermaphrodites (*e.g.*, David *et al.* 2007) or clonal from sexual reproduction in aphids (Delmotte *et al.* 2002). However, for a conclusive distinction between other breeding systems, a genome-wide approach may be essential. This is illustrated in the present paper for the distinction of self-fertilization *vs.* automictic parthenogenesis, comparing genomic data with microsatellite data and direct observations.

Self-fertilization and automictic parthenogenesis both reduce genome-wide heterozygosity among offspring compared to their parents, thereby increasing homozygosity due to identity by descent (Hartl and Clark 2007; Charlesworth and Willis 2009). Under self-fertilization, in which male and female gametes produced by the same, hermaphrodite individual fuse, the expected reduction in offspring heterozygosity for diploid, autosomal loci is 50% per generation. A similar heterozygosity reduction also occurs under some forms of automictic parthenogenesis (also called "automixis"). Automictic parthenogenesis is a common form of parthenogenetic (*i.e.*, female-only) reproduction (Bell 1982; Mogie 1986; Suomalainen *et al.* 1987), in which offspring are produced by fusion of two products of a single meiosis. Examples are intratetrad mating in fungi or fusion of an egg cell with a polar body in animals (Suomalainen *et al.* 1987; Hood and Antonovics 2004; Stenberg and Saura 2009). A more detailed account of the different processes that are summarized under automixis is given below.

The distinction between automixis and self-fertilization is subtle both in terms of the expected heterozygosity reduction among offspring as well as with respect to the processes that lead to it. Both involve the fusion of two meiotic products produced by a single individual. Self-fertilization involves fusion of products of different, independent meioses and therefore parental alleles are sampled with replacement. In contrast, automixis involves fusion of the products of a single meiosis and therefore parental alleles are sampled without replacement. Sampling of parental alleles with replacement leads to the well-known Mendelian expectations of genotype frequencies (50% heterozygotes, 25% of each homozygote) among self-fertilized offspring. However, to understand the consequences of sampling of maternal alleles without replacement during automixis, we have to distinguish two cases: Under "central fusion" two products that have been separated during meiosis I (the first meiotic division) fuse, and under "terminal fusion" two products that have been separated during meiosis II fuse. Because homologous chromosomes (carrying different alleles at heterozygous loci) are separated during meiosis I, and sister chromatids (carrying identical alleles) are separated during meiosis II, central fusion tends to retains parental heterozygosity and terminal fusion tends to lead to fully homozygous genotypes. However, because recombination reshuffles alleles between homologous chromosomes, these expectations hold only for the centromere (at which sister chromatids are attached to each other). Expected offspring heterozygosity at loci far from the centromere attains 67% of parental heterozygosity for both central and terminal fusion. This is because, far from the centromere, alleles are distributed at random across sister and nonsister chromatids due to recombination, and therefore they are sampled randomly without replacement (once one meiotic product is chosen, two of the three remaining meiotic products contain the alternate allele) (Rizet and Engelmann 1949; Barratt *et al.* 1954; Suomalainen *et al.* 1987; Pearcy *et al.* 2006, 2011; Engelstädter *et al.* 2011).

Patterns of heterozygosity reduction between parents and offspring can thus be used to distinguish self-fertilization from automixis and/or central from terminal fusion. This approach has previously been used in a few organisms to address the question of whether automixis occurs via central or terminal fusion (Pearcy *et al.* 2006; Lampert *et al.* 2007; Oldroyd *et al.* 2008). However, differences in the realized levels of heterozygosity reduction among breeding systems depend on recombination rates and may be modulated by the degree of recombination interference and, if offspring heterozygosity is assessed at any later stage than the zygote, by differential survival of heterozygotes *vs.* homozygotes (*i.e.*, viability selection, Wang and Hill 1999).

We therefore first derive two specific theoretical predictions of how to distinguish self-fertilized from automictic offspring and central from terminal fusion based on heterozygosity patterns. We then use the freshwater crustacean *Daphnia magna* to empirically assess and compare the consequences of self-fertilization and automixis for offspring heterozygosity. We use known, self-fertilized offspring as controls and compare them with offspring whose breeding system was initially unknown but could by the present study be identified as automictic. *D. magna* reproduces by cyclical parthenogenesis, in which clonal reproduction is intermitted by sexual reproduction. The clonal offspring may develop into males or females (environmental sex determination) and sexual reproduction always leads to the production of diapause stages ("ephippia": structures formed by maternal tissue, usually encapsulating two diapausing embryos). Hence, "self-fertilized" offspring in diapause can easily be generated by growing clonal cultures to high population densities and letting males mate with their genetically identical sisters. We acknowledge that within-clone mating (mating of a female with a genetically identical male) may only genetically but not ecologically be equivalent to self-fertilization (fertilization between male and female organs of a single, hermaphrodite individual), but for simplicity, we do not distinguish between these terms in the present paper.

While diapause stages can be produced clonally in some species of *Daphnia* (Hebert and Crease 1980), they were hitherto thought to be always produced by sexual reproduction in *D. magna*. However, we have previously found that some strains of *D. magna* do not produce males ("nonmale producing strains," NMP), even when stimulated with a "male-inducing" hormone (Innes and Dunbrack 1993; Galimov *et al.* 2011). In natural populations, these strains still participate in sexual reproduction, but only via the female

function, that is, by producing diapause eggs that have to be fertilized by males from other, male-producing (MP) strains (*i.e.*, strains that produce both males and females with sex determined by the environment). When grown in isolation (*i.e.*, in NMP-only cultures), females still produce the diapause capsules, but these are usually empty (*i.e.*, do not contain viable embryos). Yet, very rarely, a few offspring hatch from these ephippia, indicating that a very low percentage of them do contain viable embryos (Galimov *et al.* 2011). The offspring are diploid and show segregation of maternal alleles, indicating that they are not produced clonally (Galimov *et al.* 2011). They may thus be produced either by within-clone mating through rare and undetected male production in the maternal NMP strain or by automictic parthenogenesis (Galimov *et al.* 2011). To evaluate these possibilities, we used (i) direct testing for the presence of males by phenotypic screening of large samples, (ii) crossing attempts between different NMP strains (if rare males are present they are expected to fertilize females of other NMP strains as well as their own), and (iii) an assessment of the heterozygosity patterns among offspring by microsatellite genotyping and restriction site-associated DNA (RAD) sequencing. Our results showed that only the genomic approach (RAD sequencing) could provide conclusive evidence for the mode of reproduction by which these offspring had been produced. More generally, our study thus serves to illustrate the observed and expected genome-wide patterns of heterozygosity reduction under automixis and self-fertilization and to provide evidence for the great potential of genomic approaches for elucidating cryptic breeding systems.

Expected Heterozygosity Reduction Under Automixis

The expected heterozygosity reduction under automixis has been described before (Rizet and Engelmann 1949; Barratt *et al.* 1954; Suomalainen *et al.* 1987; Pearcy *et al.* 2006, 2011; Engelstädter *et al.* 2011). However, different aspects are discussed in different papers, and the literature on breeding systems is rather disparate from the literature on genetic mapping in fungi or on mapping of centromeres either by natural or artificial automixis. Furthermore, in addition to central and terminal fusion, a further term "random fusion" is sometimes discussed, but its definition and effects on heterozygosity reduction require clarification. Finally, the effects of recombination interference on heterozygosity reduction have only rarely been considered in the breeding systems literature (*e.g.*, Asher 1970; Nace *et al.* 1970). For these reasons, we briefly review here the literature on expected heterozygosity reduction under automixis with the focus on the comparison with self-fertilization. We identify two main predictions regarding expected heterozygosity patterns, an interchromosomal and an intrachromosomal one, which allow distinguishing automictic from self-fertilized offspring using genomic data. We also mathematically derive predictions on the intrachromosomal patterns of heterozygosity in offspring produced by terminal and central fusion, accounting for different degrees of recombination interference.

The terms central fusion and terminal fusion are derived from ordered tetrads (Tucker 1958; Suomalainen *et al.* 1987). In many fungi and algae, the four products of meiosis remain together in an envelope called "ascus," with some of them retaining a specific order (Bos 1996): The four meiotic products of a diploid parent heterozygous A1A2 at a centromeric locus are ordered along a sequence A1_A1_A2_A2, with meiosis I explaining the central division and meiosis II the two terminal divisions (each division is indicated by an underscore). Hence, fusion of neighboring meiotic products during within-tetrad mating can either be terminal (leading to homozygous centromeric regions A1A1 or A2A2) or central (leading to heterozygous centromeric regions A1A2). However, because the effects on offspring heterozygosity are identical, the term central fusion is often used to describe the fusion of any two meiotic products that have been separated during meiosis I (or where meiosis I is suppressed, Asher 1970). Equivalently, the term terminal fusion is used to describe the fusion of any products that have been separated during meiosis II (or where meiosis II is suppressed, Asher 1970), not only in ordered tetrads.

Random fusion can be defined as fusion of two randomly chosen products of a meiotic tetrad (Suomalainen *et al.* 1987; Pearcy *et al.* 2006; Lampert *et al.* 2007). Hence, with random fusion, 2/3 of the offspring are produced by central fusion and 1/3 by terminal fusion (once one meiotic product is chosen, only one of the three remaining products carries the same allele at the centromeric locus shown above, thus central fusion occurs with a probability of 2/3). Yet, in animals, meiosis typically leads to one oocyte and polar bodies, and automictic fusion usually (but not always, *e.g.*, Seiler and Schäffer 1960) occurs between the oocyte and one of the polar bodies. However, the first polar body often decays rapidly or does not undergo meiosis II (*e.g.*, in *Daphnia*, Zaffagnini and Sabelli 1972), and these details of the reproductive mode may change the proportion of offspring produced by central *vs.* terminal fusion even under random expectations (*i.e.*, without specific mechanism favoring one over the other). It may therefore be more useful to distinguish cases in which both central fusion and terminal fusion occur, possibly in different proportions (we term this "mixed fusion") from cases in which one of them is the exclusive mode of reproduction. With mixed fusion, any given offspring is produced by either central or terminal fusion (two specific meiotic products fuse or meiosis I or meiosis II is suppressed). This leads to a first general prediction, which should enable differentiating automixis from self-fertilization: Independently of whether automixis occurs by central or terminal fusion, the homozygosity of centromeric regions across different chromosomes should be 100% correlated within a given offspring (Figure 1). That is, either all centromeric regions should become homozygous (offspring produced by terminal fusion) or they should all retain parental heterozygosity (offspring produced by central fusion). In contrast, under self-fertilization, each centromeric region is expected to become homozygous or retain parental heterozygosity

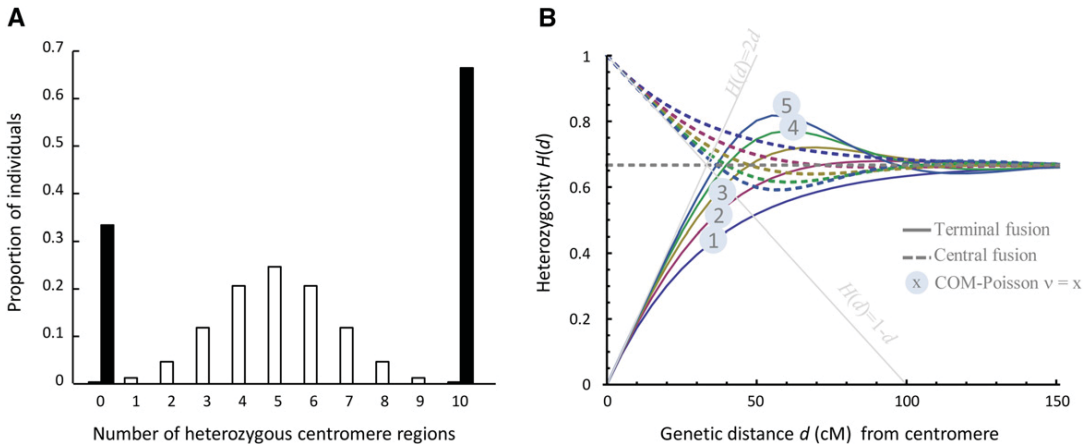


Figure 1 Expected interchromosomal (A) and intrachromosomal patterns (B) of heterozygosity reduction in automictic offspring. (A) The proportion of individuals that retain parental heterozygosity at a given number (out of 10) centromeric regions. Solid bars represent automictic offspring, which should always have either 0 or 10 heterozygous centromeric regions (the relative proportion of individuals with heterozygous *vs.* homozygous regions depends on the proportion of offspring produced by central *vs.* terminal fusion; here 2/3 central fusion is assumed). The open bars represent self-fertilized controls. (B) Expected offspring heterozygosity as a function of the genetic distance from the centromere under central (dashed) and terminal (solid) fusion and different degrees of crossover interference (File S2). $\nu = 1$ corresponds to no interference, and the two gray lines correspond to complete interference. The dashed gray line gives the expected heterozygosity for centromere–distal markers (2/3).

with an independent probability of 0.5 (*i.e.*, independently of the heterozygosity of other centromeric regions in the same individual, Figure 1). A method to determine how the interchromosomal pattern can be assessed if the centromeric positions are unknown, is outlined in Supporting Information, File S1.

Second, within each chromosome, heterozygosity is expected to gradually increase from zero (terminal fusion) or decrease from 100% (central fusion) to 67% of parental heterozygosity with increasing genetic distance from the centromere (Figure 1) (Rizet and Engelmann 1949; Engelstädter *et al.* 2011; Pearcy *et al.* 2011). The leveling off at 67% under both terminal and central fusion occurs because, at large genetic distances from the centromere, recombination effectively distributes alleles at random across the sister and nonsister chromatids. Therefore, both terminal and central fusion result in random sampling without replacement of two alleles from four chromatids and thus to the expected heterozygosity of 67% (once one chromatid is chosen, two of the three remaining chromatids carry a different allele).

The transition from zero or 100% heterozygosity at the centromere to 67% heterozygosity in centromere–distant regions depends on the genetic map distance (*i.e.*, the expected number of crossovers) and on the level of crossover interference (Figure 1, File S2) (Barratt *et al.* 1954; Nace *et al.* 1970; Zhao and Speed 1998). In File S2, we present an original derivation of this relationship, taking advantage of the flexibility of generalized Poisson distributions (Conway–Maxwell Poisson distribution, Sellers *et al.* 2012). With high degree of crossover interference, this relationship may be nonmonotonous (Figure 1). However, the initial slope of the change in heterozygosity close to the centromere is 2*d* (where *d* is the genetic distance in Morgan) under terminal fusion and –*d* under central fusion (Figure 1, File S2), irrespectively of the degree of interference. In contrast, under self-fertilization, expected heterozygosity is 50% of the parental heterozygosity and does not depend on the distance from the centromere nor on the level of crossover interference.

Several other forms of automixis are defined and discussed elsewhere (Bell 1982; Mogie 1986; Suomalainen *et al.* 1987; Stenberg and Saura 2009; Archetti 2010; Lutes *et al.* 2010; Neiman *et al.* 2014; Nougué *et al.* 2015). Their effects on genome-wide heterozygosity reduction are often very different from self-fertilization (*e.g.*, complete loss or complete retention of parental heterozygosity).

Materials and Methods

Origin of clones and outdoor experiments

We use the term "clone" to refer to a strain initiated by a single female and maintained by clonal reproduction. Clones used in this study originated from Russian populations known to contain NMP clones (Ast, BN, MZ, Vol; Galimov *et al.* 2011). They were classified as MP or NMP according to whether or not females of these clones produced male offspring during clonal reproduction when exposed to 400 nM methyl farnesoate, a juvenile hormone analog that has been shown to consistently induce male production in MP clones of *D. magna* (Olmstead and Leblanc 2002; see Galimov *et al.* 2011 for detailed methodology).

Outdoor mass cultures were carried out using two NMP treatments and two MP control treatments: (1) NMP single-clone cultures ("NMP_single") each contained a single NMP clone. Because only one maternal clone was present, the ephippial offspring produced in these cultures were the result

of within-clone mating (if rare males were present) or some form of parthenogenetic reproduction. This treatment was used (as in our earlier study, Galimov *et al.* 2011) to eliminate the possibility of clonal parthenogenesis by examining offspring for segregation of maternal alleles at microsatellite loci. Furthermore, the offspring of one culture (the culture that produced the largest number of offspring) were used to test for genome-wide patterns of heterozygosity using RAD sequencing. (2) NMP multiclone cultures ("NMP_mix") contained two to four different NMP clones (distinguishable at microsatellite loci). They were used to test for the presence of rare males by testing for the occurrence of outcrossed offspring (*i.e.*, crosses between different NMP clones). Outcrossing should occur in the presence of rare males, but not under automixis. Control treatments (3) "MP_single" and (4) "MP_mix," containing single or two to four MP clones, respectively, were used to verify male production and outcrossing under the experimental conditions as well as to assess genome-wide heterozygosity in offspring produced by self-fertilization.

The outdoor cultures were set up under ambient conditions in the botanical garden of Fribourg, Switzerland (46°48′6.00′′N, 7°8′44.04′′E) by transferring ~100 adult females of each clone into buckets containing 40 liters of artificial *Daphnia* medium (Klüttgen *et al.* 1994) as well as a 50-ml initial inoculum of natural microalgae and bacteria (50 μm filtered water from a local garden pond) as well as ~100 g of fresh horse manure to provide nutrients. Some fresh unicellular green algae, *Scenedesmus* sp., were added intermittently throughout the experiment to keep densities high, and natural rain water gradually filled the buckets to ~60 liters.

The experiment took place in two parts: A first batch of cultures was grown outside from March to November 2011, and a second batch from March/April 2013 to October 2013 (Table 1). In both batches, the clones reproduced mostly asexually during summer and fall, with intermittent production of males observed in the MP cultures and ephippia, both in MP and NMP cultures. Even though there was no systematic quantification of ephippia production in this experiment, we did not notice any obvious differences in numbers of ephippia produced between NMP and MP cultures. However, all opened ephippia from NMP cultures were empty (*i.e.*, did not contain embryos), whereas almost all ephippia from MP cultures contained embryos (several dozens of ephippia from each of the two culture types were opened). The results of the first batch suggested the possibility of clonal selection leading to substantially unequal clone frequencies in multiclone cultures and thus reduced probabilities of outcrossing (assuming presence of males and random mating). We therefore intermittently (June, July 15, August 25, and September 17, 2013) restocked all multiclone cultures of the second batch by adding up to 100 nonephippial females of the less frequent clones, after estimating clone frequencies based on microsatellite genotypes of 25 individuals of each culture. The aim of this procedure was to equilibrate clone frequencies and thus to increase the likelihood of outcrossing if rare males were present. Finally, six NMP cultures of the first batch were used to phenotypically search for rare males using large samples (~4000 individuals) taken at the end of the growing season (November 2011), with sex identified under a stereomicroscope. The same was also done for two MP control cultures.

At the end of each growing season (mid November 2011, end of October 2013 for the first and second batches, respectively), all ephippia that had accumulated at the bottom of the buckets were collected and overwintered (which is necessary for later hatching). Overwintering was done either outdoors in a small volume of water placed in the dark (first batch) or in a dark cold room at 0° (second batch). In the subsequent spring, hatching tests were carried out by transferring the ephippia to fresh *Daphnia* medium and keeping them under warm and high-light conditions (ambient Fribourg spring conditions in the first batch, ~20° greenhouse conditions in the second batch). The containers were carefully checked for hatchlings at least every 3rd day, and hatchlings were removed and stored in ethanol at −20° for later genotyping or grown in isolation to establish cultures of offspring clones. Overall, the 2011 batch yielded more hatchlings than the 2013 batch, likely due to environmental effects during growth or hatching.

DNA extraction and microsatellite analysis

Genomic DNA was extracted using the HotSHOT protocol (Montero-Pau *et al.* 2008) and nine diagnostic microsatellite loci (Table S1) were used to distinguish outcrossed from non-outcrossed offspring (the latter resulting from within-clone mating or parthenogenetic reproduction), as well as to check for segregation of markers that were heterozygous in the parent clones. We set up PCR reactions of 10 μl, using the Qiagen Multiplex PCR master mix (Qiagen, Venlo, The Netherlands). Cycling was performed following the recommendations of the manufacturer. Fragment lengths were analyzed using GeneMapper Software version 4.0 (Applied Biosystems, Foster City, CA) with GeneScan-500 LIZ as an internal size standard.

RAD sequencing

To obtain markers throughout the genome, at which heterozygosity could be assessed, we used RAD sequencing (Baird *et al.* 2008), using eight hatchlings from a single-clone NMP culture (clone AST-01-04, bucket V04), as well as 27 hatchlings from a single-clone MP culture (clone RM1-18 MP, bucket B19). Only eight offspring of an NMP clone were used because this was the highest number of offspring from a single-clone NMP culture that could successfully be grown in clonal culture in the laboratory before DNA extraction (several other hatchlings died before reproduction or were sterile). We used a RAD-sequencing protocol based on Etter *et al.* (2011) with a few modifications as specified below. Two libraries were prepared: one containing the offspring of the NMP single-clone culture, the other containing the offspring of the MP single-clone culture, with each offspring individually labeled. Each library also contained two independent

Table 1 Origins of clones, sex rations, number of hatchlings, as well as numbers of within-clone and outcrossed offspring in each of the cultures

Bucket ID	Batch	Treatment	Origin of clones	*N* males	*N* females	*N* hatchlings	*N* genotyped	*N* within-clone offspring	*N* outcrossed offspring
V02	2011	NMP_single	Vol			7	2	2	0
V03	2011	NMP_single	MZ			0			
V04	2011	NMP_single	Ast			28	14	14	0
V08	2011	NMP_single	Vol	0	4629	3	3	3	0
V10	2011	NMP_single	Ast	0	5370	3	3	3	0
V21	2011	NMP_single	MZ			8	5	5	0
B11	2013	NMP_single	Ast			11			
B12	2013	NMP_single	Vol			0			
B13	2013	NMP_single	MZ			1			
B14	2013	NMP_single	Ast			0			
B15	2013	NMP_single	Ast			0			
V01	2011	NMP_mix	MZ, Vol			13	3	3 (same parent)	0
V05	2011	NMP_mix	BN, Vol			11	7	7 (same parent)	0
V06	2011	NMP_mix	MZ, Vol			1	1	1	0
V07	2011	NMP_mix	BN, Vol	0	5105	1			
V09	2011	NMP_mix	MZ, Vol	0	4256	10	3	3 (same parent)	0
V11	2011	NMP_mix	BN (2x)	0	5550	0			
V12	2011	NMP_mix	Ast, BN, MZ, Vol			2	2	1+1 (two different parents)	0
V15	2011	NMP_mix	MZ (2x)	0	1015	1			
V17	2011	NMP_mix	MZ (2x)			4	4	4 (same parent)	0
V19	2011	NMP_mix	Ast, BN, MZ, Vol			1			
V20	2011	NMP_mix	BN (2x)			0			
B20	2013	NMP_mix	Ast, MZ, Vol			1	1	1	0
B21	2013	NMP_mix	Ast (3x)			0			
B23	2013	NMP_mix	Ast, MZ, Vol			3	3	1+2 (two different parents)	0
B24	2013	NMP_mix	Ast (3x)			0			
B26	2013	NMP_mix	Ast, MZ, Vol			1	1	1	0
B27	2013	NMP_mix	Ast (3x)			0			
B17	2013	MP_single	MZ			0			
B18	2013	MP_single	MZ			>30			
B19	2013	MP_single	MZ			>30			
D069	2011	MP_mix	BN, Vol	53	440	>30			
D096	2011	MP_mix	BN, Vol	82	232	>30			
D141	2011	MP_mix	BN, Vol	142	224	>30			
D202	2011	MP_mix	BN, Vol	109	220	>30			
B22	2013	MP_mix	MZ (4x)			>30	8	0	8
B25	2013	MP_mix	MZ (4x)			>30	8	1	7
B28	2013	MP_mix	MZ (4x)			>30	8	0	8

Empty cells indicate values that were not assessed in a given culture.

replicates of the parental clone. The details of the RAD-sequencing protocol and the analysis pipeline including quality checks, alignment, SNP calling, and genotype calling, are explained in detail in File S3.

Inter- and intrachromosomal patterns of heterozygosity reduction

Putative centromere locations were inferred from the genetic map as corresponding to large, nonrecombining regions, of which each linkage group contains exactly one, except linkage group 3 which has two such regions (*D. magna* genetic map v4.0.1, File S6; M. Dukić *et al.*, unpublished results). Centromeric regions were defined as consisting of all scaffolds (or parts of scaffolds) with the centimorgan position of these nonrecombining regions. Average heterozygosity as a function of the distance from the putative centromere was calculated for each chromosome arm separately by using a moving average, including markers within 5 cM on either side of the focal marker (but in all cases excluding markers at a distance of 0 cM from the centromeric regions). Subsequently, the averages and standard errors (SE) of these estimates were calculated across chromosome arms, and confidence limits were calculated as 1.96 SE.

To estimate the distance from the centromere of microsatellite loci, we first mapped each primer pair to the current *D. magna* assembly v2.4. Subsequently, we retrieved the position on the genetic map v4.0.1 of the closest marker on the same scaffold. In this way, we were able to obtain estimated map locations for six of the microsatellite loci (Table S2).

Probability of within-clone mating in the presence of rare males

The absence of outcrossed offspring in NMP multiclone cultures does not necessarily indicate the absence of rare males because a low number of offspring could, by chance be produced exclusively by within-clone mating. Hence, we calculated

the probability of observing zero outcrossed offspring in the presence of rare males under the assumption of random mating among the clones present at the time of resting egg production in each NMP multiclone culture. Under random mating, the probability of within-clone mating of a given clone i in a given culture j is equal to its squared frequency, f_i^2, and the overall expected frequency of within-clone mated offspring is $\Sigma(f_i^2)$, summed across all clones present in the culture. The probability of observing only offspring produced by within-clone mating among N offspring (*i.e.*, the probability that despite the presence of males not a single outcrossed offspring was observed) then equals $pr_j = [\Sigma(f_i^2)]^N$, and the combined probability across all cultures is the product $\Pi(pr_j)$.

Because the frequencies of clones at the time of resting egg production were unknown, we assumed two contrasting scenarios: First, we assumed that all original parent clones were still present at equal frequency at the time of resting egg production. This scenario maximizes the probability of outcrossing. Therefore, we also used a second, more conservative scenario: We assumed that the frequency of each parent clone at the moment of resting egg production was equal to its proportional contribution to the offspring generation. For this second scenario, we only used buckets in which offspring from more than one parent clone were present (for the other buckets, the expected frequency of within-clone mated offspring under this scenario is 100%).

Data availability

All demultiplexed read data used for genotyping were submitted to the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA): BioProject ID PRJNA279333. The reference genome used for mapping and annotation is available at http://wfleabase.org/ (dmagna_v.2.4_20100422). The full raw and corrected SNP datasets, as well as the genetic map v4.0.1 are available as supporting information (File S4, File S5, and File S6).

Results

Sex ratios

We identified the sex of 25,925 NMP individuals of *D. magna*, sampled in late season from six outdoor cultures of NMP clones, but did not find a single male (Table 1). At the same time, cultures of MP clones contained between 10.8% and 38.8% males (mean across populations = 27.2%, SE = 6.1%, total $N = 1502$). Combined with data from our earlier study (Galimov *et al.* 2011), we have now identified the sex of 33,764 NMP individuals but did not find a single male. This yields an overall upper 95% confidence limit for the true proportion of males of $\sim 10^{-4}$ (Clopper–Pearson confidence interval $1 - (\alpha/2)^{1/N}$). However, the experiment involved many more individuals than the ones that were checked ($>10^5$ individuals across the whole duration of the experiment and all NMP cultures combined). Thus the presence of rare males cannot entirely be excluded through these phenotypic observations alone.

Number of hatchlings

A total of 110 hatchlings were found in NMP cultures (between 0 and 28 per culture, Table 1). Of these, 61 were found in cultures containing just a single NMP clone and 49 in cultures containing multiple NMP clones. All MP control cultures except one yielded >30 offspring (Table 1), some of them even many more. Even though numbers of hatchlings >30 were not estimated systematically, this fits with our experience from similar experiments, where MP cultures usually yielded hundreds to thousands of hatchlings, though in rare cases only low numbers or even none (*e.g.*, Haag and Ebert 2007).

Microsatellite genotypes of offspring from cultures containing single NMP clones

In total, we investigated microsatellite genotypes of 27 offspring from cultures containing single NMP clones. In all cases, these offspring showed segregation of maternal alleles (Table S1), thus excluding clonal parthenogenesis. Average heterozygosity across all cultures and loci was 0.61 (SE = 0.04), which is significantly different from 0.5 ($N = 152$, $\chi^2 = 6.7$, $P = 0.0094$), but not from 0.67 ($\chi^2 = 2.9$, $P = 0.090$). Nonetheless, two loci had heterozygosities that were significantly lower than 0.67 (locus B008: $N = 5$, heterozygosity = 0, binomial $P = 0.005$, locus B096: $N = 21$, heterozygosity = 0.36, SE = 0.10, $\chi^2 = 9.2$, $P = 0.0022$) and, in one case, even significantly lower than 0.5 (locus B008: binomial $P = 0.031$, locus B096: $\chi^2 = 1.6$, $P = 0.20$). Indeed heterogeneity among loci was significant (generalized linear model with binomial error distribution using Firth bias correction, likelihood ratio test, $\chi^2 = 23.6$, d.f. = 6, $P = 0.0006$). In contrast, offspring from different cultures or different individuals within cultures did not significantly vary in heterozygosity (tested in the same model as the loci effects, cultures: $\chi^2 = 4.4$, d.f. = 4, $P = 0.35$, individuals nested within cultures: $\chi^2 = 20.9$, d.f. = 22, $P = 0.53$). The heterogeneity among loci was at least partly explained by the distance from the centromere: The two loci with heterozygosities significantly lower than 0.67 (loci B008 and B096) were the two loci estimated to be most closely linked to a centromere (at 25.8 and 3.6 cM, respectively). All other loci had estimated distances from the centromere of >32 cM (Table S2).

Microsatellite genotypes of offspring from cultures containing multiple clones

We obtained microsatellite genotypes of 25 offspring from cultures containing multiple NMP clones (Table 1). Among these, not a single offspring resulting from outcrossing between two of the parent clones was observed. Rather, all 25 offspring were produced by self-fertilization or automictic parthenogenesis: They showed segregation of maternal alleles, just as offspring of the single-clone cultures, but no sign of outcrossing between clones at diagnostic loci (Table 1, Table S1). In all but two of these cultures, all offspring found within the culture were produced by just one parent clone. Two cultures (V12 and B23) contained offspring from two different parent clones (Table 1), but nonetheless no

outcrossed clone was observed (*i.e.*, also these offspring were the result of self-fertilization or automictic parthenogenesis). In stark contrast, 23 of 24 genotyped offspring from cultures containing multiple MP clones were the result of outcrossing between two parent clones (Table 1, Table S1). The frequency of outcrossed offspring was even significantly higher ($\chi^2 = 5.6$, $P = 0.018$) than the 75% expected under random mating and equal frequencies of each of the four parent clones, likely due to inbreeding depression affecting hatching rates of inbred *vs.* outcrossed offspring.

Even though no outcrossed offspring was observed in the cultures containing multiple NMP clones, there is still a possibility that they were produced by the mating of a rare male with a female of the same clone. Assuming random mating of rare males with all females, the overall probability of observing zero outcrossed offspring among the 25 genotyped individuals from the cultures containing multiple NMP clones was calculated under two extreme scenarios (see *Materials and Methods*). Under the first scenario (assuming equal frequency of all introduced parent clones), this probability is very low ($\sim 10^{-9}$). The second scenario (frequency of parent clones equal to their contribution to the offspring generation) could be assessed only for the five offspring from cultures V12 and B23, in which self-fertilized/automictic offspring from more than one parent were present. Under this scenario, the probability of observing zero outcrossed offspring among the five offspring in these cultures is 0.043.

Genome-wide patterns of heterozygosity assessed by RAD sequencing

RAD sequencing was carried out on eight offspring of the AST-01-04 NMP clone (the remaining offspring of this clone died or did not reproduce and therefore it was not possible to obtain sufficient amounts of DNA). Average genome-wide heterozygosity of these eight offspring was 0.54 ($N = 2523$ loci). It ranged from 0.24 to 0.67 among linkage groups and from 0.40 to 0.73 among individuals (Table S3). The relatively high variation among individuals and linkage groups was expected because only few recombination events occur per meiosis and chromosome (Routtu *et al.* 2010, 2014), so that many linked markers show identical inheritance patterns.

As a control, RAD sequencing was also carried out on 27 offspring of the RM1-18 MP clone. Average genome-wide heterozygosity among these offspring was 0.60 ($N = 1610$ loci), varying among linkage groups between 0.46 and 0.71 (Table S3). This was significantly higher than 0.5 (linkage groups as independent replicates, $t = 3.7$, d.f. $= 9$, $P = 0.005$), but also significantly lower than 2/3 ($t = -2.7$, d.f. $= 9$, $P = 0.023$, though not quite significantly so when linkage groups were weighed according to the number of loci: $P = 0.067$).

Interchromosomal patterns of heterozygosity at putative centromere regions

The analysis carried out here requires knowledge of centromere regions. We use putative centromere regions (large, nonrecombining regions as identified on each linkage group by the genetic map v4.0.1, which maps most scaffolds of the current *D. magna* assembly; M. Dukić *et al.*, unpublished results). An equivalent analysis that does not require assumptions on putative centromere regions is presented in File S1.

The putative centromere regions were either consistently homozygous (seven offspring) or heterozygous (one offspring: V04_04) across all 10 linkage groups (Figure 2). One of the linkage groups (LG3) contained two such regions, but only the region at 90.8 cM showed the same heterozygosity as the putative centromere regions of the remaining linkage groups. Hence, this, rather than the region at 62.5 cM, is the likely centromere region of LG3 (and only this region was considered for all other analyses). The interchromosomal pattern thus strongly suggests that V04_04 was produced by central fusion and the other seven offspring by terminal fusion.

In contrast, the putative centromere regions were not consistently homozygous or heterozygous across all linkage groups within individual offspring of the RM1-18 MP clone, except for one individual, in which all 10 centromere regions were heterozygous (Figure 2). Using the observed average heterozygosity of 0.60, the probability of observing this at least one time by chance among 27 offspring is ~ 0.15 ($p = 0.6^{10}$ is the probability that an individual is heterozygous for the 10 centromeric regions, $(1 - p)^{27}$ that none of the 27 offspring is heterozygous for the 10 centromeric regions, and thus $1 - (1 - p)^{27}$ is the probability that at least 1 is heterozygous for the 10 centromeric regions).

Intrachromosomal patterns of heterozygosity

Among the seven offspring of the NMP clone that were presumably produced by terminal fusion, heterozygosity gradually increased with distance from the centromere (Figure 3) and reached an average heterozygosity of clearly >0.5 already at a distance of 50 cM. Heterozygosity in distal regions was even somewhat higher than the expected 0.67, though this was only marginally significant (the lower 95% confidence interval calculated by using chromosome arms as independent replicates mostly included 0.67). Heterozygosity of the individual produced by central fusion (offspring V04_04) averaged across all markers and all chromosome arms was 0.73, and 0.67 if only markers located >50 cM from the centromere were considered. In contrast, heterozygosity among the 27 offspring of the RM1-18 MP clone did not vary in any systematic way along the chromosomes nor according to the distance from the centromere (Figure 3).

Discussion

Reliable distinction of automixis from self-fertilization requires genomic data

Our results demonstrate that the ephippial offspring of the *D. magna* NMP clone investigated by RAD sequencing were produced by automixis, mostly but not exclusively by terminal fusion. The microsatellite results on the offspring of the other clones strongly suggest that this was also the case for the offspring of the other clones (lower heterozygosity at

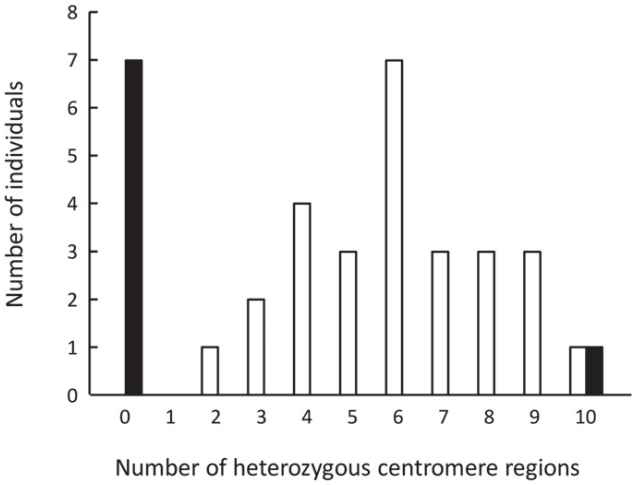


Figure 2 Observed number of individuals that retained parental heterozygosity at a given number (out of 10) of centromeric regions. Solid bars represent offspring of the AST-01-04 NMP clone, open bars, offspring of the RM1-18 MP clone. For LG3 only the region at 90.8 cM was considered.

centromere–proximal markers). However, an unequivocal distinction from self-fertilization was only possible owing to the use of genomic data, which also allowed confirmation of the centromere locations. While the phenotypic results were suggestive of automixis (no males detected in NMP cultures, no cross-breeding observed in cultures containing multiple NMP clones), they were nonetheless not entirely conclusive. Very low male frequencies could not be excluded, despite large sample sizes, and the expectation that the presence of males would lead to outcrossed offspring was based on specific assumptions that could not be verified. Moreover, the average microsatellite heterozygosity was similar to the average genome-wide heterozygosity of the self-fertilized controls and therefore also inconclusive. Also at RAD loci, automictic offspring retained 54% of parental heterozygosity, due to a predominance of terminal rather central fusion and due to inclusion of both centromere–distal and centromere–proximal markers. If only average heterozygosity had been assessed, this could easily have been mistaken as consistent with self-fertilization rather than automixis (average observed heterozygosity in self-fertilized offspring was 60%). This shows that average offspring heterozygosity is not necessarily a reliable indicator of the breeding system.

Unequivocal evidence for automixis was obtained only when the interchromosomal and intrachromosomal patterns of genome-wide heterozygosity were analyzed. These patterns are clearly inconsistent with self-fertilization, as shown by our parallel analysis of self-fertilized controls. In addition, the interchromosomal patterns provide a direct estimate of the proportion of offspring produced by terminal *vs.* central fusion: Offspring produced by terminal fusion are homozygous at the centromeric regions of all chromosomes; offspring produced by central fusion retain full parental heterozygosity at all these regions (see also Oldroyd *et al.* 2008). Even if the positions of the centromeres are unknown, the interchromosomal patterns can be analyzed by investigating if specific segregation patterns (among individuals) occur consistently on all chromosomes (File S1).

The intrachromosomal patterns of heterozygosity in automictic offspring can be used to map the centromeres, an approach that has been used both in natural automicts (Barratt *et al.* 1954, 2004) and in organisms in which automixis can be induced artificially (or meiotic tetrads or half-tetrads can be recovered by other means) (Lindsley *et al.* 1956; Eppig and Eicher 1983; Johnson *et al.* 1996; Zhu *et al.* 2013). Also in our study, the intrachromosome heterozygosity patterns among offspring produced by terminal fusion confirm the presumed locations of centromeres in *D. magna* and indicate that all *D. magna* chromosomes are metacentric. Furthermore, our results also indicate that the centromere on LG3 is located at 90.8 cM rather than 62.5 cM (two large, nonrecombining regions are found on this linkage group; Dukić *et al.*, unpublished results).

Even though each centromeric region did contain markers that were homozygous in all individuals produced by terminal fusion, there were also, in each of these regions, some markers that were not fully homozygous (average heterozygosity across centromeric regions was 12%, Figure 3). These 12% of unexpected genotypes are likely due to a combination of genotyping error (false heterozygote calls), erroneous mapping of reads from paralogous loci to single loci (*e.g.*, centromeric markers that were heterozygous in all individuals, as found on several linkage groups, Figure S1), errors in the genetic map and/or noncollinearity between chromosomes in our study population compared to the clones on which the *D. magna* map is based. Errors in the genetic map and noncollinearity would have the effect that loci that were mapped by us to the centromeric regions are in reality not in these regions, which could explain their nonzero heterozygosity.

Automixis and diversity of breeding systems in Daphnia

While most *Daphnia* species have so far been thought to produce diapause stages exclusively by sexual reproduction, a few species regularly produce parthenogenetic diapause stages (*e.g.*, obligate parthenogenetic strains of *D. pulex*; Hebert and Crease 1980). Yet, in these cases, offspring do not show segregation of maternal alleles and are therefore believed to be clonal offspring, just as the offspring resulting from parthenogenetic production of subitaneous (directly developing) eggs during the regular asexual part of the life cycle in *Daphnia* (Hebert and Ward 1972, but see comments on clonality below). Hence our results constitute the first demonstration of classical automixis in *Daphnia*.

The finding of rare automixis in NMP clones of *D. magna* is important for our understanding of the NMP/MP polymorphism. Nonmale producing clones of *D. magna* participate in sexual reproduction only through their female function (Galimov *et al.* 2011) and were therefore speculated to be unable to colonize new populations on their own. However, rare automixis may allow these populations to persist through the first period of diapause and may therefore allow

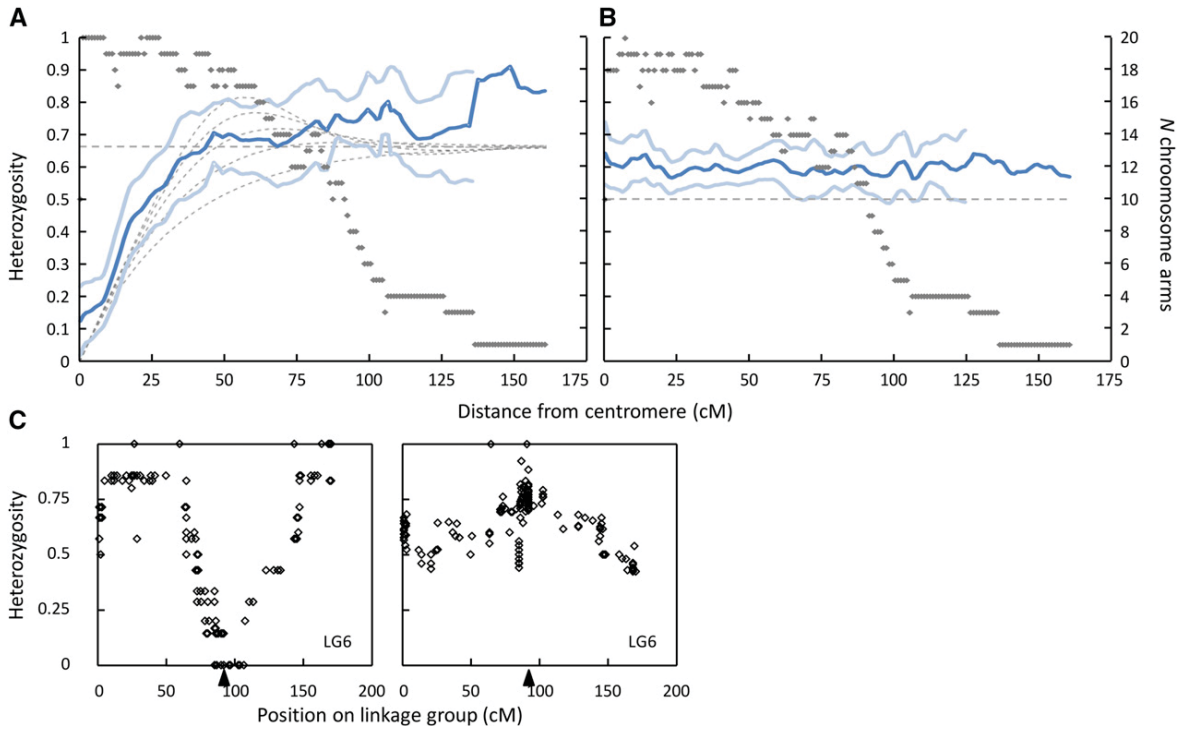


Figure 3 Heterozygosity as a function of the distance from the centromere under (A) automixis (terminal fusion only, $N = 7$ offspring) and (B) self-fertilization ($N = 27$ offspring). Dark blue lines represent averages across all chromosome arms with N chromosome arms (gray dots) according to the secondary y-axis. Light blue lines represent the 95% confidence limits, and the dashed lines, the expected heterozygosity and asymptotes under different degrees of recombination interference (see Figure 1). (C) Realized heterozygosity along linkage group 6 (automictic offspring, left; self-fertilized offspring, right) for illustration. The black triangle shows the presumed centromere position. The patterns of all linkage groups are shown in Figure S1. All heterozygosities are expressed in percentage of parental heterozygosity.

single NMP clones to colonize new populations. This may be especially important in environments with frequent extinction–colonization dynamics.

However, there is no reason to believe that rare automixis is limited to NMP clones. Production of empty ephippia (shells of resting stages not containing embryos) is frequently observed also in MP clones, when females are grown in the laboratory in isolation from males and even in nature (Ebert 2005). Yet, possible rare automictic reproduction in these cultures would be much more difficult to detect than in NMP clones because MP clones regularly produce males under the conditions needed to stimulate ephippia production. Indeed, rare automictic parthenogenesis occurs in a large number of organisms in the form of rare, spontaneous hatching of unfertilized eggs ("tychoparthenogenesis") with diploidy restored via automixis (Bell 1982; Schwander *et al.* 2010; Neiman *et al.* 2014).

In other species, automixis is a regular form of reproduction, but often exclusively or almost exclusively with central fusion. Examples are many fungal species, including yeast, in which central fusion during within-tetrad mating (= automixis) is assured by a mating type locus (unless there is a mating type switch or recombination between the centromere and the mating type locus) (Antonovics and Abrams 2004). Another example comes from several social insects, which can reproduce parthenogenetically by central fusion, and, additionally, show very low rates of crossover within the chromosomes (Baudry *et al.* 2004; Oldroyd *et al.* 2008; Rey *et al.* 2011). With very low recombination rates, central fusion effectively approaches clonality because central fusion assures that parental heterozygosity is retained at the centromeres, and low rates of crossover result only in a slow decay of heterozygosity in the centromere–distal regions.

Automixis has been hypothesized to represent an intermediate step in the evolutionary pathway from sexual to clonal reproduction (Schwander *et al.* 2010). According to this hypothesis, rare automictic reproduction (tychoparthenogenesis) with mixed fusion may become more frequent, with subsequent selection for increased rates of central fusion and repression of recombination. Parthenogenetic reproduction in *Daphnia* has indeed been termed an intermediate between clonal and automictic reproduction because subitaneous, parthenogenetic eggs are produced by a modified meiosis rather than by mitosis (Hiruta *et al.* 2010; Hiruta and Tochinai 2012). Specifically, the homologs pair and start to separate, but meiosis I is not completed, and sister chromatids of a diploid set of chromosomes are separated during meiosis II (Hiruta *et al.* 2010). In other words, meiosis I is suppressed, which is identical to central fusion (Asher 1970), but it is also indistinguishable from purely clonal reproduction as long as no recombination occurs during the paring of

homologs (a low degree of exchange occurs in *D. pulex*, Omilian *et al.* 2006; Tucker *et al.* 2013). More generally, this suggests that the mechanism of clonal parthenogenesis may, also in other diploid organisms, be an extreme form of automixis (central fusion with no or very low levels of recombination) and thus a meiosis-derived process rather than production of fertile eggs by mitosis. This distinction is important because it implies that the evolution of asexuality in these taxa may have involved strong heterozygosity-reducing processes (terminal fusion, central fusion with recombination). Hence the classical view that clonal diploids maintain their heterozygosity at least on short evolutionary timescales (thus for instance avoiding inbreeding, *e.g.*, Haag and Ebert 2004) may not have been true during the initial evolution of asexuality. Hence it may be necessary to more explicitly account for the mechanism of this transition to fully understand the selection pressures acting during the evolution of parthenogenesis from sexuality. Furthermore, the same processes may also be important for understanding the maintenance of asexuality: If clonal parthenogenesis is indeed meiosis derived, there may be residual rates of recombination during homolog pairing (Hiruta *et al.* 2010), such that transitions to homozygosity and loss of complementation may occur at higher rates than under purely mitotic parthenogenesis (Archetti 2004, 2010; Nougué *et al.* 2015).

Inbreeding depression in self-fertilized and automictic offspring

In the self-fertilized offspring, observed heterozygosities were higher than the expected 50% for the majority of the linkage groups. This suggests that the parent clone carried loci contributing to inbreeding depression, that is, loci with recessive or partly recessive deleterious alleles on these linkage groups (Fu and Ritland 1994b). Indeed, the realized heterozygosities can deviate from the expected ones in inbred individuals due to selection, and such deviations are a form of inbreeding depression (Fu and Ritland 1994a; Wang and Hill 1999). Also the higher than expected number of outcrossed offspring in the cultures containing multiple MP clones is evidence for inbreeding depression in the control cultures.

The automictic offspring also showed signs of inbreeding depression: Only few hatchlings survived to adulthood and were sufficiently fecund so that they could successfully be taken into clonal culture (Table 1). Furthermore, observed offspring heterozygosities also tended to be higher than the expected ones, even after accounting for high levels of crossover interference. A closer examination of the contribution of selection to the genome-wide patterns of observed heterozygosity is not possible due to the low number of automictic offspring investigated, and also due to complicating effects of possible genotyping errors and other possible errors (alignment, mapping, collinearity, see above) in our analysis. Due to these uncertainties, our prediction that the initial increase in heterozygosity at short distances from the centromere should be $2d$ (where d is the genetic distance in morgans), if it is not influenced by selection, could not be evaluated with the present data. Nonetheless, the strong initial increase in heterozygosity at distances up to 100 cM from the centromere is inconsistent with the absence of both recombination interference and selection, but rather indicates the action of one or both of these processes. If a larger number of offspring is analyzed and selection is estimated independently (*e.g.*, by analyzing loci at $>$100 cM from the centromere) or can be excluded (*e.g.*, by investigating zygotes), the analysis of heterozygosity patterns among automictic offspring may be used to investigate the degree of crossover interference.

Conclusions

Overall, our study shows that the mode of reproduction in automictic *vs.* self-fertilizing species can be inferred from the heterozygosity patterns among offspring. However, our study also illustrates that it was only due to the availability of genomic rather than sparse marker data that these inferences were robust to the complicating effects of recombination interference and selection. The same applies to the distinction between terminal and central fusion in species that use a mix of these two modes of reproduction (not necessarily in the ratios corresponding to random fusion). More generally, our findings support the idea that obtaining genome-wide heterozygosity data from mothers and a limited number of offspring may be a widely applicable and accessible approach to study breeding systems in species with cryptic or mixed modes of reproduction.

Acknowledgments

We thank the Moscow Zoo and N. I. Skuratov for sampling permits and support and the Botanical Garden of Fribourg, J. Schöpfer, G. Jacob, B. Walser, J. Lohr, and D. Frey for help with outdoor cultures. We thank the Department of Biosystem Science and Engineering of the Eidgenössische Technische Hochschule Zurich, in particular C. Beisel and I. Nissen for Illumina sequencing of the RADseq libraries, and we gratefully acknowledge support by T. Mathieu, D. Deguedre, and the platform Terrain d'Expériences of the Laboratoire d'excellence (LabEx) Centre Méditerranéen de l'Environnement et de la Biodiversité (CeMEB), as well as M.-P. Dubois and the platform Service des marqueurs génétiques en écologie (SMGE) at CEFE. We also thank S. Gonzalez-Martinez and two anonymous reviewers for their constructive comments on an earlier version of the paper. This work was supported by the Swiss National Science Foundation (grant no. 31003A_138203), the Russian Foundation of Basic Research, Agence nationale de la recherche (project SilentAdapt), and by the European Union (Marie Curie career integration grant PCIG13-GA-2013-618961, DamaNMP). The authors declare no competing financial interests.

Literature Cited

Antonovics, J., and J. Y. Abrams, 2004 Intratetrad mating and the evolution of linkage relationships. Evolution 58: 702–709.

Archetti, M., 2004 Recombination and loss of complementation: a more than two-fold cost for parthenogenesis. J. Evol. Biol. 17: 1084–1097.

Archetti, M., 2010 Complementation, genetic conflict, and the evolution of sex and recombination. J. Hered. 101: S21–S33.

Asher, J. H., 1970 Parthenogenesis and genetic variability. II. One-locus models for various diploid populations. Genetics 66: 369–391.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

Barratt, R. W., D. Newmeyer, D. D. Perkins, and L. Garnjobst, 1954 Map construction in *Neurospora crassa*. Adv. Genet. 6: 1–93.

Baudry, E., P. Kryger, M. Allsopp, N. Koeniger, D. Vautrin *et al.*, 2004 Whole-genome scan in thelytokous-laying workers of the Cape honeybee (*Apis mellifera capensis*): Central fusion, reduced recombination rates and centromere mapping using half-tetrad analysis. Genetics 167: 243–252.

Bell, G., 1982 *The Masterpiece of Nature: The Evolution and Genetics of Sexuality.* University of California Press, Berkeley.

Bos, C. J. E. (Editor), 1996 *Fungal Genetics: Principles and Practice*. Marcel Dekker, New York.

Charlesworth, D., and J. H. Willis, 2009 The genetics of inbreeding depression. Nat. Rev. Genet. 10: 783–796.

David, P., B. Pujol, F. Viard, V. Castella, and J. Goudet, 2007 Reliable selfing rate estimates from imperfect population genetic data. Mol. Ecol. 16: 2474–2487.

Delmotte, F., N. Leterme, J. P. Gauthier, C. Rispe, and J. C. Simon, 2002 Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. Mol. Ecol. 11: 711–723.

Ebert, D., 2005 *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia. Bethesda: National Library of Medicine*. National Center for Biotechnology Information, Bethesda, MD. Available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books.

Engelstädter, J., C. Sandrock, and C. Vorburger, 2011 Contagious parthenogenesis, automixis, and a sex determination meltdown. Evolution 65: 501–511.

Eppig, J. T., and E. M. Eicher, 1983 Application of the ovarian teratoma mapping method in the mouse. Genetics 103: 797–812.

Etter, P. D., J. L. Preston, S. Bassham, W. A. Cresko, and E. A. Johnson, 2011 Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. PLoS One 6: e18561.

Fu, Y.-B., and K. Ritland, 1994a Evidence for the partial dominance of viability genes in *Mimulus guttatus*. Genetics 136: 323–331.

Fu, Y.-B., and K. Ritland, 1994b On estimating the linkage of marker genes to viability genes controlling inbreeding depression. Theor. Appl. Genet. 88: 925–932.

Galimov, Y., B. Walser, and C. R. Haag, 2011 Frequency and inheritance of non male producing clones in *Daphnia magna*: Evolution towards sex specialization in a cyclical parthenogen? J. Evol. Biol. 24: 1572–1583.

Haag, C. R., and D. Ebert, 2004 A new hypothesis to explain geographic parthenogenesis. Ann. Zool. Fenn. 41: 539–544.

Haag, C. R., and D. Ebert, 2007 Genotypic selection in *Daphnia* populations consisting of inbred sibships. J. Evol. Biol. 20: 881–891.

Hartl, D. L., and A. G. Clark, 2007 *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.

Hebert, P. D. N., and R. D. Ward, 1972 Inheritance during parthenogenesis in *Daphnia magna*. Genetics 71: 639–642.

Hebert, P. D. N., and T. Crease, 1980 Clonal coexistence in *Daphnia pulex* (Leydig): another planktonic paradox. Science 207: 1363–1365.

Hiruta, C., and S. Tochinai, 2012 Spindle assembly and spatial distribution of γ-tubulin during abortive meiosis and cleavage division in the parthenogenetic water flea *Daphnia pulex*. Zoolog. Sci. 29: 733–737.

Hiruta, C., C. Nishida, and S. Tochinai, 2010 Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. Chromosome Res. 18: 833–840.

Hood, M. E., and J. Antonovics, 2004 Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. Genetics 166: 1751–1759.

Innes, D. J., and R. L. Dunbrack, 1993 Sex allocation variation in *Daphnia pulex*. J. Evol. Biol. 6: 559–575.

Johnson, S. L., M. A. Gates, M. Johnson, W. S. Talbot, S. Horne *et al.*, 1996 Centromere-linkage analysis and consolidation of the Zebrafish genetic map. Genetics 142: 1277–1288.

Klüttgen, B., U. Dulmer, M. Engels, and H. T. Ratte, 1994 ADaM, an artificial freshwater for the culture of zooplankton. Water Res. 28: 743–746.

Lampert, K. P., D. K. Lamatsch, P. Fischer, J. T. Epplen, I. Nanda *et al.*, 2007 Automictic reproduction in interspecific hybrids of poeciliid fish. Curr. Biol. 17: 1948–1953.

Lindsley, D. L., G. Fankhauser, and R. R. Humphrey, 1956 Mapping centromeres in the axolotl. Genetics 41: 58–64.

Lutes, A. A., W. B. Neaves, D. P. Baumann, W. Wiegraebe, and P. Baumann, 2010 Sister chromosome pairing maintains heterozygosity in parthenogenetic lizards. Nature 464: 283–286.

Mogie, M., 1986 Automixis: its distribution and status. Biol. J. Linn. Soc. Lond. 28: 321–329.

Montero-Pau, J., A. Gómez, and J. Muñoz, 2008 Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. Limnol. Oceanogr. Methods 6: 218–222.

Nace, G. W., C. M. Richards, and J. H. Asher, 1970 Parthenogenesis and genetic variability. I. Linkage and inbreeding estimations in the frog, *Rana pipiens*. Genetics 66: 349–368.

Neiman, M., T. F. Sharbel, and T. Schwander, 2014 Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. J. Evol. Biol. 6(27): 1346–1359.

Normark, B. B., 2003 The evolution of alternative genetic systems in insects. Annu. Rev. Entomol. 48: 397–423.

Nougué, O., N. O. Rode, R. Zahab, A. Ségard, L.-M. Chevin *et al.*, 2015 Automixis in *Artemia*: solving a century-old controversy. J. Evol. Biol., in revision (doi: 10.1111/jeb.12757).

Oldroyd, B. P., M. H. Allsopp, R. S. Gloag, J. Lim, L. A. Jordan *et al.*, 2008 Thelytokous parthenogenesis in unmated queen honeybees (*Apis mellifera capensis*): central fusion and high recombination rates. Genetics 180: 359–366.

Olmstead, A. W., and G. A. Leblanc, 2002 Juvenoid hormone methyl farnesoate is a sex determinant in the crustacean *Daphnia magna*. J. Exp. Zool. 293: 736–739.

Omilian, A. R., M. E. A. Cristescu, J. L. Dudycha, and M. Lynch, 2006 Ameiotic recombination in asexual lineages of *Daphnia*. Proc. Natl. Acad. Sci. USA 103: 18638–18643.

Pearcy, M., O. Hardy, and S. Aron, 2006 Thelytokous parthenogenesis and its consequences on inbreeding in an ant. Heredity 96: 377–382.

Pearcy, M., O. J. Hardy, and S. Aron, 2011 Automictic parthenogenesis and rate of transition to homozygosity. Heredity 107: 187–188.

Rey, O., A. Loiseau, B. Facon, J. Foucaud, J. Orivel *et al.*, 2011 Meiotic recombination dramatically decreased in thelytokous queens of the little fire ant and their sexually produced workers. Mol. Biol. Evol. 28: 2591–2601.

Rizet, G., and C. Engelmann, 1949 Contribution à l'étude génétique d'un Ascomycète tétrasporé: *Podospora anserina* (Ces.). Rehm. Rev. Cytol. Biol. Veg. 11: 201–304.

Routtu, J., B. Jansen, I. Colson, L. De Meester, and D. Ebert, 2010 The first-generation *Daphnia magna* linkage map. BMC Genomics 11: 508.

Routtu, J., M. Hall, B. Albere, C. Beisel, R. Bergeron *et al.*, 2014 An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. BMC Genomics 15: 1033.

Schwander, T., S. Vuilleumier, J. Dubman, and B. J. Crespi, 2010 Positive feedback in the transition from sexual reproduction to parthenogenesis. Proc. Biol. Sci. 277: 1435–1442.

Seiler, J., and K. Schäffer, 1960 Untersuchungen über die Entstehung der Parthenogenese bei Solenobia triquetrella F. R. (Lepidoptera, Psychidae). Chromosoma 11: 29–102.

Sellers, K. F., S. Borle, and G. Shmueli, 2012 The COM-Poisson model for count data: a survey of methods and applications. Appl. Stochastic Models Data Anal. 28: 104–116.

Stenberg, P., and A. Saura, 2009 Cytology of asexual animals, pp. 63–74 in *Lost Sex. The Evolutionary Biology of Parthenogenesis*, edited by I. Schön, K. Martens, and P. van Dijk. Springer Science, Dordrecht, The Netherlands.

Suomalainen, E., A. Saura, and J. Lokki, 1987 *Cytology and Evolution in Parthenogenesis*, CRC Press, Boca Raton, FL.

Tucker, A. E., M. S. Ackerman, B. D. Eads, S. Xu, and M. Lynch, 2013 Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. Proc. Natl. Acad. Sci. USA 110: 15740–15745.

Tucker, K. W., 1958 Automictic parthenogenesis in the honey bee. Genetics 43: 299–316.

Wang, J. L., and W. G. Hill, 1999 Effect of selection against deleterious mutations on the decline in heterozygosity at neutral loci in closely inbreeding populations. Genetics 153: 1475–1489.

Zaffagnini, F., and B. Sabelli, 1972 Karyologic observations on the maturation of the summer and winter eggs of *Daphnia pulex* and *Daphnia middendorffiana*. Chromosoma 36: 193–203.

Zhao, H., and T. P. Speed, 1998 Statistical analysis of ordered tetrads. Genetics 150: 459–472.

Zhu, C., Y. Sun, X. Yu, and J. Tong, 2013 Centromere localization for bighead carp (*Aristichthys nobilis*) through half-tetrad analysis in diploid gynogenetic families. PLoS One 8: e82950.

Communicating editor: S. Gonzalez-Martinez

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.179879/-/DC1

Uncovering Cryptic Asexuality in *Daphnia magna* by RAD Sequencing

Nils Svendsen, Celine M. O. Reisser, Marinela Dukić, Virginie Thuillier, Adeline Ségard, Cathy Liautard-Haag, Dominique Fasel, Evelin Hürlimann, Thomas Lenormand, Yan Galimov, and Christoph R. Haag

DOI: 10.1534/genetics.115.179879

Uncovering cryptic asexuality in *Daphnia magna* by RAD-sequencing

Nils Svendsen, Celine M. O. Reisser, Marinela Dukić, Virginie Thuillier, Adeline Ségard, Cathy Liautard-Haag, Dominique Fasel, Evelin Hürlimann, Thomas Lenormand, Yan Galimov, and Christoph R. Haag

Supporting information

Supporting File S1

Assessing the inter-chromosomal pattern when centromere locations are unknown

The inter-chromosomal prediction for offspring heterozygosity under automixis is that all chromosomes in a given offspring should either retain 100% of parental heterozygosity or become fully homozygous at markers in the centromere regions. If centromere locations are unknown, this prediction cannot directly be assessed. However, if mapped markers are available, it is possible to test for specific "segregation patterns" by tabulating, for each marker, the individuals in which the marker becomes homozygous and in which it retains parental heterozygosity. If offspring are produced by central fusion, one would expect to find on each chromosome markers that retain parental heterozygosity in all offspring. Similarly, with pure terminal fusion, one would expect to find on each chromosome markers that become homozygous in all individuals. A sufficient number of markers is needed so that it can be assumed that each chromosome contains at least one marker that is in full linked with the centromere. If some offspring are produced by a terminal fusion and some by central fusion, one would expect to find on each chromosome markers that are heterozygous in a given set of offspring (those produced by central fusion) and homozygous in the rest (those produced by terminal fusion), with the important point being that it should be the same set of individuals that retain heterozygosity for all these markers and each chromosome should contain at least one of these markers.

We illustrate this with using a reduced set of loci with complete information (no missing genotypes) for all eight automictic offspring (N = 1693 loci). With eight offspring, there are 2^8 = 256 possible segregation patterns, each of which can be represented binary string for offspring1 to offspring8 (zero: homozygous, 1: heterozygous). For instance 00011000 is a marker, which is heterozygous in offspring4 and offspring5, and homozygous in all other offspring. We identified the segregation pattern for each of the 1693 markers and counted how many times and on how many linkage groups each specific segregation pattern occurred. Only one segregation pattern occurred on all ten linkage groups: homozygous in all individuals except individual V04_04. This pattern was shown by a total of 332 loci, with between 12 and 106 loci per linkage group. Moreover, on each linkage group, these markers were located in just one region. The ten other most common segregation patterns (Supporting Table S4) include loci that were heterozygous in all offspring (found on eight linkage groups) and loci homozygous in all offspring (found six linkage groups), but they did not occur in just

a single region in these linkage groups and probably contain some error (genotyping error alignment error, etc.). The only other pattern that was observed on more than four linkage groups is a pattern that is very similar to the presumed centromeric one (Supporting Table S4), and indeed was found in many pericentromeric regions. None of the segregation patterns among the 27 self-fertilized offspring occurred on more than three linkage groups (a total of 769 markers were investigated), except for 34 loci distributed across nine linkage groups that were heterozygous in all individuals. Within each linkage group, these loci did not occur in a single region, and show strongly differing segregation patterns compared to adjacent markers, which suggests that they may be explained by alignment errors (e.g., false mapping of paralogous loci to a single position).

Overall these results show that even without information on the centromere locations it is possible to conclusively infer the mode of reproduction, given a sufficient number of mapped markers. Conversely, the results also show that mapping of centromeres can be achieved and even if some offspring are produced by terminal fusion and others by central fusion, and that the proportion of offspring produced by terminal vs. central fusion can be directly estimated from the same data.

File S2:

Expected offspring heterozygosity under central vs. terminal fusion

Expected heterozygosity $H(d)$ at a distance d (in Morgan) from the centromere can be computed in two steps. The first step is to derive expected heterozygosity $H(x)$ for any fixed number x of crossovers between the marker and the centromere. This can be obtained by recurrence. Under terminal fusion, we have

$$H(x+1) = 1 - H(x) + H(x)/2 \qquad \text{(A1)}$$

Indeed, if the marker was homozygous (1-$H(x)$), it becomes heterozygous with an additional crossing over, and if it was already heterozygous, there is only one chance over two that it will remain heterozygous with an additional crossing over ($H(x)/2$). Hence, with $H(0) = 0$ (i.e., terminal fusion), we obtain

$$H(x) = \frac{2}{3}\left(1 - \left(-\frac{1}{2}\right)^x\right) \qquad \text{(A2)}$$

This function oscillates (0, 1, $\frac{1}{2}, \frac{3}{4}, \frac{5}{8}, \frac{11}{16}, \frac{21}{32}$, ...) and stabilizes at 2/3 after many cross-overs. (Note that heterozygosity under central fusion can be obtained from the result under terminal fusion noting that $H_{cf} = 1 - H_{tf}/2$ and that $H_{cf}(0) = 1$; Engelstädter *et al.* 2011). The second step is to assume that, in absence of interference, the number of crossovers X over a distance d follows a Poisson distribution with mean $2d$ (recalling that 0.5 Morgan corresponds to one cross-over). We obtain

$$H(d) = \sum_{x=0}^{\infty} P(X = x)\frac{2}{3}\left(1 - \left(-\frac{1}{2}\right)^x\right) \qquad \text{(A3)}$$

where $P(X = x)$ is given by the Poisson distribution. We find

$$H(d) = \frac{2}{3}(1 - e^{-3d}) \qquad \text{(A4)}$$

(Engelstädter *et al.* 2011). The equivalent result under central fusion is

$$H(d) = 1 - \frac{1}{3}(1 - e^{-3d}) \quad \text{(A5)}$$

(Rizet and Engelmann 1949; Barratt *et al.* 1954). In order to compute $H(d)$ in presence of interference, we propose here to use Conway-Maxwell Poisson distribution (Sellers *et al.* 2012) that generalizes the Poisson distribution allowing for over or underdispersion (positive interference corresponding to underdispersion). This distribution adds a parameter ν to control for the level of dispersion. Its probability density function is

$$P(X = x) = \frac{\lambda^x}{Z(\lambda, \nu)(x!)^\nu} \quad \text{(A6)}$$

where $Z(\lambda, \nu)$ is a normalization equal to $\sum_x \lambda^x/(x!)^\nu$, which can be expressed using the generalized hypergeometric function

$$Z(\lambda, \nu) = {}_0F_{\nu-1}(\emptyset, \mathbf{1}, \lambda), \quad \text{(A7)}$$

where **1** is a vector of 1 of dimension ν-1. Using the probability density (A7) in Eq. (A6) yields an heterozygosity function $H(d)$ for various degree of interference. This is illustrated in Figure 1. Strong interference leads to a non-monotonic mapping function as more evenly spaced cross over events will cause $H(d)$ to reflect the oscillatory behavior of $H(x)$ (Eq. A2). All mapping functions have a slope of two at d=0 and tend to 2/3 for large d. Non monotonicity arises as soon as there is interference, but it becomes noticeably large for $\nu \geq 2$. This method can also be applied to obtain a standard mapping function $M(d)$ expressing the recombination fraction as a function of the genetic distance. For instance using the Mather formula (Mather 1935)

$$M(d) = \frac{1}{2}\big(1 - P(X = 0)\big) = \frac{1}{2}(1 - Z(\lambda, \nu)^{-1}) \quad \text{(A8)}$$

In both cases, the mapping requires to express $H(d)$ or $M(d)$ not in terms of λ the parameter of the COM-Poisson distribution, but in terms of d (which is half the expected number of cross over, i.e. half the mean of the COM-Poisson distribution). Here again, the mean of the COM-

Poisson can be expressed in terms of generalized hypergeometric functions, but a simpler approximation is sufficient for most purposes:

$$E(X) = 2d = (1 - e^{-2\lambda})\left(\lambda^{1/\nu} - \frac{\nu - 1}{2\nu}\right) + \lambda e^{-4\lambda} \qquad \text{(A9)}$$

Supporting Figure S2 illustrates this mapping. The case $\nu = 1$ corresponds to Haldane mapping, while $\nu = 3$ is close to the Kosambi mapping used in *Drosophila* (Chen 2013). Note that heterozygosity with interference has already been treated by Barratt *et al.* (1954) for the case of central fusion, however using a less general model (necessitating more restrictive assumptions) than the models based on the COM-Poisson distribution (see also, Nace *et al.* 1970; Zhao and Speed 1998). The latter and other count models (e.g., Zhao *et al.* 1995) are increasingly used also to model interference in classical genetic mapping (e.g., Choi *et al.* 2013).

References

Barratt, R. W., D. Newmeyer, D. D. Perkins, and L. Garnjobst, 1954 Map construction in *Neurospora crassa*. Adv. Genet. 6**:** 1-93.

Chen, Z., 2013 *Statistical methods for QTL mapping*. Chapman and Hall/CRC, New York.

Choi, K., X. Zhao, K. A. Kelly, O. Venn, J. D. Higgins *et al.*, 2013 *Arabidopsis* meiotic crossover hotspots overlap with H2A.Z nucleosomes at gene promoters. Nat. Genet. 45**:** 1327-1336.

Engelstädter, J., C. Sandrock, and C. Vorburger, 2011 Contagious parthenogenesis, automixis, and a sex determination meltdown. Evolution 65**:** 501-511.

Mather, K., 1935 Reductional and equational separation of the chromosomes in bivalents and multivalents. J. Genet. 30**:** 53-78.

Nace, G. W., C. M. Richards, and J. H. Asher, 1970 Parthenogenesis and genetic variability. I. Linkage and inbreeding estimations in the frog, *Rana pipiens*. Genetics 66**:** 349-368.

Rizet, G., and C. Engelmann, 1949 Contribution à l'étude génétique d'un Ascomycète tétrasporé: *Podospora anserina* (Ces.) Rehm. Rev. Cytol. Biol. Veg. 11**:** 201–304.

Sellers, K. F., S. Borle, and G. Shmueli, 2012 The COM-Poisson model for count data: a survey of methods and applications. Appl. Stoch. Models Bus. Ind. 28**:** 104-116.

Zhao, H., and T. P. Speed, 1998 Statistical analysis of ordered tetrads. Genetics 150**:** 459-472.

Zhao, H., T. P. T. P. Speed, and M. S. McPeek, 1995 Statistical analysis of crossover interference using the chi-square model. Genetics 139**:** 1045-1056.

File S3:

Detailed RAD-sequencing protocol and analysis of RAD-sequencing data

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/100 mL. The aim of this treatment was to minimize contaminant DNA (i.e., bacterial DNA or algal DNA) in in the gut and on the surface of the body. Genomic DNA was extracted using the Qiagen Blood and Tissue kit following manufacturer's instructions and digested with *Pst*I (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 were 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 *Pst*I cut site. The reads were sorted individually by barcode and filtered to remove reads with uncalled bases or an overall base-call quality score of less than 25. The last five bases of each read were trimmed due to a decrease in base-calling quality. Reads were subsequently aligned to the *Daphnia magna* genome (V2.4, 20100422; *Daphnia* Genomic Consortium, WFleaBase) 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 remaining reads were filtered according to mapping quality (reads that did not map end-to-end, had a mapping quality score of less than 25, or more than eight substitutions compared to the reference genome were discarded).

Assignment of reads to RAD loci (defined by unique 90 bp locations on the reference genome) and genotype calling was performed in Stacks V1.19 (Catchen *et al.* 2011), with a bounded SNP model in *pstacks* (--bound_high of 0.04, according to the base call error rate provided by the sequencing facility). We only retained loci with a maximum of two high frequency haplotypes (i.e. alleles) per locus per individual (maximally two alleles are expected in a diploid individual). Low-frequency haplotypes (i.e., representing less than 2% of the number of reads per locus in a given individual) were discarded due to the possibility of sequencing error. Routines *cstacks* and *sstacks* were operated with default settings and with the -g option to use genomic location as method to group reads. We also used the option –n with a parameter of 2 in *cstacks* (i.e., allowing a maximum of two mismatches between individuals) to reduce the risk of considering paralogous loci as alleles. For genotype calling, the distribution of the minor allele frequency indicated that a large majority of heterozygous loci had a minor allele frequency between 0.2 and 0.5 within individuals. We thus fixed the max_het_seq parameter to 0.2 in the routine *genotypes*. Consequently, 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 data. Loci were also filtered according to sequencing depth: Loci with less than 20 reads were discarded (to reduce uncertainty in genotype calls, Han *et al.* 2014), as were reads with a more than five times higher depth than the average depth across all RAD-loci within a given individual (to reduce the risk of including repetitive elements / multi –copy genes). Finally, we used the automated correction procedure in Stacks to correct potentially miscalled offspring genotypes through a reassessment of the likelihood of genotype calls taking parental genotypes into consideration (Catchen *et al.* 2011). Only loci that were consistently called heterozygous in both replicates of the parental individual were retained.

After genotype calling, loci were mapped to the *Daphnia magna* genetic map v4.0.1 (M. Dukić et al., unpubl., deposited on Dryad). 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. Missing genotypes were inferred only if (i) two other RAD-loci were present on both sides of the missing marker on the same scaffold, (ii) these four other loci indicated that no crossing over had occurred in this region in this particular individual, and (iii) the genotypes of the other individuals for that locus were consistent with correct mapping of the locus (no more than two recombination events compared to loci on either side across the eight offspring of the AST-01-04 clone and no more than four

recombination events across the 27 offspring of the RM1-18 clone). Similarly, suspect genotypes suggesting one crossover immediately before and a second crossover immediately after the locus were removed and treated as missing if (i) the two loci immediately before and the two loci immediately after on the same scaffold suggested no crossover in that region in this particular individual (without considering loci with missing data) and (ii) if the genotypes of the other individuals for that locus were consistent with correct mapping of the locus (using the same criteria as above). We refrained from additional inference of missing genotypes or removal of suspect genotypes because the *D. magna* genetic map was based on a different population (thus some re-arrangement may be possible) and also because the scaffolding of the current assembly may contain some errors.

After all filtering and correcting, we retained 2523 loci for the analysis of the AST-01-04 family (corresponding to the number of heterozygous loci in the parent clone) and 1610 loci for analysis of the RM1-18 family. Considering suspect individual genotypes as missing (0.5% in the AST-01-04 family and 6.4% in the RM1-18 family), 12.6% and 22.9% of all individual genotypes were missing in the two families, respectively, but this could be reduced to 5.7% and 17.8% by inferring missing genotypes according to above criteria. The proportion of missing and suspect genotypes in the RM1-18 family was higher than in the AST-01-04 family, likely due to the lower average sequencing depth (43.3 reads per locus and individual in the RM1-18 library vs. 54.8 reads per locus and individual in the AST-01-04 library). Both the original and the corrected data set will be deposited on Dryad, but only the corrected one was used in the analyses.

References

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait, 2011 Stacks: building and genotyping loci *de novo* from short-read sequences. G3 (Bethesda) 1**:** 171-182.

Han, E., J. S. Sinsheimer, and J. Novembre, 2014 Characterizing bias in population genetic inferences from low-coverage sequencing data. Mol. Biol. Evol. 31**:** 723-735.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25**:** 1754-1760.

Files S4-S6:

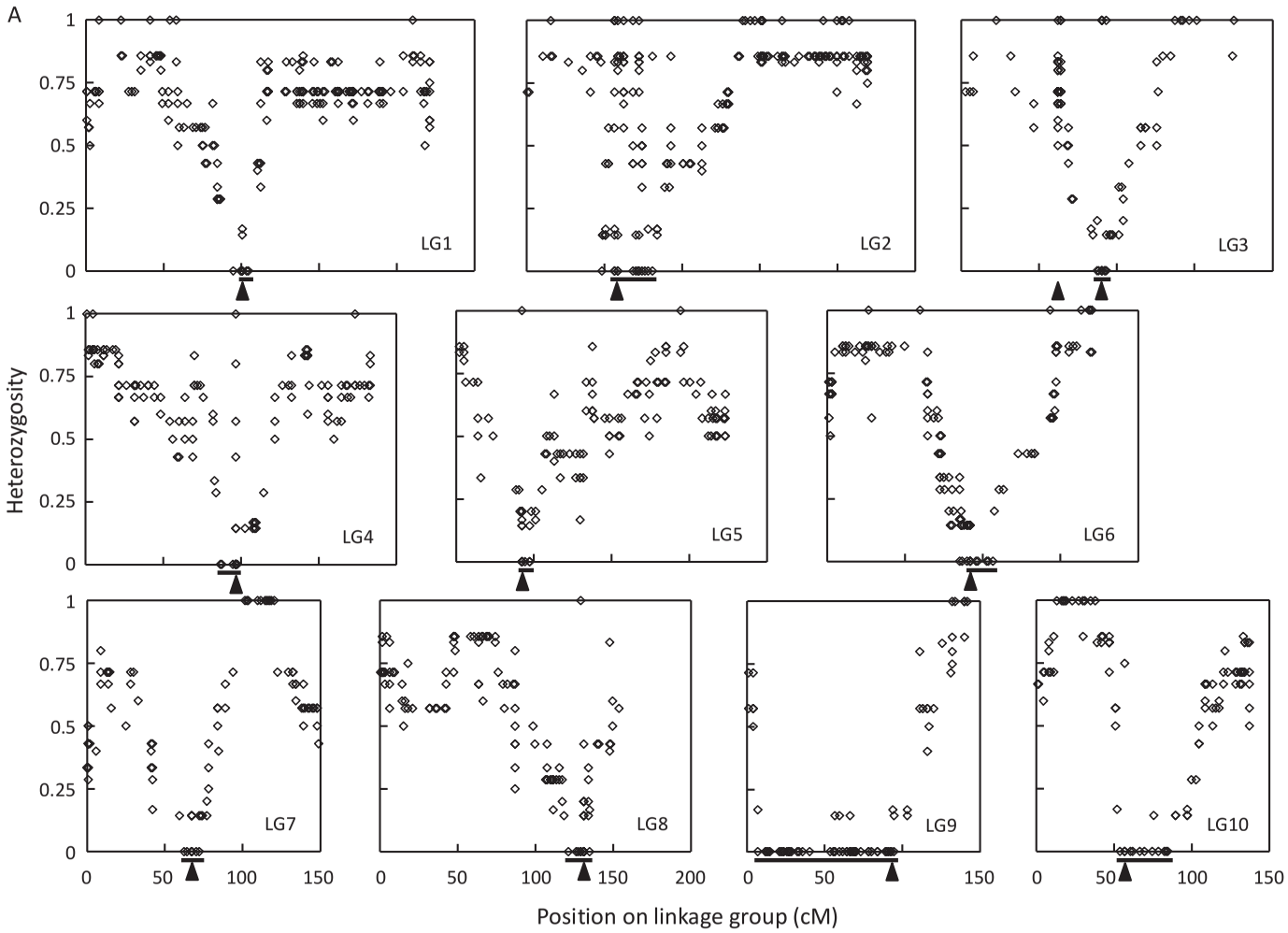
Available for download at

www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.179879/-/

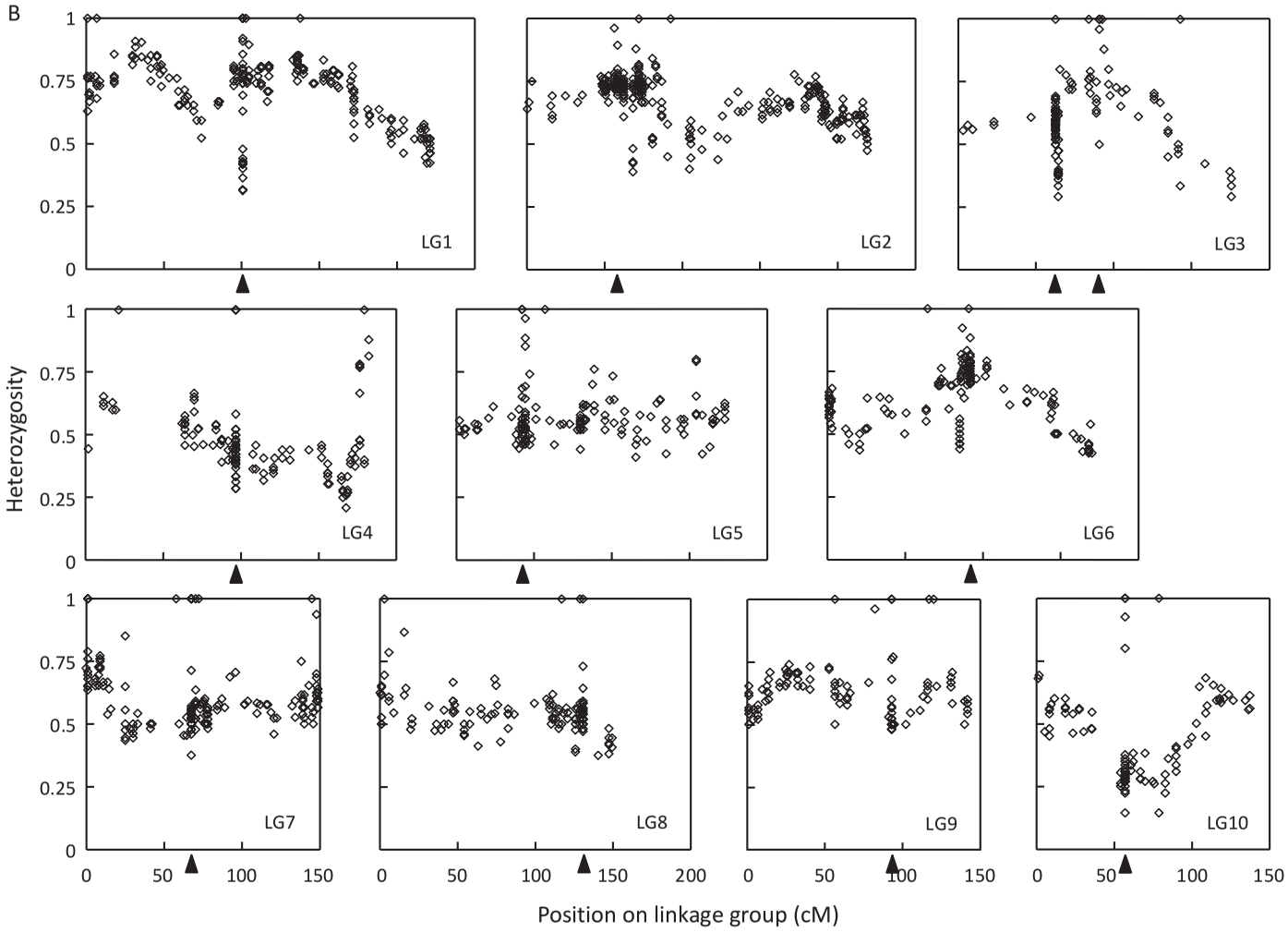
File S4: .xlsx file containing the raw and corrected SNP data set for all offspring of the AST-01-04 NMP clone.

File S5: .xlsx file containing the raw and corrected SNP data set for all offspring of the RM-1-18 MP clone.

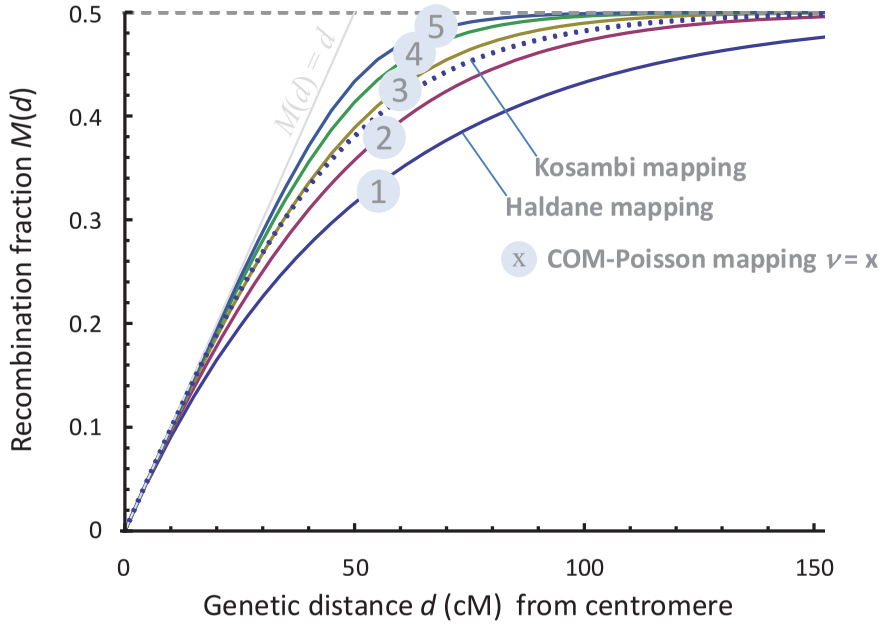
File S6: .xlsx file containing the genetic map v4.0.1.



Supporting Figure S1: (A) Heterozygosity among the seven offspring of the AST-01-04 NMP clone produced by terminal fusion, depending on marker position on the linkage groups of *D. magna*. The bar below each linkage group shows the region around the presumed centromere in which offspring V04-04 was heterozygous, and the triangle shows the presumed centromere location (two in case of LG3). (Figure continued on next page)



Supporting Figure S1: (B) Heterozygosity among the 27 offspring of the RM1-18 MP clone.



Supporting Figure S2: Mapping functions under different degrees of interference. $M(d)$ is the recombination fraction as a function of the distance d (in Morgan) from the centromere and different degrees of interference as measured by the level of underdispersion ν of the COM-Poisson distribution describing the number of crossing-over per unit of genetic map distance. The case $\nu = 1$ corresponds to zero interference (Haldane mapping), while Kosambi mapping is close to the case $\nu = 3$.

Supporting Table S1: Microsatellite genotypes of parents and offspring as well as inferred parents (for offspring only). Potential parents indicate the parental clones that were placed in the buckets from which the offspring was obtained. B008 to B135 are the nine microsatellite loci used in this study (MOLECULAR ECOLOGY RESOURCES PRIMER DEVELOPMENT CONSORTIUM *et al.* 2011), with genotypes indicated by fragment lengths. (Table continued on next pages).

A. Microsatellite genotypes of parent clones.

Clone ID	MP/NMP	Bucket(s)	B008	B030	B045	B050	B064	B074	B096	B107	B135
1MDM6	NMP	V01, V06	165/167	157/160	120/122	234/240	142/146	198/198	243/245	262/270	
AST-01-04	NMP	B20, B23, B26, V04, V10, V12	163/163	155/157	122/122	237/240	138/144	198/200	239/245	272/278	187/191
BN-08	NMP	V12	167/167	157/157	122/122	228/242		198/200	241/241	266/266	
BN-48	NMP	V05	163/179	159/159	120/122	228/245	140/140	200/200	239/241	270/272	
MOS-01-02	NMP	V09, V17	163/172	155/163	122/122	237/240	138/138	196/198	241/241	270/270	
MOS-01-04	NMP	B20, B23, B26, V12, V18	167/172	160/160	122/122	232/232	144/144	198/198	241/241	270/278	189/191
RM1-02	NMP	V21	167/167	155/160	122/122	235/235	144/146	198/200	243/243	268/280	
VV2	NMP	B20, B23, B26, V01, V02, V05, V06, V08, V09, V12	157/165	153/157	122/122	226/244	138/144	198/202	239/243	270/270	187/189
RM1-18	MP	B22, B25, B28	167/172	155/155	127/127	237/237					187/191
RM1-22	MP	B22, B25, B28	161/167	155/160	122/122	232/232					191/191
RM1-35	MP	B22, B25, B28	172/172	155/160	120/122	237/237					189/189
RM1-39	MP	B22, B25, B28	167/169	155/155	122/122	232/242					189/189

B. Microsatellite genotypes of offspring.

Hatchling ID	Treatment	Bucket ID	Potential parents	B008 25 cM	B030 33 cM	B045	B050	B064	B074	B096	B107	B135	Parent1	Parent2	Outcrossed
V02-01	NMP_single	V02	VV2	165/165	153/157	122/122	226/244	138/138	198/202	239/243	270/270		VV2		No
V02-03	NMP_single	V02	VV2	157/157	153/157	122/122	226/244	144/144	198/198	243/243	270/270		VV2		No
V04-01	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	138/144	198/200	239/239	272/278		AST-01-04		No
V04-02	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	138/138	198/200	239/245	278/278		AST-01-04		No
V04-03	NMP_single	V04	AST-01-04	163/163	157/157	122/122	237/240	138/144	198/200	245/245	272/272		AST-01-04		No
V04-05	NMP_single	V04	AST-01-04	163/163	157/157	122/122	237/240	138/144	198/200	239/239	272/278		AST-01-04		No
V04-06	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	138/144	198/200	245/245	272/278		AST-01-04		No
V04-11	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/237	144/144	198/200	245/245			AST-01-04		No
V04-14	NMP_single	V04	AST-01-04	163/163	157/157	122/122	237/240	138/144			272/278		AST-01-04		No
V04-18	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	138/138	198/200	239/239	272/272		AST-01-04		No
V04-19	NMP_single	V04	AST-01-04	163/163	157/157	122/122	237/240	144/144	198/200	245/245			AST-01-04		No
V04-22	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	144/144	198/200	239/239	272/272		AST-01-04		No
V04-23	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	138/138	198/200	239/245	272/278		AST-01-04		No
V04-24	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	138/144	198/200	245/245	272/278		AST-01-04		No
V04-25	NMP_single	V04	AST-01-04	163/163	155/157	122/122	237/240	138/138	198/200	239/239	272/278		AST-01-04		No
V04-27	NMP_single	V04	AST-01-04	163/163	157/157	122/122	237/240	138/144	198/200	239/239	272/278		AST-01-04		No
V08-01	NMP_single	V08	VV2	165/165	153/153	122/122	244/244	138/144	198/202	239/243	270/270		VV2		No
V08-02	NMP_single	V08	VV2	157/157	157/157	122/122	226/244	138/144	198/198	239/239	270/270		VV2		No
V08-03	NMP_single	V08	VV2	165/165	153/157	122/122	226/244	138/144	198/202	239/243	270/270		VV2		No
V10-01	NMP_single	V10	AST-01-04	163/163	155/157	122/122	240/240	138/144	198/198	239/245	272/278		AST-01-04		No
V10-02	NMP_single	V10	AST-01-04	163/163	155/157	122/122	237/237	138/144	198/200	239/245	272/278		AST-01-04		No
V10-03	NMP_single	V10	AST-01-04	163/163	155/157	122/122	237/237	138/144	198/200	239/245	272/278		AST-01-04		No
V21-03	NMP_single	V21	RM1-02	167/167	160/160	122/122	235/235	146/146	198/200	243/243	268/280		RM1-02		No
V21-04	NMP_single	V21	RM1-02	167/167	160/160	122/122	235/235	144/146	200/200	243/243	268/268		RM1-02		No

Table S1B continued.

Hatchling ID	Treatment	Bucket ID	Potential parents	B008	B030	B045	B050	B064	B074	B096	B107	B135	Parent1	Parent2	Outcrossed
V21-05	NMP_single	V21	RM1-02	167/167	160/160	122/122	235/235	144/146	200/200	243/243	268/280		RM1-02		No
V21-07	NMP_single	V21	RM1-02	167/167	155/160	122/122	235/235	144/144	198/200	243/243	268/268		RM1-02		No
V21-08	NMP_single	V21	RM1-02	167/167	160/160	122/122	235/235	144/146	198/200	243/243	280/280		RM1-02		No
V01-02	NMP_mix	V01	1MDM6, VV2	157/165	153/153	122/122	226/244	138/144	198/202	239/243	270/270		VV2		No
V01-06	NMP_mix	V01	1MDM6, VV2	165/165	157/157	122/122	226/244	138/138	198/198	239/243			VV2		No
V01-07	NMP_mix	V01	1MDM6, VV2	157/165	153/157	122/122		138/144					VV2		No
V05-01	NMP_mix	V05	BN-48, VV2	165/165	153/157	122/122	226/244	138/138	198/202	239/243	270/270		VV2		No
V05-02	NMP_mix	V05	BN-48, VV2	157/165	153/157	122/122	244/244	138/138	198/202	239/239	270/270		VV2		No
V05-04	NMP_mix	V05	BN-48, VV2	165/165	157/157	122/122	226/244	138/144	198/202		270/270		VV2		No
V05-05	NMP_mix	V05	BN-48, VV2	165/165	153/153	122/122	226/244	138/138	198/202		270/270		VV2		No
V05-07	NMP_mix	V05	BN-48, VV2	165/165	153/153	122/122	226/226	144/144	198/202	239/243	270/270		VV2		No
V05-09	NMP_mix	V05	BN-48, VV2	157/165	153/157	122/122	226/244	144/144	198/202		270/270		VV2		No
V05-10	NMP_mix	V05	BN-48, VV2	165/165	153/153	122/122	226/244	138/144	198/198	243/243	270/270		VV2		No
V06-01	NMP_mix	V06	1MDM6, VV2	165/167	157/160	122/122		142/146	198/198				1MDM6		No
V09-06	NMP_mix	V09	MOS-01-02, VV2	157/165	153/157	122/122	226/244	144/144	202/202	243/243	270/270		VV2		No
V09-09	NMP_mix	V09	MOS-01-02, VV2	157/165	153/157	122/122	244/244	138/144	198/202	239/243	270/270		VV2		No
V09-10	NMP_mix	V09	MOS-01-02, VV2	157/165	153/157	122/122	226/226	138/144	198/198	239/239	270/270		VV2		No
V12-01	NMP_mix	V12	AST-01-04, BN-08, MOS-01-04, VV2	157/165	153/157	122/122	244/244	138/138	198/202	239/243	270/270		VV2		No
V12-02	NMP_mix	V12	AST-01-04, BN-08, MOS-01-04, VV2	167/167	160/160	122/122	232/232	144/144	198/198	241/241	270/270		MOS-01-04		No
V17-01	NMP_mix	V17	MOS-01-02, MOS-01-04	167/172	160/160	122/122	232/232	144/144	198/198	241/241	270/270		MOS-01-04		No
V17-02	NMP_mix	V17	MOS-01-02, MOS-01-04	167/167	160/160	122/122	232/232	144/144	198/198	241/241	270/278		MOS-01-04		No
V17-03	NMP_mix	V17	MOS-01-02, MOS-01-04	167/172	160/160	122/122	232/232	144/144	198/198	241/241	270/278		MOS-01-04		No
V17-04	NMP_mix	V17	MOS-01-02, MOS-01-04	172/172	160/160	122/122	232/232	144/144	198/198	241/241			MOS-01-04		No

Table S1B continued.

Hatchling ID	Treatment	Bucket ID	Potential parents	B008	B030	B045	B050	B064	B074	B096	B107	B135	Parent1	Parent2	Outcrossed
B20-01	NMP_mix	B20	AST-01-04, MOS-01-04, VV2	167/172	160/160	122/122	232/232					189/191	MOS-01-04		No
B23-01	NMP_mix	B23	AST-01-04, MOS-01-04, VV2	167/167	160/160	122/122	232/232					189/189	MOS-01-04		No
B23-02	NMP_mix	B23	AST-01-04, MOS-01-04, VV2	157/165	153/157	122/122	226/244					187/189	VV2		No
B23-03	NMP_mix	B23	AST-01-04, MOS-01-04, VV2	172/172	160/160	122/122	232/232					191/191	MOS-01-04		No
B26-01	NMP_mix	B26	AST-01-04, MOS-01-04, VV2	163/163	157/157	122/122	237/240					187/191	AST-01-04		No
B22-11	MP_mix	B22	RM1-18, RM1-22, RM1-35, RM1-39	167/172	155/155	122/127	232/232					187/191	RM1-18	RM1-22	Yes
B22-12	MP_mix	B22	RM1-18, RM1-22, RM1-35, RM1-39	167/169	155/155	122/122	232/232					189/191	RM1-22	RM1-39	Yes
B22-13	MP_mix	B22	RM1-18, RM1-22, RM1-35, RM1-39	167/169	155/160	122/122	232/232					189/191	RM1-22	RM1-39	Yes
B22-14	MP_mix	B22	RM1-18, RM1-22, RM1-35, RM1-39	167/172	155/155	122/122	232/237					189/191	RM1-22	RM1-35	Yes
B22-15	MP_mix	B22	RM1-18, RM1-22, RM1-35, RM1-39	167/169		122/122						189/191	RM1-22	RM1-39	Yes
B22-16	MP_mix	B22	RM1-18, RM1-22, RM1-35, RM1-39	161/172		122/127						187/191	RM1-18	RM1-22	Yes
B22-17	MP_mix	B22	RM1-18, RM1-22, RM1-35	167/167		122/127						191/191	RM1-18	RM1-22	Yes
B22-18	MP_mix	B22	RM1-18, RM1-22, RM1-35	167/172		120/127						189/191	RM1-18	RM1-35	Yes
B25-11	MP_mix	B25	RM1-18, RM1-22, RM1-35	167/169	155/155	122/122						189/191	RM1-22	RM1-39	Yes
B25-12	MP_mix	B25	RM1-18, RM1-22, RM1-35	167/169	155/155	122/127	237/242					187/189	RM1-18	RM1-39	Yes
B25-13	MP_mix	B25	RM1-18, RM1-22, RM1-35	167/169	155/155	122/122	232/232					189/191	RM1-22	RM1-39	Yes
B25-14	MP_mix	B25	RM1-18, RM1-22, RM1-35		155/160							189/189	RM1-35	RM1-39	Yes
B25-15	MP_mix	B25	RM1-18, RM1-22, RM1-35	161/167		122/122						191/191	RM1-22		No
B25-16	MP_mix	B25	RM1-18, RM1-22, RM1-35	167/169		122/122						189/191	RM1-22	RM1-39	Yes

Table S1B continued.

Hatchling ID	Treatment	Bucket ID	Potential parents	B008	B030	B045	B050	B064	B074	B096	B107	B135	Parent1	Parent2	Outcrossed
B25-17	MP_mix	B25	RM1-18, RM1-22, RM1-35	167/169		122/122						189/191	RM1-22	RM1-39	Yes
B25-18	MP_mix	B25	RM1-18, RM1-22, RM1-35			120/122						189/191	RM1-22	RM1-35	Yes
B28-11	MP_mix	B28	RM1-18, RM1-22, RM1-35	161/172	155/155	122/127	232/237					191/191	RM1-18	RM1-22	Yes
B28-12	MP_mix	B28	RM1-18, RM1-22, RM1-35	161/172	155/160	122/127	232/237						RM1-18	RM1-22	Yes
B28-13	MP_mix	B28	RM1-18, RM1-22, RM1-35	169/172	155/155	122/122	237/242						RM1-35	RM1-39	Yes
B28-14	MP_mix	B28	RM1-18, RM1-22, RM1-35	161/167	155/155	122/127	232/237					187/191	RM1-18	RM1-22	Yes
B28-17	MP_mix	B28	RM1-18, RM1-22, RM1-35	167/172		122/127						189/191	RM1-18	RM1-35 or RM1-39	Yes
B28-18	MP_mix	B28	RM1-18, RM1-22, RM1-35	167/167		122/127						191/191	RM1-18	RM1-22	Yes
B28-19	MP_mix	B28	RM1-18, RM1-22, RM1-35	161/172		122/122						189/191	RM1-22	RM1-35	Yes
B28-20	MP_mix	B28	RM1-18, RM1-22, RM1-35	167/172		122/127						191/191	RM1-18	RM1-22	Yes

Supporting Table S2: Average heterozygosity at microsatellite loci among offspring from single clone cultures and the position in of the microsatellites with respect to the centromere

Locus	Scaffold (*D. magna* assembly 2.4)	Position on scaffold (base pairs)	Linkage group (v4.0.1)	Distance from centromere (cM)	*N*[a]	Heterozygosity (confidence limita)
B008	scaffold03124	1781052	6	25.8	5	0 (0 – 0.40)
B030	scaffold00243	463057	3	32.9	27	0.59 (0.41 – 0.76)
B050	scaffold02066	483524	3	77.6	22	0.77 (0.56 – 0.90)
B064	scaffold00443	151077	1	37.9	27	0.63 (0.44 – 0.79)
B074	contig23904	1149	unmapped	NA	27	0.78 (0.59 – 0.90)
B096	scaffold01005	1506200	9	3.6	22	0.36 (0.20 – 0.57)
B107	scaffold00763	136826	1	100.0	22	0.59 (0.30 – 0.77)

[a]*N* refers to the number of genotyped offspring from heterozygous parents

Supporting Table S3: Number of loci and average heterozygosity per linkage group in automictic and self-fertilized offspring

Linkage group	*N* loci automicitc offspring	Average heterozygosity	*N* loci selfed offspring	Average heterozygosity
LG1	386	0.64	222	0.71
LG2	355	0.69	318	0.68
LG3	282	0.58	131	0.63
LG4	263	0.56	135	0.49
LG5	209	0.51	151	0.56
LG6	250	0.54	141	0.64
LG7	200	0.45	168	0.62
LG8	192	0.48	137	0.54
LG9	212	0.25	107	0.63
LG10	174	0.54	100	0.46
Average		**0.54**		**0.60**
Sum	**2523**		**1610**	

Supporting Table S4: The ten most frequently observed segregation patterns among the eight offspring of the AST-01-04 NMP clone. All other segregation patterns occurred for fewer loci or on a smaller number of linkage groups.

Linkage group	*N* loci total[a]	*N* loci (11111111)[b]	*N* loci (00000000)[b]	*N* loci (00000001)[b]	*N* loci (01000001)[b]	*N* loci (11011100)[b]	*N* loci (10011111)[b]	*N* loci (11111101)[b]	*N* loci (11111011)[b]	*N* loci (11110110)[b]	*N* loci (11110011)[b]
LG1	275	1	10	22	1		108		28	5	4
LG2	246	17	1	24	13			81		4	6
LG3	199	21	2	21	1	109					
LG4	174	2	1	37	2				19		8
LG5	122	2		12	6				1		
LG6	173	2		33	9					11	8
LG7	119			35	5						
LG8	131	4		18						23	
LG9	150		2	106	5						
LG10	104	15	1	24	6			5			
Sum	**1693**	**64**	**17**	**332**	**48**	**109**	**108**	**86**	**48**	**43**	**26**

[a]Only without missing values for any of the eight offspring

[b]Offspring are ordered in the following way: V10_03, V04_01, V04_16, V04_18, V04_22, V04_25, V04_27, V04_04, 1 = heterozygous, 0 = homozygous.

Chapter IV

GENES MIRROR GEOGRAPHY IN *Daphnia magna*

Manuscript published in Molecular Ecology as**:**
Fields, P. D., C. Reisser, M. Dukić, C. R. Haag, D. Ebert. 2015. Genes mirror geography in *Daphnia magna*. Molecular Ecology 24:4521–4536.

CONTRIBUTION: I was involved in study design, preparing RAD-sequencing libraries and data analysis

MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 4521–4536 doi: 10.1111/mec.13324

Genes mirror geography in *Daphnia magna*

PETER D. FIELDS,* CÉLINE REISSER,†‡ MARINELA DUKIĆ,* CHRISTOPH R. HAAG†‡ and DIETER EBERT*

**Zoological Institute, University of Basel, Vesalgasse 1, Basel CH-4051, Switzerland, †Centre d'Ecologie Fonctionnelle et Evolutive – UMR 5175, campus CNRS, 1919, route de Mende, 34293 Montpellier Cedex 5, France, ‡Department of Biology, Ecology and Evolution, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland*

Abstract

Identifying the presence and magnitude of population genetic structure remains a major consideration in evolutionary biology as doing so allows one to understand the demographic history of a species as well as make predictions of how the evolutionary process will proceed. Next-generation sequencing methods allow us to reconsider previous ideas and conclusions concerning the distribution of genetic variation, and what this distribution implies about a given species evolutionary history. A previous phylogeographic study of the crustacean *Daphnia magna* suggested that, despite strong genetic differentiation among populations at a local scale, the species shows only moderate genetic structure across its European range, with a spatially patchy occurrence of individual lineages. We apply RAD sequencing to a sample of *D. magna* collected across a wide swath of the species' Eurasian range and analyse the data using principle component analysis (PCA) of genetic variation and Procrustes analytical approaches, to quantify spatial genetic structure. We find remarkable consistency between the first two PCA axes and the geographic coordinates of individual sampling points, suggesting that, on a continent-wide scale, genetic differentiation is driven to a large extent by geographic distance. The observed pattern is consistent with unimpeded (i.e. no barriers, landscape or otherwise) migration at large spatial scales, despite the fragmented and patchy nature of favourable habitats at local scales. With high-resolution genetic data similar patterns may be uncovered for other species with wide geographic distributions, allowing an increased understanding of how genetic drift and selection have shaped their evolutionary history.

Keywords: *Daphnia magna*, isolation by distance, population genetic structure, principle component analysis, Procrustes analysis

Received 14 March 2015; revision received 13 July 2015; accepted 14 July 2015

Introduction

Genetic structuring of populations exists in nearly every plant or animal species (Holsinger & Weir 2009). The origins of population genetic structure are multifarious include a mixture of neutral (genetic drift) and selective (local adaptation) dynamics operating simultaneously (Whitlock & McCauley 1999) and being moderated through ecological and evolutionary forces such as gene flow, mutation and population size. Population genetic analyses aim to identify and disentangle the origins of population genetic structure and in so doing provide an understanding of a species' demographic and evolutionary history (Rosenberg & Nordborg 2002). A particularly interesting form of population genetic structure, in part due to its explanatory potential of both historical and future evolutionary processes such as the scale, frequency and spatiotemporal dynamics of gene flow, can arise when genetic distances correlate with geographic distances, called isolation by distance (= IBD; not to be confused with identity by descent). IBD may result from a long-term dynamic of distance-dependent migration/dispersal (Wright 1943) coupled with spatially restricted (as opposed to global) mating (Wright 1946). Identifying the presence and strength of IBD allows for a clearer understanding of the types of ecological and evolutionary processes that led to the current species

Correspondence: Peter D. Fields, Fax: +41(0)61 267 03 62; E-mail: peter.fields@unibas.ch

distribution (Slatkin 1985; although see Meirmans 2012 for a review of problems arising from the interpretation of IBD estimates). In some cases, accounting for IBD can help (or may even be necessary) to disentangle selective and nonselective allelic variance when attempting to identify loci associated with local adaptation (Coop *et al.* 2010; Pannell & Fields 2014).

Numerous methodologies exist for determining the occurrence of population genetic structure in a species. F_{ST} is arguably the most widely employed measure of population genetic structure (Whitlock 2011), although caveats exist concerning the exact interpretation of this summary (Jost 2008; Jakobsson *et al.* 2013). Estimation of F_{ST} is possible for even the largest genomic data sets, although the interpretation of the summary can prove problematic as both neutral and selective forces will act to generate outliers in a large distribution of values across a focal genome (Holsinger & Weir 2009; Lotterhos & Whitlock 2014). Additionally, robust estimates of F_{ST} require multiple samples per population (Willing *et al.* 2012; Reitzel *et al.* 2013; Robinson *et al.* 2014). A number of model-based summaries attempt to obviate these difficulties, and in particular those built around the *F*-model (Gaggiotti & Foll 2010) have shown great promise in both estimating population genetic structure and identifying the factors that lead to its increase/decrease (Foll & Gaggiotti 2006). These approaches are computationally intensive and make analysis of large data sets (thousands of loci and individuals) difficult or impossible to analyse (although see Novembre 2014; Raj *et al.* 2014 for a description of recent advances in estimating the underlying model of STRUCTURE in such data sets). An alternative method for detecting population genetic structure that is both more computationally attainable for even the largest data sets and requires fewer samples per population employs principle component analysis (PCA). Cavalli-Sforza and colleagues (Menozzi *et al.* 1978; Cavalli-Sforza & Feldman 2003) first applied PCA to understand population genetic structure in human populations. A number of conclusions derived from these early analyses have been recently questioned (Novembre & Stephens 2008), in particular the appropriateness of PCA for historical demography inference. Importantly, further statistical genetic advances have placed PCA on a firmer biological basis (McVean 2009) and confirmed the consistency of results obtained by PCA with those identified by other methods, such as STRUCTURE (Reich *et al.* 2008).

Combining PCA with other statistical methods was suggested as an effective method for the identification and estimation of IBD (Wang *et al.* 2012). Novembre *et al.* (2008) used PCA to identify spatial population genetic structure in a large human SNP data set collected from 3000 Europeans genotyped at more than half a million sites. Subsequently, Wang *et al.* (2012) were able to apply PCA and Procrustes rotation, wherein a matrix of coordinates is rotated in a manner to maximize the similarity to a target matrix. This method shows the correspondence between PCA-based analysis of variation in allele frequencies and the geographic origin of the DNA samples and ultimately estimates the role of spatial isolation (as opposed to other mechanisms such as local adaptation or landscape features which may promote/impede gene flow) in generating population genetic structure. Importantly, the insights provided by Novembre *et al.* (2008) and Wang *et al.* (2012) were only made possible by the very large number of molecular markers compared to previous population genetic analyses of human populations. This contrasts with earlier studies using much smaller data sets or non-nuclear-based genetic markers, which suggested little to no population genetic structure in European human populations (Cavalli-Sforza *et al.* 1994). Few other biological systems have been similarly analysed at a continental scale in order to determine to what degree population genetic differentiation recapitulates the geography of the location of sampling.

Continued and drastic reduction in the cost of DNA sequencing is allowing more systems of basic and applied biological interest to attain similar levels of genomic resources as is available for studying human population genetic structure. Additionally, the demography and ecology of human populations is quite distinct from most biological systems, and so the questions arises just how prevalent patterns of genetic diversity observed in human populations might be in species with similar large-scale ranges. *Daphnia magna* is a freshwater crustacean that has come to be recognized as a model system for eco-evolutionary and physiological research (Stollewerk 2010; Colbourne *et al.* 2011; Orsini *et al.* 2012, 2013). Multiple studies have suggested population genetic structure is strong in *D. magna* at a local level, in large part due to founder events (Whitlock & McCauley 1990; De Gelas & De Meester 2005; Haag *et al.* 2005, 2006; Vanoverbeke *et al.* 2007; Walser & Haag 2012; Orsini *et al.* 2013). A study by De Gelas & De Meester (2005) which used large-scale sampling of *D. magna* isolates on a 609-bp sequence of the cytochrome oxidase subunit I mitochondrial gene concluded that there was little signal of spatial population genetic structure in *D. magna* within Europe and a patchy distribution of individual lineages. Walser & Haag (2012) analysed European clones using microsatellites found clear population genetic structure (as measured with F_{ST}), but evidence for IBD only at the regional scale. Similarly, Haag *et al.* (2005) found IBD only within islands but not across islands in a Finnish metapopulation of *D. magna*. In addition, all previously identified patterns of IBD in *Daphnia* showed a large scatter,

which was attributed to the large degree of stochasticity involved in the colonization process (Innes 1991; Lynch & Spitze 1994; Vanoverbeke & De Meester 1997; Orsini *et al.* 2013). In this study, we attempt to resolve these differing observations by applying whole-genome RAD sequencing, identifying about 50 thousand polymorphic sites. We apply the method described by Wang *et al.* (2012), finding strong support for the congruence between PCA analysis of allelic variation and the geographic origins of the clones under study.

Methods

Clone collection

The *D. magna* genotypes (clones; *D. magna* can be maintained as stable genotypes under laboratory conditions due to their cyclical parthenogenetic life cycle) used in this study originated either from field-collected plankton samples (15 clones) were hatched from field-collected resting eggs (seven clones) or resulted from inbred crosses in the laboratory (two clones). Field-collected planktonic females were brought to the laboratory, and individual females were allowed to reproduce asexually. These isofemale lines were kept in the laboratory under conditions of continuous asexual reproduction. Resting eggs (ephippia) collected on the surface of pond sediments were washed and stimulated to hatch by exposure to continuous light under room temperature in well-oxygenated medium. Hatchlings were isolated and isofemale lines were produced and kept under conditions of continuous asexual reproduction. Two clones were obtained by selfing of field-collected females (clones produced by means of parthenogenesis lead to male offspring, which can fertilize sexual eggs of their clonal sisters). These two selfed clones are the parents of a standing *D. magna* QTL panel (Routtu *et al.* 2010; Roulin *et al.* 2013; Routtu & Ebert 2014). One clone (from Southern Germany; DE-Iinb1) is the result of one round of selfing, and the other clone (from Finland, FI-Xinb3) resulted from three rounds of selfing. The Finnish clone had also been used for the *D. magna* reference genome (V 2.4; Daphnia Genome Consortium).

RAD library preparation and sequencing

We used a restriction site-associated DNA approach (RAD; Baird *et al.* 2008) to obtain genetic markers, following the protocol developed in Etter *et al.* 2011, with modifications. The 24 individuals used in this study were part of a larger project consisting of three libraries of 30 individuals. Individuals were treated for 72 h with three antibiotics (streptomycin, tetracycline, ampicillin) at a concentration of 50 mg/L for each antibiotic and were fed with dextran beads (Sephadex 'Small' by Sigma Aldrich: 50 μm diameter) at a concentration of 0.5 g/100 mL. This treatment was used to remove contaminant DNA (i.e. bacterial DNA or algal DNA from the gut). The DNA was extracted with a Qiagen Blood and Tissue kit following manufacturer's instructions and digested with PstI (New England Biolabs). Digested DNA was barcoded with genotype-specific P1 adapters and pooled to create a library containing 2100 ng DNA. The pooled library was sheared on a Bioruptor using 2 times 3 cycles (1 cycle 30 s ON, 1 min OFF), and fragments between 300 and 500 bp were selected through agarose gel electrophoresis. DNA fragments were blunted and prepared for P2 adapter ligation. The library was amplified through PCR (30 s at 98 °C, followed by 18 cycles of 10 s at 98 °C, 30 s at 65 °C and 30 s 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 600 bp. Single-end 100 cycle sequencing was performed by the Quantitative Genomics Facility service platform at the Department of Biosystem Science and Engineering (D-BSSE, ETH), in Basel, Switzerland, on an Illumina HiSeq 2000. All demultiplexed read data used for genotyping were submitted to NCBI SRA: BioProject ID PRJNA288911.

Quality control, demultiplexing, mapping and SNP identification

Library quality and per-base quality was controlled with FastQC (Patel & Jain 2012), and reads were checked for barcode integrity, absence of adapter sequences within the reads and integrity of PstI cut site. The reads were sorted individually by barcode and filtered to remove reads with uncalled bases and an overall quality score under 24. The last six bases of each read were trimmed to prevent false positives due to a slight decrease in base quality. Reads were subsequently aligned to the *D. magna* genome (V2.4; Daphnia Genomic Consortium, wFleaBase) using BWA v.0.7.10 (Li 2013), and SAM files were converted to BAM files, sorted and indexed using SAMtools v.0.1.19 (Li *et al.* 2009). To identify SNP polymorphisms, we applied GATK v.3.1 (McKenna *et al.* 2010) base quality score recalibration, indel realignment, performed SNP and INDEL discovery and genotyped all 24 samples simultaneously using standard hard filtering parameters or variant quality score recalibration according to GATK Best Practices recommendations (DePristo *et al.* 2011; Van der Auwera *et al.* 2013). Scripts for running GATK were adapted from De Wit *et al.* (2012), and following Peterson *et al.* (2012), only SNP polymorphisms with parameters QD (quality by depth) ≥6 and GQ (genotype quality) ≥20 were retained for downstream analyses.

PCA and total population structure

We performed principle component analysis using SMARTPCA v.5.1 (Patterson *et al.* 2006) to estimate principle components of SNP allelic variation across the 24 clones. The GATK generated VCF file was converted to SmartPCA format using the vcf2smartpca.py script from De Wit *et al.* (2012). PCA analysis was conducted only for those loci for which all individuals could be genotyped to avoid potential bias associated with RAD sequencing approaches (Arnold *et al.* 2013). To obtain an estimate of the total population structure, we estimated the observed (H_I) and expected (H_T) heterozygosity of using VCFtools (Danecek *et al.* 2011) function –hardy and then estimated total population structure as $F_{IT} = 1 - \frac{H_I}{H_T}$ across the nuclear genome. We estimated F_{IT} rather than F_{ST}, because the sample sizes per population were too small for meaningful estimates of expected within-population heterozygosity H_I. Nonetheless, assuming no strong, systematic deviations from Hardy–Weinberg equilibrium within populations (the two inbred individuals were not considered in this analysis), our estimate of F_{IT} allows us to still estimate a composite value of within- and among-population divergence (Fields *et al.* 2014). In addition, 1-F_{IT} estimates the proportion of the total genetic variation that is present within rather than among individuals, and hence, F_{IT} corresponds to the maximum proportion of the total variation that can potentially be explained by geography (as this necessarily involves only the among-individual component of genetic variation).

Procrustes rotation

We used Procrustes analysis to compare the geographic coordinates of clone origins to the first two components (PC1 and PC2) of the PCA performed on the genetic data (Cox & Cox 2000; Wang *et al.* 2010, 2012). Procrustes analysis minimizes the sum of squared Euclidean distances between two sets of points by transforming, or 'rotating', one set of points to match the other, while preserving the relative distances among all points within the map (Wang *et al.* 2012). The similarity of the two maps is quantified using the Procrustes similarity statistic $t_0 = \sqrt{1 - D}$, where D is the minimum sum of the squared Euclidean distances between the two maps, scaled to range between 0 and 1 (Wang *et al.* (2010, 2012). We used the geographic coordinates of the sampling locations for individual clones as the fixed matrix and Procrustes-transformed PCA coordinates to superimpose the PCA maps on the geographic maps. We used the 'procrustes' function from the vegan package (Oksanen *et al.* 2013) in the statistical software R v. 3.1.1 (R Development Core Team 2014) to conduct the Procrustes rotation and estimate t_0. The rotation matrix estimated by vegan was used to calculate the rotation angle θ of the PCA map via the Procrustes analysis in degrees counterclockwise. We used the 'protest' function, which applies the method of Peres-Neto & Jackson (2001), in the vegan package to test the statistical significance of t_0 using 100 000 permutations. As in Wang *et al.* (2012), the resulting P-value tests $\Pr(t > t_0)$, specifically the probability of the observing a similarity statistic higher than t_0 under the null hypothesis that no geographic pattern exists in the population structure. The Procrustes analysis was carried out for the entire data set and for two subsets, excluding either of the two most distant localities, the clone from Mongolia (MN) and the clone from Israel (IL).

Correlations between pairwise relatedness and distance

We estimated pairwise relatedness via kinship coefficients using the VCFtools (Danecek *et al.* 2011) function –relatedness, which implements the kinship coefficient estimate of Yang *et al.* (2010). Under the model of Yang *et al.* (2010), a random individual sampled from a single large population should have a relatedness value of 0 and 1 for an individual with themselves, respectively. As deviations from panmixia within a sample arise, the relatedness value of Yang *et al.* (2010) should take on values <0, wherein more negative values suggest greater deviations from panmixia and more distant relationships. To determine whether a pattern of IBD exists among the sampled clones, we used a Mantel test to compare the matrix of geographic distance and pairwise relatedness. We used the R package ecodist (Goslee & Urban 2007) and tested the specific null hypothesis that the estimated Mantel r is significantly <0 (one-sided, significantly negative) using 100 000 permutation tests.

Marker number sensitivity analysis

To examine how robust the presented analysis of correspondence between genes and geography is with respect to variation in data set size (number of loci), we generated random subsets of our SNP data set, from 50 to 10 000 loci. Estimates of PCA1 and PCA2, as well as Procrustes similarity scores, were obtained for each subset (50, 100, 250, 500–10 000 loci, with increments of 500 loci) using SmartPCA and the vegan function 'procrustes', respectively. The relationship between pairwise distance and relatedness was also assessed for these subsets using the R package ecodist and testing the null hypothesis that the estimated Mantel r is significantly <0 using 100 000 permutation tests.

Results

Sequencing results

We obtained 135 878 437 reads from the 24 individuals sequenced, but 1 603 591 did not show a full restriction site and 3 709 500 were of low quality (quality score <24). On average, 89.5% bases had a quality of >30. The mean quality was 34.9 and 74.6% of the reads passed the Illumina filtering, with individual library coverage ranging between 68 and 137X. Variant identification via GATK resulted in a total of 52 682 polymorphic loci segregating among the 24 *D. magna* clones that could be called across all clones (D1). Removal of the MN (D2) and then MN and IL (D3) clones reduced this number to 45 374 and 43 463 SNPs, respectively.

PCA, total population structure and procrustes analysis

The individual loadings of on the first two axes of the PCA for each of the data sets and summarization of variance explained by the PCA can be found in Table 2 and Table 3, respectively. Percentage of the total variation explained by the first two axes of the PCA (i.e., the eigenvalues of PCA1 and PCA2, Table 3) for the full sample (D1) was 4.324 and 1.553, respectively (with loadings of the individual samples). The PCA plot indicated that the Mongolian (MN) clone is a distinct outlier from the other samples (Fig. 1A–B; Table 1). Excluding the MN clone (sample D2) resulted in 1.832; 1.708 eigenvalues for PCA1; PCA2 and their respective eigenvectors (Tables 2 and 3). In the new partitioning, the IL clone was now a single outlier (Fig. 1C–D). We therefore estimated PCA1; PCA2 on a second reduced data set (D3), excluding the MN and IL clones, and obtained eigenvalues 1.888; 1.576 and their respective eigenvectors (Tables 2 and 3). Plots of PCA1; PCA2 of D3 (Fig. 1E–F) were largely consistent with D2. The PCA plots of each data set exhibited a remarkable similarity to the geographic coordinates of individual clone sampling locations, with a large degree of correspondence of PCA1 of the genetic data with the geographic east–west axis and of PCA2 with the north–south axis, especially in the D2 and D3 data sets.

Table 1 Sampling locations and collection type of individual clones. ID is an abbreviated name used for each population (the two letters indicated the country of origin), Clone ID is the formal identification used by the Ebert laboratory for individual clones (including country code, population acronym and clone number), and life stage collected refers to the life stage at which the clone was collected (i.e. adult individuals, resting eggs or if the clone was manipulated within the Ebert laboratory)

ID	Clone ID	Country	Latitude	Longitude	Life stage collected
BA	BY-G-9	Belarus	52.421464	31.013781	Field-collected plankton
BL	BE-KN2-1	Belgium	51.355731	3.334453	Field-collected plankton
CH	CH-H-876	Switzerland	47.557563	8.861583	Field-collected plankton
CZ	CZ-N1-1	Czech Republic	48.775317	16.723528	Field-collected plankton
DE	DE-Iinb1	Germany	48.206375	11.709727	Selfed clone (laboratory produced)*
FI	FI-Xinb3	Finland	59.833183	23.260387	Selfed clone (laboratory produced)†
FR	FR-C1-1	France	43.591583	4.591517	Field-collected plankton
GB	GB-EL75-69	United Kingdom	51.527556	-0.158147	Field-collected plankton
GR	GR-K-1	Greece	40.703997	23.144222	Field-collected plankton
HU	HU-HO-2	Hungary	46.800000	19.133333	Field-collected plankton
IL	IL-M1-8	Israel	31.714561	35.05099	Hatched from ephippium
IR	IR-GG1-1	Iran	37.918978	46.707003	Hatched from ephippium
MN	MN-DM-1	Mongolia	45.032708	100.660481	Field-collected plankton
RU-1	RU-AST-1	Russia	45.903611	47.656389	Hatched from ephippium
RU-2	RU-BN-2	Russia	50.088135	43.292921	Hatched from ephippium
RU-3	RU-BN-6	Russia	50.088135	43.292921	Hatched from ephippium
RU-4	RU-RM1-1	Russia	55.763514	37.581667	Field-collected plankton
RU-5	RU-RM1-2	Russia	55.763514	37.581667	Field-collected plankton
RU-6	RU-SPB-09	Russia	59.811111	30.133056	Field-collected plankton
RU-7	RU-SPB-35	Russia	59.811111	30.133056	Field-collected plankton
RU-8	RU-VOL-39	Russia	48.530000	44.486944	Hatched from ephippium
RU-9	RU-VOL-56	Russia	48.530000	44.486944	Hatched from ephippium
RU-10	RU-VOL-2	Russia	48.530000	44.486944	Field-collected plankton
SE	SE-G4-20	Sweden	60.253040	18.306090	Field-collected plankton

*Field collected and then once selfed.
†Field collected and then three times selfed.

Table 2 SmartPCA estimated principle component loadings for the first two eigenvectors, PC1 and PC2, for three different data sets. $PC1_{D1}$ and $PC2_{D1}$ describe the full set of clones, $PC1_{D2}$ and $PC2_{D2}$ derive from the data set excluding the MN clone, and $PC1_{D3}$ and $PC2_{D3}$ derive from the data set excluding the MN and IL clones

Population ID	$PC1_{D1}$	$PC2_{D1}$	$PC1_{D2}$	$PC2_{D2}$	$PC1_{D3}$	$PC2_{D3}$
BA	−0.0469	−0.0086	−0.0114	0.0527	−0.0007	0.0705
BL	−0.0446	−0.3814	−0.3814	−0.0496	−0.3788	−0.0787
CH	−0.0496	−0.2447	−0.2427	−0.0108	−0.2379	−0.0258
CZ	−0.0461	−0.1044	−0.1055	0.0257	−0.0975	0.0155
GE	−0.0486	−0.1337	−0.1347	0.0246	−0.126	0.0197
FI	−0.0486	−0.1337	−0.0715	0.174	−0.0428	0.495
FR	−0.0401	−0.3976	−0.3988	−0.1754	−0.419	−0.4326
GB	−0.047	−0.3289	−0.3259	−0.0286	−0.3213	−0.0635
GR	−0.035	−0.0224	−0.0258	−0.0756	−0.0373	−0.1702
HU	−0.0445	−0.0766	−0.0775	−0.013	−0.0774	−0.0359
IL	−0.0419	0.2455	0.2143	−0.894	–	–
IR	−0.0348	0.2539	0.2547	−0.1004	0.2471	−0.4527
MN	0.9786	−0.0162	–	–	–	–
RU-1	−0.0356	0.1715	0.1735	0.047	0.1829	−0.0347
RU-2	−0.0381	0.1855	0.1864	0.0781	0.2003	−0.0235
RU-3	−0.0423	0.1805	0.1793	0.0868	0.1946	−0.0005
RU-4	−0.0448	0.122	0.1197	0.0978	0.1366	0.0558
RU-5	−0.045	0.1207	0.1183	0.0967	0.1347	0.0599
RU-6	−0.0457	−0.1041	−0.1059	0.0651	−0.0929	0.1553
RU-7	−0.0466	−0.1023	−0.1038	0.0676	−0.09	0.154
RU-8	−0.0351	0.2544	0.2623	0.1065	0.2793	−0.0701
RU-9	−0.0309	0.2759	0.2874	0.1283	0.3085	−0.0735
RU-10	−0.0358	0.2621	0.2703	0.1258	0.2905	−0.0598
SE	−0.0506	−0.0806	−0.0813	0.1707	−0.0529	0.496

Table 3 Summary of PCA partitioning and Procrustes analysis for three different data sets. θ (° counterclockwise) is the rotation angle for the PCA map that optimizes the Procrustes similarity with the geographic map. *P*-values derive from 100 000 permutations of population labels

Dataset	Variance explained by PC1 (%)	Variance explained by PC2 (%)	Rotation Angle θ (°)	Procrustes Similarity t_0	*P*-value
All Samples	4.324	1.553	−52.48	0.6748	$<10^{-5}$
MN Excluded	1.832	1.708	99.34	0.8389	$<10^{-5}$
MN, IL Excluded	1.888	1.576	89.96	0.8489	$<10^{-5}$

Estimates of mean whole-genome H_I for D1, D2 and D3 were 0.108, 0.124 and 0.131, respectively. Estimates of mean whole-genome H_T for D1, D2 and D3 were 0.158, 0.172 and 0.178, respectively. Total population genetic structure as estimated by F_{IT} was 0.267, 0.226 and 0.212, for D1, D2, and D3.

We applied Procrustes analyses to quantify how well the qualitative similarities observed in the PCA plots matched the map of geographic origins of the *D. magna* clones. The estimated Procrustes similarity (t_0) between D1 sampling locations and the estimated PCA1; PCA2 coordinates was 0.6748, with a counterclockwise rotation angle (θ) of −52.48. Between D2 sampling locations and the estimated PCA1; PCA2 coordinates, the estimates were 0.8389, with a θ of 99.34. Finally, the correlation between D3 sampling locations and the estimated PCA1; PCA2 coordinates was 0.8489, with a θ of 89.96 (Table 3; Fig. 2). The angle of approximately 90° indicates that PCA1 roughly corresponds to longitude and PCA2 to latitude (rather than the other way around). We tested the significance of t_0 ($t_0 > t$) using 100 000 permutations with the *protest* function. For D1-3, $t_0 > t$ with a *P*-value $<10^{-5}$ (Table 3; Fig. 3). Importantly, removal of the IL clone did not lead to qualitatively distinct conclusions. Removal of clones deriving from the same sampling location did not lead to qualitatively different results (data not shown), as well.

To quantify the relationship between pairwise relatedness and pairwise distance of sampling location for each clone, we applied Mantel tests. The Mantel *R* provides an estimate of correlation between pairs of matrices, which was estimated as −0.77, −0.646 and −0.645, respectively, for data sets D1-3. We used 100 000

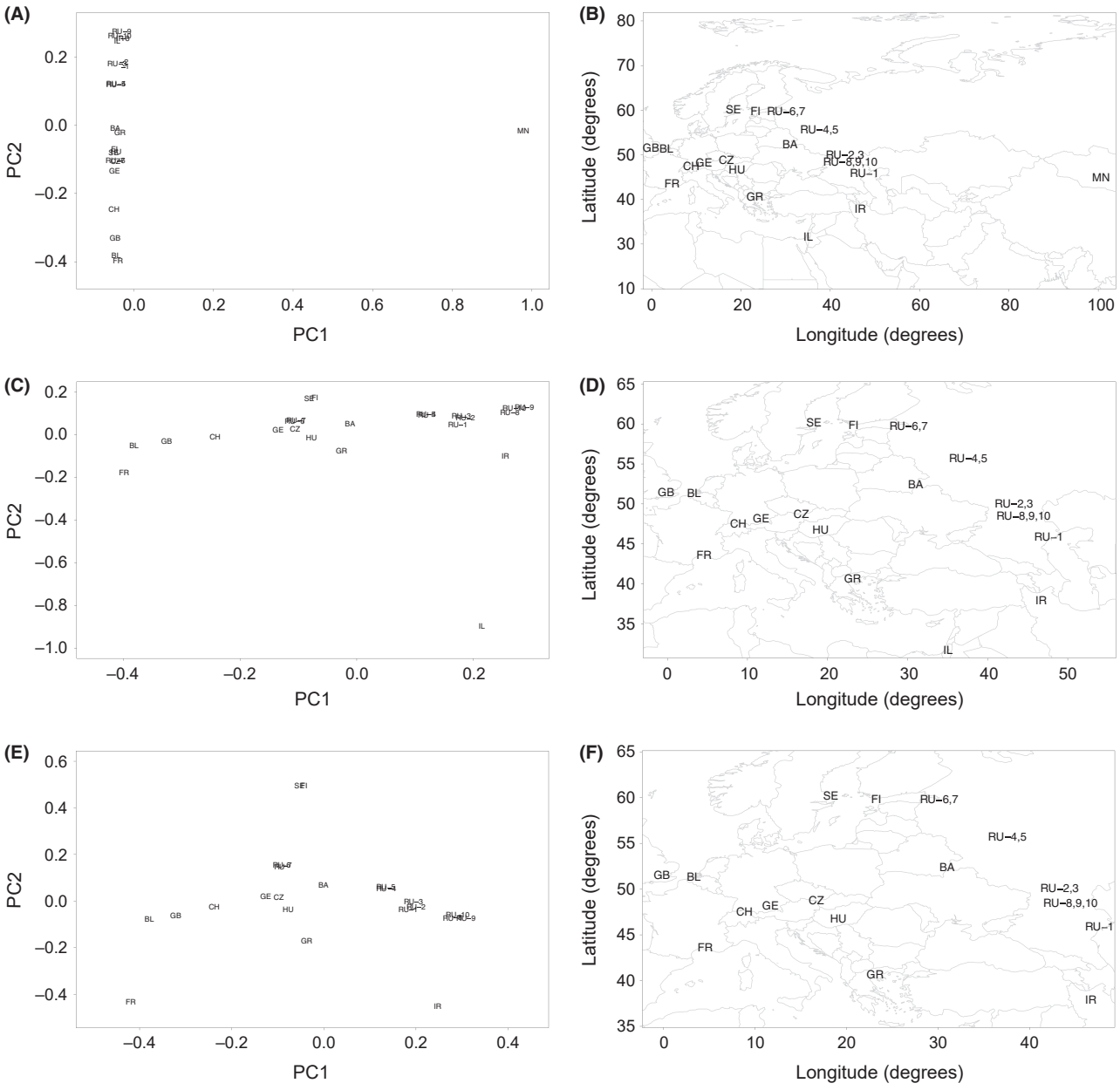


Fig. 1 Plot of principal component axis one (PC1) and axis two (PC2) and sampling locations for SNP polymorphism deriving from *Daphnia magna* clones sampled across its range. Individual points and their sampling locations are labelled with their individual clone names as described in Table 1. (A) and (B) PCA plot when including all genotyped clones and their sampling locations, (C) and (D) excluding the MN clone, and (E) and (F) excluding the MN and IL clones, respectively.

permutation tests to determine if R is significantly <0. For each data sets, D1-3, the resultant P-value was $<10^{-4}$, with 95% confidence intervals for R of [−0.811, −0.641], [−0.686, −0.615] and [−0.687, −0.611], respectively.

We explored the relationship of pairwise relatedness and pairwise distance for three subsets of focal clones to consider how specific outliers affected the observed IBD pattern. When the MN clone was used as a focal individual and pairwise distance and relatedness was estimated between it and all other clones, we found a mean correlation of −0.350 (P-value = 0.101; R^2 = 0.081); when the MN clone was excluded and the IL clone was used as the focal individual, we found a mean correlation of −0.793 (P-value $<10^{-4}$; R^2 = 0.610); and when the MN and IL clone were excluded and the IR clone was used as the focal individual, we found a mean correlation of −0.930 (P-value $<10^{-4}$; R^2 = 0.860) (Fig. 4). Thus, by removing outlier individuals from the pairwise

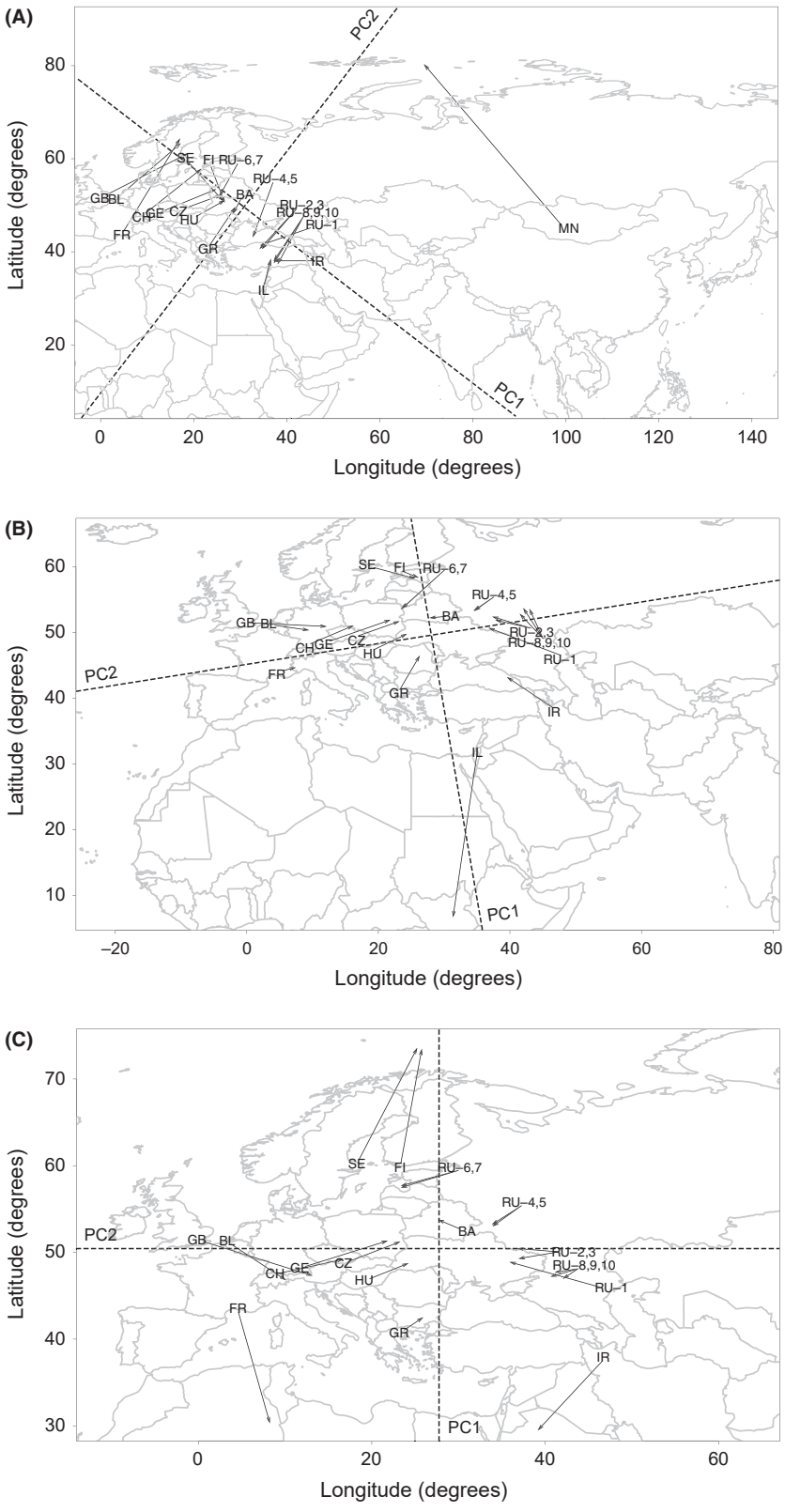


Fig. 2 Procrustes analysis of genetic and geographic coordinates *Daphnia magna* clones collected across the species' range in Eurasia. Clone sampling locations are indicated by the population symbol described in Table 1, and the arrows point to the Procrustes-transformed PCA (PC1 and PC2) values for (A) all collected clones ($t_0 = 0.6748$; $P < 10^{-5}$; $\theta = -52.48$), (B) the MN clone excluded ($t_0 = 0.8389$; $P < 10^{-5}$; $\theta = 99.34$) and (C) the MN and IL clones excluded ($t_0 = 0.8489$; $P < 10^{-5}$; $\theta = 89.96$).

comparison in the same manner as was performed in the Procrustes analysis, we found a pattern consistent with the IBD moving away from the Middle East.

We determined the robustness of our inference of gene geography correspondence and the observed pattern of IBD to inference based on a smaller number of

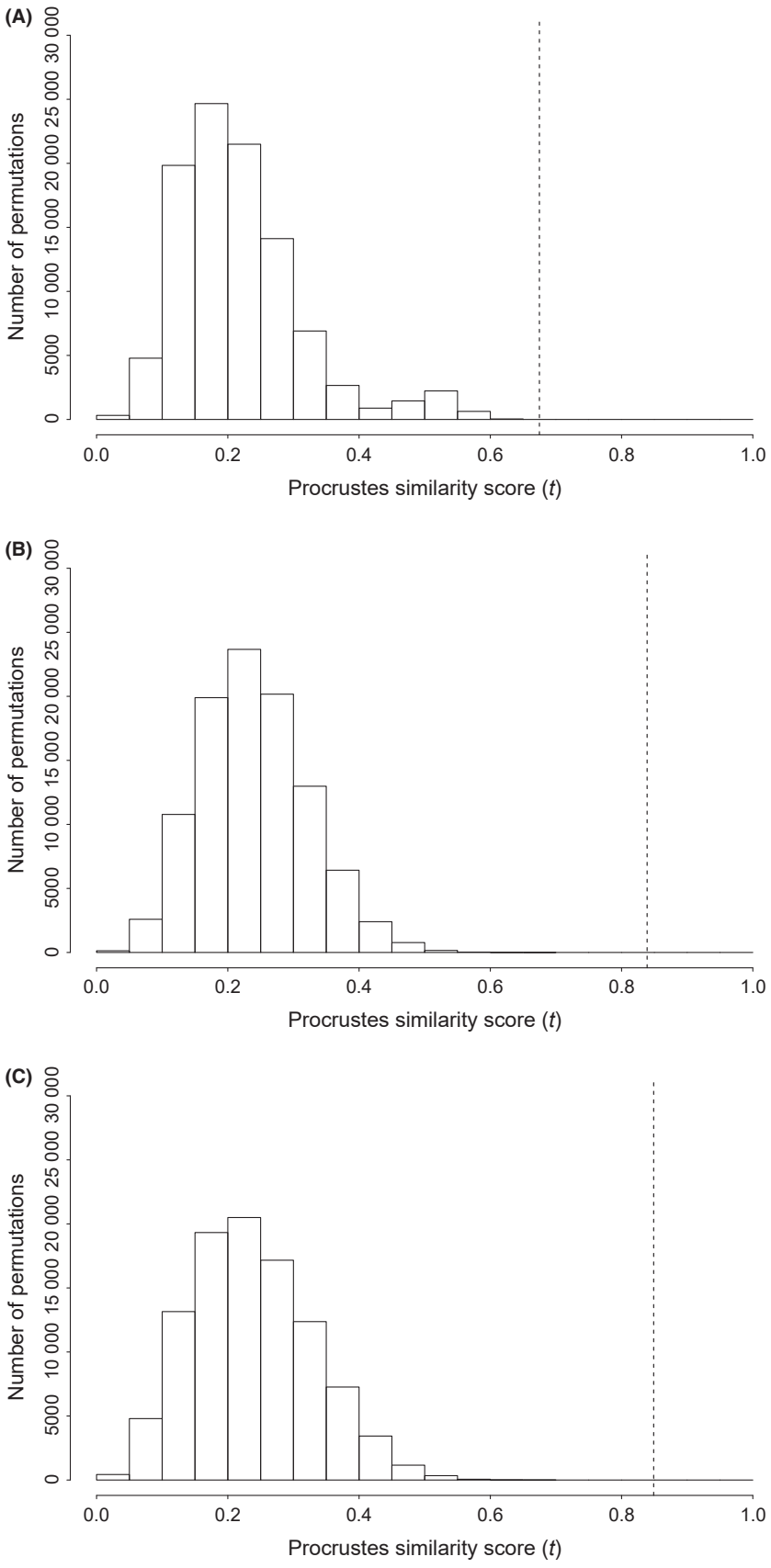


Fig. 3 Histograms of the Procrustes similarity t of 100 000 permutations in Fig. 2A,B,C. The black dotted vertical line indicates the observed Procrustes similarity score, t_0. (A) The full data set in Fig. 2A ($t_0 = 0.6748$, $P < 10^{-5}$), (B) exclusion of the MN clone ($t_0 = 0.8389$, $P < 10^{-5}$) and C) exclusion of the MN and IL clones ($t_0 = 0.8489$, $P < 10^{-5}$).

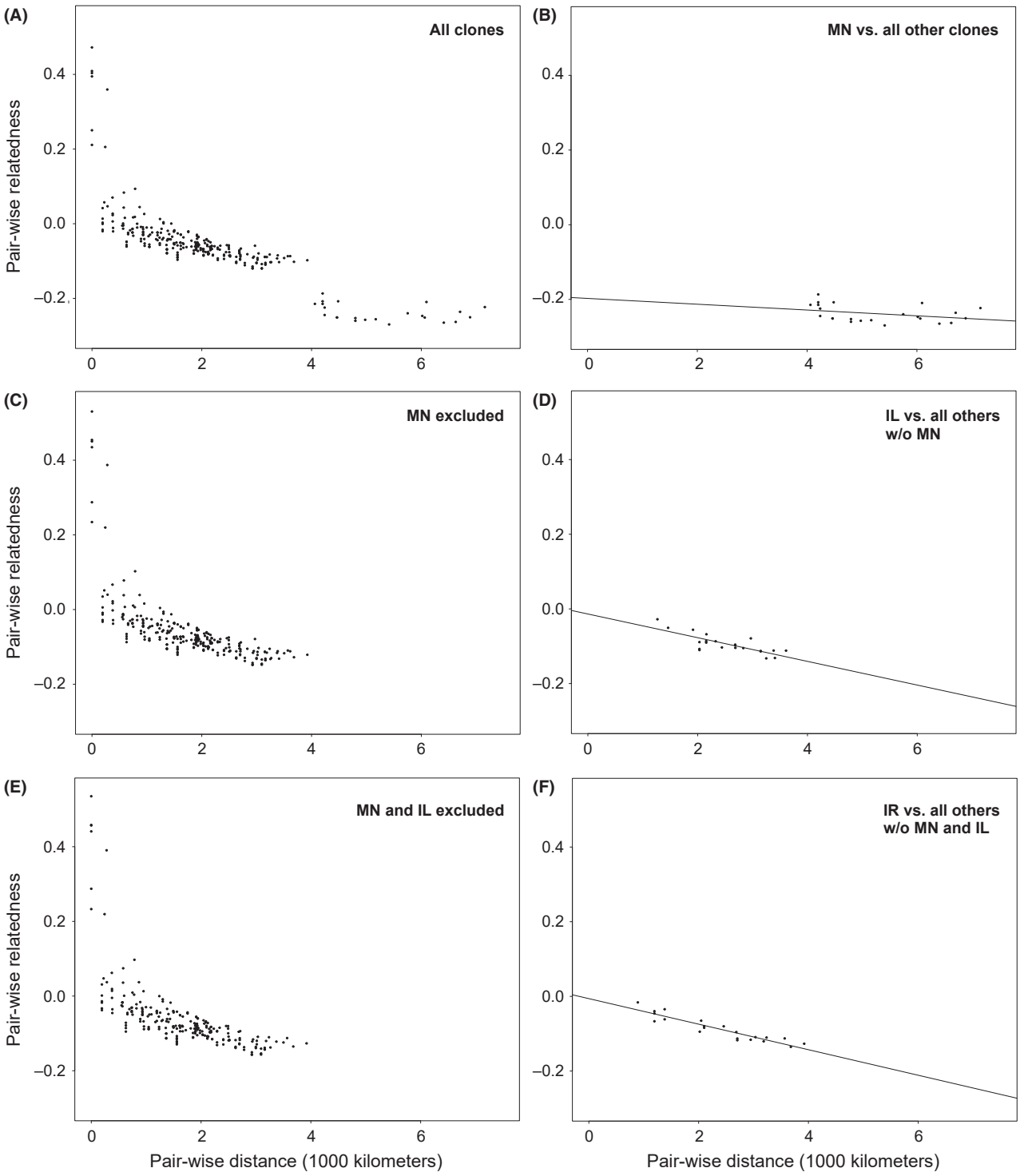


Fig. 4 Plot of the relationship between pairwise distance and pairwise whole-genome relatedness. (A) The relationship between pairwise distance and whole-genome relatedness with all clones included within the estimate ($P < 10^{-4}$; Mantel $r = -0.77$), (B) pairwise distance and whole-genome relatedness between the MN clone and all other sampled clones ($P = 0.101$; $R^2 = 0.081$), (C) pairwise distance and whole-genome relatedness excluding the MN clone ($P < 10^{-4}$; Mantel $r = -0.65$), D) pairwise distance and whole-genome relatedness between the IL clone and all other clones excluding the MN clone ($P < 10^{-4}$; $R^2 = 0.61$), (E) pairwise distance and whole-genome relatedness with the MN and IL clone excluded ($P < 10^{-4}$; Mantel $r = -0.65$), and (F) pairwise distance and whole-genome relatedness between the IR clone and all other clones excluding the MN and IL clone ($P < 10^{-8}$; $R^2 = 0.86$) .

markers. There was a noticeable decline in the t_0 value when <100 randomly selected loci from each data set were included in the PCA analysis and Procrustes, with a minimum $t_0 = 0.50$, 0.51, and 0.67, respectively, for D1, D2 and D3. Adding additional loci resulted in an asymptote of t_0 values at around 1500 loci (Fig. 5A), suggesting results become robust at much larger sample sizes than have previously been derived for elucidating population genetic structure in *D. magna*. Our analysis of overall IBD as determined by Mantel test remained significant (*P*-value $<10^{-4}$) at all loci sampling sizes. Finally, we explored the effect loci sample size on the relationship of pairwise distance and pairwise relatedness for the three outlier clones by assessing the R^2 value of the linear relationship of the two variables. The correlation of pairwise distance and relatedness remained insignificant (*P*-value > 0.05) for all subsamples of SNP loci when the MN clone acted as the focal clone. The relationship of pairwise distance and relatedness became significant at sample sizes ≥1500 (*P*-value <0.05) when the IL and IR clone acted as focal individuals, while the R^2 value showed a steady increase from minimum values of 0.0004, 0.046, when the IL and IR clones were used as focal individuals, respectively (Fig. 5B), although no asymptote was observed at 10 000 loci, again suggesting the utility of increasing the level of sampling of loci across the focal species' genome. The Mantel tests remain significant (*P*-value $<10^{-4}$) at all loci sample sizes for both D2 and D3.

Discussion

In the present analysis, we show a remarkable consistency between a PCA partitioning of genetic variance into its first two principal coordinates and the geographic sampling coordinates in the *D. magna* system (Fig. 1). Previous theoretical studies have suggested a set of conditions, such as nonimpeded migration/gene flow over short to intermediate distances, are needed to generate such a pattern (McVean 2009). Distinct empirical examples showing a similar consistency between PCA partitioning of genetic variance and geographic sampling coordinates in humans across the globe (Novembre *et al.* 2008; McVean 2009; Wang *et al.* 2012). As was also noted by Wang *et al.* (2012) for their analysis of human data sets, the total variance explained by PC1 and PC2 in our *D. magna* data set is generally low (below 6%; Table 3). The present sampling scheme did not allow for a full hierarchical decomposition of genetic variance into individual and among-population components. We estimated total population genetic structure with F_{IT}, which is a composite measure of both within- and among-population divergence. F_{IT} varied slightly, with the full data set showing the highest level of total population structure, 0.267, consistent with larger divergence from panmixia when the MN clone was included. F_{IT} for D2 and D3 was slightly lower, 0.226 and 0.212, respectively. Although we lack data on multiple individuals for most populations, and so were unable to estimate the within-population variance, the variation within individuals provides a lower-bound estimate. The data show that 75% of the variance occurs already within individuals. Hence, the relative small amount of the overall variance explained by PC1 and PC2 has to be roughly multiplied by 4 to obtain an estimate of what proportion of the among-population variance was explained by PC1 and PC2. The overall estimate of F_{IT} is approximately 0.25, suggesting that approximately 0.06 of 25% or approximately 0.24 of 100% of the among-individual variation is explained by the two major PC dimensions. Significantly, the Procrustes analysis of PC1 and PC2 and the geographic sampling locations is a way how one can derive a quantitative estimate of the degree to which geographic isolation leads to genetic

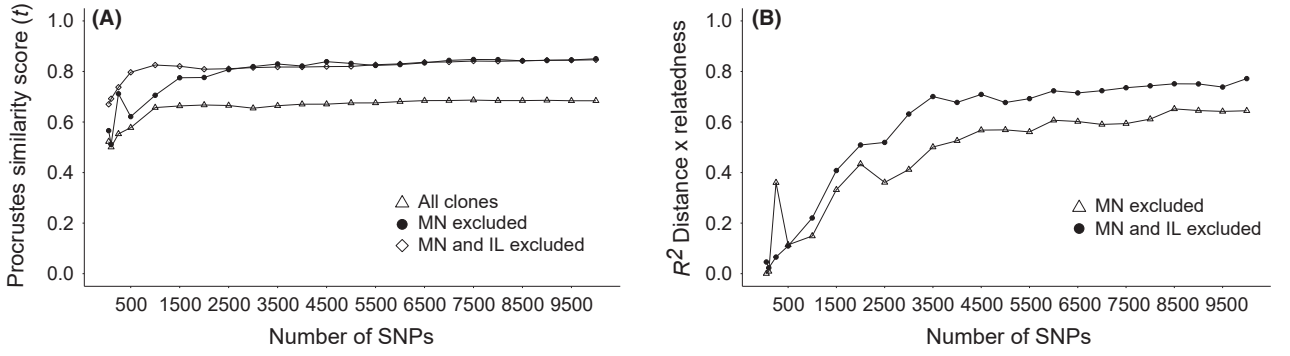


Fig. 5 Procrustes analysis and the relationship of pairwise distance/relatedness using increasing numbers of loci. (A) A random subset of *L* randomly selected markers were used to generate Procrustes similarity scores and (B) estimate the relationship (R^2) between pairwise distance and pairwise relatedness using the IL and IR clones as focal individuals. The dataset which included the MN clone as a focal individual was excluded from the plot as the relationship between geographic distance and relatedness remained insignificant with the range of *L*. *L* = 50, 100, 250, 500, 1000, 1500 . . . 10 000.

differentiation, or conversely the similarity between genes and geography which decreases with distance in *D. magna* (Wang *et al.* 2012). Our approach allows us to suggest that *D. magna* has indeed, contrary to earlier studies, a population genetic structure within Europe consistent with IBD (Fig. 2). Our estimate of F_{IT} may also suggest a role for within-population genetic variation, as approximated by variation in individual heterozygosity, in decreasing the quantity of variance explained by the PC1 and PC2 of genetic variance (Novembre *et al.* 2008; Wang *et al.* 2012).

To substantiate our interpretation of IBD, we also compared the relationship of pairwise genomic relatedness and distance between sampling locations of individual *D. magna* clones. These correlations were always significantly negative, a result consistent with the expectation of IBD where gene flow and migration becomes more limited as distance increases (Fig. 4). Because most of our sampled populations are composed of only a single clone, we cannot eliminate the possibility that an individual clone may be the result of a recent and long-distance migration event. However, under the utilized framework, identifying any other stochastic process that would have led to the conclusion that the distribution of geographic variation at a continent-wide scale is strongly related to geographic distance is unlikely, that is the pattern of gene geography correspondence should have dissipated rather than strengthened. Additionally, in cases where two individuals were sampled from the same population, similar profiles in the PCA plot were observed, suggesting that using just one of these individuals would have well represented those populations and that results were both quantitatively and qualitatively similar for the Procrustes analysis and tests of IBD (Figs 1 and 2). Two of the utilized clones, FI-Xinb3 and DE-Iinb1, derive from three and one generations of selfing, respectively. Inbreeding in a natural population will increase the total homozygosity in a population although not necessarily change overall allele frequencies (Hartl & Clark 1997). Under the present framework for determining IBD based upon whole-genome allele frequencies, additional inbreeding may increase the variance in the relationship between geographic distance and genetic divergence, but the overall expectation of the relationship will remain.

By analysing the present data set using both the full sample and removing both the most distantly sampled genotypes, from Mongolia (MN) and Israel (IL), we can infer a number of additional insights. The MN clone is clearly the most divergent sample. The MN clone is geographically the most distant from all other clones, but also the Himalayan mountain range might contribute as a barrier influencing gene flow and migration, but this cannot be investigated here with so few samples in the region. A clear linear pattern of IBD is seen when either the IR or IL clones act as focal individuals in correlating pairwise distance and relatedness (Fig. 4).

An earlier study by De Gelas & De Meester (2005) suggested there may be little population genetic structure in European *D. magna*. A potential explanation for the discrepancy between their observations and our results is the increase in the total power to detect genetic differentiation as our study encompasses tens of thousands of SNPs, while the previous study relied on a single gene, and furthermore, we used nuclear markers as opposed to a mitochondrial gene (cytochrome oxidase subunit 1, or COI). Importantly, population genetic inference deriving from the organelle or nuclear genome may not always be concordant (Jezkova *et al.* 2015). The studies by Novembre *et al.* (2008) and Wang *et al.* (2012) which displayed gene geography correspondence similarly relied on large-scale NGS data. Many earlier phylogeographical analyses relying on mitochondrial gene sequences alone have recently been called into question (Galtier *et al.* 2009). By conducting a sensitivity analysis wherein we subsampled the total number of loci, we show that signals of IBD become much more robust as the number of loci greatly exceeds that previously applied to phylogeographic analysis of *D. magna* (Fig. 5).

Short-distance dispersal of *D. magna* happens via abiotic factors such as water and wind, and biotic factors such as aquatic and terrestrial invertebrates, humans and birds (van de Meutter *et al.* 2008). The exact mechanism by which long-distance gene flow and migration takes place in *D. magna* is unknown, although one of the most likely mechanisms is via anthropomorphic interactions and those with avian species and fresh water invertebrates (Brown 1929; Proctor 1964; Proctor & Malone 1965; Mellors 1975; Figuerola & Green 2002; Figuerola *et al.* 2003, 2005; van de Meutter *et al.* 2008). If bird-assisted dispersal is the main long-term dispersal mechanism for *D. magna*, geographic barriers, such as seas or mountain ranges, are less likely to act as dispersal barrier, although, some bird species follow stereotypic dispersal routes (e.g. along coastlines). Previous work on a well-characterized coastal metapopulation of *D. magna* in Finland has suggested that most colonizations derive from approximately 1.7 individuals, although dispersal capacity overall was high (Haag *et al.* 2005). A recent analysis by Walser & Haag (2012) used microsatellite markers and six nuclear gene sequences to compare patterns of IBD in the northern populations of *D. magna* such as those studied in Haag *et al.* (2005), which are known to experience more population turnover and increased genetic drift (in particular dynamics associated with local extinction and recolonization; see Ebert *et al.* 2013), and central European samples which are believed to experience less turnover

[see De Meester *et al.* 2002; the monopolization hypothesis (MH)]. MH, wherein a clone's arrival precedence and population growth rate limits effective local establishment, considers dynamics of local adaptation and population connectivity on a scale that does not allow for direct interrogation with the present data set. Orsini *et al.* (2013), focusing on a much more local scale of 19 Belgium ponds, find genomewide signatures of local adaptation due to biotic and abiotic factors at the focal scale and a lack of IBD, which they consider as evidence for the MH.

The geographic origin of the *D. magna* species is not yet known, although the closely related species, *D. similis* and *D. lumholtzi*, co-occurs with *D. magna* within a much smaller range (Middle East and Africa), so limited evidence exists that these regions may in fact be the origin of the species (pers. comm. Adam Petrusek; De Gelas & De Meester 2005; Popova & Kotov 2013). A potential reason that the IL clone is such a distinct outlier in the Procrustes analysis (Fig. 2B) and also a slightly less good fit to the IBD pattern could be its location within the city of Jerusalem, which may act as a barrier to gene flow and migration from more natural populations within the region following a much earlier colonization event. Samples from the surrounding and coastal regions of Israel may allow for better assessment of this hypothesis. Additional insights will be available by reconstructing the demographic history of the species, and in so doing, we might (i) determine the origin of the species, which may include more than a single refugium (Hewitt 1996, 1999) and (ii) ascertain the number of expansion events that may have led to the current distribution of allelic diversity, neither of which can be determined with the application of PCA-based analysis (Novembre & Stephens 2008).

Conclusion

Many previous studies have investigated genetic differentiation in *Daphnia*, although mostly at a local or regional scale. Invariantly, these studies found high genetic differentiation among populations and some found evidence for IBD, while others not. The studies that did find evidence for IBD also found a strong scatter in the relationship between genetic differentiation and geographic distance, which is in contrast to our findings. This scatter has been interpreted as evidence for the strong stochasticity involved in the colonization process, which is a main determinant of the genetic population structure in *Daphnia*. In addition, subsequent gene flow between populations is augmented by hybrid vigour in some parts of the distribution range, but not in others, and the amount of local gene flow may also be influenced by local adaptation to specific environmental features of the ponds. Our study differs in two important respects from these earlier studies, the use of a large number of marker loci and a continent-wide sample. The results clearly indicate that, at a larger scale, most of the stochastic effects that occur at a local scale disappear and genetic similarity between populations becomes almost strictly dependent on geographic distance. In the present study, we describe a distinct pattern of population genetic differentiation, wherein partitioning of genetic variation at ten thousands of SNPs into its first two principle components followed by Procrustes rotation results in a remarkable similarity to the two dimensional grid of geographic sampling locations. To the authors' knowledge, this distinct pattern has only previously been documented in human populations (Novembre *et al.* 2008; Wang *et al.* 2012). While the level of gene flow and migration is unquestionably larger in humans, the pattern that might arise between PCAs and sampling locations, suggested by McVean (2009), may still be a reasonable expectation for species that migrate via mechanisms that are less impeded by particular landscape features. Additionally, the inferred rate of migration or gene flow may at times be much greater than that expected from counting individual migrants due to the role of hybrid vigour (Ingvarsson & Whitlock 2000; Whitlock *et al.* 2000; Haag & Ebert 2007) in *D. magna*, generating effective migration rates much higher than would be expected otherwise (Ebert *et al.* 2001). Providing better estimates of the type and magnitude of population genetic structure will allow for future studies of *D. magna*, a model system for both applied and basic research, to better disentangle signals of historical demography and natural selection.

Acknowledgements

We would like to thank Frida Ben-Ami, Luc De Meester, Andrei Papkou, Jason Andras, Sandra Lass, Jarkko Routtu, Loukas Theodosiou, Alexey Kotov and Yan Galimov for providing materials used in the present study. We would also like to thank Andrea Berardi, Kimberly Gilbert and the Ebert and Haag laboratories for helpful comments and conversations.

References

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Brown LA (1929) The natural history of cladocerans in relation to temperature-III. Preadaptation and dispersal. *The American Naturalist*, **63**, 443–454.

Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nature Genetics*, **33**, 266–275.

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes*. Princeton University Press, Princeton, New Jersey.

Colbourne JK, Pfrender ME, Gilbert D *et al.* (2011) The ecoresponsive genome of daphnia pulex. *Science*, **331**, 555–561.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.

Cox TF, Cox MAA (2000) *Multidimensional Scaling*, 2 edn. Chapman & Hall, Boca Raton, Florida.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

De Gelas K, De Meester L (2005) Phylogeography of Daphnia magna in Europe. *Molecular Ecology*, **14**, 753–764.

De Meester L, Gómez A, Okamura B, Schwenk K (2002) The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica*, **23**, 121–135.

De Wit P, Pespeni MH, Ladner JT *et al.* (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, **12**, 1058–1067.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

R Development Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ebert D, Hottinger J, Pajunen V (2001) Temporal and spatial dynamics of parasite richness in a Daphnia metapopulation. *Ecology*, **82**, 3417–3434.

Ebert D, Hottinger JW, Pajunen VI (2013) Unsuitable habitat patches lead to severe underestimation of dynamics and gene flow in a zooplankton metapopulation. *Journal of Animal Ecology*, **82**, 759–769.

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011) Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE*, **6**, e18561.

Fields PD, McCauley DE, McAssey EV, Taylor DR (2014) Patterns of cyto-nuclear linkage disequilibrium in Silene latifolia: genomic heterogeneity and temporal stability. *Heredity*, **112**, 99–104.

Figuerola J, Green AJ (2002) Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater Biology*, **47**, 483–494.

Figuerola J, Green AJ, Santamaría L (2003) Passive internal transport of aquatic organism.

Figuerola J, Green AJ, Michot TC (2005) Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. *The American Naturalist*, **165**, 274–280.

Foll M, Gaggiotti OE (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.

Gaggiotti OE, Foll M (2010) Quantifying population structure using the F-model. *Molecular Ecology Resources*, **10**, 821–830.

Galtier N, Nabholz B, Glemin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Haag CR, Ebert D (2007) Genotypic selection in Daphnia populations consisting of inbred sibships. *Journal of Evolutionary Biology*, **20**, 881–891.

Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D (2005) Genetic diversity and genetic differentiation in Daphnia metapopulations with subpopulations of known age. *Genetics*, **170**, 1809–1820.

Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D (2006) Founder events as determinants of within-island and among-island genetic structure of Daphnia metapopulations. *Heredity*, **96**, 150–158.

Hartl DL, Clark AG (1997) *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics*, **10**, 639–650.

Ingvarsson PK, Whitlock MC (2000) Heterosis increases the effective migration rate. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267**, 1321–1326.

Innes DJ (1991) Geographic patterns of genetic differentiation among sexual populations of *Daphnia pulex*. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **69**, 995–1003.

Jakobsson M, Edge MD, Rosenberg NA (2013) The Relationship Between FST and the Frequency of the Most Frequent Allele. *Genetics*, **193**, 515–528.

Jezkova T, Riddle BR, Card DC *et al.* (2015) Genetic consequences of postglacial range expansion in two codistributed rodents (genus Dipodomys) depend on ecology and genetic locus. *Molecular Ecology*, **24**, 83–97.

Jost LOU (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Lynch M, Spitze K (1994) Evolutionary genetics of *Daphnia*. In: *Ecological Genetics* (ed. Real L), pp. 86–108. Princeton Univ. Press, Princeton, New Jersey.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

McVean G (2009) A genealogical interpretation of principal components analysis. *PLoS Genetics*, **5**, e1000686.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.

Mellors WK (1975) Selective predation of ephippal *Daphnia* and the resistance of ephippal eggs to digestion. *Ecology*, **56**, 974–980.

Menozzi P, Piazza A, Cavallisforza L (1978) Synthetic maps of human gene-frequencies in Europeans. *Science*, **201**, 786–792.

van de Meutter F, Stoks R, de Meester L (2008) Size-selective dispersal of *Daphnia* resting eggs by backswimmers (*Notonecta maculata*). *Biology Letters*, **4**, 494–496.

Novembre J (2014) Variations on a common STRUCTURE: new algorithms for a valuable model. *Genetics*, **197**, 809–811.

Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, **40**, 646–649.

Novembre J, Johnson T, Bryc K *et al.* (2008) Genes mirror geography within Europe. *Nature*, **456**, 98–101.

Oksanen J, Blanchet FG, Kindt R *et al.* (2013) vegan: Community Ecology Package.

Orsini L, Spanier KI, De Meester L (2012) Genomic signature of natural and anthropogenic stress in wild populations of the waterflea Daphnia magna: validation in space, time and experimental evolution. *Molecular Ecology*, **21**, 2160–2175.

Orsini L, Mergeay J, Vanoverbeke J, De Meester L (2013) The role of selection in driving landscape genomic structure of the waterflea Daphnia magna. *Molecular Ecology*, **22**, 583–601.

Pannell JR, Fields PD (2014) Evolution in subdivided plant populations: concepts, recent advances and future directions. *New Phytologist*, **201**, 417–432.

Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, **7**, e30619.

Patterson N, Price AL, Reich D (2006) Population structure and Eigenanalysis. *PLoS Genetics*, **2**, e190.

Peres-Neto PR, Jackson DA (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, **129**, 169–178.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, **7**, e37135.

Popova EY, Kotov AA (2013) Latitudinal patterns in the diversity of two subgenera of the genus Daphnia O.F Müller (Crustacea: Cladocera: Daphniidae). *Zootaxa*, **3736**, 16.

Proctor VW (1964) Viability of crustacean eggs recovered from ducks. *Ecology*, **45**, 656–658.

Proctor VW, Malone CR (1965) Further evidence of the passive dispersal of small aquatic organisms via the intestinal tract of birds. *Ecology*, **46**, 728–729.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.

Reich D, Price AL, Patterson N (2008) Principal component analysis of genetic data. *Nature Genetics*, **40**, 491–492.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, **22**, 2953–2970.

Robinson JD, Bunnefeld L, Hearn J, Stone GN, Hickerson MJ (2014) ABC inference of multi-population divergence with admixture from unphased population genomic data. *Molecular Ecology*, **23**, 4458–4471.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380–390.

Roulin AC, Routtu J, Hall MD *et al.* (2013) Local adaptation of sex induction in a facultative sexual crustacean: insights from QTL mapping and natural populations of Daphnia magna. *Molecular Ecology*, **22**, 3567–3579.

Routtu J, Ebert D (2014) Genetic architecture of resistance in Daphnia hosts against two species of host-specific parasites. *Heredity*, **114**, 241–248.

Routtu J, Jansen B, Colson I, De Meester L, Ebert D (2010) The first-generation Daphnia magna linkage map. *BMC Genomics*, **11**, 508.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics*, **16**, 393–430.

Stollewerk A (2010) The water flea Daphnia–a 'new' model system for ecology and evolution? *Journal of Biology*, **9**, 21.

Van der Auwera GA, Carneiro MO, Hartl C *et al.* (2013) From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, **2013**, 11.10.1–11.10.33.

Vanoverbeke J, De Meester L (1997) Among-populational genetic differentiation in the cyclical parthenogen Daphnia magna (Crustacea, Anomopoda) and its relation to geographic distance and clonal diversity. *Hydrobiologia*, **360**, 135–142.

Vanoverbeke J, De Gelas K, De Meester L (2007) Habitat size and the genetic structure of a cyclical parthenogen, Daphnia magna. *Heredity*, **98**, 419–426.

Walser B, Haag CR (2012) Strong intraspecific variation in genetic diversity and genetic differentiation in Daphnia magna: the effects of population turnover and population size. *Molecular Ecology*, **21**, 851–861.

Wang C, Szpiech ZA, Degnan JH *et al.* (2010) Comparing spatial maps of human population-genetic variation using procrustes analysis. *Statistical Applications in Genetics and Molecular Biology*, **9**, 13.

Wang C, Zöllner S, Rosenberg NA (2012) A Quantitative Comparison of the Similarity between Genes and Geography in Worldwide Human Populations. *PLoS Genetics*, **8**, e1002886.

Whitlock M (2011) G 'st and D do not replace Fst. *Molecular Ecology*, **20**, 1083–1091.

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST} not equal $1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

Whitlock MC, Ingvarsson PK, Hatfield T (2000) Local drift load and the heterosis of interconnected populations. *Heredity*, **84**, 452–457.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, **7**, e42649.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.

Yang J, Benyamin B, McEvoy BP *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, **42**, 565–569.

P.F., C.R., M.D., C.H. and D.E. conceived, designed, and collected the data for the present analysis. P.F., C.R., M.D., C.H. and D.E. analyzed the data. P.F., C.R., C.H. and D.E. wrote the paper.

Data accessibility

All analysis scripts as well as raw and processed data are available on Dryad (doi:10.5061/dryad.bv3gm) and the NCBI SRA database (BioProject ID PRJNA288911), respectively. These data include the following:

1. 24 RADseq-based Illumina short-read files.
2. One VCF file.
3. SmartPCA input files.
4. R scripts describing Procrustes rotation and permutation tests.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The key contribution of my PhD work was the enrichment of genomic resources available for *Daphnia magna*. *D. magna* is one of the most studied species of *Daphnia* genus and it is increasingly recognized as a model organism for environmental (ecological) genomics and physiological research. I have investigated the genetic consequences of sexual and asexual reproduction in *D. magna*, both of which are important for understanding many aspects of the genome evolution and adaptive success of this crustacean, but also for the design and interpretation of future genomic studies on this valuable model organism.

Implementing a Restriction site associated DNA (RAD) sequencing protocol has enabled me to construct a high-density genetic map for *D. magna* (Chapter I). Due to the high number of markers that were incorporated in this map, it became possible to unite an ascertained order of markers with the available physical genome draft of *D. magna* at an unprecedented level of consistency. Thus, the map presented in the Chapter I of my thesis can provide a basis for the on-going improvements of the assembly of the genome that should become available to a wide community of *Daphnia* researchers in the near future. However, the major achievement of this first part of my thesis is the characterization of the genome-wide recombination landscape in *D. magna*. This is important for two reasons: (i) describing regions within chromosomes where recombination is rare or absent indicated approximate positions of centromeres, which were not previously known, and (ii) the knowledge of the frequency and distribution of meiotic recombination along chromosomes is essential for understanding how selection interacts with genetic linkage in modelling the pattern of genomic variation within and between species (Cutter and Payseur 2013). The map presented in Chapter I shows variation in recombination rate on a broad-scale (measured in Mb) allowing us to differentiate between recombining and non-recombining regions. However, a higher resolution map might be obtained by including more F2 individuals in the linkage analysis. This, in parallel with the increasing quality of genome assembly, would enable the analysis of recombination rates at a finer scale and the identification of recombination hotspots (if existing) in the recombining regions of each chromosome.

Following the improvements in genome annotations of *D. magna* it will also be possible to explore if the variation in recombination rate correlates with genome organization. Is there a correlation between gene density and recombination rate? What is the correlation between gene expression and recombination rates, and can we see any evidence of codon bias usage in *Daphnia*? How are essential genes distributed across the genome of *D. magna* and are they clustering in regions of low recombination as it was shown for baker's yeast, *Candida albicans* and *Caenorhabditis elegans* (Pál and Hurst 2003)? Is there a correlation between transposable elements and the variation in recombination rate? Answering these questions is crucial for understanding the interplay of recombination, mutation and selection in shaping genome evolution in *Daphnia* and furthering our knowledge of the mechanisms responsible for the adaptive success of the genus and its cosmopolitan distribution.

Furthermore, *Daphnia*'s ability to produce resting eggs that accumulate in lake sediments and may remain viable for several decades (Limburg and Weider 2002; Jankowski and Straile 2003) offers a new exciting venue for studying the evolution of recombination rates in response to different environmental pressures. Since ancient *Daphnia* populations can be resurrected, the recombinational histories of different genomic regions can be compared, and

temporal differences in recombination rates inferred. Moreover, it would be interesting to see if those differences correlate with variation in biotic and abiotic stressors documented for a given habitat.

While RAD sequencing proved to be very useful for the robust genotyping of an experimental cross (Chapter I), the assessment of a genome-wide heterozygosity patterns (Chapter III) and even detecting geography-driven genetic differentiation of *Daphnia* populations (Chapter IV), it gave inconclusive results when used for the quest of a rare phenomenon, such as ameiotic recombination (Chapter II). Following the reports of recombination during asexual reproduction that resulted in long genomic tracts where ancestral heterozygosity was lost (loss of heterozygosity - LOH) in *D. pulex* and *D. obtusa*, I have used RAD-sequencing to inspect if the same phenomena could be detected in *D. magna*. RAD-sequencing was my method of choice for several reasons. First, it provided a mean for much denser genome sampling than it would be the case even with a large number of microsatellites. Second, it allowed me to asses LOH in more individuals than it would be possible by using whole-genome sequencing at the same cost. And third was the high sequencing depth that could be achieved with a reduced representation genome sequencing approach, such as RAD-sequencing, at much lower cost when compared to whole-genome sequencing. Having a dense distribution of markers across the genome was important for estimating the size of a genomic portion that would become homozygous after a bout of asexual reproduction if there would be any cross-over (CO) recombination. This is a crucial parameter of the Loss of Complementation (LOC) hypothesis proposing the evolutionary advantage of sexual reproduction (Archetti 2004b, 2010), which I aimed to evaluate during my PhD work. The same could have been achieved with a whole-genome sequencing approach, but my reasoning behind choosing the reduced representation genome sequencing on many individuals was following: (i) if recombination during asexual reproduction was rare and occurring only in a small number of asexual offspring, I would have had better chances to detect it by analysing larger number of individuals and (ii) having high sequencing depth (what is very expensive to achieve with whole-genome sequencing) should have enable me to differentiate between homozygosity and hemizygosity (deletion where only one copy of the allele remains in the genome) which was also reported to give a signal of LOH in *D. pulex* (Xu et al. 2011).

One of the aims of Chapter II of my thesis was to experimentally estimate the parametric values of the LOC hypothesis (Archetti 2004a,b, 2010) which argues that the immediate advantage of sex lies in the preservation of complementation (masking of recessive deleterious mutations) which is lost due to genome homogenization (i.e., LOH) by CO recombination during asexual reproduction.

Although I cannot exclude the possibility that LOH indeed occurs in genomic regions that were not covered through RAD-sequencing, the inability to detect any LOH events with a high number of markers (on average 4357 per individual) in forty asexual daughters originating from four lines of *D. magna*, suggests that LOH is less frequent than it was previously assumed (Omilian et al. 2006; Xu et al. 2011) or highly localized in small genomic regions as it was recently showed in *D. pulex* using whole-genome sequencing approach (Keith et al. 2015). Even though my results indicate that the portion of the genome

becoming homozygous in a single generation of asexual reproduction was largely overestimated in a model proposed by Archetti (Archetti 2004b), one could argue that the logics of his hypothesis is still supported by the absence of evidence for CO recombination during parthenogenesis of *Daphnia*.

For example, karyological studies showed that the mechanism of parthenogenesis in *D. pulex* is meiosis with the suppression of the first division occurring at an early anaphase I (Hiruta et al. 2010; Hiruta and Tochinai 2012). Such a mechanism opens the possibility for CO recombination to occur during bivalent formation (pairing of homologous chromosomes) in prophase I. However, the most detailed study of genetic consequences of parthenogenesis in *D. pulex,* using mutation accumulation lines and whole-genome sequencing, reported tracts of LOH that rarely exceeded 200 bp, what is indicative of meiotic gene-conversions (Keith et al. 2015). Long tracts of LOH that would be caused by COs were not detected. One of the possible explanations for this finding, in the light of LOC hypothesis, is the selection against unmasking of recessive deleterious alleles: only individuals with minimum levels of LOH would be able to survive and propagate in mutation accumulation lines as COs would induce high mortality due to LOC.

An alternative explanation would be that the transition to asexuality has to associated with the evolution of mechanisms to suppress COs during parthenogenesis, as it was proposed by Schurko et al. (2009). Based on the similar expression patterns of meiosis specific genes during sexual (meiotic) and asexual (ameiotic) reproduction in *D. pulex*, the authors concluded that parthenogenetic oogenesis in *Daphnia* is indeed orchestrated by a meiotic machinery. Interestingly though, the majority of genes involved in the process of homologous recombination are present in a single copy within the *D. pulex* genome while seven homologs of RECQ2 (BLM) helicase were found (Schurko et al. 2009). This enzyme is involved in directing recombination towards an early formation of non-COs and represents the main barrier for the CO resolution (Wu and Hickson 2003; Bernstein et al. 2010). Similarly, in the *D. magna* genome, eight homologs of BLM helicase are present. This peculiarity prompts the assumption that even though parthenogenetic oogenesis may follow the meiotic processes, the formation of CO structures is probably channelled to alternative mechanisms for DSB repair in order to reduce negative consequences of LOH. As it was reported in Chapter II of my thesis, evidence of ameiotic COs could not be detected even when the influence of selection is minimized, thus making me more in favour of this second explanation for the absence of large scale LOH during asexual reproduction in *Daphnia*.

Chapter III of my thesis provides an illustration of the flexibility of reproductive strategies in *Daphnia*. More precisely, in the absence of males, *D. magna* can engage meiotic asexual reproduction (automixis) as it was revealed by the assessment of genome-wide heterozygosity using RAD-sequencing. This was the very first demonstration of automictic reproduction in *Daphnia* and an exemplar study of how reduced genome sequencing methodology can serve for unrevealing cryptic breeding systems. Due to its ability to reproduce both sexually and asexually, *Daphnia* has already been used for studying the evolutionary benefits of sex. However, the discovery of automixis in *Daphnia* adds another layer to these investigations since automixis has been hypothesised to be an intermediate step in a transition from meiotic to ameiotic reproduction (Schwander et al. 2010).

The final chapter of my thesis additionally proves the power of RAD-sequencing in studies aiming to understanding the roles of evolutionary processes that influence variation across genomes and populations. RAD-sequencing of European populations of *D. magna* permitted a high resolution analysis for linking genetic variation with the geographic location of individual samples, which was not possible in previous studies using a small number of markers. This study provided a better insight into the population genetic structure of *D. magna* and suggested that genetic differentiation is driven to a large extent by geographic distance. Furthermore, improved estimates of type and magnitude of genetic structure will set a foundation for future studies to disentangle signals of historical demography and selection in *D. magna*.

In conclusion, I believe that the implementation of RAD-sequencing is a valuable addition to an array of genomic techniques that can be used for the investigations of many aspects of genome biology in *D. magna*. I hope that the results presented in my thesis will contribute to the further establishment of *Daphnia* as a model system for linking ecology and genome evolution, and finally shading more light on the genetic basis of phenotypic diversity and adaptive capacity of this astonishing critter.

REFERENCES:

Archetti, M. 2010. Complementation, genetic conflict, and the evolution of sex and recombination. J. Hered. 101 Suppl :S21–33.

Archetti, M. 2004a. Loss of complementation and the logic of two-step meiosis. J. Evol. Biol. 17:1098–105.

Archetti, M. 2004b. Recombination and loss of complementation: a more than two-fold cost for parthenogenesis. J. Evol. Biol. 17:1084–97.

Bernstein, K. A., S. Gangloff, and R. Rothstein. 2010. The RecQ DNA helicases in DNA repair. Annu. Rev. Genet. 44:393–417.

Cutter, A. D., and B. A. Payseur. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14:262–74.

Hiruta, C., C. Nishida, and S. Tochinai. 2010. Abortive meiosis in the oogenesis of parthenogenetic *Daphnia pulex*. Chromosome Res. 18:833–40.

Hiruta, C., and S. Tochinai. 2012. How Does the Alteration of Meiosis Evolve to Parthenogenesis? - Case Study in a Water Flea , *Daphnia pulex* - Meiosis - Mol. Mech. Cytogenet. Divers. 109–122.

Jankowski, T., and D. Straile. 2003. A comparison of egg-bank and long-term plankton dynamics of two *Daphnia* species, *D. hyalina* and *D. galeata*: Potentials and limits of reconstruction. Limnol. Oceanogr. 48:1948–1955.

Keith, N., A. E. Tucker, C. E. Jackson, W. Sung, J. I. Lucas, D. R. Schrider, and A. J. Younge. 2015. High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. Genome Res. 26:60–69.

Limburg, P. A., and L. J. Weider. 2002. "Ancient" DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. Proc. Biol. Sci. 269:281–287.

Omilian, A. R., M. E. A. Cristescu, J. L. Dudycha, and M. Lynch. 2006. Ameiotic recombination in asexual lineages of *Daphnia*. Proc. Natl. Acad. Sci. U. S. A. 103:18638–43.

Pál, C., and L. D. Hurst. 2003. Evidence for co-evolution of gene order and recombination rate. Nat. Genet. 33:392–395.

Schurko, A. M., J. M. Logsdon, and B. D. Eads. 2009. Meiosis genes in *Daphnia pulex* and the role of parthenogenesis in genome evolution. BMC Evol. Biol. 9:78.

Schwander, T., J. Dubman, B. J. Crespi, S. Vuilleumier, J. Dubman, and B. J. Crespi. 2010. Positive feedback in the transition from sexual reproduction to parthenogenesis. Proc. Biol. Sci. 277:1435–1442.

Wu, L., and I. D. Hickson. 2003. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. Nature 426:870–874.

Xu, S., A. R. Omilian, and M. E. Cristescu. 2011. High rate of large-scale hemizygous deletions in asexually propagating *Daphnia*: implications for the evolution of sex. Mol. Biol. Evol. 28:335–42.

Acknowledgements

First and foremost, I am deeply grateful to Prof. Dieter Ebert and Dr. Christoph Haag for believing in me and giving me an opportunity to be a part of their research groups. This turned to be a life changing experience for me in many ways. I am thankful for their guidance, patience, support and even the challenges they have placed in front of me. Their enthusiasm, creativity and critical thinking will serve as an exemplar model in all of my future endeavours.

A special gratitude goes to all of my collaborators and funding agencies (Swiss National Science Foundation and Nikolaus und Bertha Burckhardt-Bürgin-stiftung) that have made this work possible.

I am extremely grateful to Daniel Berner and Marius Roesti for their unconditional help in all of the aspects of my research. They were the teachers one could only wish for, and I cannot express how much I have benefited from our collaboration. Both scientifically and personally, I will always be in debt to them for all the knowledge and the support I have received.

I would like to thank to Peter Damian Fields, for many fruitful discussions, endless readings of my writing, encouragements and new insights into scientific reality that he has provided me.

I am immensely grateful to Walter Salzburger for being a friend, a pusher, a role model and often a life-saver.

I am thankful to all the members of the Ebert, Haag, Schärer and Salzburger lab, both past and present, for making my long working hours enjoyable, providing help in the times of panic, memorable moments in Swiss Alps, relaxing lunch breaks, cultural exchanges and making me feel welcomed in every sense of the word.

Being a part of the Doctoral Program for Population Genomics has enabled me to meet and exchange ideas with many fellow students through our joint workshops and I would like to express my special gratitude to Ute Friedrich for making it all so easy and perfectly organized.

I would also like to thank to Lukas Zimmerman, Yasmin Picton, Jürgen Hottinger, Urs Stiefel, Brigitte Aeschbach and other members of the administrative support at the University of Basel, that have made my research feasible.

I need to thank to my friends and family, my heroes and dreamers, since without them all of this would become sort of irrelevant. I am especially grateful to Roberto Arbore. The list of reasons would exceed the volume of this thesis and I cannot imagine how my time in Basel would have been without him. Lucas Marie-Orleach and Nikolas Vellnow made our office a rewarding place to be, and I will always be thankful for having them around in both cheerful and challenging times of my PhD. The amazing Zoological Institute also brought Christian Felber to my life, an alien fellow who has motivated and distracted me when it was the most needed.

I want to thank Marija Podolski, Lidija Kobak and Jelena Rajkov for their amazing personality, the level of care and shelter they have provided me and all the past and future adventures, nail-polishing sessions or scientific discussions that will bring me another step closer in figuring out the meaning of life.

In addition, I am extremely grateful to Janko Gospočić, Duje Burić, Sasha Mushegian, Robert Bakarić, Tom Robbins, Jason Andras, Tim Janicke, Kate Ballantine, Karen Haag, Kurt Vonnegut, Louis Du Pasquier, Melanija and Marko Marijanović, Dimitri von Arx, Pete the Dog, Anne Roulin, Lauryn Hill, Andrea Kaufmann, Dario Moser, Anđela Jukić, Maja Batić, Pepijn Luijcx, Nicolas Boileau, Gillberto Bento, Clara von Basel, Catia Fonseca, Elham Sheikh-Jabbari, Marco Colombo, KRS One, Kendrick Lamar, Kemistry and Storm, Hugo Grante, Maria Popova, Ivana Sušilović, Irena Delonga, Damir Omerbašć, Sassja, Dita Vizoso, Lukas Schärer, Fabrizia Ronco, Attila Rüegg, Ivo and Ivana Lovrić and many more who have shared with me the pleasures of my PhD life.

Lastly, but maybe most importantly, I want to express my deepest gratitude to "majka" Davorka, "ćako" Jakov, Tino and Martina for setting a foundation of everything I am and about to become.