

# Transition from Environmental to Partial Genetic Sex Determination in *Daphnia* through the Evolution of a Female-Determining Incipient W Chromosome

Céline M.O. Reisser,<sup>\*,1,2,3</sup> Dominique Fasel,<sup>2</sup> Evelin Hürlimann,<sup>2</sup> Marinela Dukič,<sup>4</sup> Cathy Haag-Liautard,<sup>2</sup> Virginie Thuillier,<sup>2</sup> Yan Galimov,<sup>5</sup> and Christoph R. Haag<sup>1,2</sup>

<sup>1</sup>Centre d'Ecologie Fonctionnelle et Evolutive CEFE UMR 5175, CNRS Université de Montpellier Université Paul-Valéry Montpellier EPHE, Montpellier, France

<sup>2</sup>Université de Fribourg, Ecology and Evolution, Fribourg, Switzerland

<sup>3</sup>IFREMER Centre du Pacifique, Taravao, Tahiti, Polynésie Française

<sup>4</sup>Universität Basel, Zoology Institute, Evolutionary Biology, Basel, Switzerland

<sup>5</sup>Koltsov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia

\*Corresponding author: E-mail: Celine.Reisser@ifremer.fr.

Associate editor: John Parsch

## Abstract

Sex chromosomes can evolve during the evolution of genetic sex determination (GSD) from environmental sex determination (ESD). Despite theoretical attention, early mechanisms involved in the transition from ESD to GSD have yet to be studied in nature. No mixed ESD-GSD animal species have been reported, except for some species of *Daphnia*, small freshwater crustaceans in which sex is usually determined solely by the environment, but in which a dominant female sex-determining locus is present in some populations. This locus follows Mendelian single-locus inheritance, but has otherwise not been characterized genetically. We now show that the sex-determining genomic region maps to the same low-recombining peri-centromeric region of linkage group 3 (LG3) in three highly divergent populations of *D. magna*, and spans 3.6 Mb. Despite low levels of recombination, the associated region contains signs of historical recombination, suggesting a role for selection acting on several genes thereby maintaining linkage disequilibrium among the 36 associated SNPs. The region carries numerous genes involved in sex differentiation in other taxa, including *transformer2* and *sox9*. Taken together, the region determining the genetic females shows characteristics of a sex-related supergene, suggesting that LG3 is potentially an incipient W chromosome despite the lack of significant additional restriction of recombination between Z and W. The occurrence of the female-determining locus in a pre-existing low recombining region illustrates one possible form of recombination suppression in sex chromosomes. *D. magna* is a promising model for studying the evolutionary transitions from ESD to GSD and early sex chromosome evolution.

**Key words:** sex determination, sex chromosome, turnovers, gynodioecy, male sterility mutation, *Daphnia magna*.

## Introduction

Sex chromosomes have evolved independently multiple times in many taxa (Miura 2008; Pokorná and Kratochvíl 2009; Stöck et al. 2011; Cheng et al. 2013; The Tree of Sex Consortium 2014). Two evolutionary routes are thought to lead to the evolution of genetic sex determination, and sometimes of sex chromosomes. First, separate sexes may evolve from hermaphroditism, most likely through a male-sterility (i.e., female determining) mutation creating a breeding system called gynodioecy, with genetic females co-occurring with functional hermaphrodites. An initial male-determining mutation is also possible, but is less likely (Charlesworth and Charlesworth 1978). The sex-determining mutation can be favored through causing obligate outcrossing, if inbreeding depression is severe, or if there is a fitness disadvantage to investing resources in both male and female functions compared to investing in only one sex

(Charlesworth and Charlesworth 1978; Innes and Dunbrack 1993). Additional mutations in genes linked to the sex-determining locus may then be favored if their effects are sexually antagonistic (i.e. have opposite effects in the two sexes; Rice 1987; Ellegren 2011). In the case of a female sex-determining locus, additional mutations include ones benefiting males, and deleterious in females (including female-sterility mutations creating males), which will often be eliminated unless they are linked to the femaleness-determining gene. Closer linkage will then be favored in this genomic region (Bull 1983), which may lead to suppression of recombination. Eventually, this may result in a system with “proto-sex chromosomes” carrying linked genes that determine both sexes genetically (with male and female determining mutations on opposite homologs).

The second route towards evolving GSD and sex chromosomes may start from environmental sex determination (ESD), which could be the ancestral state in several major

animal groups (Ohno 1967; Pokorná and Kratochvíl 2016). Although transition from ESD to GSD may be gradual, involving shifting genotype-specific thresholds for male vs. female development under fluctuating environmental conditions (Van Dooren and Leimar 2003), a scenario similar to that of the transition from hermaphroditism to separate sexes is also plausible: a female-determining mutation in a population with pure ESD could lead to a system in which females are genetically determined, while other individuals have ESD (the route through an initial male-determining mutation is also possible). Such a sex-determining mutation can be favored if it restores a 1:1 sex ratio of the population, for example, after environmental conditions change (Edwards 1998), or through other advantages like those outlined for the first pathway above.

The evolution of suppressed recombination between sex chromosomes and more generally, the early stages of sex chromosome evolution remain poorly understood (Wright et al. 2016). Because of their intermediate position between hermaphroditic and dioecious species, gynodioecious plants might have been good candidates to study these early stages. However, they proved not to be very informative in this respect, because gynodioecy is often controlled by cyto-nuclear interactions between mitochondrial male sterility mutations and nuclear “restorer” genes (McCauley and Bailey 2009; Beukeboom and Perrin 2014). Only a few species appear to have pure nuclear control and may then conform to the above theory of sex chromosome evolution (Kohn 1988; Connor and Charlesworth 1989; Weller and Sakai 1991; Spigler et al. 2008). Species in transition from ESD to GSD are therefore of particular interest, although the transition may be rapid (Pokorná and Kratochvíl 2009), so that few systems are available for study.

In some populations of *Daphnia*, pure ESD individuals co-occur with genetically determined females (Innes and Dunbrack 1993; Tessier and Caceres 2004; Galimov et al. 2011). *Daphnia* are cyclical parthenogens, in which clonal reproduction with live born offspring is interspersed with sexual reproduction phases producing diapause stages. Sex determination is usually environmental (Hobaek and Larson 1990; Zhang and Baer 2000; LeBlanc and Medlock 2015), with cues differing among populations (Roulin et al. 2013). In nature, male development of the clonal offspring present in the ovaries may be elicited by the mother emitting a juvenile hormone (JH) or a JH pathway-related molecule (Olmstead and Leblanc 2002). Male production can also be artificially induced by adding hormone analogs to the culture medium (Olmstead and Leblanc 2002). However, some *Daphnia* individuals never produce males, neither under natural conditions nor when artificially exposed to hormone analogs (Innes and Dunbrack 1993; Tessier and Caceres 2004; Galimov et al. 2011). This nonmale-producer “NMP” trait segregates as a single Mendelian locus (or single region) with a dominant female-determining allele called “W”. Heterozygous genotypes (ZW) are genetically determined females called NMP individuals, which take part in sexual diapause stage production exclusively as females. Homozygous genotypes (ZZ) are cyclical parthenogens with ESD. They produce diapause stages either through

male or female function, and are called male producers (“MP” individuals; Galimov et al. 2011).

*Daphnia* populations harboring NMP as well as MP individuals offer an opportunity to study the transition from ESD to GSD, and potentially the early stages of sex chromosome evolution. At present, female is the only genetically determined sex in *D. magna* populations, so that the situation resembles gynodioecy, with no genetically determined males present. Nonetheless, the early steps of sex chromosome evolution may have taken place on both homologs. For instance, sexually antagonistic mutations may have occurred in genes linked to the female-determining locus, and suppressed recombination may have evolved in response to such two-loci polymorphisms in the region.

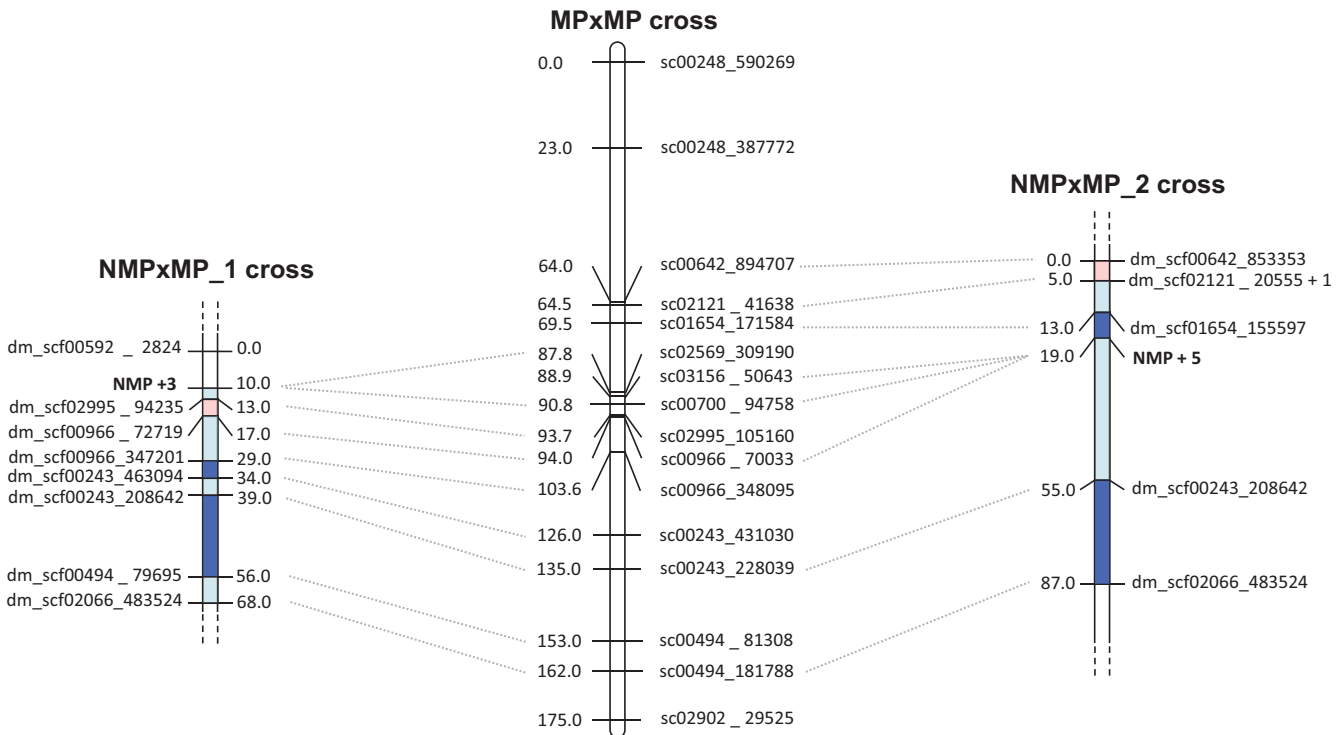
We used genetic linkage mapping, association mapping, and comparative genomics methods to genetically characterize the sex-determining locus and its surrounding region in *D. magna*. Specifically, we first investigated whether the sex-determining locus maps to the same genomic location in crosses involving NMP females from populations with highly divergent mitochondrial lineages (Galimov et al. 2011; Svendsen et al. 2015). This is relevant because the W allele is expected to get co-transmitted with the mitochondrial haplotype (as both have female-limited transmission). Hence the existence of NMP in these divergent lineages might be due to recurrent evolution of a W allele, or, alternatively, might hint at long-term maintenance of the polymorphism within this species. The recombinant frequencies among the offspring of the experimental crosses were also used to investigate the hypothesis of reduced recombination between W and Z. Second, we used association mapping within a single population to identify female-associated SNPs and test if the association among SNPs is maintained purely through physical linkage or also by selection. For this, we estimated levels of linkage disequilibrium and searched for traces of historical recombination between different associated SNPs in the region around the sex-determining locus. Finally, we screened the NMP associated region of *D. magna* for genes involved in sex-determination in this or other species to identify candidate female determining genes. The overall aim of these approaches was to gain a genetic understanding of the transition from ESD to GSD and to assess the possible evolution of an incipient ZW sex chromosome system in *D. magna*.

## Results

### Genetic Linkage Mapping of the NMP-Determining Region Using Microsatellites Markers

We performed three NMPxMP crosses involving NMP females from three populations with highly divergent mitochondrial haplotypes. In each cross, F1 lines were phenotyped (MP vs. NMP) and genotyped at a total of 81 microsatellite loci to perform linkage mapping of the genomic regions underlying the NMP phenotype. The resulting maps were compared with the *Daphnia* genetic map (Dukić et al. 2016), which is based on an F2 panel of an MPxMP cross.

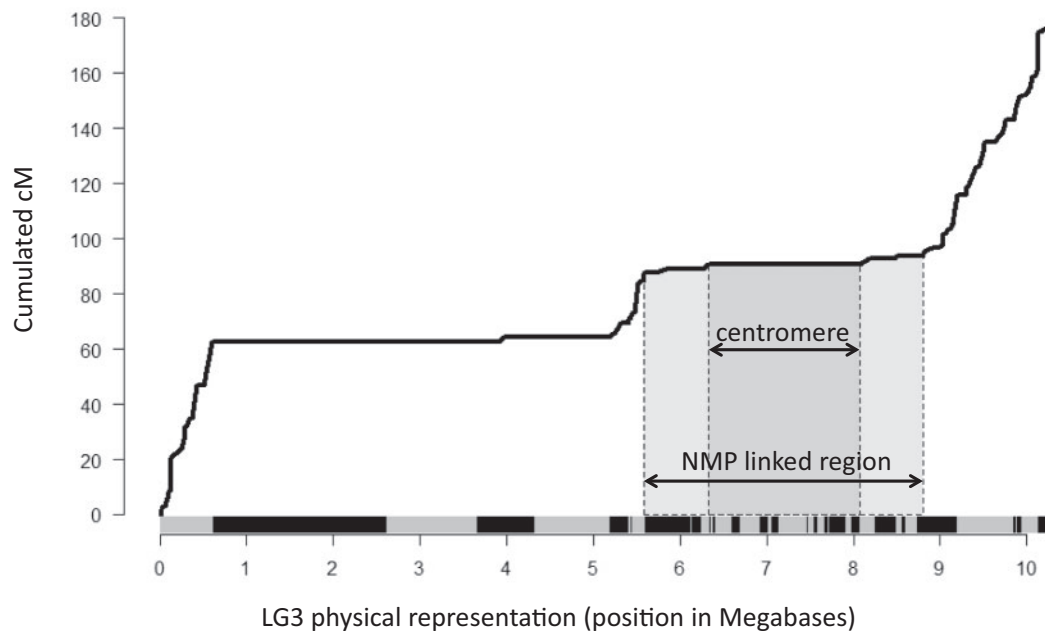
In the NMPxMP\_1 and NMPxMP\_2 crosses, partially overlapping sets of thirteen markers were significantly linked to



**Fig. 1.** Genetic map of the two NMP  $\times$  MP crosses (microsatellite markers) and of the MP $\times$ MP cross used to create the *Daphnia magna* reference genetic map (SNP markers, Dukič et al. in revision). Only Linkage Group 3 (LG3) is shown. Map distances are in centiMorgans, calculated with the Kosambi mapping function in R/qtI. Areas in light blue/light red show a nonsignificant reduction/expansion of recombination by comparison to the MP cross whereas areas in bright blue indicates a significant reduction of recombination. For the two NMP $\times$ MP crosses, one marker per position is represented. In NMP $\times$ MP<sub>1</sub>, NMP + 3 indicates that three markers were in full linkage with the NMP locus. In NMP $\times$ MP<sub>2</sub>, dm\_scf00532\_1398 was fully linked with dm\_scf02121\_20555; also, NMP + 5 indicates that five markers were fully linked with the NMP locus.

the NMP phenotype (or to other markers linked to it). In each cross, eleven of these markers were successfully genotyped in >70% of offspring and were thus used for mapping (see [supplementary material S1, Supplementary Material](#) online). They all mapped to linkage group 3 (LG3) of the reference genetic map ([fig. 1](#)). In the NMP $\times$ MP<sub>1</sub> cross, three markers (dm\_scf02569\_310402, dm\_scf00933\_2550 and dm\_scf00700\_81490) were completely linked with the NMP phenotype, as were five markers in the NMP $\times$ MP<sub>2</sub> cross (dm\_scf02569\_317703, dm\_scf01492\_1407, dm\_scf00933\_2550, dm\_scf03156\_57375, and dm\_scf00966\_75426). Only two of these markers were genotyped in the NMP $\times$ MP<sub>3</sub> cross, but they too showed complete linkage with the NMP phenotype (a third tested marker was not polymorphic in this cross, see [supplementary material S1, Supplementary Material](#) online). In all three crosses, the fully linked markers mapped between cM positions 87.8 and 94.0 cM of LG3 in the reference genetic map. This region also contains the centromere (at 90.8 cM). We call this region the “NMP region” ([fig. 2](#)). A Marey map of LG3 shows that the NMP-region corresponds to a large (~3 Mb) nonrecombining region around the centromere ([fig. 2](#)). Nonrecombining regions around the centromeres are found on all linkage groups of *D. magna*, and are not a sign of reduced recombination between sex chromosomes, since they also occur on all other linkage groups in the MP $\times$ MP cross of the reference genetic map ([Dukič et al. 2016](#)).

We investigated the possibility of additional recombination suppression in NMP $\times$ MP crosses as an indication of reduced recombination between the (putative) Z and W chromosomes compared to between two Z chromosomes. Indeed, genetic map distances between adjacent markers in and around the NMP region were on average lower in the NMP $\times$ MP crosses than in the reference genetic map ([fig. 1](#)). However, this pattern is not statistically supported. The reduction was significant only for two, nonidentical intervals in each of the two crosses ([table 1](#)). Furthermore, the two crosses are not entirely independent. Both involved, at a different stage of the experiment, outcrossing of an NMP-female (different in the two crosses) to a male from the same distant population. In one cross this was done to create the actual mapping population, in the other it was done one generation earlier to obtain a highly heterozygous, hybrid NMP female with the intention to increase the number of markers available for mapping the W. Moreover, it is possible that the slight reduction in recombination frequencies was not specific to the NMP region (we did not have sufficient genotype data on linked markers in other genomic regions to test for this possibility). Overall, our results are clearly inconsistent with strongly reduced recombination across large parts of the putative incipient Z and incipient W chromosomes ([table 1](#)), suggesting that if such a reduction has happened (compared to recombination in MP $\times$ MP crosses), it concerns only a small portion of the chromosome.



**Fig. 2.** Marey map of LG3 in the MP  $\times$  MP cross, showing the evolution of the recombination rate along the physical position on LG3. The x axis shows the position in Megabases. Alternated grey and black color bar on the x axis represents the length of the different scaffolds that compose LG3. The light grey shaded area corresponds to the genomic region containing markers fully linked with the NMP phenotype (from 5.58 to 8.81 Mb). This linked region encompasses the centromere, represented by the dark grey shaded area (from 6.32 to 8.08 Mb).

**Table 1.** Comparison between Crosses with One Parent Heterozygous (NMP $\times$ MP) and Both Parents Homozygous for the MP Factor

(a)	NMP $\times$ MP_1 versus MP $\times$ MP	P-value	% recombinant NMP $\times$ MP_1	% recombinant MP $\times$ MP
	nmp–dm_scf02995_94235	1.000	1.515	1.905
	dm_scf02995_94235–dm_scf00966_72719	0.300	4.545	0.952
	dm_scf00966_72719–dm_scf00966_347201	0.655	12.121	15.238
	dm_scf00966_347201–dm_scf00243_463094	>0.001*	4.545	30.476
	dm_scf00243_463094–dm_scf00243_208642	0.070	4.545	13.333
	dm_scf00243_208642–dm_scf00494_79695	0.016*	13.636	30.476
	dm_scf00494_79695–dm_scf02066_483524	1.000	12.121	13.333
(b)	NMP $\times$ MP_2 versus MP $\times$ MP	P-value	% Recombinant NMP $\times$ MP_2	% Recombinant MP $\times$ MP
	dm_scf00642_853353–dm_scf02121_20555	0.115	3.704	0
	dm_scf02121_20555–dm_scf01654_155597	1.000	7.404	7.692
	dm_scf01654_155597–nmp	0.015*	7.404	23.077
	nmp–dm_scf00243_208642	0.474	27.778	34.615
	dm_scf00243_208642–dm_scf02066_483524	0.016*	25.926	46.154

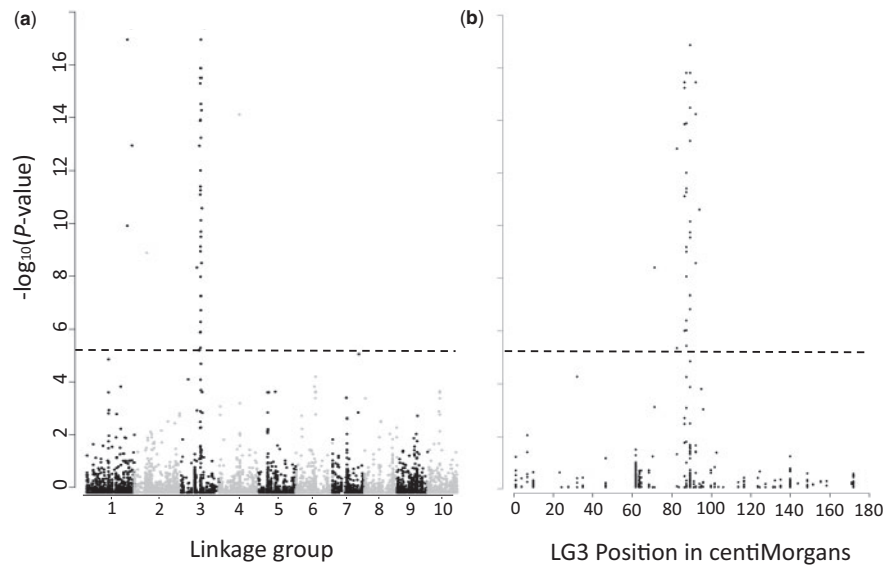
NOTE.—The table shows Fisher's Exact Test P-values for all pairs of markers studied in (a) the NMP $\times$ MP\_1 cross and (b) the NMP $\times$ MP\_2 cross. Table also gives the proportion of recombinants for each cross, and for the reference cross (MP $\times$ MP, Dukič et al. 2016).

\*Statistical significance at alpha = 0.05.

### Association Mapping of SNPs in the NMP Region

A genome-wide association analysis using RAD-sequencing of 72 individuals from a single population revealed 43 SNPs significantly associated ( $FDR < 10^{-5}$ ) with the NMP phenotype (fig. 3a, supplementary material S2, Supplementary Material online). Of these, 36 were contained in a region of LG3 between 72.3 and 95.7 cM, which is only slightly larger than the NMP region defined above (fig. 3b). Of the other seven significantly associated SNPs, five mapped to two scaffolds on LG1, and one SNP to each of LG2 and LG4 (table 2; fig. 3). The number of associated SNPs per scaffold did not correlate with

the scaffold size (Pearson's correlation coefficient  $r^2 = -0.117$ ). On the physical map of LG3 (see Methods and Dukič et al. 2016), the main associations were distributed across 3.6 Mb, between positions 5.4 and 9.0 (fig. 4c; LG3 is 10.1 Mb in total). The 36 highly associated SNPs are distributed across 15 noncontiguous scaffolds whose combined length is 2.42 Mb (about a quarter of LG3's size). Thus, even though the physical map based on linkage disequilibrium (LD) mapping may be incorrect, there is strong evidence that significantly associated SNPs are distributed across a large proportion of LG3. Interestingly, 25% of the significantly associated SNPs occurred



**Fig. 3.** Genome-wide association results. Association of SNP loci with the NMP phenotype in a sample of 53 MP and 17 NMP individuals (a) across the entire genome and (b) on LG3. On LG3, markers between 72.3 and 95.7cM show significant associations with the NMP phenotype (the dashed line shows significance, with FDR-corrected  $P$ -values  $< 10^{-3}$ ).

**Table 2.** List of Scaffolds Containing SNPs Significantly Associated to NMP (Chi square test;  $P < 10^{-5}$ )

LG	Scaffold	Size (bp)	Scaffold Position (cM)	No. SNPs	Position of SNPs on Scaffold	P-Values
1	scf00512	3,718,170	188.7	2	1898250; 1898257	2.04E-17; 1.48E-10
1	scf00205	87,516	210.8	3	7803; 7820; 7827	1.63E-13; 1.63E-13; 1.63E-13
2	scf02190	2,111,488	58.2	1	49642	1.46E-09
3	scf02000	31,222	72.3	1	4181	5.10E-09
3	scf00848	142,519	83.9	2	34647; 68264	5.62E-06; 1.67E-13
3	scf02569	397,658	87.9–88.8	9	384651; 381715; 268729; 232532; 193880; 80640; 80555; 77438; 77432	1.25E-06; 5.24E-16; 1.99E-14; 8.42E-16; 1.04E-11; 1.35E-12; 8.62E-10; 1.29E-09; 5.30E-12
3	scf03156	106,208	88.8	4	4433; 50649; 50655; 79120	2.30E-16; 1.20E-06; 1.20E-06; 5.14E-07
3	scf00027	84,679	88.8	4	34526; 29953; 26878; 12813	1.11E-08; 7.27E-12; 4.61E-06; 1.84E-14
3	scf02723	46,460	90.8	5	2032; 2194; 13416; 13482	8.44E-14; 2.04E-17; 2.30E-16; 4.78E-15
3	scf01727	97,879	90.8	1	6345	5.62E-08
3	scf02895	190,495	90.8	2	190140; 190147	3.76E-10; 3.76E-10
3	scf01943	93,640	90.8	2	13984; 67370	5.79E-08; 2.42E-10
3	scf02003	324,342	90.8	1	27223	9.30E-11
3	scf03194	167,565	90.8	1	124008	1.91E-07
3	scf00378	73,438	93.7	1	22714	5.24E-16
3	scf00452	41,028	93.7	2	34895; 34916	8.23E-15; 8.23E-15
3	scf02995	126,150	93.7	1	108378	3.47E-09
3	scf00966	460,511	95.7	1	174527	3.33E-11
4	scf00311	941,766	95.0	1	669922	1.18E-14

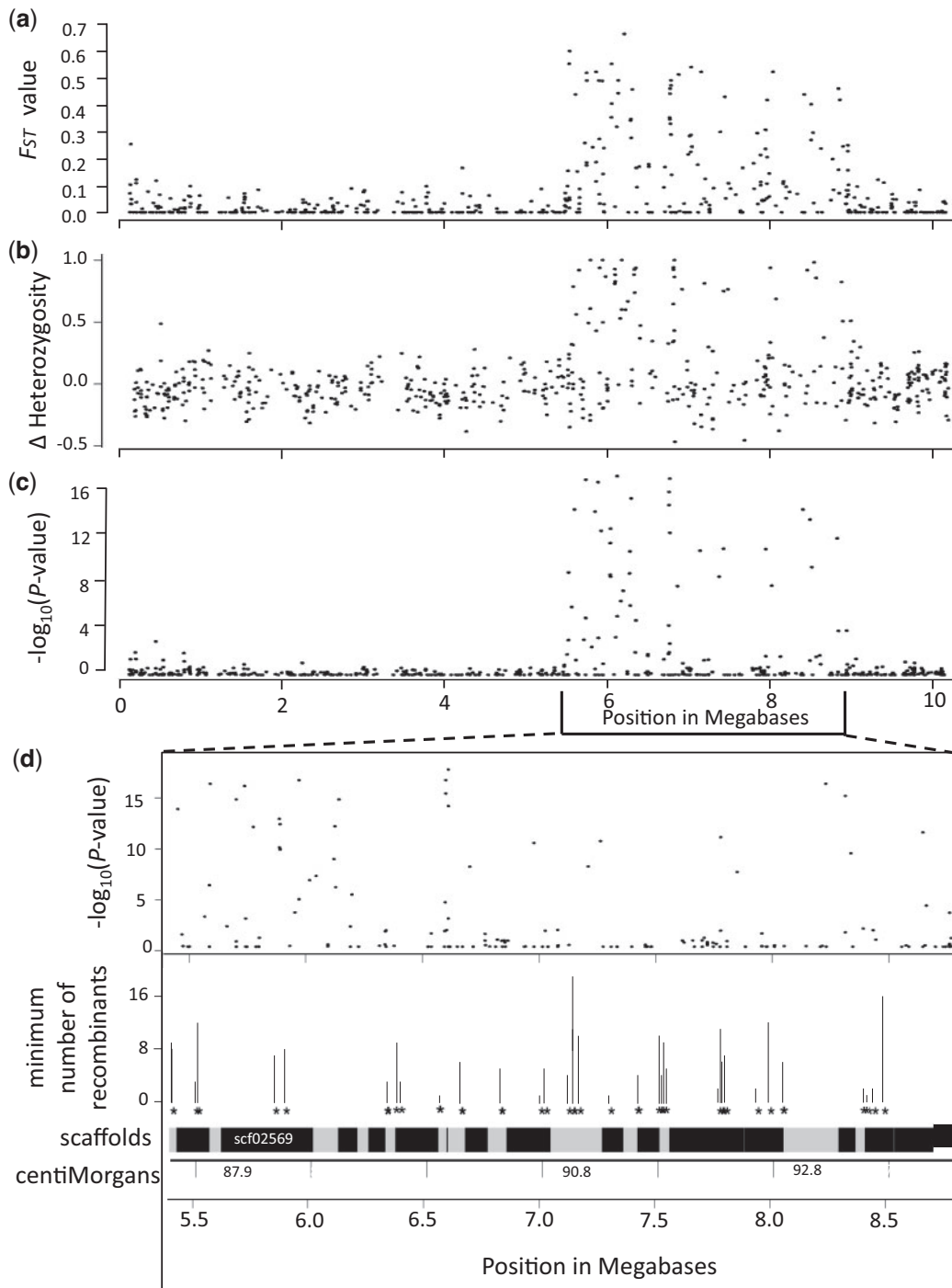
NOTE.—LG: linkage group; Size: total size of the scaffold (in basepair); No. SNPs: number of associated SNPs on the scaffold; P-values: resulting Chi square P-value.

on just a single 400 kb scaffold. This scaffold, scf02569, is therefore a good candidate for the location of the initial, female-determining mutation.

### Differentiation, Heterozygosity, Linkage Disequilibrium, and Historical Recombination in the NMP Region

To investigate the levels of differentiation between the NMP and the MP individuals in the NMP region, we performed several correlated analyses (to test for consistency of the results): we estimated  $F_{ST}$ , compared levels of heterozygosity

between both phenotypes, and performed a LD analysis. The NMP region showed strong differentiation between NMP and MP individuals for the associated SNPs, reflected in  $F_{ST}$ -values for individual markers as high as 0.7 (fig. 4a). Equally, there was strong LD between the 36 associated SNPs (fig. 5), although interspersed with areas of low LD, among non-associated SNPs. In concordance with the association results, we also found higher levels of heterozygosity in NMP individuals than in MP individuals (average heterozygosity in the NMP region of 0.50 vs. 0.33 respectively,  $P < 0.0001$ ; fig. 4b). This difference in heterozygosity is specific

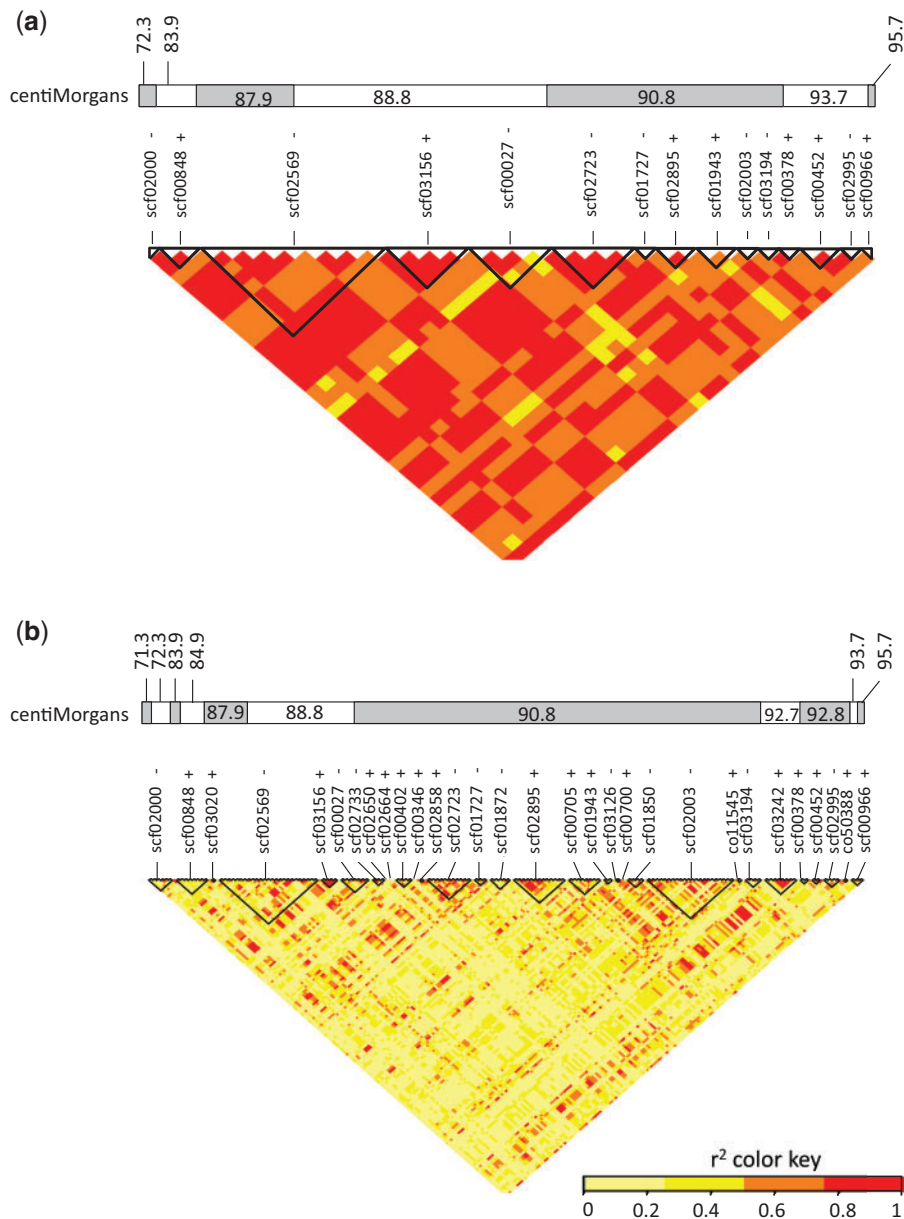


**Fig. 4.** Subdivision between NMP and MP individuals ( $F_{ST}$  values, in panel a), the difference in heterozygote frequencies (panel b), and log-transformed  $P$ -values for associations on LG3 (panel c). Panel (d), for the NMP and flanking regions, shows the associations and the minimum numbers of recombinant haplotypes at different positions represented as both the genetic map locations in centiMorgans, and the physical positions of scaffolds.

to the NMP region: in the rest of the genome, the NMP and MP individuals did not differ in heterozygosity (the means were 0.34 and 0.32 respectively,  $P = 0.149$ ). However, heterozygosity of four, supposedly inbred MP individuals was only half of the population average (supplementary material S3, Supplementary Material online).

Second, to investigate whether the occurrence of multiple, highly associated SNPs distributed across a large region can be

explained by a lack of recombination, we tested for historical recombination in the region using Hudson's four gamete test (see "Materials and Methods" section). This test detected the presence of recombination in 37 pairs of adjacent polymorphic sites in the NMP region. These historical recombination events were distributed across the entire NMP region, and were found both between and within scaffolds (fig. 4d). This implies that a low level of recombination or gene



**Fig. 5.** Linkage Disequilibrium Heatmap ( $r^2$  coefficients) in the NMP region. Results are shown for (a) the NMP associated SNPs, (b) all SNPs mapping to the NMP and flanking region corresponding to the NMP region using the 72 individuals (55 MP and 17 NMP). Black triangles represent scaffolds, and the bi-colored band represents the genetic map positions of each SNP, in centiMorgans. + and – represent the orientations of the scaffolds (forward and reverse respectively).

conversion occurs between the incipient Z and the incipient W chromosomes, or has at least occurred historically.

### Identification of Candidate Genes Involved in Sex Determination, and Prediction of the Effect of SNP Substitutions on Amino Acid Identity.

To search the NMP region for genes known to be involved in sex determination/differentiation in other taxa, we analyzed the 283 nonannotated protein sequences that are located in the NMP region. Of these, 184 returned a BLAST result, but 39 were described as hypothetical proteins in the *D. pulex* draft transcriptome DAPPUDRAFT and did not reliably match any proteins in the NCBI non-redundant protein database. Among the remaining 145 successfully annotated sequences,

121 (82%) had a top hit on *D. pulex* or *D. magna* sequences, while the others matched various arthropods (15 sequences), and other invertebrates (10 sequences). We compared the 145 annotated genes with the NCBI list of 601 genes known to be involved in sex determination or sex differentiation in other invertebrates, and identified 14 candidate genes (table 3). These include genes involved in sex-specific endocrine signaling pathways, such as a membrane androgen receptors (*zip9*), a tissue specific modulator of the ecdysone response in *Drosophila* (*Broad* complex; Karim et al. 1993), as well as a member of the aldo-keto reductase family (Penning et al. 2000). Scf02569 harbors *transformer 2* (*tra2*), a well-known splicing factor of the sex determination pathway in insects. Scf02569 also harbors a gene resembling transcription factor

**Table 3.** Results of the Blast Run Showing the 14 Candidate Genes and their Location in the *D. magna* Genome

Scaffold	Start position	End position	Size (aa)	NCBI Name	Species	Classification	E-value	Similarity
scf00027	2877	6078	316	Serine arginine-rich splicing factor 7	<i>Harpegnatos saltator</i>	Insecta	4.0E–50	82.8%
scf00848	96321	97283	136	Aldo-keto reductase family 1, member C4	<i>Riptortus pedestris</i>	Insecta	4.1E–59	72.5%
scf02003	35289	35935	136	Poly-U-binding splicing factor Half Pint	<i>Acyrtosiphon pisum</i>	Insecta	5.9E–67	95.1%
scf02003	213333	214454	115	Cytochrome P450 314 family	<i>Daphnia magna</i>	Crustacea	3.8E–46	88.3%
scf02569	3227	4315	108	Zinc transporter zip11	<i>Tribolium castaneum</i>	Insecta	3.9E–29	80.2%
scf02569	9179	10907	300	Zinc transporter zip9	<i>Poecilia formosa</i>	Vertebrata	2.1E–75	74.8%
scf02569	35151	44725	606	SOX-9-like transcription factor	<i>Acromyrmex echinatio</i>	Insecta	5.0E–48	89.8%
scf02569	218892	220701	292	Dnaj homolog dnaj-5	<i>Acromyrmex echinatio</i>	Insecta	4.4E–109	72.1%
scf02569	334258	337000	462	Broad-complex	<i>Oncopeltus fasciatus</i>	Insecta	8.3E–50	85.2%
scf02569	340469	342584	281	Transformer2	<i>Daphnia pulex</i>	Crustacea	2.9E–119	88.7%
scf02569	76814	79370	558	Protein SPT2 homolog	<i>Acyrtosiphon pisum</i>	Insecta	2.00E–33	63.9%
scf02569	228772	229714	158	Histone deacetylase complex subunit sap18	<i>Metaseiulus occidentalis</i>	Arachnida	1.1E–55	82.1%
scf02723	1124	6033	287	Epidermal growth factor receptor kinase	<i>Zootermopsis nevadensis</i>	Insecta	1.7E–30	71.2%
scf03156	4200	8559	794	Lysine-specific histone demethylase 1A	<i>Stegodyphus mimosarum</i>	Arachnida	0.0	85.3%

NOTE.—The table reports the start and end position of the gene on the scaffold it maps to, the size of the expected protein (number of amino-acid), the NCBI attributed gene name, the taxon with the best blast hit and its classification, the corresponding E-value, and the percentage of sequence similarity.

*sox9*, which acts to inactivate the female differentiation pathway and promote spermatogenesis in males in mammals. In addition, genes involved in chromatin remodeling (lysine-specific histone demethylases, histone deacetylases) are also present in the region, and could be involved in the usual environmental aspect of sex determination in *Daphnia*. The 14 genes are located on 6 different scaffolds, with scf02569 containing eight of these genes, whereas other scaffolds harbor at most two of these genes (table 3).

Nine of the 176 SNPs identified by RAD-sequencing and mapping in the NMP region were located in genes within which the exon-intron structure could not be determined (very likely because of errors in the assembly or gene structure definition), 69 were in intergenic regions, 17 in 5' UTR regions, 8 in 3' UTR regions, 19 in introns, and 53 within a gene (supplementary material S2, Supplementary Material online). Of these 53 coding SNPs, 32 of are in non-annotated genes, including one (on scaffold02003 in position 98755) that introduced a stop codon into a gene which was annotated as “hypothetical protein” in the *D. pulex* draft genome. 23 SNPs were synonymous and 30 were nonsynonymous mutations. Only one significantly associated SNP induced a nonsynonymous mutation in a candidate gene: the SNP at position 4433 on scf03156, inducing a change from valine to leucine in the lysine specific histone demethylase. However, RAD-sequencing covers only a small fraction of the genome (here 154745 loci were mapped, with reads of 95 bp, representing an estimated 6.12% of the genome). Hence, additional non-synonymous SNPs may be present in parts of the NMP region not covered by RAD-sequencing reads. The same applies for potential regulatory SNPs.

## Discussion

Our results suggest that the NMP phenotype is determined by a single, large genomic region located on LG3. The few

significantly linked SNPs on other linkage groups may be in linkage disequilibrium with the major region on LG3, maintained by pleiotropic effects. Alternatively, they may be explained by errors in the genetic map or in the genome assembly, and might, in fact, be variants within the NMP-region. The NMP phenotype mapped to the same LG3 region in all three crosses involving populations with divergent mitochondrial lineages (Galimov et al. 2011). This indicates either a single evolutionary origin of NMP in *D. magna* or parallel evolution repeatedly implicating the same genomic region. Given the divergent mitochondrial sequences and the co-inheritance of mtDNA with the female-determining allele, a single evolutionary origin of NMP in *D. magna* would suggest that the female-determining mutation is old. However, parallel (convergent) evolution remains possible, particularly as the NMP region contains more than one gene potentially involved in sex determination and sex differentiation. These genes could represent different mutational targets for NMP-inducing mutations, and hence the mutation may not be the same in all populations, despite occurring in the same genomic region. Finally, rare paternal transmission of mitochondria or transmission of the female-determining mutation through rare males cannot be ruled out, but are unlikely (Galimov et al. 2011; Svendsen et al 2015). Interestingly, a dominant NMP phenotype has also been described in *D. pulex* (Innes and Dunbrack 1993) and in *D. pulicaria* (Tessier and Caceres 2004). Considering that the separation of *D. magna* and *D. pulex* was estimated to be 150 Ma (Kotov and Taylor 2011), parallel evolution of the NMP phenotype seems more likely for these two species, rather than a femaleness allele having been maintained for such a long evolutionary time.

According to the classical model of sex chromosome evolution discussed in the “Introduction” section, the establishment of a female-determining mutation is followed by



additional mutations with sex antagonistic (SA) effects occurring at nearby loci. Recombination between these loci and the female sex-determining locus is deleterious, and hence SA selection is thought to favor a reduction of recombination between the incipient Z and W (or X and Y). Such recombination reduction in the heterogametic sex (compared to recombination in the homogametic one) is considered to be a hallmark of early sex chromosome evolution. Alternatively, the sex-determining mutation may occur in a region with already low recombination. Such regions occur on autosomes, for instance due to inversions or in the vicinity to the centromere (Hoffmann and Rieseberg 2008; Ironside 2010; Joron et al. 2011). If such a region contains, from the outset, multiple linked potential target loci for SA mutations, then a further reduction of recombination after the occurrence of the sex-determining mutation might not be favored by selection, at least not initially. In our study, the sex-determining mutation mapped to the peri-centromeric region of LG3, which, as expected, has a low recombination rate not only in MPxNMP crosses, but also in MPxMP crosses. Interestingly, the sex-determining mutations of *Papaya* and *Populus* are also found in peri-centromeric locations (Yu et al. 2007; Kersten et al. 2014). In addition, we did not find strong evidence for a further reduction of recombination between the incipient Z and W chromosomes, suggesting that SA selection did not act strongly (if at all) to further extend the low-recombination region containing the female-determining mutation. Yet, SA selection may still play a major role in determining the NMP phenotype. Under the hypothesis that SA selection occurs, multiple loci should contribute to the NMP phenotype, not only for the determination of femaleness, but also for the expression or fitness of the female phenotype, or for enhancing maleness of ZZ individuals. Here, we found highly associated SNPs distributed on a total of 3.6 Mb, but separated by regions of low LD in which there is strong evidence for historical recombination or gene conversion. This suggests a role for selection to be acting on several genes in the NMP region. This selection is likely sex-antagonistic, as it appears to maintain different W-linked vs. Z-linked alleles in several sub-regions independently (and thereby maintain the high LD between associated SNPs despite historical recombination). We identified 14 genes in the NMP region that are known to be involved in sex determination/sex differentiation in other taxa. These genes might be potential target for SA mutations. Hence, the NMP region shows all characteristics of a supergene (Joron et al. 2011), with different alleles appearing to be selectively maintained between Z and W in several genes throughout the region. Such a “sex-supergene” is nothing else than the first step towards the evolution of a sex chromosome, even though low recombination in this region is not due to secondary suppression of recombination subsequent to the occurrence of the sex determining mutation, but due to a chance event (incidence of the sex-determining mutation in a large region with pre-existing low levels of recombination). Nonetheless, such a chance event may in fact be one of the possible mechanisms leading to initial suppression of recombination between incipient sex chromosomes.

The transition from ESD to GSD is thought to happen rapidly in nature. However, intermediate stages (such as gynodioecy, or partial GSD) can be evolutionary stable, depending on their origin and genetic control (Charlesworth and Charlesworth 1978). The NMP mutation might first have been favored by selection because it leads to obligate outcrossing (Innes and Dunbrack 1993). This is supported by the fact that inbreeding depression in *Daphnia* is strong (Lohr and Haag 2015) and that within-clone mating occurs at an appreciable frequency (e.g., 5% of MP individuals in our association study were offspring of within-clone mating). Once it attained an intermediate frequency, it was probably also affected by negative frequency-dependent sex-ratio selection, which may have set the stage for additional mutations to gradually improve maleness of ZZ individuals (Innes and Dunbrack 1993; Galimov et al. 2011). The coexistence of ESD and GSD individuals in some populations of *D. magna* might be evolutionary stable. This stability could be due to their reproductive cycle, called cyclical parthenogenesis, which might prevent further evolution towards full GSD. Indeed, only females participate in parthenogenetic reproduction. Hence, in order to profit from parthenogenetic multiplication, all genotypes have to be females at some point during the seasonal cycle. ZZ females might increase their male function during sexual reproduction by prolonging their investment in parthenogenetic clutches (late season parthenogenetic clutches are usually male; Galimov et al. 2011). It is also possible that the maleness of ZZ individuals may have been quantitatively improved by linked genes with SA effects (see above). Nonetheless, at the very end of the season, it may still be more profitable for these females to engage in sex rather than to abandon reproduction or to parthenogenetically produce additional males which might not have the time to develop into adults. If, however, *Daphnia* was to complete the transition from ESD to pure GSD (as envisaged by the theory outset in the “Introduction” section), the incipient Z homolog of the NMP region would need to acquire a mutation determining the male sex (i.e., a recessive female-sterility mutation), and the two chromosomes would then be called proto-sex chromosomes. This has not happened in *D. magna*, as the ZZ individuals still have ESD and still contribute to the diapause stage production through both male and female functions. However, while hatchlings from diapause stages always seem to be female in *Daphnia*, male hatchlings do occur in the related *Daphniopsis ephemeralis*, living in short-lived environments (Schwartz and Hebert 1987). Hence, it is not inconceivable that some *Daphnia* populations or species might evolve full GSD, perhaps especially under conditions that reduce the importance of the parthenogenetic phase.

The molecular mechanisms underlying male differentiation in *D. magna* and its relationship with ESD are the object of much research (Kato et al. 2008, 2010, 2011; LeBlanc and Medlock 2015). To date, we know that a homolog of *dsx* is found in *D. magna* (*dapmadsx*) and that it is a major (possibly terminal) effector regulating the male phenotype (Kato et al. 2011; Salz 2011; Beukeboom and Perrin 2014). The regulation of its expression appears, however, to be different in *Daphnia*

compared to other arthropods, as no evidence for sex-specific expression nor sex-specific splicing was identified for the *transformer* (*tra*) and the *dapmadsx* genes in male versus female embryos (Kato et al. 2010, 2011; Verhulst et al. 2010). Instead, the expression of *dapmadsx* might be controlled by an “on/off” mechanism whose activation may involve elements responsive to juvenile hormone (Kato et al. 2011). Our results show that the NMP region does not contain *dapmadsx* nor *tra*, and hence neither of these genes is the likely location of the female-determining mutation. However, the NMP mutation does contain multiple genes that are potentially members of the same pathways. These genes include *tra2* on scf02569, a splicing regulator known to interact with *tra* to control the female sex-specific splicing of *dsx* in insects. It will be interesting to investigate if and how *tra2* interacts with *dapmadsx*. Moreover, we found four genes involved in hormonal pathways (*zip9*, *zip11*, *Broad* complex, aldo-keto reductase), which might be relevant here as *dsx* expression might be controlled by hormone-responsive elements. Furthermore, scf02569 contains a gene resembling transcription factor *sox9*, which inactivates the female differentiation pathway and promotes spermatogenesis in male mammals. Overall, the presence of these 14 genes in the NMP region is congruent with a major role of this region in sex determination/differentiation. In addition, scf02569 is a strong candidate for carrying the original sex-determining mutation. Not only does it contain 25% of the significantly associated SNPs and eight of the 14 candidate genes, but one of these candidate genes (*tra2*) appears to be one of the most promising ones, based on its function in other organisms, and on what is currently known about the mechanisms underlying male differentiation in *D. magna*.

In the present study, we determined that the dominant female sex-determining locus observed in multiple populations of *D. magna* is located within a single genomic region, in the peri-centromere of LG3. With a pre-existing low recombination rate, the region contains multiple genomic segments in high LD spanning a total of 3.6 Mb, separated by regions of low LD due to historical recombination or gene conversion. The strong LD is likely maintained by sex antagonistic selection, and multiple genes involved in both males and female sex differentiation were found throughout the region. We conclude that *D. magna*'s LG3 carries a sex-related supergene, and is an incipient W chromosome, the youngest stage that might be possible to empirically observe in sex chromosome evolution. *D. magna* is thus a very promising model for studying the evolutionary transitions from ESD to GSD and early stages of sex chromosome evolution.

## Materials and Methods

### Genetic Linkage Mapping of the NMP-Determining Region, and Assessment of Recombination Rates

In order to map the genomic region responsible for the NMP phenotype, we performed experimental crosses between known NMP and MP genotypes. Because the NMP locus is believed to be heterozygous dominant in females, the goal was to maximize heterozygosity of the mother NMP clone,

while ensuring the MP father was homozygous or had different alleles. All three crosses involved outcrossing between two populations to ensure that a sufficient number of markers had different genotypes between fathers and mothers. One of the crosses also involved a NMP mother that was already a hybrid between two populations, again in order to maximize heterozygosity. The first cross called “NMPxMP\_1” involved a NMP female from Volgograd, Russia (48.53 N, 44.486944 E) and a male from Orog-Nur, Mongolia (45.032708 N, 100.718133 E) as well as 66 of their F1 offspring. The second cross (NMPxMP\_2) used a hybrid NMP female, which was produced by crossing a NMP female from Moscow, Russia, (55.763514 N, 37.581667 E), with a male from Orog-Nur. The cross then involved this hybrid female and a male from Vääränmaanruskia, Finland (60.271617 N, 21.896317 E), as well as 54 of their offspring. The third cross (NMPxMP\_3) involved a NMP female from Yakutsk, Russia (61.964047 N, 129.630956 E) and a male from Rybnoye, Russia (56.425003 N, 37.602672 E), as well as 22 of their offspring. The three NMP females had divergent mitochondrial haplotypes (based on a 534 bp alignment of the COI sequences, the average number of substitutions per site was 0.037 between Yakutsk and Moscow, 0.041 between Yakutsk and Volgograd, and 0.010 between Moscow and Volgograd; Galimov et al. 2011; Genbank accession number JF750770.1 for Moscow, AY803073.1 for Volgograd, and KY082572 for Yakutsk). Microsatellite markers distributed across the genome were used to investigate the parental lines. Markers that were heterozygous in the NMP mother and for which father and mother had different genotypes were selected and genotyped in the offspring. For all these markers, it could unambiguously be determined which of two maternal alleles was transmitted to a given offspring. Hence, linkage of markers to the NMP-determining region could be assessed, by assaying co-transmission of maternal alleles with the phenotype (i.e. with the W vs. Z chromosome). Before genotyping, the offspring were thus phenotyped using the juvenile hormone Methyl Farnesoate, which triggers the production of males in MP strains but not in NMP strains of *Daphnia* (Galimov et al. 2011).

Microsatellite loci were amplified using the M13-protocol (Shuelke 2000). For each locus, unlabeled forward and reverse primers were used together with fluorescently labeled, universal M13 primer. The forward primer consisted of a locus-specific part as well as an overhang complementary to M13. PCR reactions were carried out using the Type-it Microsatellite PCR Kit (Qiagen) according to the manufacturer's protocol with an annealing temperature of 60 °C. After 22 cycles, the annealing temperature was lowered to 53 °C for another 20 cycles in order to allow for proper M13 annealing. The resulting PCR products were diluted four times and mixed with a LIZ5000 size ladder (Applied Biosystems). Samples were genotyped using ABI 3730 capillary sequencer and GENEMAPPER software v. 3.0 (Applied Biosystems). A total of 81 microsatellite loci (see [supplementary material S1](#), [Supplementary Material](#) online) were tested in the parents. Of these, 60 were polymorphic in one or both parents and thus genotyped in the offspring (47 in NMPxMP\_1 and 21 in

NMPxMP\_2, partially overlapping). Linkage to the NMP phenotype was assessed with a Fisher's Exact tests (two-tailed). Some of the markers were specifically designed in regions for which linkage to the NMP phenotype was suspected based on information from an earlier version of the genetic map (Routtu et al. 2010, 2014) and the initial finding of weak but significant linkage of one marker (dm\_scf00243\_208642) in the NMPxMP\_1 cross. Therefore, the markers do not represent a random sample throughout the genome. The NMPxMP\_3 cross, which included the NMP female with the most divergent mitochondrial haplotype, was done at a later stage. It was only used to test whether NMP maps to the same genomic region as in the two other crosses. Hence, only three loci closely linked to NMP in the first two crosses were also genotyped in the offspring of this cross.

Genetic linkage mapping of the NMP region was carried out in R/qtl (Broman et al. 2003). It soon became evident that NMP mapped to a region of LG3 of the *D. magna* reference genetic map (Dukič et al. 2016), of which a first version called v4.0.1 was published in Svendsen et al. (2015). Hence, map construction was done using markers that either showed significant linkage with the NMP phenotype ( $P < 0.01$  in pairwise Fisher's exact tests) or were found on scaffolds of the *D. magna* genome v2.4 (bioproject reference PRJNA298946, on the NCBI repository: <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA298946>) that had been mapped to LG3. Markers that had more than one third of missing genotypes (amplification failures, etc.) were discarded.

The mapping procedure consisted in first positioning and ordering the markers according to previously available data (Dukič et al. 2016), if possible, and secondly by estimating genetic distances among the ordered markers according to standard procedures. Specifically, we ordered our markers by using the cM position of the nearest mapped SNP from the same scaffold in the reference genetic map v4.0.1. Microsatellite markers on scaffolds that were not mapped in the reference genetic map were positioned according to the estimated recombination fraction in our crosses between these markers and already mapped markers. The only exception to this procedure was done for microsatellite marker scf02066\_483524, which is located on a mis-assembled part of scf02066, closely linked to the end of scf00494 (Dukič et al. 2016), and thus was ordered according to this position. Once ordered, Kosambi-corrected genetic map distances among all markers were recalculated from the offspring genotypes of our crosses using R/qtl (with the option `sliding-window = 8` markers).

To test for a reduction of recombination around the NMP region in the MPxNMP crosses versus the reference MPxMP cross of the genetic map v4.0.1, we compared the genetic distances for intervals of adjacent markers between the crosses. Specifically, for each interval, we assessed the number of recombinant vs. non-recombinant individuals in each of the NMPxMP crosses versus the reference cross, and tested for significant differences using Fisher's exact test (two-tailed) implemented in R core package stats (R Development Core Team 2008).

### Association Mapping of SNPs in the NMP Region

To further characterize the NMP region, we used single-nucleotide polymorphism (SNP) data obtained by RAD-sequencing of a random sample of 72 individuals (17 NMP and 55 MP) obtained by hatching resting stages from the Moscow population. BioSample metadata and FASTQ files are available in the NCBI SRA database under the BioProject accession number PRJNA352441, and BioSamples accession number SAMN05980834 and SAMN05980835 for the NMP and MP individuals respectively. The details of the RAD-sequencing protocols, alignment to the *D. magna* reference genome, and SNP calling are given in [supplementary material S4, Supplementary Material](#) online. We obtained 7376 SNPs to be analyzed ([supplementary material S2, Supplementary Material](#) online for an Excel document containing all SNPs). We first performed a genome-wide association study, using the expectation that any bi-allelic SNP functionally related to NMP or tightly linked to it should be heterozygous in NMP individuals and homozygous for the more frequent of the two alleles in MP individuals, corresponding to the ZW and ZZ genotypes, respectively. Only sites with a minor allele frequency larger than 0.1 and less than one third of the individuals with missing genotypes were used in the analysis. For each retained bi-allelic site, we grouped individuals into four categories: Heterozygous NMP individuals, homozygous NMP individuals, homozygous MP individuals for the major allele, and MP individuals nonhomozygous for the major allele (either heterozygous or homozygous for the minor allele). For each site, we counted the number of individuals in each of the four categories and calculated the expected number of individuals (under the null-hypothesis of no association) using standard Hardy-Weinberg proportions with allele frequencies estimated across all individuals. We then used Pearson's Chi-square tests with two degrees of freedom to evaluate the genotype-phenotype association at each site. However, in order to only test the hypothesis specified above, any excess that went in the direction opposite the hypothesis (for instance an excess of non-heterozygous individuals among NMP) was discarded (i.e., was not taken into account for the overall Chi-square value). The R script of this association analysis is given in [supplementary material S5, Supplementary Material](#) online. The significance of the association was assessed by correcting the  $P$ -value of the Chi-square test according to an overall false discovery rate (FDR) of  $10^{-5}$  using the `p.adjust` function of the R core stats package.

### Differentiation, Heterozygosity, Linkage Disequilibrium, and Historical Recombination in the NMP Region

To confirm that the NMP associated region is highly differentiated between NMP and the MP individuals, we estimated  $F_{ST}$  for each SNP with the R package PopGenome (Pfeifer et al. 2014). We also performed a linkage disequilibrium (LD) analysis of the NMP region and investigated relative levels of heterozygosity for MP and NMP individuals. Note that all three analyses test essentially for the same thing (Charlesworth et al. 1997): If there are many strongly

associated SNPs, we would expect  $F_{ST}$  to be high, heterozygosity to be high in NMP individuals, and LD to be high, at least between associated SNPs. All three analyses were run to check for consistency, and in the case of heterozygosity also as a comparison with other genomic regions. The subsequent analyses were restricted to cM positions between 85 and 95 of the genetic map, which includes the NMP region as well as some flanking regions on either side. However, due to the dearth of recombination in this region, the relative position and orientation of many of the scaffolds is unknown (several entire scaffolds having the exact same cM position). Hence, we first inferred the likely relative position and orientation of these scaffolds by LD mapping in MP individuals, assuming no structural rearrangement of those scaffold between MP and NMP individuals (see [supplementary material S6, Supplementary Material](#) online for details of the LD mapping procedures). Once the inferred physical map was established, we used it to plot the genotype-phenotype association in the region as well as  $F_{ST}$ , and heterozygosity.

Linkage disequilibrium for all individuals (i.e., including NMP) was then estimated via pairwise  $r^2$  for each pair of SNPs (see [supplementary material S7, Supplementary Material](#) online), using the program MCLD (Zaykin et al. 2008). Significance was tested using 9999 permutations, and the extent of LD was visualized using a heatmap constructed in R using the LDheatmap package (Shin et al. 2006). To test for signatures of historical recombination between W and Z haplotypes, we used a filtered dataset composed of 140 polymorphic sites in the region, retaining just one site per read (a maximum of two polymorphic sites on the same read were present in the whole data set, but SNPs on the same read were always in full linkage). We then phased these data using the GERBIL program implemented in the package GEVALT V2.0 (Davidovich et al. 2007). We did not allow the program to infer missing genotypes, since the algorithm favors the use of the more common allele, here the Z allele (only 17 out of 144 haplotypes are W haplotypes). The phasing resulted in two Z haplotypes for each MP individual and in one W and one Z haplotype for each NMP individual (Z haplotypes were identified according to the allelic state of the two most strongly associated SNPs named scf2723\_2194 and scf2723\_13482). Because we could not exclude genotyping nor phenotyping error (the latter only in the direction of falsely identifying an individual as NMP), we used conservative criteria for the test: we first removed the NMP individual (Rm1-01) which had a high genotypic resemblance to MP and thus resulted the highest evidence for recombination (this individual was most likely mis-phenotyped during the hormonal exposure). Furthermore, before carrying out the test we corrected singleton variants within each haplotype group: if a variant was present in only one haplotype in the group, we reverted its state to the majority allele in this group. Overall, 27 loci and 6 loci out of 140 loci were reverted in W and Z, respectively. This conservatively assumes that all these singleton variants were due to genotyping error (note that loci with a minor allele frequency of  $< 0.1$  across both groups had already been excluded during the initial filtering; see [supplementary material S8, Supplementary Material](#) online for the list of

uncorrected and corrected haplotypes). Finally, to test only for recombination between Z and W haplotypes (as opposed to recombination within Z), we performed the Hudson's four-gamete test, which stipulates that the presence of the four possible gametes in a pair of segregating bi-allelic sites is a sign of historical recombination or gene conversion (Hudson 1985). We inspected all instances where recombination was detected by the test and retained only those where the inferred recombination had occurred between the Z and the W haplotypes.

### SNP Effect and Identification of Candidate Genes Involved in Sex Determination

To assess whether the NMP region contains candidate genes with known functions related to sex differentiation or sex determination, we extracted all 1,306 protein sequences corresponding to transcript sequences located on the scaffolds in the NMP region, and reduced isoform redundancy using BlastClust (<http://toolkit.tuebingen.mpg.de/blatclust>) with the following parameters: minimum length coverage of 60%, minimum identity of 90%, minimum transcript size of 100 amino acids. This resulted in a set of 361 protein sequences, which we trimmed by hand to remove further the redundancy (we only kept one transcript for each gene). The retained 283 protein sequences (see [supplementary material S9, Supplementary Material](#) online for the complete list) were blasted against the NCBI nr database with Blast2GO (Conesa and Götz 2008), using blastp and a maximum  $e$ -value of  $10^{-10}$ . Annotated genes were then compared with a list of 601 genes obtained from the NCBI gene data base using the keywords "sex determination" and "sex differentiation". We also used the GFF file (also available on NCBI, bioproject reference PRJNA298946), which contains gene features of *D. magna*, to classify each SNP in the NMP region according to whether it induces a synonymous or a non-synonymous change. This analysis was done using the software tool IGV (Robinson et al. 2011).

### Data Accessibility

BioSample metadata and FASTQ files are available in the NCBI SRA database under the BioProject accession number PRJNA352441, and BioSamples accession number SAMN05980834 and SAMN05980835 for the NMP and MP individuals respectively.

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

We thank the Zoo of Moscow and N. I. Skuratov for sampling permits, and David Frey for help with culture maintenance indoors. We thank the Department of Biosystem Science and Engineering of the ETH Zurich, in particular C. Beisel and I. Nissen for Illumina sequencing, and we gratefully acknowledge support by M-P. Dubois, the platform Service des Marqueurs Génétiques en Ecologie at CEFÉ, the genotyping

and sequencing facilities of the Institut des Sciences de l'Evolution Montpellier (ISEM) and the Labex Centre Méditerranéen Environnement Biodiversité (CeMEB). We thank M. Roesti for the modified RAD-seq protocol and for the discussions and scripts on the linkage disequilibrium analysis. We thank the University of Fribourg, the Montpellier Bioinformatic Biodiversity platform and the Labex CeMEB for access to high-performance computing clusters. The sequence data for the *D. magna* genome project V2.4 were produced by The Center for Genomics and Bioinformatics at Indiana University and distributed via the SRA database in collaboration with the *Daphnia* Genomics Consortium (project supported in part by NIH award 5R24GM078274-02 "Daphnia Functional Genomics Resources", BioProject reference PRJNA298946). We also thank Peter Fields and Deborah Charlesworth for their constructive comments at various stages of this paper, as well as David Innes and two other anonymous reviewers who contributed to its great improvement during the reviewing process. This work was supported by the Swiss National Science Foundation (Grant no. 31003A\_138203), the Russian Foundation of Basic Research Grant 16-04-01579, the European Union (Marie Curie Career Integration Grant PCIG13-GA-2013-618961, DamaNMP), and the Council on Grants from the President of the Russian Federation in support of leading scientific schools (NSH-9249.2016.5).

## References

- Beukeboom L, Perrin N. 2014. The evolution of sex determination. Oxford: Oxford University Press.
- Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889–890.
- Bull JJ. 1983. Evolution of Sex determining Mechanisms. San Francisco/Menlo Park (CA): Benjamin/Cummings Pub. Co., Advanced Book Program.
- Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. *Am Nat.* 112:975–997.
- Charlesworth B, Nordborg M, Charlesworth D. 1997. The effect of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res Camb.* 70:155–174.
- Cheng J, Czyponka T, Nolte AW. 2013. The genomics of incompatibility factors and sex determination in hybridizing species of *Cottus* (Pisces). *Heredity* 111:520–529.
- Conesa A, Götz S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* 2008:619832.
- Connor HE, Charlesworth D. 1989. Genetics of male sterility in gynodioecious *Cortaderia* (Gramineae). *Heredity* 63:373–382.
- Davidovich O, Kimmel G, Shamir R. 2007. GEVALT: an integrated software tool for genotype analysis. *BMC Bioinformatics* 8:36–43.
- Dukić M, Berner D, Roesti M, Haag CR, Ebert D. 2016. A high-density genetic map reveals variation in recombination rate across the genome of *Daphnia magna*. *BMC Genet.* 17:137.
- Edwards AWF. 1998. Selection and the sex ratio: Fisher's sources. *Am. Nat* 151:564–569.
- Ellegren H. 2011. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet.* 12:157–166.
- Galimov Y, Walser B, Haag CR. 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.
- Hobaek A, Larson P. 1990. Sex determination in *Daphnia magna*. *Ecology* 71:2255–2268.
- Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation?. *Annu Rev Ecol Evol.* 39:21–42.
- Hudson RK. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164.
- Innes DG, Dunbrack RL. 1993. Sex allocation variation in *Daphnia pulex*. *J Evol Biol.* 6:559–575.
- Ironside JE. 2010. No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays* 32:718–726.
- Joron M, Frezal L, Jones RT, Chamberlain N, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477:203–206.
- Karim FD, Guild GM, Thummel CS. 1993. The *Drosophila Broad-Complex* plays a key role in controlling ecdysone-regulated gene expression at the onset of metamorphosis. *Development* 118:977–988.
- Kato Y, Kobayashi K, Oda S, Colbourne JK, Tatarazako N, Watanabe H, Igushi T. 2008. Molecular cloning and sexually dimorphic expression of DM-domain genes in *Daphnia magna*. *Genomics* 91:94–101.
- Kato Y, Kobayashi K, Oda S, Tatarazako N, Watanabe H, Igushi T. 2010. Sequence divergence and expression of a transformer gene in the branchiopod crustacean *Daphnia magna*. *Genomics* 95:160–165.
- Kato Y, Kobayashi K, Watanabe H, Iguchi T. 2011. Environmental sex determination in the branchiopod crustacean *Daphnia magna*: deep conservation of a doublesex gene in the sex-determining pathway. *PLoS Genet.* 7:e1001345.
- Kersten B, Pakull B, Groppe K, Lueneburg J, Fladung M. 2014. The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. *Plant Biol.* 16:411–418.
- Kohn J. 1988. Why be female?. *Nature* 335:431–433.
- Kotov A, Taylor DJ. 2011. Mesozoic fossils (>145MYA) suggest the antiquity of the subgenera of *Daphnia* and their coevolution with chaoborid predators. *BMC Evol Biol.* 11:129–138.
- LeBlanc GA, Medlock EK. 2015. Males on demand: the environmental-neuro-endocrine control of male sex determination in daphnids. *FEBS J.* 282:4080–4093.
- Lohr JN, Haag CR. 2015. Genetic load, inbreeding depression, and hybrid vigor covary with population size: an empirical evaluation of theoretical predictions. *Evolution* 69:3109–3122.
- McCauley DE, Bailey MF. 2009. Recent advances in the study of gynodioecy: the interface of theory and empiricism. *Ann Bot.* 104:611–620.
- Miura I. 2008. An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. *Sex Dev.* 1:23–331.
- Ohno S. 1967. Sex chromosomes and sex-linked genes. Berlin, Heidelberg, New York: Springer-Verlag.
- Olmstead AW, Leblanc GA. 2002. Juvenoid hormone methyl farnesoate is a sex determinant in the crustacean *Daphnia magna*. *J Exp Zool.* 293:736–739.
- Penning TM, Burczynski ME, Jez JM, Hung CF, Lin HK, Ma H, Moore M, Palackal N, Ratnam K. 2000. Human 3 $\alpha$ -hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochem J.* 351:67–77.
- Pfeifer B, Wittelsberger U, Ramos Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses. *Mol Biol Evol.* 31:1929–1936.
- Pokorná M, Kratochvíl L. 2009. Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap? *Zool J Linn Soc.* 156:168–183.
- Pokorná M, Kratochvíl L. 2016. What was the ancestral sex-determining mechanism in amniote vertebrates? *Biol Rev.* 91:1–12.

- R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://R-project.org>.
- Rice WR. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 41:911–914.
- Robinson JT, Thorvaldsdóttir H, Winkler W, Guttman M, Lander ES, Getz C, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotech* 29:24–26.
- Roulin AC, Routtu J, Hall MD, Janicke T, Colson I, Haag CR, Ebert D. 2013. Local adaptation of sex induction in facultative sexual crustacean: insights from QTL mapping in natural populations of *Daphnia magna*. *Mol Ecol* 22:3567–3579.
- Routtu J, Hall MD, Albere B, Beisel C, Bergeron RD, Chaturvedi A, Choi JH, Colbourne J, De Meester L, Stephens MT, et al. 2014. An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. *BMC Genomics* 15:1033–1048.
- Routtu J, Jansen B, Colson I, De Meester L, Ebert D. 2010. The first-generation *Daphnia magna* linkage map. *BMC Genomics* 11:508–514.
- Salz HK. 2011. Sex determination in insects: a binary decision based on alternative splicing. *Curr Opin Genet Dev* 21:395–400.
- Schwartz SS, Hebert PDN. 1987. Breeding system of *Daphniopsis ephemeralis*: adaptations to a transient environment. *Hydrobiologia* 145:195–200.
- Shin JH, Blay S, McNeney B, Graham J. 2006. LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw* 16:Code Snippet 3.
- Shuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotech* 18:233–234.
- Spigler RB, Lewers KS, Main DS, Ashman TL. 2008. Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity* 101:507–517.
- Stöck M, Horn A, Grossen C, Lindtke D, Sermier R, Betto-Colliard C, et al. 2011. Ever-young sex chromosomes in European tree frogs. *PLoS Biol* 9:e1001062.
- Svendsen N, Reisser CMO, Dukić M, Thuillier V, Ségard A, Liautard-Haag C, Fasel D, Hürlimann E, Lenormand T, Galimov Y, et al. 2015. Identification of cryptic asexuality in *Daphnia magna* by RAD-sequencing. *Genetics* 201:1143–1155.
- Tessier AJ, Caceres CE. 2004. Differentiation in sex investment by clones and populations of *Daphnia*. *Ecol Lett* 7:695–703.
- The Tree of Sex Consortium. 2014. Tree of sex: a database of sexual systems. *Sci Data* 1:140015.
- Van Dooren TJM, Leimar O. 2003. The evolution of environmental and genetic sex determination in fluctuating environments. *Evolution* 57:2667–2677.
- Verlhulst EC, Van de Zande L, Beukeboom LW. 2010. Insect sex determination: it all evolves around *transformer*. *Curr Opin Genet Dev* 20:376–383.
- Weller SG, Sakai AK. 1991. The genetic basis of male sterility in *Schiedea* (Caryophyllaceae), an endemic Hawaiian genus. *Heredity* 67:265–273.
- Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. *Nat Commun* 7:12087.
- Yu Q, Hou S, Hobza R, Feltus FA, Wang X, Jin W. 2007. Chromosomal location and gene paucity of the male specific region on *papaya* chromosome. *Mol Genet Genomics* 278:177–185.
- Zaykin DV, Pudovkin AI, Weir BS. 2008. Correlation-based inference for linkage disequilibrium with multiple alleles. *Genetics* 180:533–545.
- Zhang L, Baer KN. 2000. The influence of feeding, photoperiod and selected solvents on the reproductive strategies of the water flea, *Daphnia magna*. *Environ Pollut* 110:425–430.