Journal of
Plankton Research

plankt.oxfordjournals.org

J. Plankton Res. (2015) 37(3): 626–635. First published online April 29, 2015 doi:10.1093/plankt/fbv028

New possibilities arise for studies of hybridization: SNP-based markers for the multi-species *Daphnia longispina* complex derived from transcriptome data

JAKUB RUSEK[1]*, GÖKCE B. AYAN[1,2], PATRICK TURKO[3,4], CHRISTOPH TELLENBACH[3,5], SABINE GIESSLER[1], PIET SPAAK[3,4] AND JUSTYNA WOLINSKA[6]

[1]DEPARTMENT OF BIOLOGY II, LUDWIG MAXIMILIANS UNIVERSITY, GROSSHADERNER STRASSE 2, PLANEGG-MARTINSRIED 82512, GERMANY, [2]DEPARTMENT OF EVOLUTIONARY ECOLOGY, MAX PLANCK INSTITUTE FOR EVOLUTIONARY BIOLOGY, AUGUST-THIENEMANN STRASSE 2, PLÖN 24306, GERMANY, [3]DEPARTMENT OF AQUATIC ECOLOGY, EAWAG, SWISS FEDERAL INSTITUTE OF AQUATIC SCIENCE AND TECHNOLOGY, DÜBENDORF 8600, SWITZERLAND, [4]INSTITUTE OF INTEGRATIVE BIOLOGY, ETH ZURICH, ZURICH 8092, SWITZERLAND, [5]SCHOOL OF BIOSCIENCES, UNIVERSITY OF BIRMINGHAM, EDGBASTON, BIRMINGHAM B15 2TT, UK AND [6]DEPARTMENT OF ECOSYSTEM RESEARCH, LEIBNIZ-INSTITUTE OF FRESHWATER ECOLOGY AND INLAND FISHERIES, MÜGGELSEEDAMM 301, BERLIN 12587, GERMANY

*CORRESPONDING AUTHOR: rusek.jakub@gmail.com

Received September 9, 2014; accepted March 25, 2015

Corresponding editor: Beatrix E. Beisner

In order to trace community dynamics and reticulate evolution in hybrid species complexes, long-term comparative studies of natural populations are necessary. Such studies require the development of tools for fine-scale genetic analyses. In the present study, we developed species-diagnostic SNP-based markers for hybridizing freshwater crustaceans: the multispecies *Daphnia longispina* complex. Specifically, we took advantage of transcriptome data from a key species of this hybrid complex, the annotated genome of a related *Daphnia* species and well-defined reference genotypes from three parental species. Altogether eleven nuclear loci with several species-specific SNP sites were identified in sequence alignments of these reference genotypes from three parental species and their interspecific hybrids. A PCR-RFLP assay was developed for cost-efficient large population screening by SNP-based genotyping. Taxon assignment by RFLP patterns was nearly perfectly concordant with microsatellite genotyping across several screened populations from Europe. Finally, we were able to amplify two short regions of these loci in formaldehyde-preserved samples dating back to the year 1960. The species-specific SNP-based markers developed here provide valuable tools to study hybridization over time, including the long-term impact of various environmental factors on hybridization and biodiversity changes. SNP-based genotyping will finally allow eco-evolutionary dynamics to be revealed at different time scales.

KEYWORDS: low-quality DNA; formaldehyde; SNP genotyping; interspecific hybridization; species-specific SNPs

available online at www.plankt.oxfordjournals.org

INTRODUCTION

Hybridization between species is a common phenomenon, often triggered by changing ecological conditions (Grant and Grant, 2008), and is especially frequent in recently diverged groups (Seehausen, 2004). Sometimes, hybridization allows the introgression of novel genomic variants from one species to another (Mallet, 2007; Abbott *et al.*, 2013). At least 10% of all animal and 25% of all plant species form natural hybrids (Mallet, 2005). Hybrid and parental co-existence is largely determined by interaction of multiple ecological and genetic factors, including types and strengths of reproductive barriers operating in the systems, the vigour and fertility of the hybrids, pathogen pressure or selection by predation (e.g. Wolf *et al.*, 2001; Hall *et al.*, 2006; Duenez-Guzman *et al.*, 2009; Daum *et al.*, 2012).

Several hybrid species complexes have become model systems in evolutionary biology to study the role of hybridization in speciation and adaptation. These include sunflowers (Rieseberg *et al.*, 2007), irises (Arnold and Meyer, 2006), cichlids (Seehausen *et al.*, 1997), Darwin's finches (Grant and Grant, 1996; Lamichhaney *et al.*, 2015) and members of the genus *Daphnia* (Crustacea: Anomopoda). *Daphnia* reproduce clonally but switch to sexual reproduction when conditions become harsh (cyclical parthenogenesis). During these sexual phases, hybrids can be created if parental species co-exist. Sexually produced eggs are enveloped by a rigid chitinous structure forming a resting stage known as an ephippium. Ephippia can endure harsh conditions such as freezing and desiccation over long periods of time, allowing populations to persist in temporary uninhabitable basins by hatching when conditions improve. However, in some temperate and tropical permanent lakes and ponds, conditions remain favourable for asexual reproduction year-round, allowing sexually incompetent hybrids to achieve long persistence of large populations. Consequently, even a few hybridization events might be sufficient to establish a hybrid population that will be maintained by clonal reproduction (Spaak, 1997). Such long-term maintenance of hybrid populations has been frequently described in the *Daphnia longispina* complex in Europe, involving a number of species, such as *D. cucullata*, *D. galeata* and *D. longispina* (e.g. Keller *et al.*, 2008).

Although hybridization is a frequent phenomenon in nature, its role in adaptation and evolution of natural populations remains ambiguous. One major limiting factor for such studies is the lack of tools for fine-scale genetic analyses as well as of long-term studies of natural populations to reliably assess the frequency of hybridization events and study evolutionary dynamics. Hybridizing species of the *D. longispina* complex provide a valuable model system to study long-term evolutionary dynamics as several resources may be used. For example, zooplankton samples have been routinely archived over the last decades in several research institutes. These hold valuable historical information but were often stored in DNA-damaging preservatives such as formaldehyde or denatured ethanol. Also, ephippia deposited in lake sediments may remain hatchable for decades (reviewed in Brendonck and De Meester, 2003; Frisch *et al.*, 2014), and ephippial eggs from dated sediment cores can be directly genotyped (Duffy *et al.*, 2000; Cousyn *et al.*, 2001; Reid *et al.*, 2002).

Multiple marker systems have been developed for species identification in the *D. longispina* complex, and thus to study hybridization over time. Since 1986 until relatively recently, allozymes have been used for this purpose (e.g. Wolf and Mort, 1986; Haag *et al.*, 2005; Seda *et al.*, 2007), but these markers are resource-demanding in terms of the sample quantity and quality required (Taylor *et al.*, 1996; Gießler, 1997). Moreover, a sufficiently high number of molecular markers with species-specific alleles are needed in order to discriminate between different classes of hybrids and backcrosses (Nason and Ellstrand, 1993; Boecklen and Howard, 1997; Sovic *et al.*, 2014), but the limited availability of tissue constrains the number of allozyme loci that can be scored (typically four, e.g. Spaak, 1996; Wolinska *et al.*, 2006; Petrusek *et al.*, 2013). Thus, the discriminatory power of allozymes is often insufficient when applied to single individuals. Moreover, only two of these loci have been found to be species-specific in the *D. longispina* complex (Wolf and Mort, 1986; Gießler, 1997), further restricting their utility for hybrid detection. The first available nuclear markers for the *D. longispina* complex were the two internal transcribed spacers (ITS; Billiones *et al.*, 2004; Taylor *et al.*, 2005; Petrusek *et al.*, 2008), but these unfortunately did not allow full resolution between all coexisting species (Giessler and Englbrecht, 2009). Furthermore, a single diploid marker can only discriminate parental species and F1 hybrids, but is unable to detect more complex hybrid classes. Recently, microsatellite (MS) markers have been established (Brede *et al.*, 2006) which address allelic variation at up to 15 loci per individual and, since then, have been widely used for taxon identification in the *D. longispina* complex (e.g. Brede *et al.*, 2009; Thielsch *et al.*, 2009; Yin *et al.*, 2010). Although the possibility of obtaining large numbers of MS loci makes them a powerful marker system, they pose several problems for species identification. Firstly, almost no species-specific MS alleles have been described in the *D. longispina* complex (Dlouha *et al.*, 2010), making taxon identification reliant on the joint information from allele

frequencies at all loci (Selkoe and Toonen, 2006; Dlouha *et al.*, 2010). Moreover, MS markers are quite long (i.e. 80–250 bp), limiting their applicability to fragmented DNA samples, such as degenerated *Daphnia* diapausing eggs (Brede *et al.*, 2009) or samples preserved in DNA-damaging chemicals (e.g. formaldehyde).

The above problems can be avoided by using single-nucleotide polymorphism (SNP)-based markers. Because of the usually dense SNP abundance throughout the genome (Clarke *et al.*, 2014; Lapegue *et al.*, 2014), a high number of diagnostic markers can be developed. The most important advantage compared with MSs is that, at the nucleotide level, each base is unambiguously identified allowing identical alleles to be addressed in different studies. Moreover, as the detection of only one base at a given site is needed, short DNA fragments are sufficient to target the genomic region of interest. Therefore, SNP-based genotyping is the method of choice for species and hybrid assignment via diagnostic alleles (Pujolar *et al.*, 2014). Species assignment by SNP-based markers is deterministic and thus superior compared with probabilistic MS-genotyping (Anderson and Garza, 2006; Hauser *et al.*, 2011). Additionally, a low amount of genomic DNA is sufficient for SNP-based genotyping and, because only small DNA fragments are needed, even low-quality DNA samples can be accessed.

The goal of the present study was thus to identify unlinked species-specific SNP-based markers enabling a quick and reliable identification of three parental species and interspecific hybrids in the *D. longispina* complex for large-scale community analyses, and long-term evolutionary studies. When species inference is based on a low number of loci, then three species-diagnostic alleles per gene locus are necessary for successful discrimination of species and hybrids. Because SNPs are biallelic markers, a single SNP site does not allow for direct discrimination among more than two species. Therefore, we selected short fragments with two complementary informative SNP sites. The resulting species-diagnostic contigs at chosen marker loci were then labelled "SNP-based markers." We mapped available transcriptome data (see Acknowledgements) from one of the parental species to the already annotated *D. pulex* genome (Colbourne *et al.*, 2011) to identify candidate genome regions for SNP markers on different chromosomes. Firstly, to look for species-specific SNP-based markers, we sequenced several putatively unlinked nuclear gene loci from three target species (*D. cucullata*, *D. galeata*, *D. longispina*) and their hybrids using well-defined reference genotypes by different marker sets. Secondly, from the sequence alignment, we chose appropriate short contigs with complementary informative SNP sites to identify species-diagnostic alleles in this system. Thirdly, we compared the taxon assignment of these newly developed SNP-based markers with the taxon assignment of previously established MS markers and verified the marker association in population samples. Finally, we tested the SNP-based markers in formaldehyde-preserved samples dating back to 1960, to explore the accessibility of old historical zooplankton samples for DNA analyses.

METHOD

Development of species-specific SNP-based markers

Identification of candidate genomic regions

Candidate regions were identified by mapping the *D. galeata* transcriptome (available on request, see Acknowledgements) to the *D. pulex* genome (Colbourne *et al.*, 2011). Mapping enabled the identification of putative intron locations that are not present in the transcripts, so that the designed primers could be located on exons. This also helped us to deliberately choose markers located on putatively different chromosomes, in order to ensure linkage independence. Mapping was carried out in Geneious 6.1.7 (Biomatters, Auckland, New Zealand). 5191 *D. pulex* scaffolds were sorted by chromosomal origin (Colbourne *et al.*, 2005), and scaffolds of unknown origin were discarded. The remaining 84 scaffolds were grouped, and the 29 500 *D. galeata* transcripts were then mapped to the reference scaffolds [parameters used: maximum gap size of 100 bp, maximum gaps per alignment of 20%, word length of 12 bp, Index word length of 11, maximum mismatches per alignment of 40%, no iteration (i.e. transcripts were individually mapped to the reference rather than to a new alignment)]. Scaffold alignments of $<$800 bp were discarded, leaving 214 *D. galeata* contigs mapped to 47 *D. pulex* scaffolds. All mapped contigs were inspected visually to determine the type of genomic region that they mapped to, according to the annotated *D. pulex* genome: coding sequence, intron, exon, 5′ or 3′ untranslated region (UTR), or intergenic region. From each chromosome, three *D. galeata* contigs were chosen, which were expected to differ in their degree of conservation and therefore in their potential to bear species-specific SNPs. Thus, per chromosome, one coding sequence, one intergenic sequence, and (where possible) one UTR or intron was chosen. Each *D. galeata* contig chosen was mapped on a different *D. pulex* scaffold.

Another alignment was performed to discover the location of excised *D. galeata* introns. The approach was similar to the one above, with the exception that only *D. galeata* contigs between 800 and 1200 bp in length were used, and that mapping was performed with the

maximum gap size widened to 500 bp. One match in the *D. galeata* transcript was chosen per *D. pulex* chromosome (criteria used: exact matching of a gap in a *D. galeata* transcript to a *D. pulex* "intron" annotation (which we interpreted as the signature of an excised intron); total intron length of at least 100 bp, flanked by exons, which presumably would increase the probability of successful priming; and presence on a scaffold which has not yet been used). After locating the introns, primers were designed on exons. Primer binding sites that would be split by an intron were excluded, since the resultant PCR amplicon would be too long.

Reference genotypes

To identify putative species-specific SNP-based markers, 22 well-defined reference genotypes from the *Daphnia longispina* complex, representing three species (*D. cucullata*, *D. galeata*, *D. longispina*) and two interspecific hybrids were sequenced at each target region. These genotypes were collected across 15 locations in Europe, between 1985 and 2007, and maintained in the laboratory as clonal lineages (for detailed information about origin and taxon assignment, see Supplementary data, Table SI). For an initial search for putative species-specific SNP-based markers, a subset of eight reference genotypes (i.e. two *D. cucullata*, two *D. galeata* and four *D. longispina* genotypes; see column "Initial screening" in Supplementary data, Table SI) was sequenced. Putative SNPs obtained by aligning the sequences of the candidate loci (Table I) were then verified in 14 additional reference genotypes (see column "Sequencing" in Supplementary data, Table SI). Species specificity of a subset of such identified SNP-based markers was further verified with altogether 165 genotypes (see below).

SNP discovery

Total genomic DNA was extracted from single *Daphnia* individuals according to the protocol described elsewhere (Giessler and Wolinska, 2013). DNA was then dissolved in 50 μL of sterile water. Primers were designed using PerlPrimer v1.1.21 (Marshall, 2004) and their annealing temperature was determined by applying a thermal gradient (Table I). Each PCR was carried out using 1.5 μL of genomic DNA, 0.25 mM of deoxynucleoside triphosphates, 0.5 μM each of forward and reverse primer, 1× buffer, and 1.5 U of a DreamTaq polymerase (Thermo Fisher scientific, Waltham, MA, USA), with a final reaction volume of 20 μL. Capillary sequencing of purified products was carried out on an ABI 3730 DNA Analyzer using the BigDye 1.1 Terminator Sequencing Kit (both Applied Biosystems, Foster City, CA, USA). The resulting electropherograms were carefully checked by eye and corrected if necessary in MEGA5 (Tamura *et al.*, 2011). In cases of failed direct sequencing (e.g. due to indel polymorphisms), PCR products were cloned using StrataClone PCR Cloning Kit (Agilent Technologies, La Jolla, CA, USA). Two to four positive clones were sequenced as described above. For each of the selected nuclear loci, sequence alignment was carried out in MEGA5 using the MUSCLE algorithm (Edgar, 2004) and corrected manually if necessary.

Consistency of taxon assignment by SNP-based markers and microsatellites

PCR-RFLP assay

We developed a PCR-RFLP assay (for a subset of nine loci chosen in the way that they could be recognized by restriction enzymes) allowing fast and reliable SNP screening of large numbers of samples. For each diagnostic locus (SNP-based marker), two complementary species-specific SNPs were detected by two restriction enzymes chosen by NEBcutter (Vincze *et al.*, 2003). Enzymes were selected to provide a different restriction fragment pattern for each one of the three species studied here (Table II). To test the RFLP assay, 29 reference

Table I: Summary of the loci with species-specific SNP-based markers used in the present study

Name[a]	Chromosome	Locus type	Primer F	Primer R	Ta	Total alignment length (bp)
SPEC	1	CDS	TGACCCATAGCAAGCACAG	CGCCGTATGGAATTGGATCAG	63	613–744
1237BB	3	Intergenic/5′UTR/CDS	ACACGCTAGACCAATGGG	GGCAAAGAACGACTCCTC	60	497–1074
IG3	3	Intergenic	CGCCCAAGTGTATTATAGACGA	TCTACTGCATCATATTCCCTCC	60	218–1062
TW	4	CDS/intron	ACGGCATTCACAACCCAG	GCGAAATATCCATGTTCACAGAG	64	373–808
TLR	5	CDS	ATCTCGTAACCGCTCCCA	CGATTTCAACCCATCACCAG	63	544–712
DSTPRK	5	Intergenic/5′UTR	GGTCAGCAGGAAGTAGTGG	AAAGACAGATTCGGAGGAGAG	61.5	346–757
N-SMA	6	CDS/intergenic	TTTCCGCACATTACTACATCAC	CTTCTTTACCACTCACTTCGCT	60	233–956
IG6	6	Intergenic	GCATTGAACCAAATCGCCC	GTTGAGAAAGTTGTACCGCAG	60	409–538
IPO	12	Intergenic	ACCTCCCAGCAATCAGCA	ATCCGGCTCTCAAAGTTCC	60	387–738
GIDAP	Scaffold 116	NA	GTGGAAAGTAAAGTGGCGA	AAGGTGATAGAATGTGCTGG	61.5	869–1083
GPDCP	Scaffold 88	NA	TTTCCGTGTTTAGCACCCA	GTCGTTGTAGCTCATGATCCA	64	751–803

CDS, coding sequence; Ta, primer annealing temperature.
[a]For the locus name origin see Supplementary data, Table SII.

Table II: Summary of the SNP-based markers used in RFLP assay for nine loci

Locus	SNP[a]	Taxon specificity[b]	Position of SNP in PCR product	Restriction enzyme
SPEC	C/T	cuc	144	TaqI
	T/C	gal	517	NcoI
1237BB	T/C	lon	523	BpuEI
	C/G	cuc(gal)[c]	524	TaqI
IG3	G/A	cuc	132	AluI
	C/T	gal	303	SspI
TW	G/A	gal	143	Hpy188I
	C/T	cuc	350	HpaI
TLR	A/C	gal	205	BsmAI
	A/G	lon	589	AseI
DSTPRK	G/C	cuc	312	TaqI
	C/T	gal	314	Hpy188III
N-SMA	A/G	gal	184	Hpy166II
	A/G;T/G	all	184;227	ApeKI
IPO	G/T;C/T	cuc	237;292	HaeIII
	G/A	lon	279	BccI
GIDAP	A/C-T/C	all	219–268	TaqI
	A/T	cuc	423	FspI

[a]First letter indicates the base specific for one of the three species; second letter indicates the other variant of the polymorphism.
[b]Column indicates, which of three species is identified by the restriction reaction (cuc: *D. cucullata*; gal: *D. galeata*; lon: *D. longispina*; all: enzyme distinguishes all three species, due to multiple restriction sites).
[c]SNP site is specific for *D. cucullata*; however, the restriction enzyme provides a specific pattern for *D. galeata*, due to the presence of another SNP in close proximity.

genotypes were used (see column "PCR-RFLP" in Supplementary data, Table SI), partially overlapping with the previous selection of genotypes. The resulting fragments were resolved by agarose electrophoresis. In case of low resolution of the restriction fragment pattern due to similar size of fragments, primers were re-designed to shift the lengths of resulting bands. As a last step, a large-scale PCR-RFLP assay was run for four loci (SPEC, GIDAP, IG3 and TW), on 165 *Daphnia* individuals isolated from field collections across 47 sites in Europe [i.e. 29 reference genotypes, 39 additional laboratory clones and 97 additional *Daphnia* individuals (see column "Large-scale" in Supplementary data, Table SI)]. Each of these 165 individuals had a unique MS multi-locus genotype (MLG). All clones used in SNP-based genotyping by PCR-RFLP were anonymous to the experimenter.

Consistency of SNP-based and microsatellite markers for taxon assignment

To allow comparison between the marker types, all 165 DNA samples used in SNP analyses were additionally genotyped by MSs (in the case of the reference genotypes, the MS genotypes were already known from previous studies; see Yin *et al.*, 2010 and Supplementary data, Table SI). Fifteen MS markers (Brede *et al.*, 2006) were used in two sets of multiplex polymerase chain reaction (MP1: Dgm105, Dgm112, SwiD5, SwiD7, SwiD8; and MP2: Dgm109, Dp196, Dp281, Dp512, SwiD1, SwiD2, SwiD10, SwiD12, SwiD14, SwiD15), using the Qiagen Multiplex PCR Kit (Venlo, Netherlands) and following protocols described elsewhere (Yin *et al.*, 2010). A standard genotype (G100; Supplementary data, Table SI) was used to allow adjustment of alleles from different runs.

The assignment of all 165 samples assessed by MS markers to taxonomic groups was done by a two-step analysis. Firstly, factorial correspondence analysis (FCA) based on MLGs was used to extract all axes in GENETIX 4.05 (Belkhir *et al.*, 2004). The clustering of samples in the bi-plot of the first two FCA axis loadings served to predefine groups for subsequent discriminant analysis in SPSS 20.0 (stepwise method, Mahalanobis distance). Secondly, discriminant analysis on FCA scores (DFCA) from all extracted FCA axes allowed the derivation of group centroids and the probability of each MLG's membership in a certain group (taxon). In both steps, the bi-plot of the scores from the first two axes/functions allowed the visualization of intermediate hybrid groups. The same procedure was run with MLGs as identified by SNP-based markers. To test for the consistency between taxon assignment by different marker systems, DFCAs were run on FCA-scores derived from one marker system and groups predefined by the other markers. This was done for all individual SNP-based markers as well as for MLGs composed of altogether four SNP-based markers (comprising eight SNP sites). Respective DFCAs allowed testing of the consistency in taxon assignment using the two marker systems.

Application of SNP-based markers to historical samples

Daphnia individuals belonging to the *D. longispina* complex were randomly selected from formaldehyde-preserved zooplankton samples collected from Greifensee, Switzerland (16 individuals from 1960 and 5 from 1990 samples). Total genomic DNA was extracted following a protocol by Turko (see Supplementary data).

Primers flanking two complementary species-specific SNPs at two loci (SPEC: F:GCATTAGCAGCTAACT CGG, R:AGAGCTATCGTGACCCTG and TW: F:CG CACTTATTATTTGGACTGAC, R:TTGAAGGTGCT AAAGAGGGT), resulting in ca. 80 bp amplicons, were designed using PerlPrimer v1.1.21, and their annealing temperature was determined using a thermal gradient. DNA from one randomly selected *Daphnia* individual per year (1990 and 1960) was then amplified at both SNP loci. The reaction mix consisted of 1.5 μL gDNA, 0.4 μM of each forward and reverse primer and 12.5 μL of the PCR mastermix (Promega, Fitchburg, WI, USA)

for a total volume of 25 μL. Amplicons of both SNP-based markers from the older sample (individual from 1960) were cloned using the CloneJET PCR cloning kit (Thermo Fisher Scientific, Waltham, MA). In order to verify the specificity of the amplified region and the presence of the diagnostic SNPs, one (SPEC) and three (TW) positive colonies were sequenced and aligned with the previously obtained results. Finally, PCR for the SPEC locus was carried out for all 21 individuals and the success of the amplification was checked by agarose gel electrophoresis.

RESULTS

Development of species-specific SNP-based markers

In total, 48 nuclear loci (mapped on all 12 chromosomes of *D. pulex*) were tested for the presence of species-specific SNPs in a subset of eight reference genotypes (see column "Initial screening" in Supplementary data, Table SI) representing the three species: *D. cucullata*, *D. galeata*, and *D. longispina*. Of those nuclear loci, 37 were excluded from further tests due to the presence of non-specific PCR products, failure in direct sequencing, or the absence of candidate species-specific SNPs (data not shown). To attain specificity for all three species each of the chosen 11 markers consists of two bi-allelic SNP sites (Table I). In the example of locus TW: SNP1 = G/A, SNP2 = C/T. The specific base of SNP1 for *D. galeata* is G (genotype is GG). The unspecific base for two other species is A (genotype AA). In the SNP2, C is specific for *D. cucullata*, leading to a CC genotype (for two other species: genotype TT). The complementary information from the two SNPs at an individual locus can therefore be combined to yield deterministic identities for all three species (*D. cucullata*: AA-CC, *D. galeata*:GG-TT, *D. longispina*:AA-TT). Similarly, F1 hybrids will be heterozygous at the respective SNP site (e.g. *D. galeata* × *longispina*: GA-TT). For all 11 loci, the candidate SNPs were verified in altogether 22 reference genotypes (see column "Sequencing" in Supplementary data, Table SI). All sequences have been deposited in GenBank (see Data Archiving). Sequence information from hybrid genotypes revealed expected parental alleles at respective SNP sites (Supplementary data, Table SI). The 11 loci with diagnostic SNPs were distributed over at least six chromosomes, according to mapping to the *D. pulex* genome (for more details see Table I). Four loci (SPEC, GIDAP, TW and GPDCP) were developed by re-sequencing of random transcripts prior to mapping. Therefore, the bioinformatic analysis and their mapping to chromosomes were done afterwards. Due to an incomplete annotation of the *D. pulex* genome, or possibly different *D. galeata* genome architecture, it was not possible to assign two of the loci (GIDAP and GPDCP) to chromosomes.

Consistency of taxon assignment by SNP-based markers and microsatellites

The taxon assignment by MS markers was derived from DFCA on 15-locus genotypes (MLGs with missing loci were allowed) where the first two discriminant functions explained 99.4% of the variance. Altogether six groups were identified and 99.2% of the 165 individuals were reclassified to the same taxonomic group as predefined from clustering in the bi-plot of the first two FCA axes. Group centroids of the three parental species and three hybrid groups are shown in Fig. 1A. The taxon assignment by DFCA using 4-locus-SNP genotypes resulted only in five distinct groups (Fig. 1B), because there was no *D. cucullata* × *longispina* hybrid among the reference clones, and the only *D. cucullata* × *longispina* hybrid

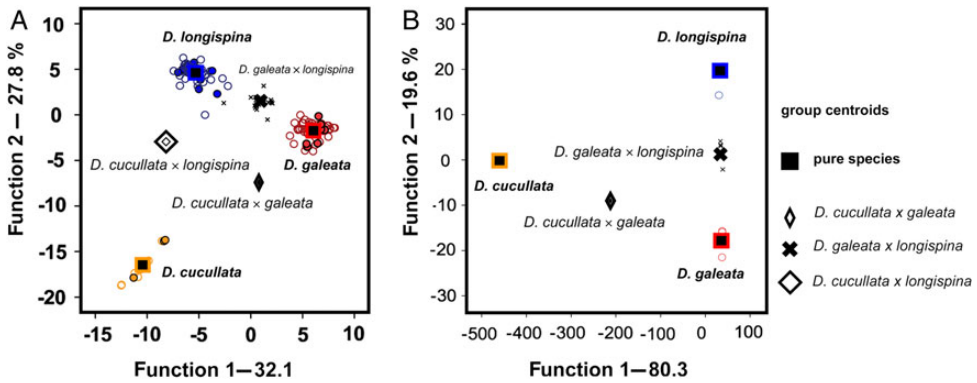

Fig. 1. Taxon assignment of 165 individuals from the *D. longispina* complex by discriminant analysis on FCA scores (10 axis). Assignment based on (**A**) 15 microsatellite loci. Filled circles represent species-specific genotypes among 29 reference clones from reference genotypes; crosses and diamond represent two reference hybrid taxa; (**B**) on 4-locus-SNP genotypes. Group centroids specify within-group means for all variables in the multivariate space. Loci with missing values were allowed in the analyses of both marker types.

individual as identified from field samples (by 15 MSs) had an incomplete 4-locus-SNP genotype. The consistency in taxon assignments by MSs and SNP-based markers from nine associated loci is shown in Table III for the 29 reference clones used for PCR-RFLP development. The consistency was 100% for eight of the nine SNP-based marker loci, and 96.6% in the remaining 1237BB locus, as emerged from DFCA. Four SNP-based marker loci (i.e. SPEC, IG3, GIDAP and TW) were further analysed in altogether 165 samples. Also, in the larger data set, the association between marker types was high: ranging between 87 and 95% correspondence in taxon assignment, when DFCA was run on single SNP-based marker loci (Table III). When DFCA was applied on 4-locus-SNP genotypes, with groups predefined by MSs, the consistency was slightly improved to 96%.

Application of SNPs to historical samples

Two randomly chosen DNA samples from formaldehyde-preserved individuals (in 1960 and 1990) were amplified by using primers flanking the SNP sites at two SNP loci: SPEC and TW. Amplicons resulted in 80 bp (SPEC) and 78 bp (TW) long fragments. Sequences from cloned fragments of amplicons of the 1960 individual confirmed the presence of the expected informative SNPs (see alignments in PANGAEA; Data Archiving). To verify the accessibility of formaldehyde-preserved individuals for SNP-based genotyping, a PCR reaction was then run for 21 individuals (16 from 1960 and 5 from 1990), for the locus SPEC, including positive and negative controls. The amplification was successful for 20 of 21 samples tested.

Table III: Consistency in group assignments of individuals using SNP-based and microsatellite markers

SNP genotype	Number of samples	Resolved groups	% Consistently classified by DFCA
	Reference clones		
SPEC	29	5	100
IG3	29	5	100
GIDAP	29	5	100
TW	29	5	100
TLR	29	5	100
Nsma	29	5	100
DSTPRK	29	5	100
Locus_1237bb	29	5	96.6
ipo	29	5	100
	Data set for validation		
SPEC	160	5	93.9
IG3	164	5	95.7
GIDAP	161	6	87.0
TW	163	5	92.6
4-locus-SNP genotypes	154	5	96.1
3 or 4-locus-SNP-genotypes (with missing values)	165	5	95.8

Discriminant analyses on FCA scores derived from individual SNP-based markers or 4-SNP-locus genotypes served to test for the association with the taxon assignment by the second marker type in individuals from 29 reference clones (from three species and two hybrid groups) and a larger data set including field samples where at maximum six groups could be resolved by microsatellites.

DISCUSSION

In order to study the significance of hybridization events over time, reliable methods for detecting species, hybrids and backcrosses are needed. The set of 11 SNP-based markers developed in this study enabled reliable discrimination among three species of the *D. longispina* complex and their hybrids. This taxon assignment was concordant with other molecular markers; there was almost a perfect correspondence between MS- and SNP-based marker assignments. The gene loci with species-specific SNP-based markers developed here are distributed over multiple chromosomes, as shown by the mapping against the annotated genome of *D. pulex* which suggests their linkage independence. The development strategy for species-specific SNP-based markers described here could be applied to other non-model species with prior knowledge of transcriptome data only. Specifically, we mapped *D. galeata* transcripts to the annotated *D. pulex* genome and identified types of genomic regions and their putative architecture. Thus, primers could be placed on exons, which increased the probability of successful amplification of the region (i.e. primer binding site interrupted by an intron or overly long predicted PCR amplicons were avoided). In general, transcriptomes have been widely used for SNP discovery in non-model organisms (reviewed in Ekblom and Galindo, 2011). Moreover, species-specific SNP-based markers allow, unlike MSs, direct identification of hybrid genotypes by heterozygous diagnostic allele combinations. Because of the high abundance of SNPs in a genome and their unambiguous nature, the statistical power to resolve hybrid and parental taxa is much more precise compared with allozymes or MSs. Indeed, the alignments of sequences from 11 loci revealed numerous intraspecific SNPs, opening the gateway for the transition from MS to SNP markers in future population genetic studies of the *D. longispina* species complex, including low-quality samples.

In order to apply the SNP-based markers developed for community screenings, PCR-RFLP was chosen as a method of SNP detection. This simple technique allows rapid and cost-efficient detection of SNPs, using common laboratory equipment (e.g. Laguerre *et al.*,

1994; Ota *et al.*, 2007; Rusek *et al.*, 2013). For the assay, diagnostic SNP sites were chosen in the way that they could be recognized by restriction enzymes (Sachidanandam *et al.*, 2001). The comparison between SNP-based and MS-based assignments (done separately for each locus, but also by using an MLG approach) shows high correspondence in assigning individuals even originating from distant geographical regions. This suggests fixed allelic differences for the species of the *D. longispina* complex.

We have further demonstrated that old and poorly preserved DNA samples spanning several decades can be successfully genotyped using short amplicons. This might allow for studies of hybridization over time. While, for most species, this endeavour might be difficult or impossible to achieve, *Daphnia* appears to be an ideal system, due to the natural archives of resting stages in lake sediments. Moreover, old plankton samples are stored in collections of limnological institutes and museums worldwide. Using these archives, the SNP-based markers developed here will allow genotypic changes to be traced over time. The samples explored here have been preserved in the DNA-damaging chemical formaldehyde. Genotyping of such low-quality DNA samples was possible for the first time due to the use of short PCR amplicons (so far developed for two markers). The field population samples, preserved in formaldehyde, provide better knowledge about the *Daphnia* community composition, than dormant egg banks from sediment, which represent only the sexually active part of the population (e.g. Keller and Spaak, 2004). Therefore, by concomitantly studying (formaldehyde-preserved) pelagic *Daphnia* populations and ephippial eggs from sediments, precise research on past hybridization dynamics can be conducted. Similarly, by the comparison of the pelagic (asexual) and the dormant (sexual) part of *Daphnia* communities over time, it is possible to measure directly the involvement of taxa in both types of reproduction (Keller *et al.*, 2007).

In conclusion, we have shown that a relatively small set of biallelic SNPs provides sufficient information to obtain concordant results in taxon assignment compared with MS and other markers. MLGs can be used to identify the taxon composition of natural communities of the *Daphnia longispina* complex, by applying the SNP-based genotyping developed here. Our method is especially useful for samples that contain low quality and quantity DNA such as samples preserved in formaldehyde or DNA samples from degenerated resting eggs. This will now make it possible to study the long-term impact of various environmental factors (e.g. chemicals, parasite pressure, temperature) on biodiversity changes at community and population levels.

DATA ARCHIEVING

All sequences have been deposited in GenBank (KP004346-KP004398, KR094469-KR094725). Alignments of cloned fragments originating from formaldehyde-preserved samples are stored in PANGAEA; doi:10.1594/PANGAEA. 836026.

ACKNOWLEDGEMENTS

We thank Sabine Radetzki who helped with molecular work and Johanna Griebel for providing field samples and microsatellite data. Sequencing of the *Daphnia galeata* transcriptome was made possible in 2011 by members of the *D. galeata* transcriptome project: M. Cordellier and K. Schwenk (Frankfurt, Germany), P. Spaak and N. Brede (Dübendorf, Switzerland), E. von Elert (Cologne, Germany), W. C. Kerfoot (Houghton, MI, USA), D. Ebert (Basel, Switzerland), L. De Meester (Leuven, Belgium), L. J. Weider (Oklahoma, OK, USA), C. Laforsch (Munich, Germany), M. Pfrender (Notre Dame, IN, USA), K. Mockaitis, D. Gilbert and J. Colbourne (Indiana, IN, USA). These people funded the sequencing, assembly and annotation by the Center for Genomics and Bioinformatics (CGB) at Indiana University. We thank J. Colbourne, K. Mokaitis, J. -H. Choi, D. Rusch, Z. Smith, J. Ford and A. Buechlein at the CGB. The transcriptome data are available on request from P. Spaak (spaak@eawag.ch). This work benefits from and contributes to the *Daphnia* Genomics Consortium. Finally, we thank Joachim Mergeay for his valuable comments which helped to improve our manuscript.

FUNDING

This work was supported by a joint "lead agency" grant from the German Science Foundation (WO 1587/3-1 to J.W.) and the Swiss National Science Foundation (SNF 310030L_135750 to P.S.).

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J. W., Brelsford, A. *et al.* (2013) Hybridization and speciation. *J Evol Biol*, **26**, 229–246.

Anderson, E. C. and Garza, J. C. (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, **172**, 2567–2582.

Arnold, M. L. and Meyer, A. (2006) Natural hybridization in primates: One evolutionary mechanism. *Zoology*, **109**, 261–276.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. and Bonhomme, F. (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique

des populations. *Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France)*.

Billiones, R., Brehm, M., Klee, J. and Schwenk, K. (2004) Genetic identification of *Hyalodaphnia* species and interspecific hybrids. *Hydrobiologia*, **526**, 43–53.

Boecklen, W. J. and Howard, D. J. (1997) Genetic analysis of hybrid zones: Numbers of markers and power of resolution. *Ecology*, **78**, 2611–2616.

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B. and Schwenk, K. (2009) The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proc Natl Acad Sci USA*, **106**, 4758–4763.

Brede, N., Thielsch, A., Sandrock, C., Spaak, P., Keller, B., Streit, B. and Schwenk, K. (2006) Microsatellite markers for European *Daphnia*. *Mol Ecol Notes*, **6**, 536–539.

Brendonck, L. and De Meester, L. (2003) Egg banks in freshwater zooplankton: evolutionary and ecological archives in the sediment. *Hydrobiologia*, **491**, 65–84.

Clarke, S. M., Henry, H. M., Dodds, K. G., Jowett, T. W. D., Manley, T. R., Anderson, R. M. and Mcewan, J. C. (2014) A High Throughput Single Nucleotide Polymorphism Multiplex Assay for Parentage Assignment in New Zealand Sheep. *PLoS ONE*, **9**, DOI:10.1371/journal.pone.0093392.

Colbourne, J., Singan, V. and Gilbert, D. (2005) wFleaBase: the *Daphnia* genome database. *BMC Bioinformatics*, **6**, 45.

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A. *et al.* (2011) The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**, 555–561.

Cousyn, C., De Meester, L., Colbourne, J. K., Brendonck, L., Verschuren, D. and Volckaert, F. (2001) Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences*, **98**, 6256–6260.

Daum, J. M., Davis, L. R., Bigler, L. and Woodhams, D. C. (2012) Hybrid advantage in skin peptide immune defenses of water frogs (*Pelophylax esculentus*) at risk from emerging pathogens. *Infect Genet Evol*, **12**, 1854–1864.

Dlouha, S., Thielsch, A., Kraus, R. H. S., Seda, J., Schwenk, K. and Petrusek, A. (2010) Identifying hybridizing taxa within the *Daphnia longispina* species complex: a comparison of genetic methods and phenotypic approaches. *Hydrobiologia*, **643**, 107–122.

Duenez-Guzman, E. A., Mavarez, J., Vose, M. D. and Gavrilets, S. (2009) Case studies and mathematical models of ecological speciation. 4. Hybrid speciation in butterflies in a jungle. *Evolution*, **63**, 2611–2626.

Duffy, M. A., Perry, L. J., Kearns, C. M., Weider, L. J. and Hairston, N. G. (2000) Paleogenetic evidence for a past invasion of Onondaga Lake, New York, by exotic *Daphnia curvirostris* using mtDNA from dormant eggs. *Limnol Oceanogr*, **45**, 1409–1414.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32**, 1792–7.

Ekblom, R. and Galindo, J. (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Frisch, D., Morton, P. K., Chowdhury, P. R., Culver, B. W., Colbourne, J. K., Weider, L. J. and Jeyasingh, P. D. (2014) A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecol Lett*, **17**, 360–368.

Gießler, S. (1997) Analysis of reticulate relationships within the *Daphnia longispina* species complex. Allozyme phenotype and morphology. *J Evol Biol*, **10**, 87–105.

Giessler, S. and Englbrecht, C. C. (2009) Dynamic reticulate evolution in a *Daphnia* multispecies complex. *J Exp Zool Part a*, **311A**, 530–U3.

Giessler, S. and Wolinska, J. (2013) Capturing the population structure of microparasites: using ITS-sequence data and a pooled DNA approach. *Molecular ecology resources*, **13**, 918–28.

Grant, B. R. and Grant, P. R. (1996) High survival of Darwin's finch hybrids: Effects of beak morphology and diets. *Ecology*, **77**, 500–509.

Haag, C. R., Riek, M., Hottinger, J. W., Pajunen, V. I. and Ebert, D. (2005) Genetic diversity and genetic differentiation in *Daphnia* metapopulations with subpopulations of known age. *Genetics*, **170**, 1809–1820.

Hall, R. J., Hastings, A. and Ayres, D. R. (2006) Explaining the explosion: modelling hybrid invasions. *Proc R Soc B*, **273**, 1385–1389.

Hauser, L., Baird, M., Hilborn, R., Seeb, L. W. and Seeb, J. E. (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Mol Ecol Res*, **11**, 150–161.

Keller, B. and Spaak, P. (2004) Nonrandom sexual reproduction and diapausing egg production in a *Daphnia* hybrid species complex. *Limnol Oceanogr*, **49**, 1393–1400.

Keller, B., Wolinska, J., Manca, M. and Spaak, P. (2008) Spatial, environmental and anthropogenic effects on the taxon composition of hybridizing *Daphnia*. *Philos Trans R Soc Lond B*, **363**, 2943–2952.

Keller, B., Wolinska, J., Tellenbach, C. and Spaak, P. (2007) Reproductive isolation keeps hybridizing *Daphnia* species distinct. *Limnol Oceanogr*, **52**, 984–991.

Laguerre, G., Allard, M. R., Revoy, F. and Amarger, N. (1994) Rapid identification of *Rhizobia* by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl Environ Microbiol*, **60**, 56–63.

Lamichhaney, S., Berglund, J., Almen, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerova, M., Rubin, C.-J. *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, .

Lapegue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadieu, C., Gayral, P., Ballenghien, M., Genestout, L. *et al.* (2014) Development of SNP-genotyping arrays in two shellfish species. *Mol Ecol Res*, **14**, 820–830.

Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends Ecol Evol*, **20**, 229–237.

Mallet, J. (2007) Hybrid speciation. *Nature*, **446**, 279–283.

Marshall, O. J. (2004) PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*, **20**, 2471–2472.

Nason, J. D. and Ellstrand, N. C. (1993) Estimating the frequencies of genetically distinct classes of individuals in hybridized populations. *J Hered*, **84**, 1–12.

Ota, M., Fukushima, H., Kulski, J. K. and Inoko, H. (2007) Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. *Nature Protocols*, **2**, 2857–2864.

Petrusek, A., Seda, J., Machacek, J., Riha, M., Prchalova, M. and Schwenk, K. (2013) Intraspecific spatial genetic differentiation of a

Daphnia species within a long narrow reservoir. *Hydrobiologia*, **715**, 169–180.

Petrusek, A., Seda, J., Machacek, J., Ruthova, S. and Smilauer, P. (2008) *Daphnia* hybridization along ecological gradients in pelagic environments: the potential for the presence of hybrid zones in plankton. *Philos Trans R Soc Lond B*, **363**, 2931–2941.

Pujolar, J. M., Jacobsen, M. W., Als, T. D., Frydenberg, J., Magnussen, E., Jonsson, B., Jiang, X., Cheng, L. *et al.* (2014) Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. *Heredity*, **112**, 627–637.

Reid, V., Carvalho, G., George, D. G. and Griffiths, H. (2002) A technique for the molecular genetic analysis of Daphnia resting eggs from sub-recent lake sediments. *J Paleolimnol*, **27**, 481–486.

Rieseberg, L. H., Kim, S.-C., Randell, R. A., Whitney, K. D., Gross, B. L., Lexer, C. and Clay, K. (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, **129**, 149–165.

Rusek, J., Klumplerova, M., Molinkova, D., Sedlinska, M., Dusek, L., Muzik, J., Putnova, L., Vrtkova, I. *et al.* (2013) Genetics of anti-EHV antibody responses in a horse population. *Res Vet Sci*, **95**, 137–142.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C. *et al.* (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409**, 928–933.

Seda, J., Kolarova, K., Petrusek, A. and Machacek, J. (2007) *Daphnia galeata* in the deep hypolimnion: spatial differentiation of a "typical epilimnetic" species. *Hydrobiologia*, **594**, 47–57.

Seehausen, O. (2004) Hybridization and adaptive radiation. *Trends Ecol Evol*, **19**, 198–207.

Seehausen, O., Vanalphen, J. J. M. and Witte, F. (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, **277**, 1808–1811.

Selkoe, K. A. and Toonen, R. J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett*, **9**, 615–629.

Sovic, M. G., Kubatko, L. S. and Fuerst, P. A. (2014) The effects of locus number, genetic divergence, and genotyping error on the utility of dominant markers for hybrid identification. *Ecology and evolution*, **4**, 462–473.

Spaak, P. (1996) Temporal changes in the genetic structure of the *Daphnia* species complex in Tjeukemeer, with evidence for backcrossing. *Heredity*, **76**, 539–548.

Spaak, P. (1997) Hybridization in the *Daphnia galeata* complex: are hybrids locally produced? *Hydrobiologia*, **360**, 127–133.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, **28**, 2731–2739.

Taylor, D. J., Hebert, P. D. N. and Colbourne, J. K. (1996) Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Mol Phylogenet Evol*, **5**, 495–510.

Taylor, D. J., Sprenger, H. L. and Ishida, S. (2005) Geographic and phylogenetic evidence for dispersed nuclear introgression in a daphniid with sexual propagules. *Mol Ecol*, **14**, 525–537.

Thielsch, A., Brede, N., Petrusek, A., De Meester, L. and Schwenk, K. (2009) Contribution of cyclic parthenogenesis and colonization history to population structure in *Daphnia*. *Mol Ecol*, **18**, 1616–1628.

Vincze, T., Posfai, J. and Roberts, R. J. (2003) NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res*, **31**, 3688–3691.

Wolf, D. E., Takebayashi, N. and Rieseberg, L. H. (2001) Predicting the risk of extinction through hybridization. *Conserv Biol*, **15**, 1039–1053.

Wolf, H. G. and Mort, M. A. (1986) Interspecific hybridization underlies phenotypic variability in *Daphnia* populations. *Oecologia*, **68**, 507–511.

Wolinska, J., Bittner, K., Ebert, D. and Spaak, P. (2006) The coexistence of hybrid and parental *Daphnia*: the role of parasites. *Proc R Soc B*, **273**, 1977–1983.

Yin, M., Wolinska, J. and Giessler, S. (2010) Clonal diversity, clonal persistence and rapid taxon replacement in natural populations of species and hybrids of the *Daphnia longispina* complex. *Mol Ecol*, **19**, 4168–4178.