Conservation Genet Resour (2018) 10:897–905
DOI 10.1007/s12686-017-0912-0

METHODS AND RESOURCES ARTICLE

Discovery and validation of species–specific diagnostic SNP markers for the endangered San Diego fairy shrimp (*Branchinecta sandiegonensis*) and the versatile fairy shrimp (*Branchinecta lindahli*)

Ketan V. Patel[1] · Andrew J. Bohonak[1] · Marie A. Simovich[2] · Natalie S. Goddard[1] · Nicholas S. Graige[1]

Received: 19 June 2017 / Accepted: 26 October 2017 / Published online: 3 November 2017

Abstract Because only 3–7% of historically present vernal pool habitat remains in coastal San Diego County, conservation efforts must prioritize both the maintenance of these pools and the genomic integrity of their inhabitants. Coastal vernal pools found in southern California are home to the federally endangered San Diego fairy shrimp *Branchinecta sandiegonensis*. Simovich et al. (J Crustac Biol 33:730–740, 2013) characterized hybridization between *B. sandiegonensis* and the versatile fairy shrimp (*Branchinecta lindahli*) using morphological characters, but these characters are only found in adult females. To detect adult female, male and juvenile hybrids in the wild, we developed a genomic hybrid index comprised of 20 SNP loci using 16 individuals with no morphological evidence of hybridization, collected from populations unlikely to facilitate introgression. These loci have alternatively fixed alleles between the two species. This genomic hybrid index was validated using 426 individuals from 27 localities using morphology and habitat information. Our data suggest that some artificial and disturbed pool basins harbor hybrids, and thus have the potential to be stepping-stones for the future spread of hybrids. This genomic hybrid index will be a useful tool for identifying putative *Branchinecta* hybrids from both mature and immature life history stages, and aid in the monitoring and recovery of non-admixed *B. sandiegonensis*.

✉ Ketan V. Patel
Kpatel102486@gmail.com

[1] Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4614, USA

[2] University of San Diego, 5998 Alcala Park, San Diego, CA 92110, USA

Keywords *Branchinecta* · Fairy shrimp · Hybridization · Vernal pools · Transcriptome assembly · SsSNP loci · Conservation

Introduction

Similar to habitat loss, the alteration of native habitat is often linked to the initial listing of a species as threatened or endangered, and usually decreases the likelihood of a species' recovery. Landscape homogenization can facilitate the expansion of invasive species into the ranges of endemics breaking down habitat partitions and facilitating novel competition scenarios (Mooney and Cleland 2001; Olden et al. 2004; Olden and Rooney 2006; Devictor et al. 2008a, b; Simovich et al. 2013). Along with the decreasing habitat heterogeneity, formerly unique communities will become increasingly similar and novel competition scenarios will favor the range expansion of non-native generalists. Establishment of non-natives into human-altered habitats can also impact biodiversity through a loss of local/endemic genetic and species diversity (Anderson and Stebbins 1954; Rhymer and Simberloff 1996; Ellstrand and Schierenbeck 2000). If invasive species readily hybridize with native species, genetic boundaries between both species may erode, resulting in the irrevocable loss of native genetic stock (Rhymer and Simberloff 1996).

Coastal southern California's vernal pools are ephemeral wetlands which host an array of plants and animals adapted to the bi-phasic (e.g., wet and dry) nature of the habitat. One of the most evident faunal elements consist of crustaceans. During periods of pool inundation, formerly dormant crustacean cysts hatch, develop to the adult stage, reproduce, and deposit cysts for the remainder of pool inundation (Belk 1998; Erickson and Belk 1999). The most

common large crustaceans are fairy shrimp (Anostracans) in the genus *Branchinecta*, including the San Diego fairy shrimp, *Branchinecta sandiegonensis* (Fugate 1993; USFWS 1997). This federally listed species is characterized as a narrow-range endemic that is found only in highly functional (Bauder et al. 2009) coastal vernal pools in southern California and Baja California, Mexico (Fugate 1993; Erickson and Belk 1999). As a result of urban expansion, estimates suggest that only 3–7% of the original coastal vernal pool habitat remains intact (Bauder and McMillan 1998; King 1998). Additionally, the associated construction and vehicular traffic have created countless artificial basins (e.g., deep impoundments, road ruts, ditches) that may potentially harbor invasive generalists. For example, *Branchinecta lindahli* was once thought to be restricted to inland playas, but now occurs in a variety of man-made pools in and around converted vernal pool habitat (Fugate 1998; Simovich et al. 2013). In addition to the arrival of *B. lindahli* as a competitor, *B. sandiegonensis* will hybridize with *B. lindahli* in both in situ and ex situ conditions (Fugate 1998; Erickson and Belk 1999; Simovich et al. 2013). Due to the threat posed by interspecific hybridization (Rhymer and Simberloff 1996), hybrid detection and subsequent management must play a crucial part in the conservation and recovery of *B. sandiegonensis*.

To detect hybridization between *B. sandiegonensis* and *B. lindahli*, Simovich et al. (2013) developed a morphological hybrid index based on an adult female's thoracic spine pattern (see also Patel et al. in review). Spines (dorsolateral processes) on nine thoracic segments of mature females are scored, and the resulting morphological hybrid index distinguishes *B. sandiegonensis* from *B. lindahli* and putative hybrids (Rogers 2002; Simovich et al. 2013). However, morphological identification of hybrids is limited for several reasons. First, only adult females display the diagnostic characters needed to identify putative hybrids. However, species identification keys used by those with federal permits rely on male characters. Second, the occasional presence of atypical character states in a particular individual could potentially reflect selection on that character, rather than introgression of the entire genome. Third, by relying solely on morphology, misidentification may occur due to transgressive phenotypic variation displayed in highly admixed individuals (Seehausen 2004). In these instances, hybrid offspring may display extreme phenotype variation compared to the reference phenotypes used to identify either parent species (Seehausen 2004; Arnold 2006; Hedrick 2013). In the worst-case scenario, introgressive hybridization events through multiple generations may render diagnostic morphological markers ineffective.

Genetic markers such as single nucleotide polymorphisms (SNPs) have the potential to distinguish parental species and place hybrid individuals in distinct classes based on multi-locus genotypic patterns (Pritchard et al. 2000; Anderson and Thompson 2002; Li et al. 2015). Characterized by alternatively fixed loci between parent species, species-specific SNPs (ssSNPs) are ancestry-informative markers that are easily diagnosable, highly reliable, and represent genome-wide patterns of interspecific admixture (Primmer et al. 2002; VÄHÄ and Primmer 2006; Bajec et al. 2015; Li et al. 2015). Because hybridization may occur across multiple generations, ssSNP loci may also help to infer the proportion of an individual's genome where one gene copy was inherited from each parental species (i.e. interspecific ancestry). Interspecific ancestry can be then used to distinguish between early and late-generation hybrids (Gompert and Buerkle 2009, 2010). For example, F_1 hybrids are expected to have a genomic hybrid index value of 0.5 and be heterozygous with regards to species-specific alleles across the entire genome. As subsequent genetic admixture between parental species and hybrid classes occurs, later-generation hybrids will tend to show higher variance in both heterozygosity and average interspecific ancestry (Fig. 2A). Late stage hybrids (e.g., F_{10}, F_{20}) may display little heterozygosity, but show a mosaic pattern where some loci are fixed for one parental SNP and others are fixed for the second parental SNP. In addition, the genomic profile of admixed individuals may be skewed by asymmetry in backcross frequencies with each parental species.

Here, we describe the first de novo assembled transcriptomes for *B. sandiegonensis* and *B. lindahli*, and their alignment to discover alternatively fixed species-specific SNP loci. Our objectives for this study are to (i) develop a robust genomic panel capable of detecting both male and female putative hybrids, and (ii) validate the resultant genomic panel with a dataset of morphologically characterized individuals using protocols published in Simovich et al. (2013). Identification of hybrids through SNP genotyping will aid in the recovery of *B. sandiegonensis* through detection/monitoring of hybrids across a variety of functional and disturbed pool types, and discerning between admixed and pure *B. sandiegonensis* populations.

Methods

Sample collection, library construction, and RNA-seq with poly A tail enrichment

To obtain representative species-specific genomic diversity for transcriptome assembly, we collected nine reference samples for *B. sandiegonensis* across four coastal vernal pool sites and seven reference samples of *B. lindahli* from two inland playa sites (denoted by § in Table 1). These reference sites were chosen because they represent archetypical habitat for each species, have been sampled over multiple seasons,

Table 1 Transcriptome assembly statistics, mapping, and candidate loci filtering summary statistics for *B. sandiegonensis* and *B. lindahli*

Post RNA sequencing steps	*B. sandiegonensis*	*B. lindahli*
Number of raw 100 bp paired reads	117,154,626	73,891,652
Number of contiguous sequences (contigs)	2,099,012	1,566,647
Total trinity 'genes'	49,603	39,142
Total trinity transcripts	74,667	39,142
Percent GC	45.22%	45.18%
Contig N10	5117	4881
Contig N20	3572	3718
Contig N30	2775	3013
Contig N40	2178	2457
Contig N50	1700	1979
Median contig length	429	566
Average contig	890	1071
Total assembled bases	66,434,959	84,447,374
Number of filter reads mapped	54,584,949	65,876,018
Number of properly paired reads	30,560,204	52,166,678
Alignment using NCBI blastn and MUSCLE		
Post alignment filtering		
Alignments with reciprocal matches 99.9–97% (with duplicates)	3117	3397
Alignments with a single match	932	932
Alignments exhibiting zero gaps	742	742
Alignments with a total length greater than 200 bp	457	457

and no individuals were hybrid females (based on the index of Simovich et al. 2013). Samples were immersed in separate collection vials containing RNA later® (Ambion, Austin, TX, USA) and stored at − 20 °C. Prior to RNA extraction, equal amounts of tissue from shrimp of the same species were pooled together to form a species pool (Konczal et al. 2014) and homogenized using a roto-homoginizer at − 20 °C in the presence of TRIzol™ reagent. Total RNA was extracted using the TRIzol™ extraction protocol (Chomczynski et al. 1987) followed by an RNA purification step using Ambion™ cleanup kit (Ambion, Austin, TX, USA). Total RNA concentrations for both species pools were evaluated by Qubit® flourometer. The two resulting species pools were sent to Hudson Alpha Genomic Services Lab (Huntsville, AL USA) for library preparation and subsequent RNA sequencing using poly (A) tail enrichment. Sequencing was carried out using a Illumina HiSeq 2000 with the option of 100 bp paired-end reads resulting with approximately 25 million reads per species pool.

De novo transcriptome assembly and SNP discovery

Data files containing raw sequencing reads in FASTQ format, quality scores, and paired reads information were returned from Hudson Alpha Genomic Services Lab (Huntsville, AL, USA) for the subsequent trimming and transcriptome assembly. Sequencing adapters were trimmed using Trim Galore! (Krueger 2015) and raw reads were filtered for quality control by removing reads with quality scores less than 20 and length below 30 base pairs using prinseq-lite-0.20.4 (Schmieder and Edwards 2011). Reads from each species pool were used to assemble transcriptomes for *B. sandiegonensis* and *B. lindahli* using the Trinity assembler (v.2014-04-13; Grabherr et al. 2011). Trinity employs three methods (named *Inchworm, Chrysalis* and *Butterfly*) for transcriptome assembly without a reference genome (i.e. denovo assembly). Briefly, *Inchworm* assembles raw sequencing reads by greedy k-mer extension (default is set to k-mer 25) into a single representative (i.e., contig) for a set of variant reads that share k-mers. *Chrysalis* then clusters related contigs, and constructs de Bruijn graphs for each cluster, which represent the complexity of overlaps between variant contigs. In the final step, *Butterfly* analyzes all the paths taken by sequencing reads and read pairings with respect to the corresponding de Bruijn graphs for all clusters and reports all plausible transcript sequences (Grabherr et al. 2011). Following denovo transcriptome assembly, raw reads were mapped to each respective transcriptome assembly using Bowtie 1.1.1 set to default options (Langmead et al. 2009).

To isolate SNPs that would serve as diagnosable markers for hybrid identification, we focused our efforts on the discovery of loci that would display fixed-allelic differences between *B. sandiegonensis* and *B. lindahli* (e.g.,

'T/T' in *B. sandiegonensis*, 'C/C' in *B. lindahli*, and 'T/C' in F1 hybrids). Therefore, individuals with at least one heterozygous genotype or deviation from the genotypes of either pure *B. sandiegonensis* or *B. lindahli* (Table 2) would be considered hybrids. Prior to ssSNP discovery, contig sequences containing within-species SNPs and/or likely to contain sequence variants were discovered using SAMtools (Li et al. 2009), and were manually discarded from further panel development. To identify genomic segments of high homology between species, the entire *B. sandiegonensis* transcriptome assembly was compared with the entire *B. lindahli* transcriptome assembly using NCBI nucleotide *BLAST*: blastn (Altschul et al. 1990). Initially, contigs with matches less than 97.0% (i.e. 97% of base pair matches across the entire length of the contig) were discarded. Contig alignments were then filtered to discard multiple sequence matches (i.e., hits), gaps, insertions/deletions (INDELS), and segments less than 200 base pairs in total length. The remaining contig alignments were globally aligned using MUSCLE (Edgar 2004) as implemented within the Mesquite program (Madison and Madison 2004) and assessed visually using AliView (Larsson 2014). Any contig sequence alignment that failed to match globally was subsequently discarded. Contig alignments that remained were selected as potential diagnostic markers and carried forward to primer design.

Species-specific panel development and initial testing

Contigs containing putative ssSNP loci were sent to University of Arizona Genetic Core (UAGC) for primer design. Multiplex assays were designed using the MassARRAY Assay Design® software with the goal of multiplexing of 30 SNPs. Only SNPs with at least a 100 bp flanking region on either side of the polymorphic site were selected for the assay design. Candidate primer pair sequences were returned and were subsequently compared to both transcriptome assemblies using NCBI nucleotide *BLAST*: blastn (Altschul et al. 1990). Candidate loci that had primers pairs with hits to multiple sites on either transcriptome assembly were discarded. The remaining candidate ssSNP loci were carried forward for marker validation.

To test the genotyping success of the panel, we used 30 candidate ssSNP markers to genotype 46 morphologically identified individuals. Briefly, samples were sent to the University of Arizona Genetic Core (UAGC) facility for genotyping using the Sequenom MassARRAY genotyping platform (Bradić et al. 2011). Noncalls resulting from low probability or bad spectrum were noted and resolved by eye if possible. Individuals with lower than 90% call rates were removed, and failed loci were discarded or redesigned. Primer pairs that successfully amplified target loci were formatted into a final 20-plex ssSNP panel.

Table 2 Detailed information of the 20-SNP Agena Bioscience multiplex including species-specific alleles for *B. sandiegonensis, B. lindahli*, protein information (if applicable), and genotyping failure rate (%) found in this study

ssSNP marker ID	*B. sandiegonensis* allele	*B. lindahli* allele	Protein identification via NCBI non-redundant protein database (if applicable)	Genotyping failure (%)
RDcomp25015	G	A		0.25
comp1246633	C	T		0.25
comp12974	C	T	LSM domain	0.25
comp28208	C	G		0.49
comp2628	T	C		0.74
comp1209936	A	G		1.23
RDcomp40235	T	C	Ribosomal protein L44	1.47
comp29744	T	C		1.47
RDcomp33135	G	A		1.72
comp19493	C	A		1.72
comp32848	T	A	HMG (high mobility group) box	1.72
comp19136	A	G		2.21
comp37098	G	A	3′5′-cyclic nucleotide phosphodiesterase	3.19
comp678743	G	A		5.15
comp12997	C	A	Ubiquitin-2 like Rad60 SUMO-like	5.64
comp20933	G	A	Neurotransmitter-gated ion-channel ligand binding domain	6.37
comp3767	A	C		6.37
comp31041	A	G	Ion transport protein	12.01
comp977876	G	C	Immunoglobulin I-set domain	13.24

Species-specific SNP validation

To validate the quality and performance of the final 20-plex ssSNP panel, we used a dataset of 391 adult female shrimp, morphologically identified as *B. sandiegonensis, B. lindahli*, and various interspecific hybrids using the morphological hybrid index developed by Simovich et al. (2013). The arrangement of spines on thoracic segments 3 through 11 was given one of three possible scores using the Simovich et al. (2013) criteria (see also Patel et al. in review). Character states congruent with *B. lindahli* were given a score of 1, character states congruent with *B. sandiegonensis* were given a score of 3, and character states that are atypical for both species were given a score of 2. Numeric scores were averaged across all thoracic segments, and the average score was used to categorize individuals as *B. lindahli* (1.0–1.3), hybrids (1.4–2.5), or *B. sandiegonensis* (2.6–3.0). To verify that the genomic panel could detect male hybrids, 35 males from a total of five pools were sampled within the *B. sandiegonensis* species range.

We selected seven males from Brown Parcel A, six from Proctor Valley Corral side B, nine from Palmdale pool 1, four from Palmdale pool 2, and nine from Palmdale pool 4. Pools that contained female morphological hybrids were characterized as vehicular road ruts, man-made deep impoundments, or artificial pools as a result of habitat remediation efforts. Detailed information regarding sample localities, hybrid presence, and disturbance characteristics are presented in Table 3.

Following the manufacturer's specifications, DNA was extracted and isolated from approximately 10 mg of tissue per sample using the Qiagen DNeasy kit (Qiagen). DNA concentration and purity were estimated using an Implen Nanophotometer™ Pearl. The 426 extractions were then sent to the University of Arizona Genetics Core facility for ssSNP screening using the Agena Bioscience MassARRAY genotyping platform (Bradić et al. 2011). Diagnostic ssSNP genotypes were subsequently converted into numeric format according to species-specific alleles (e.g. *B. lindahli* = 1, heterozygous loci = 2, and *B. sandiegonensis* = 3) to match assignments used in the morphological hybrid index. A scatterplot was used to compare the morphological and genomic hybrid indices. The interspecific ancestry for each individual was calculated using the functions est.h, and genomic.clines in the program INTROGRESS (Gompert and Buerkle 2009), as implemented in R (R Core Team 2016). Interspecific ancestry was visualized as a function of the genomic hybrid index using the est.h function with the triangle plot command.

Results and discussion

Transcriptome assembly and SNP discovery

Library sequencing produced 117,154,626 100 bp paired-end reads from the *B. sandiegonensis* species pool and 73,891,652 100 bp paired-end reads for *B. lindahli* pool. A total of 49,603 and 39,142 contigs (trinity genes) were recovered for *B. sandiegonensis* and *B. lindahli* respectively. Mean contig sizes and N50's were 889.75 bp and 1700 for *B. sandiegonensis*, and 1070 and 1979 for *B. lindahli* (Table 1). Bowtie mapped 54,584,949 and 6,587,601 filtered reads with a total of 30,560,204 and 52,166,678 properly paired reads to the *B. sandiegonensis* and *B. lindahli* respectively. Reciprocal blasting followed by global alignment using MUSCLE (Edgar 2004) yielded a total of 457 unique contig matches that were above 97% similarity, over 200 bp in length, and possessed neither gaps nor INDELs. Data pipeline information regarding transcriptome assembly statistics, and subsequent filtering steps are displayed in Table 1.

Species-specific SNP panel development and validation

To determine the accuracy and reliability of the ssSNP panel as a resource to identify interspecific hybrids, we initially genotyped a subset of 46 morphologically identified individuals with an initial panel consisting of 30 SNPs optimized in two multiplex panels (a 22-plex and an 8-plex). From the original 30 initial candidate markers, we selected 20 candidate loci that successfully and reliably amplify for use in the final multiplex SNP panel. The final panel was validated using a separate dataset of 426 individuals (Table 3). Over half (13/20) of the loci tested had a genotyping failure rate of less than 5%, four loci had a failure rate of less than 10%, and three loci had genotype failure rates of less than 16%. Contig ID, species-specific genotypes, and marker failure rate are provided in Table 2.

Both morphological and genomic hybrid indices showed similar values for individuals in undisturbed pools, which we assumed to contain "pure" species. Non-admixed, undisturbed *B. lindahli* localities (Fig. 1; bottom-left) show high congruence between the genomic and morphological hybrid index. Non-admixed localities of *B. sandiegonensis* (Fig. 1; top-right) also show high correlation between hybrid indices. However, variation in the female morphological hybrid index within some non-admixed *B. sandiegonensis* localities may suggest phenotypic plasticity, or unreliable character scoring due to decoupling of genetic variation for spine morphology from the rest of the genome.

The correlation between genomic and morphological hybrid indices seems to weaken in disturbed pools (Fig. 1; center). Individuals from disturbed pools that are genomically similar to non-admixed *B. lindahli* show a wide range

Table 3 Sampling locality information, presence of morphological hybrids, description of hybrid pools, number of samples collected by sex, and allele frequencies for each sampling locality as determined by SNP genotyping

Site	Complex	Pool	Samples sizes: F (M)	Latitude	Longitude	Pool category	Hybrid presence	Pool description	B.s. allele frequency	B.l. allele frequency
Ramona*	Town	Main/Hunter St	20 (0)	33.23277778	− 116.9425000	Coastal vernal pool	Yes	Road rut	1.000	0.000
Ramona§,*	Town	Ramona/Day St	21 (0)	33.08638889	− 117.0797222	Coastal vernal pool	No	Road rut	0.998	0.002
Ramona§,*	Town	Main/Kalbaugh St	22 (0)	33.02712100	− 116.890167	Coastal vernal pool	No		1.000	0.000
Brown*	Parcel	A	5 (7)	32.91981500	− 117.1725877	Coastal vernal pool	Yes	Road rut	1.000	0.000
Carmel Mountain*	Carmel Mountain	Football Pool	16 (0)	33.08055560	− 117.2905556	Coastal vernal pool	Yes	Road rut	1.000	0.000
Orange County*	Costa Mesa	A	15 (0)	33.66000000	− 117.9400000	Coastal vernal pool	Yes	Artificial pool	0.096	0.904
Del Mar Mesa*	Bowtie	A	18 (0)	33.12638889	− 117.4150000	Coastal vernal pool	Yes	Road rut	1.000	0.000
Del Mar Mesa*	Del Mar Mesa	256	24 (0)	33.18111111	− 117.3563889	Coastal vernal pool	Yes	Road rut	0.906	0.094
Del Mar Mesa*	Del Mar Mesa	55	17 (0)	33.12833333	− 117.4033333	Coastal vernal pool	Yes	Road rut	1.000	0.000
Miramar*	AA10	70.1	17 (0)	32.8763791	− 117.0997655	Coastal vernal pool	No		1.000	0.000
Miramar*	AA4-7	Cobble Pool 1	19 (0)	32.84005225	− 117.1145703	Coastal vernal pool	Yes	Road rut	1.000	0.000
Miramar*	AA9	139	19 (0)	32.87629931	− 117.1103289	Coastal vernal pool	Yes	Road rut	1.000	0.000
Miramar*	Eastgate (I7)	EG-2 (Restored Road Pool)	20 (0)	32.87760071	− 117.1911484	Coastal vernal pool	No		1.000	0.000
Miramar§	Eastgate (I7)	3 (Duck Pond)	3 (0)	32.87747827	− 117.1930795	Coastal vernal pool	No		1.000	0.000
Miramar*	FF1/2 (Flightline)	2	18 (0)	32.87621366	− 117.1204777	Coastal vernal pool	No		1.000	0.000
Miramar*	Camp Elliot	Village Rut	14 (0)	− 117.140580	32.88816000	Coastal vernal pool	Yes	Road rut	0.063	0.937
Mission Trails*	Mission Trails	Shepherd's pond	17 (0)	33.06027778	− 117.0483333	Coastal vernal pool	Yes	Deep impoundment	0.078	0.922
McAuliffe§	McAuliffe Community Park	MCR5	2 (0)	32.91428300	− 117.1601950	Coastal vernal pool	No		1.000	0.000
Nobel*	Nobel Dr	3	19 (0)	33.12500000	− 117.3961111	Coastal vernal pool	No		1.000	0.000
Otay Mesa*	Proctor Valley	17	19 (0)	32.80222222	− 116.9805556	Coastal vernal pool	Yes	Road rut	1.000	0.000
Otay Mesa*	Proctor Valley	Corral Pool (side B)	12 (6)	32.72361111	− 117.0316667	Coastal vernal pool	Yes	Road rut	0.040	0.960
Los Angeles County*	Palmdale	Pool 1	10 (9)	− 118.170558	34.82417400	Inland Desert Playa	No		0.002	0.998
Los Angeles County*	Palmdale	Pool 2	12 (4)	− 118.170943	34.83131000	Inland Desert Playa	No		0.005	0.995
Los Angeles County*	Palmdale	Pool 3	20 (0)	− 118.170865	34.83315000	Inland Desert Playa	No		0.000	1.000
Los Angeles County*	Palmdale	Pool 4	8 (9)	− 118.170731	34.82689900	Inland Desert Playa	No		0.000	1.000
San Bernadino County*	Dale Dry Lake		8 (0)	34.12994800	− 115.7082180	Inland Desert Playa	No		0.000	1.000
San Bernadino County*	Melville Dry Lake		2 (0)	34.45194444	− 115.4258333	Inland Desert Playa	No		0.000	1.000
Anza Borrego§,*	Clark Dry Lake		8 (0)	33.30451500	− 116.2461020	Inland Desert Playa	No		0.000	1.000
Anza Borrego§	DiGerogio Rd		2 (0)	33.24218100	− 116.3658700	Inland Desert Playa	No		0.000	1.000

§ denotes that samples were selected for ssSNP marker discovery

* denotes that samples were selected for ssSNP marker validation

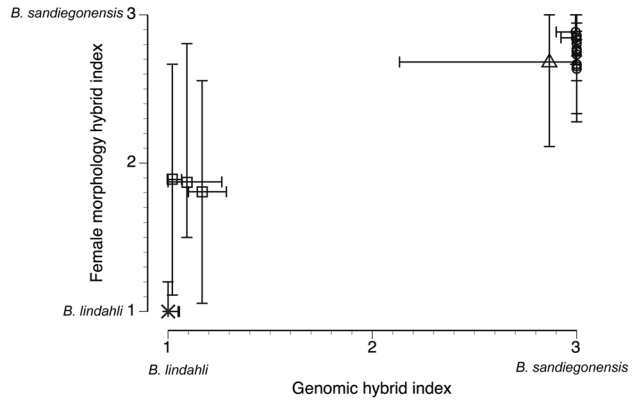

Fig. 1 Comparison of genomic hybrid index and morphological hybrid index across 24 localities: 5 (asterisk) inland playas, 3 (square) disturbed coastal pools, 1 (triangle) disturbed pool with three early stage hybrids and 17 B. sandiegonensis, and 15 (circle) disturbed and undisturbed coastal pools containing B. sandiegonensis. Site-specific female morphology (Simovich et al. 2013) and genomic scores were calculated by averaging respective scores for all shrimp sampled in each population. Populations with complete congruence between genomics and morphology are found in the bottom-left (1:1 genetically and morphologically *B. lindahli*) and top-right (3:3 genetically and morphologically *B. sandiegonensis*) corners of the plot. Each symbol represents a population mean, and the bars extend from the minimum to the maximum for each index

of variance in morphology (Fig. 1). Conversely, females from one disturbed pool (Del Mar Mesa 256) were morphologically similar to non-admixed *B. sandiegonensis* but showed high genomic variance (triangle plot in Fig. 1). This pool contained three hybrids and 17 non-admixed *B. sandiegonensis* individuals. Overall, the weak congruence between genetics and morphology in many disturbed pools and road ruts is not surprising, since the close association between genotype and morphology may dissociate as a result of genetic admixture over many generations. In some cases, repeated introgression may replace distinct species with hybrid swarms that are comprised entirely of admixed individuals (Seehausen 2004).

Focusing solely on the genomic data, the INTROGRESS plot shows "pure" *B. sandiegonensis* and *B. lindahli* restricted to opposing corners at the base of the triangle plot, with alternatively fixed alleles at the 20 ssSNP loci (Fig. 2B). Four hybrids with high interspecific heterozygosity (near the top of the triangle) are the result of ongoing interspecific hybridization in some localities (Fig. 2B). Admixed individuals at the bottom of the plot (but not in one of the two corners) are the results of past backcrosses among hybrids, or between hybrids and one of the "pure" *Branchinecta* species (Fig. 2B). Overall, the ssSNP data provide evidence for hybridization and introgression through multiple generations in certain localities. More general conclusions regarding the status of the endangered *B. sandiegonensis* will require additional sampling throughout its range.

The frequency of *B. lindahli* alleles in the coastal vernal pools we surveyed ranged from 0.000 to 0.959. The majority (17/20) of localities were shown to have very low frequencies of *B. lindahli* alleles ($0\% < q < 0.05\%$; Table 3). This suggests very limited past hybridization even in nearly-pure "native" *B. sandiegonensis* populations. However, a few putative hybrid populations exist in heavily disturbed ruts as late-stage hybrid swarms, and genomic identity in these habitats is most heavily influenced by the dominant parental species.

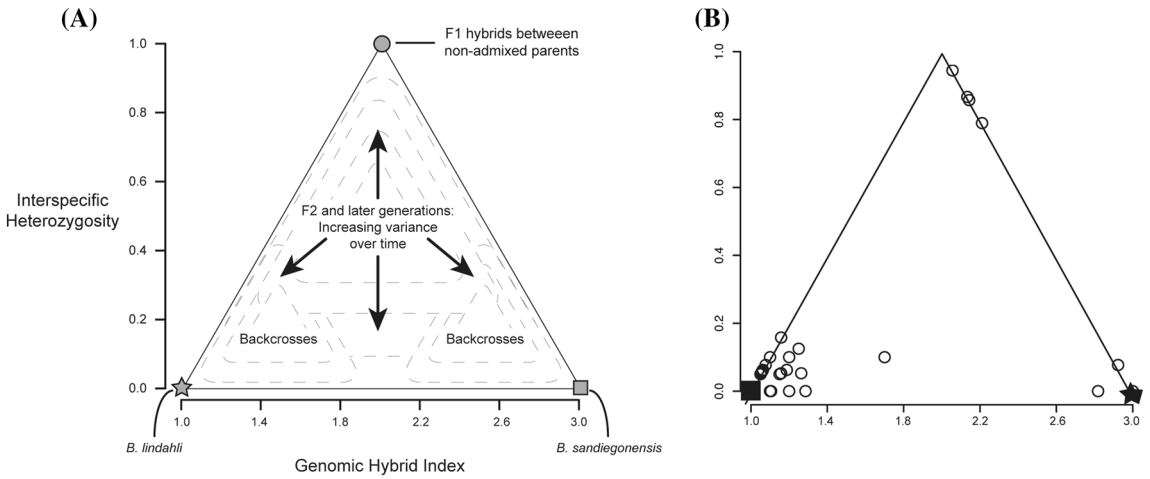

Fig. 2 Interspecific ancestry i.e. the proportion of an individual's genome where one gene copy was inherited from each parental species in admixed lineages. **A** Schematic illustration of interspecific ancestry; non-admixed parent species are found at opposite sides of the triangle base. Genetically admixed individuals can be found throughout the plot area. **B** Patterns of interspecific ancestry based on genomic hybrid index values (proportion of *B. lindahli* alleles) for individuals used in this study; non-admixed *B. sandiegonensis* (square; n = 271), non-admixed *B. lindahli* (asterisk; n = 98), hybrids/backcrosses (circle; n = 57)

In some artificial basins, hybrids resemble *B. lindahli* more than *B. sandiegonensis* both morphologically and genetically. It is conceivable that *B. sandiegonensis* was not the most common species in these populations prior to hybrid establishment. Waterkeyn et al. (2010) demonstrated that encysted embryos in pool sediments can adhere to footwear and vehicle tires, and unintentionally be dispersed. If vehicular traffic disturbance in coastal habitats both creates these habitats (e.g., deep ruts in dirt roads) and also inoculates them with *B. lindahli* propagules, the subsequent hybridization would be skewed towards *B. lindahli* (Fig. 1, Table 3). Additional studies are needed to confirm whether these patterns of hybridization generalize for various types of highly disturbed basins (e.g., vehicular road ruts, deep impoundments, artificial pools) across the entire species range. Also unresolved is whether these types of sites act as "bridgehead" populations to promote the spread of hybridization into native vernal pools (*sensu* Estoup and Guillemaud 2010).

The accuracy, time- and cost-effectiveness of ssSNPs relative to morphological markers can greatly improve the detection of putative hybrids, and quantify the degree of genetic admixture in natural populations. This study provides a new technique for the identification of both male and female putative hybrids, as well as a quantifiable metric to assess site-specific levels of interspecific ancestry. We encourage the use of this ssSNP panel for use in future studies aimed at mapping the distribution of putative hybrid populations throughout the native coastal range of *B. sandiegonensis*, as well as desert playas and various artificial basins that are more characteristic of *B. lindahli*.

The use of this tool will aid in the detection of male and female hybrids in natural populations and provide a more robust method to characterize admixed localities, thereby aiding in the mitigation of hybrid spread and overall recovery of *B. sandiegonensis* in coastal vernal pools.

Acknowledgements The authors would like to thank Charles Black for help with sample collection and his insights on fairy shrimp biology. We thank David Carlson, Shahan Derkarabetian for their assistance in transcriptome assembly and SNP discovery as well as Cheryl Hayashi, William Walton and the anonymous reviewer for providing helpful suggestions regarding manuscript preparation. We also thank the genomic services laboratory at Hudson Alpha Institute for Biotechnology for help with sample preparation and technicians at the University of Arizona Genetics Core for help with multiplex optimization. Funding for this study was granted by the San Diego Association of Governments (SANDAG), U.S. Department of the Navy (Cooperative Agreement N62473-14-2-0001), California Department of Fish and Wildlife and U.S. Fish and Wildlife Service (Cooperative Endangered Species Conservation Fund/Section 6 Grant P1382012). All ''take'' was conducted under the supervision of Marie Simovich as authorized under US Fish and Wildlife Service Permit TE-787037-4.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol 215(3):403–410

Anderson E, Stebbins GL (1954) Hybridization as an evolutionary stimulus. Evolution 8(4):378–388

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160:1217–1229

Arnold ML (2006) Evolution through genetic exchange. Oxford University Press, Oxford

Bajec SS, Pustovrh G, Jesenšek D, Snoj A (2015) Population genetic SNP analysis of marble and brown trout in a hybridization zone of the Adriatic watershed in Slovenia. Biol Conserv 184:239–250

Bauder ET, McMillan S (1998) Current distribution and historical extent of vernal pools in southern California and northern Baja California, Mexico In Ecology, Conservation, and Management of Vernal Pool Ecosystems. Proceedings from a 1996 Conference California Native Plant Society, Sacramento, California, pp 56–70

Bauder E, Bohonak AJ, Hecht B, Simovich MA, Shaw SD, Jenkins DG, Rains M (2009) A draft regional guidebook for applying the hydrogeomorphic approach to assessing wetland functions of vernal pool depressional wetlands in southern California. San Diego State University, San Diego

Belk D (1998) Global status and trends in ephemeral pool invertebrate conservation: implications for Californian fairy shrimp Witham, CW. In: Bauder ET (eds), Current distribution and historical extent of vernal pools in southern California and northern Baja California, Mexico in Ecology, Conservation, and Management of Vernal Pool Ecosystems. Proceedings from a 1996 Conference California Native Plant Society, Sacramento, California, pp 147–150

Bradić M, Costa J, Chelo IM (2011) Genotyping with sequenom. Mol Methods Evol Genet 772:193–210

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Devictor V, Julliard R, Clavel J, Jiguet F, Lee A, Couvet D (2008a) Functional biotic homogenization of bird communities in disturbed landscapes. Glob Ecol Biogeogr 17:252–261

Devictor V, Julliard R, Jiguet F (2008b) Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. Oikos 117:507–514

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci 97(13):7043–7050

Erickson CH, Belk D (1999) Fairy shrimps of California's puddles, pools, and playas. Mad River Press Inc., Eureka

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? Mol Ecol 19:4113–4130

Fugate M (1993) Branchinecta sandiegonensis, a new species of fairy shrimp (Crustacea: Anostroca) from Western North America. Proc Biol Soc Wash 106:296–304

Fugate M (1998) Branchinecta of North America: population structure and its implications for conservation practice. In: Ecology, conservation, and management of vernal pool ecosystems. California Native Plant Society, Sacramento, pp 140–146

Gompert Z, Buerkle C (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Mol Ecol 18:1207–1224

Gompert Z, Buerkle C (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. Mol Ecol Resour 10:378–384

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Chen Z (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol 22:4606–4618

King JL (1998) Loss of diversity as a consequence of habitat destruction in California vernal pools. In: Ecology, conservation, and management of vernal pool ecosystems. Proceedings from a 1996 Conference California Native Plant Society, Sacramento, pp 119–123

Konczal M, Koteja P, Stuglik MT, Radwan J, Babik W (2014) Accuracy of allele frequency estimation using pooled RNA-SEq. Mol Ecol Resour 14:381–392

Krueger F (2015) Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. https://doi.org/10.1186/gb-2009-10-3-r25

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276–3278

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li C, Gowan S, Anil A, Beck BH, Thongda W, Kucuktas H, Peatman E (2015) Discovery and validation of gene-linked diagnostic SNP markers for assessing hybridization between Largemouth bass (Micropterus salmoides) and Florida bass (M floridanus). Mol Ecol Resour 15:395–404

Madison W, Madison D (2004) Mesquite: a modular system for evolutionary analysis Version 105 http://mesquiteproject.org

Mooney HA, Cleland EE (2001) The evolutionary impact of invasive species. Proc Natl Acad Sci 98(10):5446–5451

Olden JD, Poff NL, Douglas MR, Douglas ME, Fausch KD (2004) Ecological and evolutionary consequences of biotic homogenization. Trends Ecol Evol 19(1):18–24

Olden JD, Rooney TP (2006) On defining and quantifying biotic homogenization. Glob Ecol Biogeogr 15(2):113–120

Primmer CR, Borge T, Lindell J, Sætre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Mol Ecol 11:603–612

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. Annu Rev Ecol Syst 27:83–109

Rogers DC (2002) Female-based characters for Anostracan (Crustacea: Branchiopoda) identification: a key for species of California and Oregon, USA. Hydrobiologia 486:125–132

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864

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

Simovich MA, Davis KB, Bohonak AJ (2013) Landscape homogenization threatens the genetic integrity of the endangered San Diego fairy shrimp *Branchinecta sandiegonensis* (Branchiopoda: Anostraca). J Crustac Biol 33:730–740

Team RC (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org

US Fish and Wildlife Service (1997) Endangered and threatened wildlife and plants; determination of endangered status for the San Diego fairy shrimp Federal Register Number 97-2578, pp 4925–4939

VÄHÄ JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Mol Ecol 15:63–72

Waterkeyn A, Vanschoenwinkel B, Elsen S, Anton-Pardo M, Grillas P, Brendonck L (2010) Unintentional dispersal of aquatic invertebrates via footwear and motor vehicles in a Mediterranean wetland area aquatic conservation. Mar Freshw Ecosyst 20:580–587