1 **Send correspondence to:**

2 Jeremy C Andersen
3 Department of Environmental Science, Policy and Management
4 Mulford Hall
5 University of California
6 Berkeley, CA 94720-3114, USA
7 Phone: 510-643-6481
8 Email: jandersen@berkeley.edu
9

10 Running head: Knotweed psyllid hybridization.

11
12 Development and verification of SNP arrays to monitor hybridization between two host-
13 associated strains of knotweed psyllid, *Aphalara itadori*
14

15 Jeremy C Andersen^{a,*}, Robert S Bouchier^b, Fritz S Grevstad^c, Roy Van Driesche^d, and
16 Nicholas J Mills^a
17

18 ^aDepartment of Environmental Science Policy and Management, University of California
19 Berkeley, Berkeley, CA 94720, USA

20 ^bAgriculture and AgriFood Canada-Lethbridge Research Centre, Lethbridge, AB, Canada
21 T1J 4B1

22 ^cDepartment of Botany and Plant Pathology, Oregon State University, Corvallis, OR,
23 97331, USA

24 ^dDepartment of Environmental Conservation, University of Massachusetts Amherst,
25 Amherst, MA, 01003, USA

26 ^{*}Corresponding author, e-mail: jandersen@berkeley.edu

27 **Abstract**

Three species of invasive knotweeds (*Fallopia japonica*, *F. sachalinensis*, and *F. x bohemica*) cause extensive damage to riparian and roadside habitats in North America. Currently, two strains of the psyllid *Aphalara itadori* are being evaluated for introduction into the United States and Canada for the biological control of these knotweeds following the introduction of *A. itadori* into the United Kingdom. If approved and released, hybridization between individuals from these two strains is likely and understanding whether barriers to hybridization exist could have an important impact on the sustainability of this biological control program. Here we developed two single nucleotide polymorphism (SNP) arrays and examined their utility for identifying individuals of known pure strains (Hokkaido and Kyushu) and hybrid origins. Using an array of 141 SNPs we correctly identified all individuals to pure and hybrid classes, whereas using a smaller array of 29 SNPs we were able to correctly identify pure line individuals, but not hybrids.

Keywords: Strains, biological control, Japanese knotweed, giant knotweed, Bohemian knotweed, *Fallopia* spp.

1. Introduction

Hybridization in nature was long considered a relatively rare event with both pre- and/or post-zygotic barriers preventing the transmission of genomic material between species (Kirkpatrick, 2000; Lukhtanov et al., 2005; Orr and Presgraves, 2000; Presgraves 2002). These barriers may, in part, be the result of incompatibilities among nuclear or cytoplasmic loci (Dobzhansky, 1936; Hurst and Pomiankowski, 1991) and/or incompatibilities between cytoplasmic loci and either nuclear loci or sex chromosomes (Ellison et al. 2008; Hurst and Pomiankowski, 1991). Yet, increasingly it is being recognized that hybridization between closely related organisms occurs in a broad range of taxonomic groups (Allendorf et al., 2001; Harrison and Larson, 2014) and that it has important consequences for evolution and speciation (Mallet, 2005). These hybridization events can lead to shifts in host-preference (Feder et al., 2003) and host-associated speciation events (Schwarz et al., 2005); furthermore, the rate at which hybridization events occur has undoubtedly been increased through anthropogenic activities that accidentally facilitate encounters between allopatric lineages (Allendorf et al., 2001).

Among insect taxa, hybridization appears to be a common occurrence (Schwenk et al., 2008). Given the extensive use of insect species as natural enemies in classical biological control programs (Hoddle, 2004; Van Driesche et al., 2010), hybridization has also likely played a frequent role in the sustainability of biological control services. Historically, many programs were established by collecting individuals of a selected natural enemy species from different localities and/or hosts within its region of origin. These populations have been referred to as strains, biotypes, and/or ecotypes (Heraty, 2009). This practice is much less common today, however, due to the recognition that

each strain needs to be tested separately for its host specificity (Hopper et al., 2005). From an evolutionary perspective, strains likely represent an early point in the process of speciation (i.e., before genetic isolation [Dres and Mallet, 2002]) and can be influenced by factors such as geneflow, genetic diversity, and phenotypic plasticity (Ruiz-Montoya and Nunez-Farfan, 2013). How frequently these strains hybridize, or what effect hybridization has on the sustainability of biological control services is unclear as relatively few studies have examined the effects of post-introduction hybridization. Those studies that have examined the potential effects of hybridization between strains have found evidence of both reduced (e.g. Messing and AliNiazee, 1988; Hoffmann et al., 2002) and increased (e.g., Szűcs et al., 2012) fitness for hybrids relative to their parents.

One system that provides a valuable model for testing the potential effects of hybridization on the host specialization and performance of a natural enemy in a classical biological control setting is the proposed program for the biological control of invasive knotweeds (*Fallopia* spp. [Caryophyllales: Polygonaceae]) in North America. These invasive weeds include Japanese (*F. japonica* [Houttuyn] Ronse Decraene), giant (*F. sachalinensis* Schmidt ex Maxim.), and Bohemian knotweed (*F. x bohemica* [Chrték & Chrtková] JP Bailey), and have been introduced to North America and Europe on several occasions beginning in the 1840s (Beerling et al., 1994). Invasive species of knotweed occur along riverbanks and roadways and in wetlands and disturbed areas where they invade riparian communities (Maerz et al., 2005; Siemens and Blossey, 2007) and reduce native plant and insect species diversity (Gerber et al., 2008; Murrell et al., 2011). The psyllid *Aphalara itadori* Shinji (Hemiptera: Aphalaridae) was recently released in Europe for the biological control of knotweeds (Djeddour and Shaw, 2010; Shaw et al., 2009),

91 and has also been proposed for introduction into the United States and Canada. Two
92 strains of *A. itadori* that differ in their fitness among *Fallopia* species have been proposed
93 for introduction (Grevstad et al., 2013) due to regional differences in the abundance of
94 each species of knotweed in North America (Gaskin et al., 2014). One of these strains
95 was recently released in the UK for the Japanese knotweed biological control program
96 there (Shaw et al., 2009), and originates from the Japanese island of Kyushu. This strain
97 performs best on Japanese and Bohemian knotweeds, while the other strain originates
98 from the Japanese island of Hokkaido and performs best on giant knotweed (Grevstad et
99 al., 2013). Due to overlapping distributions of the invasive species of knotweed in some
100 parts of the United States and Canada, the potential for hybridization between the
101 released strains of psyllid is expected to be high. As there are no known morphological
102 differences between the two psyllid strains, neutral molecular markers are needed to be
103 able to distinguish them post introduction and to detect the incidence of hybridization.

104 One potential barrier to hybridization between these strains may be the presence
105 of reproductive manipulating endosymbionts. Psyllids are known to harbor a diverse
106 assemblage of endosymbionts, including the maternally transmitted primary
107 endosymbiont *Candidatus Carsonella ruddii* (Gammaproteobacteria) that provides
108 essential amino acids absent from their host's diet (Baumann, 2005; Moran and Bennett,
109 2014), and a range of secondary endosymbionts that have coevolved with *C. ruddii* and
110 its insect hosts to provide additional metabolic functions (Chrudimský et al., 2012; Sloan
111 and Moran, 2012). Some secondary endosymbionts are also known to cause sex-ratio
112 distortion (Bordenstein et al., 2001; Ferrari and Vavre, 2011; Turelli, 1994), including
113 species of *Wolbachia* (Rickettsiaceae), *Cardinium* (Bacteroidaceae), *Rickettsia*

(Rickettsiaceae), *Spiroplasma* (Spiroplasmataceae), and *Arsenophonus* (Enterobacteriaceae) that are commonly found in insects (Engelstaedter and Hurst, 2009; Hilgenboecker et al., 2008). In the context of classical biological control programs, the presence of endosymbionts that either provide essential functions to their hosts or engage in reproductive manipulation may have important consequences for the establishment and impact of introduced natural enemies through the creation of barriers to gene flow (Cheyppé-Buchmann et al., 2011) and/or through influencing patterns of host specialization (Branca et al., 2011).

Therefore, the objectives of this study were 1) to develop neutral molecular markers for the two *A. itadori* strains, 2) to test the ability of these markers to successfully differentiate pure and hybrid individuals of *A. itadori*, and 3) to catalog the diversity of endosymbionts present in populations of both strains.

2. Materials and methods

2.1. Laboratory colonies

The insects in this study were sampled from laboratory populations maintained at the Insect Quarantine Facility at Oregon State University (OSU) and the Insect Microbial Containment Facility at Agriculture and Agri-Food Canada, Lethbridge (AAFC). The Hokkaido strain was originally field collected by RSB and FSG from giant knotweed on the Island of Hokkaido, Japan in July of 2007. The Kyushu strain was originally collected from Japanese knotweed on the Island of Kyushu in 2004 and was maintained in quarantine in the United Kingdom until a portion of the colony was transferred to OSU and AAFC in 2010. This strain was released into the United Kingdom in 2010 (Shaw et

al., 2009; Djeddour and Shaw, 2010). Hybrid colonies were established at AAFC by placing individual females and males from respective strains together for 24 h to ensure mating occurred before moving them into rearing cages with either giant, Bohemian, or Japanese knotweed. All rearing of parent and hybrid lines at AAFC were conducted using the following controls. Colonies were established in separate and isolated rearing rooms, with only one parent colony being accessed per quarantine visit and one-way movement from parent colonies to hybrid colonies. Re-suited was conducted for subsequent quarantine visits. For the H♀*K♂ hybrids, colonies were established in 2011 by rearing individuals for at least 25 generations before adults were placed in 95% ethanol, while for the K♀*H♂ hybrids, colonies were established in 2014 by rearing individuals for between 4-5 generations before adults were placed in 95% ethanol.

2.2. DNA extractions for identification of *A. itadori* strains

DNA was extracted from individuals of each strain of *A. itadori* from colonies maintained at OSU and AAFC, as well as from post-F1 generations of each hybrid class (Hokkaido female x Kyushu male [H♀*K♂] and Kyushu female x Hokkaido male [K♀*H♂]) from colonies reared on Japanese, giant, and Bohemian knotweed [H♀*K♂] and Japanese and giant knotweed [K♀*H♂] at AAFC. Extractions were performed using the Qiagen Gentra Puregene Kit (Qiagen Co., Venlo, Netherlands) following manufacturer's recommendations, except for the following modifications. First, the cell lysing step was performed using an incubation temperature of 55°C for a minimum of 48 h; and second, because of the low expected yields due to the small size of the psyllids, we pelleted the DNA using a prolonged centrifuge step of 13,000 rpm for 10 min, and

cleaned the DNA pellet with two separate 70% ethanol washes. DNA yields were estimated using a ND-1000 NanoDrop[®] (NanoDrop Technologies, Inc., Wilmington, Delaware) and concentrations of double stranded DNA were measured using the Qubit[®] dsDNA HS Assay kit (Life Technologies Corp., Carlsbad, California).

2.3. Mitochondrial DNA analyses

As a preliminary indicator of the amount of genetic variation present between the two strains of *A. itadori*, we examined the amount of sequence variation in the “Barcoding” region of the mitochondrial gene Cytochrome Oxidase I (COI) for two individuals of each strain. PCR reactions were carried out on a BioRad Dyad programmable thermocycler (BioRad Laboratories, Inc., Hercules, California) using the primer pair LCO and HCO (Folmer et al., 1994) and followed the conditions presented by Hebert et al. (2003). Promega GoTaq DNA polymerase, buffers, and dNTPs (Promega, Corp., Madison, Wisconsin) were used at the following concentrations; 5 µl of Promega GoTaq buffer, 0.5 µl of dNTP, 0.5 µl of both forward and reverse primers, 0.2 µl of Promega GoTaq taq polymerase, 1 µl of DNA, and sufficient HPLC grade H₂O added to bring the final volume to 25 µl. PCR products were held at 17°C before being visualized on a 1.5% agarose gel. DNA sequencing of both forward and reverse fragments was performed at the DNA Sequencing Facility at the University of California Berkeley, and sequence results were visualized and edited using Geneious Pro v. 5.6.2 (Drummond et al., 2012). The COI sequences from the two strains differ in the presence of a binding site for two commonly used restriction enzymes (HpaII and MspI). To verify whether these restriction enzymes could be used in future surveys we amplified the target

fragment of COI from 27 individuals of both the Hokkaido and Kyushu strains and 41 hybrid individuals reared on giant knotweed (20 H♀*K♂ and 21 K♀*H♂). The resulting PCR product was then digested using the MspI restriction enzyme (New England BioLabs, Inc., Ipswich, Massachusetts) following manufacturer recommendations (1 unit of enzyme for 1 ug of DNA). The resulting products were visualized by gel electrophoresis using a 1.5% agarose and stained using SYBR® Safe DNA gel stain (ThermoFisher Scientific, Co., Waltham, Massachusetts).

2.4. Next-generation sequencing

To achieve sufficient genomic material for Next-generation Sequencing (NGS), ten individuals were pooled for both the Hokkaido and Kyushu strains of *A. itadori* and placed into separate labeled 1.5ml micro-centrifuge tubes. Genomic DNA was extracted from both of the pooled samples using the Qiagen Gentra Pure Gene DNA Extraction Kit following the manufacturer's instructions. Sequencing libraries for each pooled extract were then created using the PrepX™ ILM DNA Library Kit (Wafergen Biosystems, Inc., Fremont, California) at the Functional Genomics Laboratory at the University of California Berkeley, and each library was constructed using a different Illumina (Illumina, Inc., San Diego, California) barcoding primer. Sequencing libraries were examined for fragment length distribution and concentrations using a 2100 Expert Bioanalyzer (Agilent Technologies, Santa Clara, California), and a KAPA Biosystems Library Quantification Kit (KAPA Biosystems, Wilmington, Massachusetts). The two libraries were then combined and sequenced using a single run of an Illumina HiSeq2000

sequencer at the Vincent J. Coates Genomics Sequencing Laboratory, University of California, Berkeley.

Low quality Illumina sequencing reads were removed and summary statistics were calculated using the FASTX-Toolkit (Gordon, 2014). The Illumina sequencing reads that passed the fastq quality filter were then assembled into contigs using the *de novo* assembly program Velvet 1.1.06 (Zerbino and Birney, 2008) with a k-mer length of 67. To determine the genomic origin for each of the assembled contigs we compared them to sequences in the NCBI GenBank Database (Benson et al., 2013) using the blast-n toolkit (Altschul et al., 1990). Any contig with an E-value >0.05 was considered unidentified. For contigs with an E-value ≤ 0.05 the single top-match sequence was used to classify them as either of insect, microbial (bacteria and/or archaea), viral, or “other” origin.

2.5. SNP array design and genotyping

To identify polymorphic sites for subsequent genotyping analyses, we used the software package SAMtools (Li et al., 2009) to map the individual Illumina sequence reads to our *de novo* contig assemblies, and to summarize the number, locations and quality scores for polymorphic sites. The resulting “.vcf” report was then filtered to remove contigs that included multiple polymorphic sites, polymorphic sites receiving a quality score <100 , and polymorphic sites identified as indels. Using this list of candidate SNPs, two SNP arrays were then designed using Agena Biosciences’s Online Assay Design Suite v 1.0 (Agena Bioscience, Inc., San Diego, California) following the manufacturer’s instructions with default settings, and assembled by the Genotyping

Facility at the University of Arizona. A first array (hereafter the “full array”) included 141 SNPs chosen based on their quality scores from the .vcf report, and a second array (hereafter the “reduced array”) included a subset of those SNPs from the first array (29 SNPs). The full array was then used to genotype 95 individuals (27 each of the pure Hokkaido (H) and Kyushu (K) strains reared on giant and Japanese knotweeds [OSU], respectively, and 20 H♀*K♂ and 21 K♀*H♂ hybrids reared on giant knotweed [AAFC]). The reduced array was then used to genotype an additional 95 individuals (12 pure Hokkaido reared on giant knotweed [AAFC], 23 pure Kyushu reared on Japanese knotweed [AAFC], 24 H♀*K♂ and 13 K♀*H♂ and hybrids reared on Japanese knotweed [AAFC], and 23 H♀*K♂ hybrids reared on Bohemian knotweed [AAFC]). Genotyping was conducted on an Agena Bioscience MassArray® MALDI-TOF at the Genotyping Facility at the University of Arizona where DNA concentrations were normalized before analysis. Briefly, under this platform, primer pairs from each array were used to amplify ~100 bp fragments that include a target SNP, the PCR products were then crystalized to the Agena Bioscience SpectroChip, crystals were ionized, and differences in the size of each molecule (i.e. allele) were measured by mass spectrometry. This platform is highly scalable, allowing for the analysis of 96 (up to 144 SNPs per array) to 394 individuals (up to 36 SNPs per array) in a single run, and hence the reason for testing both a full array of 141 SNPs and a reduced array of 29 SNPs.

2.6. Population structure and hybridization analyses

Because the genotyping included SNPs from the endosymbionts *C. ruddii* and *Sodalis* spp., before analysis we concatenated the SNPs for both species and rescored

them as haplotypes so that only two loci (one for each endosymbiont species) would be included in subsequent analyses. For those individuals of *A. itadori* genotyped using the full array, we then estimated their probability of assignment (Q) to distinct genetic clusters, and the optimal number of genetic clusters represented by individuals in the dataset for a potential number of clusters (K) of 1 through 10 using the program fastStructure (Raj et al., 2014). For K = 2 we assigned individuals as “pure” if they received a score of $Q \geq 0.8$ to one of the two clusters, or as a “hybrid” if they received a score of $Q < 0.8$ (Havill et al., 2012). We also estimated the probability of assignment (Z) that an individual was of parental, F1, F2, or backcrossed origin (however, we combined the assignment probabilities from all the hybrid classes to present a single Z value for hybrid status) with the software program NewHybrids (Anderson, 2008; Anderson and Thompson, 2002). Our simulations for these analyses used four independent runs of 10 million generations each, random starting values, uniform priors for the estimates of theta and pi, and a burn in period of 1 million generations. These same methods were then used to estimate the probability of assignment for individuals genotyped with the reduced array. In this case, because the SNPs used in the reduced array were also represented as a subset of those from the full array, we combined the genotyping results from both arrays and removed all SNPs that were unique to the full array before analysis. We graphed our results using code in R (R Core Team, 2014).

3. Results

3.1. Mitochondrial DNA analyses

We sequenced a 678-basepair fragment of mitochondrial locus COI from both the Hokkaido and Kyushu strains of *A. itadori*. Within strains of *A. itadori* there were no basepair differences; however, there were four basepair differences (all represented third codon position synonymous mutations) between the strains. Using the restriction enzyme *MspI* we were able to successfully differentiate the maternal lineages of the “pure” individuals tested (Fig. S1) with the product from Hokkaido individuals being cut into two fragments of ~ 250 bp and of ~ 430 bp while the product from Kyushu individuals remained uncut (~ 680 bp). Unexpectedly, hybrid individuals from both the H♀*K♂ and the K♀*H♂ crosses reared on giant knotweed displayed the two products characteristic of the Hokkaido strain. Sequences of COI from the two strains of *A. itadori* can be found on GenBank with Accession Numbers KP113670-KP113673.

3.2. Next-generation sequencing

Our Illumina HiSeq run resulted in 71,390,662 and 83,941,610 100 bp reads for the Hokkaido and Kyushu libraries, respectively, of which 66,392,260 and 77,812,449 passed the fastq quality filter. Raw sequence reads for both strains can be found at NCBI BioSample (Accession Number SUB724178). The full list of assembled contigs with E-values <0.05 can be found in Table S1, along with the single top-match sequence in the NCBI GenBank Database and its likely genomic origin (e.g., insect, microbial, viral, or other). In total we constructed 21,177 contigs with an average length of 731 bp and an average coverage of 9.5x. Of these, 13,360 contigs were unidentified, 4,961 were classified as of insect origin, 1,375 as of microbial origin (i.e., either bacteria or archaea), 39 as of viral origin, and 2,042 as of other origins. Whether the large number of

unidentified contigs represent psyllid DNA is currently unknown, and may be more easily determined once the International Psyllid Genome consortium has assembled and annotated the genome of the Asian citrus psyllid, *Diaphorina citri* Kuwayama, (NCBI BioProject: [PRJNA29447](#)). Only three of the 21,177 contigs matched a known reproductive manipulator (*Wolbachia*), therefore the presence of known reproductive manipulators in this system seems unlikely.

Among these contigs we identified 133,244 polymorphic sites using SAMtools that included 125,915 SNPs and 7,329 indels, of which 560 potential SNP markers passed our quality filters. From these 560 potential SNP markers, 141 SNPs were identified for genotyping analyses based on their quality scores and their compatibility for multiplexing. A list of these 141 SNPs, their expected polymorphic states, and 300 bp of FASTA formatted sequence data in both 5' and 3' directions are presented in Table S2.

3.3. SNP array design and genotyping

All individuals were successfully genotyped in the full array, and of the 141 SNPs included in the full array, 119 were amplified in more than 50% of the *A. itadori* samples. Thirty-eight of these represented SNPs from contigs that matched published insect genomic data, 80 from contigs that matched microbial genomic data, and one from a contig that matched viral genomic data. Of those SNPs from contigs that matched microbes, the majority (60) matched sequences from species of *Sodalis*, and four matched sequences from the primary endosymbiont *C. ruddii*. Ninety-two of the 95 individuals were successfully genotyped in the reduced array, and, of the 29 SNPs included in the reduced array, 24 were amplified in more than 90% of individuals, one was amplified in

more than 60% of individuals, and the remaining four SNPs failed to amplify. These SNPs represented a subset of those from the full array, and included 13 from contigs that matched insect genomic data, three from contigs that matched sequences for *C. ruddii*, and nine from contigs that matched sequences for *Sodalis* spp. The results of our SNP genotyping runs for the full and reduced arrays are presented in Tables S3 and S4, respectively.

3.4 Population structure and hybridization analyses

Before analysis, the genotyping results from each array were then trimmed to include only those SNPs that were successfully genotyped in >75% of individuals (Tables S3 and S4). The results from our fastStructure analyses of the full array indicated that the optimal number of genetic clusters represented in this dataset was $K = 4$. The probabilities of assignment for each individual at $K = 2$ can be seen in Fig. 1, and the results for $K = 3$ and $K = 4$ can be seen in Fig. S2. At $K = 2$, fastStructure correctly assigned all individuals as either pure or hybrids. At $K = 3$ and $K = 4$, fastStructure correctly assigned all but two individuals from the Hokkaido and Kyushu strains as pure, and all hybrid individuals were either assigned to a third genetic cluster (at $K = 3$) or two additional genetic clusters (at $K = 4$) with $Q > 0.8$. Similarly, NewHybrids correctly assigned individuals to pure and hybrid classes with $Z \geq 0.8$ (Fig. S3).

The results from our fastStructure analyses of the reduced array indicated that the optimal number of genetic clusters represented in this dataset was $K = 2$. The probability of assignment (Q) for each individual at $K = 2$ can be seen in Fig. 2, and the results for $K = 3$ and $K = 4$ can be seen in Fig. S4. At $K = 2$, 36 of the 38 individuals from the

Hokkaido strain and 42 of 50 individuals from the Kyushu strain were correctly assigned ($Q \geq 0.8$) to their respective strains, 57 of the 99 hybrids individuals were correctly classified ($Q < 0.8$) as hybrids, and one individual from the Hokkaido strain was miss-assigned to the Kyushu strain ($Q = 0.82$). As this individual was the only pure individual assigned to the wrong strain in both datasets, it likely represents a mislabeled DNA extract. In contrast to the analysis of the full array, while NewHybrids correctly assigned all pure individuals (except for our likely mislabeled extract) with the probability of assignment ($Z > 0.95$; Fig. S5) it failed to correctly identify hybrid individuals and assigned all hybrids to either the pure Hokkaido or Kyushu strains with high probabilities of assignment ($Z > 0.95$).

3.5 Endosymbiont haplotypes

In our analysis of the full array (Table S3), all individuals from the Kyushu strain had a single *C. ruddii* haplotype (labeled *Ca* haplotype 1) and all individuals from the Hokkaido strain had a different *C. ruddii* haplotype (labeled *Ca* haplotype 2), and all hybrid individuals reared from giant knotweed (including $H_{\text{♀}} * K_{\text{♂}}$ and $K_{\text{♀}} * H_{\text{♂}}$ crosses) also had *Ca* haplotype 2 (Fig. 1). For *Sodalis* sp., each strain had individuals with a strain-specific haplotype (labeled *So* haplotype 1 for Kyushu [$n = 23$] and *So* haplotype 2 for Hokkaido [$n = 12$]). In addition, four individuals from Kyushu and 14 individuals from Hokkaido had an additional *Sodalis* haplotype labeled *So* haplotype 3 (Fig. 1). For one individual from Hokkaido it was unclear whether this individual had *So* haplotype 2 or 3 because SNP062, which differentiates these two haplotypes, failed to amplify. SNP062 also failed to amplify in a large proportion of individuals (~40%) from the

reduced array (Table S4), therefore when analyzing the reduced array we combined *So* haplotypes 2 and 3 into a single haplotype group (labeled *So* shared haplotype; Fig. 2). From the reduced array, all individuals from the Kyushu strain, all K♀*H♂ hybrids from Japanese knotweed, and the likely mislabeled Hokkaido individual had *Ca* haplotype 1 (Fig. 2). All but four of these individuals (all from the Kyushu strain) also had *So* haplotype 1. The four remaining individuals from the Kyushu strain, all individuals from the Hokkaido strain and all other hybrids had *Ca* haplotype 2 and *So* shared haplotype.

4. Discussion

Long-term monitoring is typically missing from most classical biological control programs for pests and weeds (McCoy and Frank, 2010; Mills, 2000). For programs where multiple strains of a natural enemy have been introduced, the long-term effects of hybridization between the previously distinct strains is unclear, but laboratory evidence suggests the potential for either reduced fitness through hybrid breakdown (Messing and AliNiazee, 1988; Bean et al., 2013; Hoffmann et al., 2002) or increased fitness through hybrid vigor (Szűcs et al., 2012). One reason for this lack of clarity is that long-term monitoring of hybridizing populations has been difficult to implement due to the large numbers of independent loci required for accurate assignment of individuals to pure or hybrid classes (Vaha and Primmer, 2006). Using NGS technologies, we developed and validated two SNP arrays to distinguish between pure and hybrid individuals of two strains of *A. itadori*. Based on a comparison of assignment accuracy between these two arrays using frequently used assignment methods (fastStructure [Raj et al., 2014], and NewHybrids [Anderson, 2008; Anderson and Thompson, 2002]), we found that our full

array correctly identified all pure and hybrid individuals. In contrast, while the reduced array correctly identified the majority of pure individuals this array failed to identify individuals of hybrid origin using either fastStructure or NewHybrids. Therefore, post-release surveys of *A. itadori* should use, as a minimum, the SNPs found in the full array.

One curious result from these analyses was that individuals from the $K_{\text{♀}}*H_{\text{♂}}$ hybrid cross reared on giant knotweed had the maternally inherited mtDNA and endosymbiont haplotypes of the Hokkaido strain rather than the Kyushu strain. While it is possible that original source population collected in Kyushu contained individuals that had the Hokkaido mtDNA and endosymbiont haplotypes, none of the surveyed individuals from the pure Kyushu colony had mtDNA or endosymbiont haplotypes matching the Hokkaido colony. Given the large fitness differences noted for F1 adults from each strain reared on their preferred versus non-preferred hosts (Grevstad et al., 2013; Figure 1), and that colonies were allowed to establish over several generations, it is likely that this pattern is the result of contamination in the $K_{\text{♀}}*H_{\text{♂}}$ hybrid colony either by a Hokkaido female during establishment, or by a $H_{\text{♀}}*K_{\text{♂}}$ hybrid female during subsequent generations. It is interesting to note that if contamination did occur, that the descendants of the Hokkaido or $H_{\text{♀}}*K_{\text{♂}}$ female(s) must have rapidly outcompeted the descendants from the Kyushu females. This competitive advantage highlights the importance of the introduction of both strains for the biological control of invasive knotweeds, and in future studies we hope to explore whether these fitness differences are the result of maternal effects (e.g., Newcombe et al., 2015) or are conferred to their hosts from their endosymbionts (Bennett and Moran, 2015). Understanding the relative importance of these factors may help to inform the distributions and efficacies of these

two strains of *A. itadori* if they are approved and released for the biological control of invasive knotweeds in North America.

5. Conclusions

In this study we developed two SNP arrays for differentiating two strains of the knotweed psyllid, *A. itadori*, a candidate biological control agent currently being reviewed for release in North America to reduce the abundance of invasive knotweeds. We found that one of these arrays (labeled the full array) was able to accurately identify individuals of both pure and hybrid origin. If *A. itadori* is approved and released, this array will allow for the rapid identification of hybrid individuals.

Acknowledgements

We thank Raymond Wilson and Karma Tiberg for their laboratory assistance, Justin Choi and Karen Lundy for their assistance in producing the Illumina Sequencing Libraries, and Colleen Ramsower and Crystal Richt for their assistance in designing the SNP arrays and in performing the genotyping runs. We also thank Rasmus Nielsen, George Roderick and two anonymous reviewers for their constructive comments on an earlier version of this manuscript. Funding for this study was provide by USDA Forest Service Cooperative agreement grants to RVD (13-CA-11420004-049), and FSG (10-CA-11420004-255), a Washington State Department of Agriculture grant awarded to FSG (K1227), an Agriculture and AgriFoods Canada research grant awarded to RSB (#1713, #1213), and a Robert van den Bosch Memorial Fellowship awarded to JCA.

Data Archival Location

435 Illumina raw sequence reads are available through BioSample: Accession SUB724178

436 Mitochondrial sequences are available through GenBank: Accessions (KP113670-

437 KP113673)

438

439

439

440 **References**

- 441 Allendorf, F.W., Leary, R.F., Spruell, P., Wenburg, J.K., 2001. The problems with
442 hybrids: setting conservation guidelines. *Trends Ecol. Evol.* 16, 613-622. DOI:
443 10.1016/s0169-5347(01)02290-x
- 444 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local
445 alignment search tool. *J. Mol. Biol.* 215, 403-410. DOI: 10.1016/s0022-
446 2836(05)80360-2
- 447 Anderson, E.C., 2008. Bayesian inference of species hybrids using multilocus dominant
448 genetic markers. *Philos. T. Roy. Soc. B.* 363, 2841-2850. DOI:
449 10.1098/rstb.2008.0043
- 450 Anderson, E.C., Thompson, E.A., 2002. A model-based method for identifying species
451 hybrids using multilocus genetic data. *Genetics.* 160, 1217-1229.
- 452 Baumann, P., 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-
453 sucking insects. *Annu. Rev. Microbiol.* 59, 155-189. DOI:
454 10.1146/annurev.micro.59.030804.121041
- 455 Bean, D.W., Kazmer, D.J., Gardner, K., Thompson, D.C., Reynolds, B., Keller, J.C.,
456 Gaskin, J.F., 2013. Molecular genetic and hybridization studies of *Diorhabda* spp.
457 released for biological control of *Tamarix*. *Invasive Plant Sci. Manag.* 6, 1-15.
458 DOI: 10.1614/ipism-d-11-00093.1
- 459 Beerling, D.J., Bailey, J.P., Conolly, A.P., 1994. *Fallopia japonica* (Houtt) Ronse,
460 Decraene (*Reynoutria japonica* Houtt, *Polygonum cuspidatum* Sieb and Zucc). *J.*
461 *Ecol.* 82, 959-979. DOI: 10.2307/2261459

462 Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J.,
 463 Sayers, E.W., 2013. GenBank. Nucleic Acids Res. 41, D36-D42. DOI:
 464 10.1093/nar/gks1195

465 Bennett, G.M., Moran, N.A., 2015. Heritable symbiosis: the advantages and perils of an
 466 evolutionary rabbit hole. Proc. Natl. Acad. Sci. USA. 112, 10169-10176. DOI:
 467 10.1073/pnas.1421388112

468 Bordenstein, S.R., O'Hara, F.P., Werren, J.H., 2001. *Wolbachia*-induced incompatibility
 469 precedes other hybrid incompatibilities in *Nasonia*. Nature. 409, 707-710.

470 Branca, A., Le Ru, B.P., Vavre, F., Silvain, J.-F., Dupas, S., 2011. Intraspecific
 471 specialization of the generalist parasitoid *Cotesia sesamiae* revealed by
 472 polyDNAvirus polymorphism and associated with different *Wolbachia* infection.
 473 Mol. Ecol. 20, 959-971. DOI: 10.1111/j.1365-294X.2010.04977.x

474 Cheypppe-Buchmann, S., Bon, M.-C., Warot, S., Jones, W., Malausa, T., Fauvergue, X.,
 475 Ris, N., 2011. Molecular characterization of *Psytallia lounsburyi*, a candidate
 476 biocontrol agent of the olive fruit fly, and its *Wolbachia* symbionts as a pre-
 477 requisite for future intraspecific hybridization. BioControl. 56, 713-724. DOI:
 478 10.1007/s10526-011-9346-x

479 Chrudimský, T., Husník, F., Nováková, E., Hypša, V. 2012. *Candidatus Sodal*
 480 *melophagi* sp. nov.: phylogenetically independent comparative model to the
 481 Tsetse fly symbiont *Sodalis glossinidius*. PLoS One. 7, e40354. DOI:
 482 10.1371/journal.pone.0040354

483 Djeddour, D.H., Shaw, R.H., 2010. The biological control of *Fallopia japonica* in Great
484 Britain: review and current status. *Outlooks Pest Manag.* 21, 15-18. DOI:
485 10.1564/21feb04

486 Dobzhansky, T., 1936. Studies on hybrid sterility. II. Localization of sterility factors in
487 *Drosophila pseudoobscura* hybrids. *Genetics* 21,113-135.

488 Dres, M., Mallet, J., 2002. Host races in plant-feeding insects and their importance in
489 sympatric speciation. *Philos. T. Roy. Soc. B.* 357, 471-492. DOI:
490 10.1098/rstb.2002.1059

491 Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M.,
492 Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2012. Geneious
493 v 5.6.2, Available from <http://www.geneious.com>.

494 Ellison, C. K., Niehuis, O., Gadau, J., 2008. Hybrid breakdown and mitochondrial
495 dysfunction in hybrids of *Nasonia* parasitoid wasps. *J. Evol. Biol.* 21, 1844-1851.
496 DOI: 10.1111/j.1420-9101.2008.01608.x

497 Engelstaedter, J., Hurst, G.D.D., 2009. The ecology and evolution of microbes that
498 manipulate host reproduction. *Annu. Rev. Ecol.Evol. Syst.* 40, 127-149. DOI:
499 10.1146/annurev.ecolsys.110308.120206

500 Feder, J.L., Berlocher, S.H., Roethele, J.B., Dambroski, H., Smith, J.J., Perry, W.L.,
501 Gavrilovic, V., Filchak, K.E., Rull, J., Aluja, M., 2003. Allopatric genetic origins
502 for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad.*
503 *Sci. USA.* 100, 10314-10319. DOI: 10.1073/pnas.1730757100

504 Ferrari, J., Vavre, F., 2011. Bacterial symbionts in insects or the story of communities
 505 affecting communities. *Philos. T. Roy. Soc. B.* 366, 1389-1400. DOI:
 506 10.1098/rstb.2010.0226

507 Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for
 508 amplification of mitochondrial cytochrome c oxidase subunit I from diverse
 509 metazoan invertebrates. *Mol. Mar. Biol. Biotech.* 3, 294-299.

510 Gaskin, J.F., Schwarzlaender, M., Grevstad, F.S., Haverhals, M.A., Bouchier, R.S.,
 511 Miller, T.W., 2014. Extreme differences in population structure and genetic
 512 diversity for three invasive congeners: knotweeds in western North America. *Biol.*
 513 *Invasions.* 16, 2127-2136. DOI: 10.1007/s10530-014-0652-y

514 Gerber, E., Krebs, C., Murrell, C., Moretti, M., Rocklin, R., Schaffner, U., 2008. Exotic
 515 invasive knotweeds (*Fallopia* spp.) negatively affect native plant and invertebrate
 516 assemblages in European riparian habitats. *Biol. Conserv.* 141, 646-654. DOI:
 517 10.1016/j.biocon.2007.12.009

518 Gordon, A., 2014. FASTX-Toolkit, http://hannonlab.cshl.edu/fastx_toolkit/.

519 Grevstad, F., Shaw, R., Bouchier, R., Sanguankee, P., Cortat, G., Reardon, R.C., 2013.
 520 Efficacy and host specificity compared between two populations of the psyllid
 521 *Aphalara itadori*, candidates for biological control of invasive knotweeds in
 522 North America. *Biol. Control.* 65, 53-62. DOI: 10.1016/j.biocontrol.2013.01.001

523 Harrison, R.G., Larson, E.L., 2014. Hybridization, introgression, and the nature of
 524 species boundaries. *J. Hered.* 105, 795-809. DOI: 10.1093/jhered/esu033

525 Havill, N.P., Davis, G., Mausel, D.L., Klein, J., McDonald, R., Jones, C., Fischer, M.,
 526 Salom, S., Caccone, A., 2012. Hybridization between a native and introduced

527 predator of Adelgidae: An unintended result of classical biological control. Biol.
528 Control. 63, 359-369. DOI: 10.1016/j.biocontrol.2012.08.001

529 Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J. R., 2003. Biological identifications
530 through DNA barcodes. Proc. Roy. Soc. B. 270, 313-321. DOI:
531 10.1098/rspb.2002.2218

532 Heraty, J. M. 2009. Parasitoid biodiversity and insect pest management, in: Footitt, R.G.,
533 Adler, P.H. (Eds.), Insect Biodiversity: Science and Society. Wiley-Blackwell,
534 Oxford, pp. 445-462.

535 Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., Werren, J.H., 2008.
536 How many species are infected with *Wolbachia*? - a statistical analysis of current
537 data. FEMS Microbiol. Lett. 281, 215-220. DOI: 10.1111/j.1574-
538 6968.2008.01110.x

539 Hoddle, M.S., 2004. Restoring balance: using exotic species to control invasive exotic
540 species. Conserv. Biol. 18, 38-49. DOI: 10.1111/j.1523-1739.2004.00249.x

541 Hoffmann, J.H., Impson, F.A.C., Volchansky, C.R., 2002 Biological control of cactus
542 weeds: implications of hybridization between control agent biotypes. J. Appl.
543 Ecol. 39, 900-908. DOI: 10.1046/j.1365-2664.2002.00766.x

544 Hopper, K.R., de Farias, A.A.M.I., Woolley, J.B., Heraty, J.M., Britch, S.C., 2005.
545 Genetics: relation of local populations to the whole "species" - implications for
546 host range tests, in: Hoddle, M. (Ed.), Second International Symposium on
547 Biological Control of Arthropods (FHTET-2005-08). US Forest Service,
548 Washington D.C., pp. 665-670.

549 Hurst, L.D., Pomiankowski, A., 1991. Causes of sex-ratio bias may account for unisexual
550 sterility in hybrids - a new explanation of haldanes rule and related phenomena.
551 Genetics. 128, 841-858.

552 Kirkpatrick, M., 2000. Reinforcement and divergence under assortative mating. Proc.
553 Roy. Soc. B. 267, 1649-1655. DOI: 10.1098/rspb.2000.1191

554 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
555 Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009.
556 The sequence alignment/map format and SAMtools. Bioinformatics. 25, 2078-
557 2079. DOI: 10.1093/bioinformatics/btp352

558 Lukhtanov, V.A., Kandul, N.P., Plotkin, J.B., Dantchenko, A.V., Haig, D., Pierce, N.E.,
559 2005. Reinforcement of pre-zygotic isolation and karyotype evolution in
560 *Agrodiaetus* butterflies. Nature. 436, 385-389. DOI: 10.1038/nature03704

561 Maerz, J.C., Blossey, B., Nuzzo, V., 2005. Green frogs show reduced foraging success in
562 habitats invaded by Japanese knotweed. Biodivers. Conserv. 14, 2901-2911. DOI:
563 10.1007/s10531-004-0223-0

564 Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229-
565 237. DOI: 10.1016/j.tree.2005.02.010

566 McCoy, E.D., Frank, J.H., 2010. How should the risk associated with the introduction of
567 biological control agents be estimated? Agric. For. Entomol. 12, 1-8. DOI:
568 10.1111/j.1461-9563.2009.00446.x

569 Messing, R.H., AliNiazee, M.T., 1988. Hybridization and host suitability of 2 biotypes
570 of *Trioxys pallidus* (Hymenoptera, Aphidiidae). Ann. Entomol. Soc. Am. 81, 6-9.

571 Mills, N., 2000. Biological control: The need for realistic models and experimental
572 approaches to parasitoid introductions, in: Hochberg, M.E., Ives, A.R. (Eds.),
573 Parasitoid Population Biology, Princeton University Press, Princeton, pp. 217-234

574 Moran N.A., Bennett G.M., 2014. The tiniest tiny genomes. *Ann. Rev. Microbiol.* 68,
575 195-215. DOI: 10.1146/annurev-micro-091213-112901

576 Murrell, C., Gerber, E., Krebs, C., Parepa, M., Schaffner, U., Bossdorf, O., 2011.
577 Invasive knotweed affects native plants through allelopathy. *Am. J. Bot.* 98, 38-
578 43. DOI: 10.3732/ajb.1000135

579 Newcombe, D., Moore, P.J., Moore, A.J., 2015. The role of maternal effects in adaptation
580 to different diets. *Biol. J. Linn. Soc.* 114, 202-211. DOI: 10.1111/bij.12408

581 Orr, H.A., Presgraves, D.C., 2000. Speciation by postzygotic isolation: forces, genes and
582 molecules. *Bioessays.* 22, 1085-1094.

583 Presgraves, D.C., 2002. Patterns of postzygotic isolation in Lepidoptera. *Evolution.* 56,
584 1168-1183.

585 R Core Team, 2014. R: A language and environment for statistical computing. R
586 Foundation for Statistical Computing, Vienna, Austria. URL [http://www.R-](http://www.R-project.org/)
587 [project.org/](http://www.R-project.org/).

588 Raj, A., Stephens, M., Pritchard, J.K., 2014. fastSTRUCTURE: Variational inference of
589 population structure using large SNP data sets. *Genetics.* 197, 573-589. DOI:
590 10.1534/genetics.114.164350

591 Ruiz-Montoya, L., Nunez-Farfan, J., 2013. Testing local host adaptation and phenotypic
592 plasticity in a herbivore when alternative related host plants occur sympatrically.
593 *PLoS One.* 8, e79070. DOI: 10.1371/journal.pone.0079070

594 Schwarz, D., Matta, B.M., Shakir-Botteri, N.L., McPheron, B.A., 2005. Host shift to an
595 invasive plant triggers rapid animal hybrid speciation. *Nature*. 436, 546-549. DOI:
596 10.1038/nature03800

597 Schwenk, K., Brede, N., Streit, B., 2008. Introduction. Extent, processes and evolutionary
598 impact of interspecific hybridization in animals. *Philos. T. Roy. Soc. B*. 363,
599 2805-2811. DOI: 10.1098/rstb.2008.0055

600 Shaw, R.H., Bryner, S., Tanner, R., 2009. The life history and host range of the Japanese
601 knotweed psyllid, *Aphalara itadori* Shinji: potentially the first classical biological
602 weed control agent for the European Union. *Biol. Control*. 49, 105-113. DOI:
603 10.1016/j.biocontrol.2009.01.016

604 Siemens, T.J., Blossey, B., 2007. An evaluation of mechanisms preventing the growth
605 and survival of two native species in invasive Bohemian knotweed (*Fallopia x*
606 *bohemica*, Polygonaceae). *Am. J. Bot.* 94, 776-783. DOI: 10.3732/ajb.94.5.776

607 Sloan D.B., Moran N.A., 2012. Genome reduction and co-evolution between the primary
608 and secondary bacterial symbionts of psyllids. *Mol. Biol. Evol.* 29, 3781-3792.
609 DOI: 10.1093/molbev/mss180

610 Szűcs, M., Eigenbrode, S.D., Schwarzlaender, M., Schaffner, U., 2012. Hybrid vigor in
611 the biological control agent, *Longitarsus jacobaeae*. *Evol. Appl.* 5, 489-497. DOI:
612 10.1111/j.1752-4571.2012.00268.x

613 Turelli, M., 1994. Evolution of incompatibility-inducing microbes and their hosts.
614 *Evolution*. 48, 1500-1513.

615 Vaha, J.P., Primmer, C.R., 2006. Efficiency of model-based Bayesian methods for
616 detecting hybrid individuals under different hybridization scenarios and with

617 different numbers of loci. *Mol. Ecol.* 15, 63-72. DOI: 10.1111/j.1365-
618 294X.2005.02773.x

619 Van Driesche, R.G., Carruthers, R.I., Center, T., Hoddle, M.S., Hough-Goldstein, J.,
620 Morin, L., Smith, L., Wagner, D.L., Blossey, B., Brancatini, V., Casagrande, R.,
621 Causton, C.E., Coetzee, J.A., Cuda, J., Ding, J., Fowler, S.V., Frank, J.H.,
622 Fuester, R., Goolsby, J., Grodowitz, M., Heard, T.A., Hill, M.P., Hoffmann, J.H.,
623 Huber, J., Julien, M., Kairo, M.T.K., Kenis, M., Mason, P., Medal, J., Messing,
624 R., Miller, R., Moore, A., Neuenschwander, P., Newman, R., Norambuena, H.,
625 Palmer, W.A., Pemberton, R., Panduro, A.P., Pratt, P.D., Rayamajhi, M., Salom,
626 S., Sands, D., Schooler, S., Schwarzlander, M., Sheppard, A., Shaw, R., Tipping,
627 P.W., van Klinken, R.D., 2010. Classical biological control for the protection of
628 natural ecosystems. *Biol. Control* 54, S2-S33. DOI:
629 10.1016/j.biocontrol.2010.03.003

630 Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly
631 using de Bruijn graphs. *Genome Res.* 18, 821-829. DOI: 10.1101/gr.074492.107
632
633

Figure legends

Fig. 1. Probability of assignment (Q) of each individual of *A. itadori* to one of two genetic clusters using the full array. Each column represents the proportional assignment of individuals to each genetic cluster. The two rows below represent the haplotypes of both the primary endosymbiont *Candidatus Carsonella ruddi* (*Ca*) (white = *Ca* haplotype 1, black = *Ca* haplotype 2) and the secondary endosymbiont *Sodalis* sp. (*So*) (white = *So* haplotype 1, gray = *So* haplotype 2, black = *So* haplotype 3, or horizontal lines = undetermined) for each individual. We have divided the graphs to more clearly show the strains of origin.

Fig. 2. Probability of assignment (Q) of each individual of *A. itadori* to one of two genetic clusters using the reduced array. Each column represents the proportional assignment of individuals to one of up to four genetic clusters. The two rows below represent the haplotypes of both the primary endosymbiont *Candidatus Carsonella ruddii* (*Ca*) (white = *Ca* haplotype 1, black = *Ca* haplotype 2) and the secondary endosymbiont *Sodalis* sp. (*So*) (white = *So* haplotype 1, black = *So* shared haplotype) for each individual.

Figure 1

Figure 2