Using comparative genomics to develop a molecular diagnostic for the identification of an emerging pest Drosophila suzukii

Citation	Murphy, K. A., Unruh, T. R., Zhou, L. M., Zalom, F. G., Shearer, P. W., Beers, E. H., ... & Chiu, J. C. (2015). Using comparative genomics to develop a molecular diagnostic for the identification of an emerging pest Drosophila suzukii. Bulletin of entomological research, 105(3), 364-372. doi:10.1017/S0007485315000218
DOI	10.1017/S0007485315000218
Publisher	Cambridge University Press
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Bulletin of Entomological Research (2015) **105**, 364–372
doi:10.1017/S0007485315000218

Using comparative genomics to develop a molecular diagnostic for the identification of an emerging pest *Drosophila suzukii*

K.A. Murphy[1], T.R. Unruh[2], L.M. Zhou[1], F.G. Zalom[1], P.W. Shearer[3], E.H. Beers[4], V.M. Walton[5], B. Miller[5] and J.C. Chiu[1]*

[1]Department of Entomology and Nematology, College of Agricultural and Environmental Sciences, University of California, Davis, CA 95616, USA: [2]USDA-ARS, Wapato, WA 98951, USA: [3]Mid-Columbia Agricultural and Extension Center, Oregon State University, Hood River, OR 97031, USA: [4]Tree Fruit Research and Extension Center, Washington State University, Wenatchee, WA 98801, USA: [5]Department of Horticulture, Oregon State University, Corvallis, OR 97331, USA

Abstract

Drosophila suzukii (Spotted Wing Drosophila) has recently become a serious invasive pest of fruit crops in the USA, Canada, and Europe, leading to substantial economic losses. *D. suzukii* is a direct pest, ovipositing directly into ripe or ripening fruits; in contrast, other Drosophilids utilize decaying or blemished fruits and are nuisance pests at worst. Immature stages of *D. suzukii* are difficult to differentiate from other Drosophilids, posing problems for research and for meeting quarantine restrictions designed to prevent the spread of this pest in fruit exports. Here we used a combined phylogenetic and bioinformatic approach to discover genetic markers suitable for a species diagnostic protocol of this agricultural pest. We describe a molecular diagnostic for rapid identification of single *D. suzukii* larva using multiplex polymerase chain reaction. Our molecular diagnostic was validated using nine different species of *Drosophila* for specificity and 19 populations of *D. suzukii* from different geographical regions to ensure utility within species.

Keywords: spotted wing drosophila, *Drosophila suzukii*, species-specific PCR, invasive pest, comparative genomics

(Accepted 11 February 2015; First published online 25 March 2015)

Introduction

Drosophila suzukii Matsumura (Spotted Wing Drosophila [SWD]) is a recent invader to the USA, Canada, and Europe where it has become a serious economic pest of a variety of soft-skinned fruit crops, leading to substantial yearly crop losses that are in the tens of millions of dollars annually while increasing production costs (Hauser, 2011; Lee *et al.*, 2011; Walsh *et al.*, 2011; Calabria *et al.*, 2012; Cini *et al.*, 2012).

Unlike other cosmopolitan Drosophilids that oviposit in overripe and blemished fruits with no commercial value, the SWD female has evolved a serrated ovipositor that enables it to oviposit in ripe or ripening berries, e.g., caneberry and blueberry, and soft-skinned fruits, e.g., cherry, thus allowing the larvae to develop within the fruit. Adult SWD flies have distinct morphological characters, i.e., serrated ovipositors in females and spotted wings in males, which facilitate identification and differentiation from other non-pest Drosophilids that coinhabit the same agricultural settings and geographical regions. However, identification of SWD from larval stages or from poorly preserved specimens is difficult (Hauser, 2011 and M. Hauser, personal communication). Identification of SWD larvae requires researchers or fruit inspectors to rear the larvae to adulthood. This may require a week or more

*Author for correspondence
Phone: (530) 752-1839
Fax: (530) 754-9077
E-mail: jcchiu@ucdavis.edu

and such delays can result in substantial economic losses for exporters of these perishable fruits as well as delays to research on this important pest. Australia and New Zealand currently have a quarantine to safeguard against entry of SWD-infested fruits and related quarantines may be erected by other countries. Shipment delays due to the need for identification can reduce fruit quality and increase handling costs. A more efficient method to differentiate immature stages of *D. suzukii* from other non-pest Drosophilids is therefore warranted.

There are a number of commonly used polymerase chain reaction (PCR)-based molecular diagnostics that have been used for species identification, but these approaches often vary in complexity to obtain the results (Behura, 2006; Gariepy *et al*., 2007). DNA barcoding methods rely on the PCR amplification and subsequent sequencing of COI cytochrome c oxidase subunit I (COI) gene sequences to reveal species-specific polymorphisms (Hebert *et al*., 2003). Sequencing reactions may need to be outsourced to sequencing facilities, thereby adding cost and processing time, i.e., days. Alternatively, PCR-restriction fragment length polymorphism (RFLP), which involves PCR amplification of a specific gene fragment followed by restriction enzyme digestion, will yield species-specific restriction digestion patterns (Wyman & White, 1980). PCR-RFLP is a less costly method than the DNA barcoding approach since DNA sequencing is not necessary, but it still requires the extra cost of restriction enzymes as well as processing time of a day or more. A PCR-RFLP diagnostic for SWD has recently been developed (Kim *et al*., 2014), however, this diagnostic does not have the species level resolution required to distinguish between *D. suzukii* and the closely related *Drosophila subpulchrella*. Females of *D. subpulchrella* also have serrated ovipositors (Atallah *et al*., 2014) and have been shown to be a fruit pest in Asia (Atallah *et al*., 2014; Dhami & Kumarasinghe, 2014). It is therefore important to have a diagnostic marker that can differentiate between these two species to detect any future invasions of *D. subpulchrella*. Random amplified polymorphic DNA (RAPD) is another PCR-based approach for species identification, and can be designed without any prior knowledge of the genome sequence for the species of interest (Williams *et al*., 1990). Short random DNA primers are used to amplify genomic DNA, and can result in species-specific differences in amplification patterns due to polymorphisms between different genomes. Although RAPD is an easy one-step PCR method, the use of random primers can sometimes lead to complicated banding patterns that are difficult to interpret, and is generally considered to have lower resolving power (Gariepy *et al*., 2007). A better one-step PCR approach is the design of specific primer pair(s) that amplifies only the target species of concern. This approach can be used for diagnosis of a single key pest or multiple unique primers can be used for several pest species in a complex (Hebert *et al*., 2003). Finally, in addition to conventional PCR, fluorescence-based quantitative real-time PCR (Heid *et al*., 1996) has also been used for species diagnostic (Walsh *et al*., 2005). In fact, such an assay has recently been developed for SWD (Dhami & Kumarasinghe, 2014). It relies on high-resolution melt (HRM) analysis to detect sequence polymorphism between SWD and other *Drosophila* species within the COI gene sequence, and does not require gel electrophoresis for result readout. However, this type of assay requires a quantitative real-time PCR apparatus that is more expensive than conventional PCR machine and is generally not available in common laboratories or export sorting facilities.

Here, we describe an efficient PCR-based molecular diagnostic test that can be used to quickly (i.e., 1–2 h) identify all of the life stages of SWD and differentiate them from other common Drosophilids. Our molecular diagnostic test relies on a one-step PCR amplification using SWD-specific primers that were designed using the recently sequenced SWD genome (Chiu *et al*., 2013) and is not dependent on sequencing, e.g., as in DNA barcoding, restriction enzyme digestion, or the use of quantitative real-time PCR apparatus. In addition, we have optimized our assay to be compatible with the use of crude larval extract without genomic DNA extraction to increase efficiency and ease of use.

Materials and methods

Fly strains and rearing

All *Drosophila* species and strains tested in our studies as well as their sources and original collection sites are listed in table 1. For collections in CA, OR, WA, and MD, flies were either collected at locations (public land) for which specific permission was not required; or they were collected on private grower properties for which permission was obtained. Identities of growers were omitted as a courtesy. Fly collections did not involve endangered or protected species. All lines, once obtained from the sources, were maintained in Fisherbrand square, polyethylene, 6 oz. stock bottles (Fisher Scientific, Pittsburgh, PA) containing 50 ml of Bloomington stock center fly food recipe. Colonies were kept between 22 and 25°C in a cabinet incubation chamber (Percival Scientific, Inc., Perry, IA) with a 12:12 h light:dark cycle.

Identification of conserved and diverged genes through comparative genomic analysis of Drosophilids for the design of diagnostic primers

The most conserved and diverged *Drosophila* genes were determined by comparative genomic analysis of *D. suzukii* and 14 other *Drosophila* species spanning multiple groups in the subgenus *Sophophora* (*Drosophila melanogaster*, *Drosophila simulans*, *Drosophila sechellia*, *Drosophila yakuba*, *Drosophila erecta*, *Drosophila ananassae*, *Drosophila pseudoobscura*, *Drosophila persimilis*, *Drosophila willistoni*, *Drosophila takahashii*, and *Drosophila biarmipes*) and subgenus Drosophila (*Drosophila virilis*, *Drosophila mojavensis*, and *Drosophila grimshawi*). Sequence data to establish orthology using OrthologID (Chiu *et al*., 2006), a phylogenomic tool developed to identify orthologous genes from whole genomes, was originally retrieved from SpottedWingFlybase (Chiu *et al*., 2013) (*D. suzukii*), FlyBase (Marygold *et al*., 2013) (all species except *D. suzukii*, *D. takahashii*, *D. biarmipes*), and the *Drosophila* modENCODE Project (modENCODE Consortium, 2010) (*D. takahashii* and *D. biarmipes*). Whereas complete gene sets (amino acid sequences) can be retrieved for most species, they were not available for *D. takahashii* and *D. biarmipes*, the two species that are most closely related to *D. suzukii* as compared to the other *Drosophila* species included in our comparative genomics analysis. As part of a previous comparative analysis to examine the genomes of *D. suzukii* and all the aforementioned *Drosophila* species (Chiu *et al*., 2013), we annotated the complete gene sets for *D. takahashii* and *D. biarmipes*. The procedures for annotation and comparative genomic analysis using OrthologID (Chiu *et al*., 2006) have been described previously in Chiu *et al*. (2013).

Table 1. Species and strains used for testing the *D. suzukii* molecular diagnostic.

Species	Strain	Location of collection	Collector/source
D. suzukii	Colony #8	Bray's Landing, WA, USA N47.738425 W120.167644	Beers Laboratory, WSU[1]
D. suzukii	Colony #9	Daroga Park, WA, USA N47.705933 W120.19128	Beers Laboratory, WSU
D. suzukii	Colony #10	Royal City, WA, USA N46.837533 W119.5099	Beers Laboratory, WSU
D. suzukii	Colony #11	Rock Island, WA, USA N47.2350 W120.0727	Beers Laboratory, WSU
D. suzukii	Lime	Davis, CA, USA N38.55 W121.78	Begun Laboratory, UCD[2]
D. suzukii	Grape	Davis, CA, USA N38.55 W121.78	Begun Laboratory, UCD
D. suzukii	Genome strain, WT3 F10	Watsonville, CA, USA N36.94 W121.76	Begun Laboratory, UCD
D. suzukii	Mark Bolda, WAT	Watsonville, CA, USA N36.94 W121.76	Zalom Laboratory, UCD
D. suzukii	Wolfskill IFL WO-2	Winters, CA, USA N38.49 W121.98	Begun Laboratory, UCD
D. suzukii	HR3 F4	Hood River, OR, USA N45.410860 W121.321011	Shearer Laboratory, OSU[3]
D. suzukii	TD3 F5	The Dalles, OR, USA N45.351738 W121.131167	Shearer Laboratory, OSU
D. suzukii	PD3 F5	Parkdale, OR, USA N45.310333 W121.351362	Shearer Laboratory, OSU
D. suzukii	ARS	Corvallis, OR, USA N45.010035 W122.564377	Shearer Laboratory, OSU
D. suzukii	OS1	Corvallis, OR, USA N45.010035 W122.564377	Shearer Laboratory, OSU
D. suzukii	Wild population	South Korea	Betsey Miller, OSU
D. suzukii	Wild population	Scurelle, Trentino, Italy	Claudio loriatti, FEM-IASMA
D. suzukii	MTY	Ehime, Japan	Kopp Laboratory, UCD[2]
D. suzukii	Wild population	Oahu, Hawaii	Kopp Laboratory, UCD
D. suzukii	Wild population	MD, USA	Hamby Laboratory, UM[4]
D. biarmipes	Genome strain, 361.0-isol e-11	Ari Ksatr, Cambodia	Kopp Lab, UCD
D. erecta	14021-0224.01	Tucson Stock Center	Begun Laboratory, UCD
D. melanogaster	Oregon R	Roseburg, Oregon	Zalom Laboratory, UCD
D. subpulchrella	NGN5	Nagano, Japan	Begun Laboratory, UCD
D. simulans	W501	Genome strain	Begun Laboratory, UCD
D takahashii	Genome strain, 311.5-iso4	Yun Shui, Taiwan	Kopp Laboratory, UCD
D. willistoni	14030-0814-10	Guadeloupe Island, France	Begun Laboratory, UCD
D. yakuba	CY28	Cameroon, Africa	Begun Laboratory, UCD

[1]Washington State University.
[2]University of California, Davis.
[3]Oregon State University.
[4]University of Maryland.

For the identification of conserved genes, 2661 single copy orthologous gene sets present across all 15 *Drosophila* species were extracted from the entire collection of gene families generated by OrthologID using custom scripts. Maximum likelihood (ML) inference of the gene tree was performed for each orthologous set using RAxML version 7.4.2 (Stamatakis, 2006, 2012). The best of 36 protein substitution models under Γ-distributed rate heterogeneity was determined using the 'ProteinModelSelection.pl' script (Stamatakis, 2012) for each ortholog set. The average branch length was computed for each ML tree, and genes with shorter tree lengths were inferred as more conserved. This entire gene list, including average branch lengths, FlyBase annotation for *D. melanogaster* orthologs, and SpottedWingFlybase (http://spottedwingflybase.oregonstate.edu/) annotation for *D. suzukii* orthologs, is presented in Supplementary Table 1.

We used a similar approach to identify diverged genes. A total of 4781 orthologous gene sets present in single copies in both the *melanogaster* subgroup (*D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, and *D. melanogaster)* and the *suzukii-takahashii* subgroup (*D. suzukii*, *D. biarmipes*, and *D. takahashii*) were identified. These gene sets may include zero or multiple copies of orthologs from the other seven *Drosophila* species. This selection criterion was designed to be more relaxed than in the identification of conserved genes as the divergence of a gene in *D. suzukii* from its close relatives is more relevant than its overall divergence across all *Drosophila* species. ML phylogenetic inference was performed as above for each set of orthologs. To enable the calculation of average branch lengths for gene trees in the *melanogaster* and the *suzukii-takahashii* subgroups both independently and as a combined dataset, only gene sets with trees containing monophyletic *melanogaster* and *suzukii-takahashii* subgroups were retained. To reduce the chance of including inaccurately annotated genes or misidentified orthologs, which may lead to excessively long branches, we also heuristically excluded trees with a branch longer than the sum of all other branches in the same tree. Average branch lengths for the *melanogaster* and *suzukii-takahashii* clades of the resulting list of 2815 orthologous gene sets were calculated. Genes with longer lengths were identified as more diverged (Supplementary Table 2).

Primer design for multiplex PCR

The multiplex PCR diagnostic test relies on the use of two primer sets in a multiplex PCR reaction. The first set of primers was designed to amplify a product from any Drosophilid to confirm the presence of good quality DNA in the reaction and to verify the success of the PCR reaction. This internal control primer pair was designed from the coding region of a gene (*D. suzukii* ortholog [SpottedWingFlybase ID: DS10_00001395] of *sec61alpha* in *D. melanogaster* [FlyBase ID: FBgn0086357 and FBpp0078896]) that is highly conserved among the 15 *Drosophila* species we used for our comparative genomic analysis (Supplementary Table 1) to yield a 1248 base pair (bp) product (forward primer: 5′- ATCCCTTCTACTGGATC CGTG-3′ and reverse primer: 5′- ACAGCAGCGTGCCC ATG-3′) (fig. 1a).

The second set of primers was designed from a gene (*D. suzukii* ortholog [SpottedWingFlybase ID: DS10_00004458] of

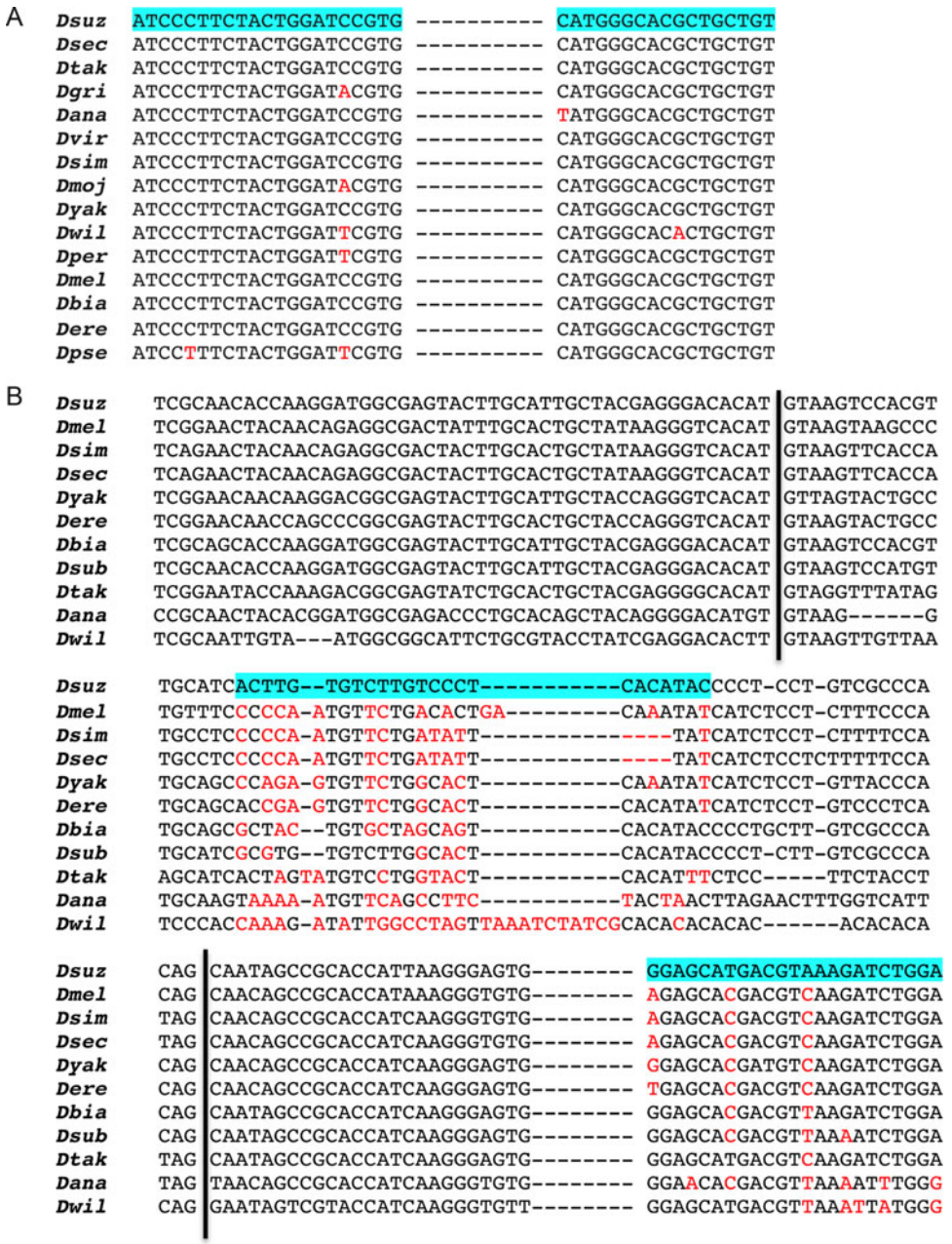

Fig. 1. Alignments for the gene regions in which the conserved and diverged primer sets are located. (a) Alignment of *D. suzukii sec61alpha* (DS10_00001395; annotation of SpottedWingFlybase (Chiu *et al*., 2013)) to orthologs from 14 other *Drosophila* species spanning multiple groups in the subgenus *Sophophora* (*D. melanogaster* (FBgn0086357), *D. simulans* (FBgn0193973), *D. sechellia* (FBgn0172841), *D. yakuba* (FBgn0235854), *D. erecta* (FBgn0115759), *D. ananassae* (FBgn0092730), *D. pseudoobscura* (FBgn0081850), *D. persimilis* (FBgn0163685), *D. willistoni* (FBgn0220690), *D. takahashii (*KB461656.1), *D. biarmipes* (KB462833.1)) and subgenus *Drosophila* (*D. virilis* (FBgn0205065), *D. mojavensis* (FBgn0140032), *D. grimshawi* (FBgn0120726)). Only the gene sequences corresponding to the location of the forward and reverse primers were shown, and sequence polymorphisms as compared to *D. suzukii sec61alpha* were indicated in red. *D. suzukii* primer sequences are highlighted in blue. (b) Alignment of *D. suzukii* DS10_00004458 (annotation in SpottedWingFlybase (Chiu *et al*., 2013)) to orthologs from 10 other *Drosophila* species: *D. melanogaster* (FBgn0035268), *D. simulans* (FBgn0185353), *D. sechellia* (FBgn0169366), *D. ananassae* (FBgn0101826), *D. yakuba* (FBgn0238459), *D. erecta* (FBgn0107076), *D. takahashii* (KB461143.1), *D. biarmipes* (KB462838.1), *D. willistoni* (FBgn0218640), and *D. subpulchrella* (KM208658). *Drosophila* species including *D. grimshawi, D. virilis, D. mojavensis*, and *D. persimilis* have a larger and more diverged intron and were excluded from this figure. 66-bp are shown upstream of the forward primer and 52-bp are shown downstream to anchor the alignment and indicate the polymorphic nature of the intronic region used for the forward primer. Vertical lines mark exon–intron boundaries. The eight dashes immediately before the reverse primer sequence indicate the portion of the sequence that is not shown in the alignment. Red color highlights polymorphic base pairs or deletions in the other *Drosophila* sequences as compared to *D. suzukii* primer sequences, which are highlighted in blue.

D. melanogaster FBgn0035268 and FBpp0072657) that is more divergent (Supplementary Table 2). Since there is currently no publicly available genome or transcriptome data available for the closely related species *D. subpulchrella,* for which females also have serrated ovipositors (Atallah *et al.*, 2014), we sequenced a gene region within the *D. subpulchrella* ortholog of *D. suzukii* DS10_00004458 (*D. melanogaster* FBgn0035268) to locate primer sequences that are sufficiently diverged between *D. suzukii* and *D. subpulchrella* to differentiate these two species in addition to the other *Drosophila* species included in our comparative genomics analysis. To obtain the sequence for the *D. subpulchrella* ortholog of *D. suzukii* DS10_00004458, PCR was performed using *D. subpulchrella* genomic DNA as a template with forward primer 5′-AGTTTTGCGTCAGCGG ATCC -3′ and reverse primer 5′-TCGTCGTCGGAGCTGT TG -3′. These primer sequences were designed based on *D. suzukii* sequences, but were sufficiently similar to amplify *D. subpulchrella* genomic DNA. Genomic DNA extraction and amplification conditions are described below. Amplified PCR products were purified using the PCR Purification Kit (Qiagen, Valencia, CA) following manufacturer's instructions. Sanger sequencing of the amplified gene region was performed by the UC Davis sequencing facility using the same set of primers (NCBI GenBank KM208658). Alignment of the resulting *D. subpulchrella* gene region with its orthologs from other *Drosophila* species were performed using MAFFT (Katoh & Standley, 2013) to identify primer sequences with sufficient sequence polymorphisms to yield a species-specific primer set (fig. 1b). We settled on a primer pair that amplifies a 263 bp product. The forward primer (5′-ACTTGTGTCTTGT CCCTCACATAC-3′) is located within an intron and the reverse primer (5′-TCCAGATCTTTACGTCATGCTCC-3′) is located within the coding region. All primers were ordered as desalted oligonucleotides from Integrated DNA Technologies (Coralville, IA).

Direct larval tissue PCR without genomic DNA extraction

Individual larvae were separated from the fly food media, wiped clean, and cut in half with a sterile razor blade. Immediately after cutting, a small portion of larval tissue was placed into a 1.5 ml microcentrifuge tube on ice prefilled with 19.5 µl of dilution buffer and 0.5 µl of DNARelease Additive from the Phire Animal Tissue Direct PCR Kit (Fisher Scientific, Pittsburgh, PA). The samples were vortexed briefly and incubated at room temperature for 5 min, placed on a 98°C heat block for 2 min, then kept on ice. PCR was prepared in a 50 µl reaction volume on ice as follows: 1 µl of diluted tissue mixture, 5 µl of 10× AccuPrime Taq DNA Polymerase Buffer 1 (Life Technologies, Grand Island, NY), 1 µl of each of the 2 primer pairs (10 µM), 0.5 µl of AccuPrime Taq DNA Polymerase, and 39.5 µl of nuclease free water. PCR was performed in a 96-well Mastercycler Pro (Eppendorf, Hauppauge, NY) and was carried out with an initial 2-min denaturation step at 95°C, followed by 35 cycles consisting of 15 s at 95°C, 30 s at 55°C, and 1 min and 30 s at 68°C. A final elongation step of 68°C for 5 min was included in the program. The amplified products were resolved on 1% TBE agarose gel for 60 min at 100 V in 1× TBE buffer (0.089 M Tris, 0.089 M boric acid and 0.002 M EDTA) and DNA was visualized under UV light using GelStar™ Nucleic Acid Gel Stain at 0.7× (Lonza, Switzerland) and the Alpha-imager mini (Protein Simple, Santa Clara, CA). The 1 kb DNA ladder (Promega, Madison, WI) was used to determine the size of the bands.

Genomic DNA extraction

Ten adult flies were used for each genomic DNA extraction. Flies were homogenized by motorized pestle in 250 µl of extraction buffer (80 mM NaCl, 160 mM sucrose, 50 mM EDTA, 125 mM,Tris-HCl (pH8.5), 0.5% SDS) in a 1.5 ml microcentrifuge tube at room temperature. The samples were incubated at 65°C for 30 min, followed by a 35-min incubation on ice. 40 µl of 8 M potassium acetate was added to the reaction and the homogenate was briefly vortexed and subsequently centrifuged at 13,000 rpm for 10 min at 4°C. The supernatant was removed and aliquoted into clean microcentrifuge tubes and centrifuged again to remove remaining debris. The supernatant was transferred to a new microcentrifuge tube and an equal volume of Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Sigma Aldrich, St Louis, MO) was added to the supernatant and vortexed before spinning at 13,000 rpm for 15 min. The top layer containing the genomic DNA was removed, mixed with a 2× volume of 100% ethanol, and left overnight to precipitate at −20°C. The precipitated DNA was pelleted by spinning at 13,000 rpm for 15 min. The pellet was washed with 70% ethanol and the ethanol was removed after a 5-min centrifugation step at 4°C. The pellet was allowed to dry before resuspending in 50 µl of nuclease free water. DNA concentrations were quantified using a Biophotometer (Eppendorf, Hauppauge, NY). The DNA was diluted to 50 ng μl^{-1} and stored at −20°C prior to use.

PCR conditions using genomic DNA as template

PCR was prepared in a 50 µl volume on ice as follows: 2 µl of genomic DNA (100 ng total), 5 µl of 10× AccuPrime Taq DNA Polymerase Buffer 1 (Life Technologies, Grand Island, NY), 1 µl of each of the 4 primers (10 µM), 0.5 µl of AccuPrime Taq DNA Polymerase, and 38.5 µl of nuclease free water. The cycling conditions and visualization methods were the same as described above.

Tissue PCR with ethanol or RNAlater-preserved specimens

Live larvae, pupae, and adult flies were placed into 95% ethanol and stored at room temperature, or in the case of the *D. suzukii* from South Korea, placed into RNA*later* (Sigma Aldrich, St Louis, MO) and stored based on manufacturer's instructions before use. The specimens were removed from the solution and dried on a clean sterile surface for 5 min. Larvae and pupae were cut in half with a sterile razor blade, while adult flies were lightly crushed and kept whole. The PCR reactions were prepared as described above in direct PCR without genomic DNA extraction.

Results and discussion

Comparative genomic analysis facilities primer design

To select the appropriate primer sequences for the *D. suzukii* molecular diagnostic, we used a comparative genomics approach to identify candidate genes by examining the most conserved and diverged genes across 15 species in the subgenus *Sophophora* and the subgenus *Drosophila*. Average branch lengths of gene trees of orthologs were used to measure

the degree of conservation and divergence. Primer sequences that can potentially amplify across all *Drosophila* were designed from conserved regions in one of the genes with the shortest branch lengths. *Sec61alpha* is the fourth most conserved in 2661 single copy ortholog groups present across the 15 *Drosophila* species we included in our comparative genomic analysis (Supplementary Table 1; highlighted in yellow). Alignment of *D. suzukii Sec61alpha* gene regions representing the forward and reverse primer sequences to corresponding regions for 14 other *Drosophila* species indicates high level of conservation with either no mismatches in most primer sequences to two mismatches in the case of *D. pseudoobscura* (fig. 1a). Since there is currently no genome or transcriptome available for *D. subpulchrella*, it was not included in our sequence analysis for *Sec61alpha*. However, since *Sec61alpha* is one of the most conserved single copy *Drosophila* genes and the *Sec61alpha* primer pair was able to amplify a gene fragment of predicted size, we assumed that the conservation of *Sec61alpha* is maintained in *D. subpulchrella*. Theoretically, primers can also be designed from many other conserved genes listed in Supplementary Table 1, but the *Sec61alpha* primers were validated in our studies presented here.

While sequences from the most diverged genes may be good candidates for designing primers that would only amplify from *D. suzukii* DNA, we initially tested a number of primers generated from *D. suzukii* genes that are in ortholog sets with gene trees that have the longest branch lengths, but found that they did not amplify consistently across *D. suzukii* populations collected from different geographical sites in the USA (data not shown). This is likely due to high levels of sequence polymorphisms. This suggests that the most divergent genes may be inappropriate as diagnostic marker as they have the potential to produce false negative test results. We then tested a candidate marker that is relatively less divergent, but still in the top 15% among the 2515 single copy genes in the *melanogaster* and the *suzukii-takahashii* subgroups, *D. suzukii* DS10_00004458 (*D. melanogaster* FBgn0035268 and FBpp0072657) (Supplementary Table 2; highlighted in yellow). We deemed this *D. suzukii* gene a suitable candidate for testing as it shows high divergence as compared to orthologs from species within the *suzukii-takahashii* as well as the *melanogaster* subgroups based on branch lengths (Supplementary Table 2; row 350 highlighted in yellow). On the other hand, although other genes, e.g., DS10_00003207 and DS10_00002465 (Supplementary Table 2; rows 345 and 347 highlighted in blue), have overall average branch lengths that are comparable to *D. suzukii* DS10_00004458 (around 0.068) when the calculation were performed using *D. suzukii* and species from both the *suzukii-takahashii* and *melanogaster* subgroups, they are not predicted to be effective as species-specific diagnostic due to the short branch lengths when computed using specific *Drosophila* subgroups. This suggests high sequence conservation between these *D. suzukii* genes to specific *Drosophila* subgroups. For example, the branch length of the gene tree comprising *D. suzukii* DS10_00003207 and orthologs in the *melanogaster* subgroup is only 0.028 (Supplementary Table 2; row 345 highlighted in blue), indicating high conservation. Similarly, the branch length of the gene tree comprising *D. suzukii* DS10_00002465 and orthologs in the *melanogaster* subgroup is only 0.01455 (Supplementary Table 2; row 347 highlighted in blue), suggesting high sequence conservation as well.

Based on multiple sequence alignment, *D. suzukii* DS10_00004458 contains variable exonic regions that are appropriate as interspecies diagnostic between *D. suzukii* and the fourteen *Drosophila* species included in our comparative genomic analysis (fig. 1b). To verify that this gene contains diagnostic sequences that can effectively differentiate between *D. suzukii* and the closely related *D. subpulchrella*, for which genome and transcriptome data are not available and is therefore not included in our bioinformatic analysis, we used Sanger sequencing to generate sequence data for the *D. subpulchrella* ortholog of *D. suzukii* DS10_00004458. As we were restricted by the limited amount of *D. subpulchrella* sequence we generated specifically for this study, we designed our forward diagnostic primer to be located within an intron, whereas the reverse primer is located within an exon. To ensure that species diagnostic primers have widespread utility and can amplify successfully across specimens from different populations, the use of exonic sequences for primer design is highly recommended because the rate of divergence in exons is generally lower as compared to that for introns and intergenic sequences due to constraints imposed by selection against deleterious mutations. There is a delicate balance between choosing diagnostic primer sequence that is sufficiently divergent between different *Drosophila* species, but yet retained appropriate level of conservation within species to ensure utility across *D. suzukii* populations. Nevertheless, subsequent PCR validation (see below) indicated that SWD-specific primers that were designed based on our bioinformatic analysis retained sufficient conservation across *D. suzukii* from different geographical regions to enable amplification in all *D. suzukii* populations tested.

Comparative genomic analysis provides a more robust framework for primer design and the theoretical foundation for the reliability of our PCR test. More importantly, since fly populations are continuously evolving, our comprehensive analysis provides the basis for the design of additional diagnostic markers in the case that DNA sequence changes render our diagnostic marker ineffective for specific fly populations. SpottedWingFlybase IDs for *D. suzukii genes* as well as FlyBase IDs for their *D. melanogaster* orthologs are provided in Supplementary Tables 1 and 2 to facilitate the retrieval of appropriate *Drosophila* sequences for sequence alignment and primer design. In fact, by entering the SpottedWingFlybase ID, e.g., DS10_00004458, into the 'Gene Query' search box on SpottedWingFlybase (http://spottedwingflybase.oregonstate.edu/) (Chiu *et al.*, 2013), users will be able to identify the gene family a particular gene belongs to and retrieve the FlyBase IDs (FBpp numbers) of corresponding orthologs from other *Drosophila* species in addition to *D. melanogaster*.

Multiplex PCR diagnostic differentiates D. suzukii *from other Drosophila species*

The internal control primer pair amplifies a 1248 bp product that is conserved in all Drosophilids for quality control purposes. The presence of the control band in non-SWD samples signals that DNA quality and amplification condition are acceptable, and the lack of amplification for the species-specific primers is not due to suboptimal PCR conditions. The SWD species-specific primer pair amplifies a 263 bp product. Thus, a positive test should result in the amplification of two DNA fragments, one at 263 bp, and another at 1248 bp, visible upon DNA gel electrophoresis.

As we aim to develop an efficient assay that can be performed using limited starting materials, we tested our PCR

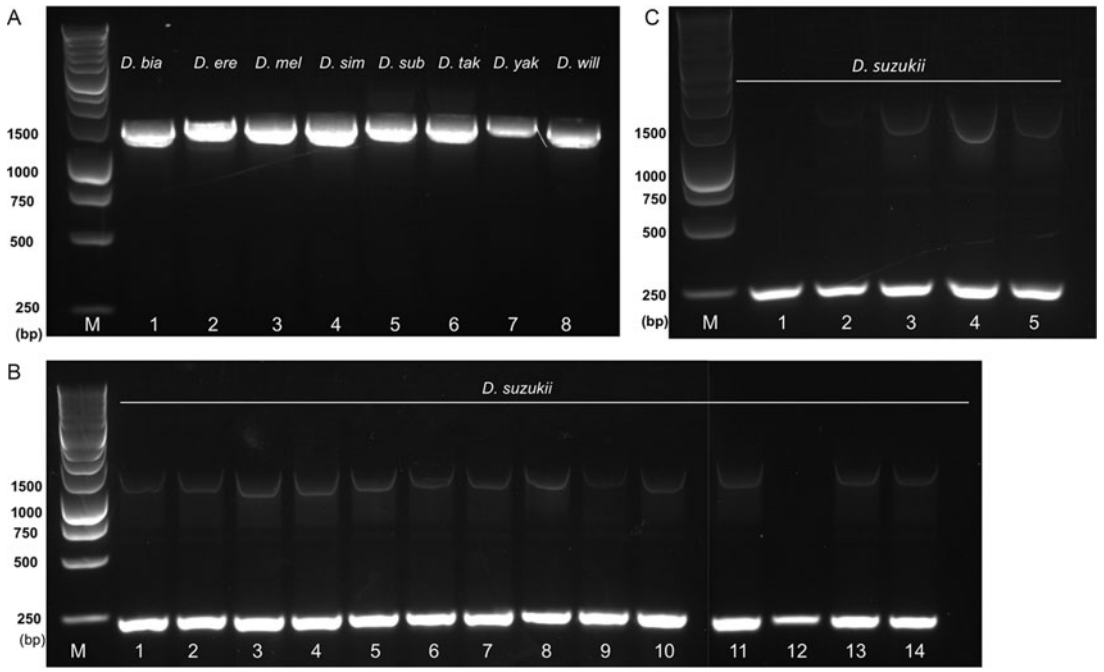

Fig. 2. DNA agarose gel electrophoresis showing results of multiplex PCR. Multiplex PCR was performed using (a) multiple *Drosophila* species (see table 1 for strain information) and (b) *D. suzukii* collected from different regions in the Western USA (lanes 1–14: genome strain WT3 F10, lime, grape, Mark Bolda WAT, Wolfskill IFL WO-2 [from CA], HR3 F4, TD3 F5, PD3 F5, ARS, OS1 [from OR], colony #8, colony #9, colony #10, and colony #11 [from WA]). PCR amplification for (a) and (b) was performed using crude larval extract isolated from a single larva that was freshly sacrificed. All crude DNA extractions and corresponding PCR reactions for (a) and (b) were repeated at least five times using biological replicate samples. Representative results are shown here. (c) Multiplex PCR was performed using crude extract from *D. suzukii* specimens collected from regions outside of the western United States. Adult flies were collected in South Korea and Italy (table 1) and preserved in RNA*later* for storage (lane 1: South Korea, lane 2: Italy). Freshly sacrificed larva from strains collected in Japan, Hawaii, and Maryland, USA (lane 1: MTY3, Japan, lane 2: Oahu, Hawaii, lane 5: Maryland).

diagnostic using crude extract isolated from either single larva or preserved adult specimen (when larva is not available) without any prior genomic extraction steps. To ensure reproducibility, at least five biological replicates were performed for each SWD and non-SWD population. A total of eight non-SWD *Drosophila* species were tested (fig. 2a and table 1). There were no false positives in that all non-SWD larvae showed robust amplification only for the control 1248 bp product. The size of the internal control band in the different species showed slight variation indicating possible in/del within the regions spanned by the control primer set. As sequence polymorphisms might have accumulated over time as *D. suzukii* populations spread, we tested the utility of our PCR diagnostic using larvae or preserved adult specimens from nineteen *D. suzukii* populations (fig. 2b, c, table 1). Freshly sacrificed larvae were used to assay *D. suzukii* populations from the continental USA. including California, Oregon, Washington and Maryland, as well as populations from Hawaii and Japan, as live cultures of these populations were available. On the other hand, preserved adult specimens were used to assay *D. suzukii* populations from Italy and South Korea because we only have access to RNAlater-preserved specimens for these sites. All of the *D. suzukii* larvae originated from different geographical regions in the USA as well as from Hawaii and Japan tested positive and showed amplification at the 263 bp (SWD-specific band) (fig. 2b, c). Moreover, adult *D. suzukii* specimens collected in South Korea and Italy and preserved in RNA*later* also tested positive (fig. 2c). This represents a 100% success rate.

As our diagnostic relies on multiplex PCR, there is competition between the primer sets for reaction components such as ATP and deoxynucleotide triphosphates (dNTPs). When one product is favored, it can outcompete the other reaction resulting in uneven amplification of the two products. Short amplicons, e.g., the 263 bp SWD-specific products, are often amplified with higher efficiency than long amplicons, e.g., the 1248 bp control product, because the polymerase is more likely to fully extend a larger percentage of the short products as the reaction proceeds. Thus, we designed the control primers to produce a longer amplicon than the SWD-specific primers so that when *D. suzukii* DNA is provided as the template, the SWD-specific product will be favored. This design increases the sensitivity of our assay, yet retains the advantage of having an internal control to illustrate the difference between robust vs. weak non-specific amplification. As shown in fig. 2b, the SWD-specific 263 bp product is clearly more robust than the 1248 bp control amplified fragment in all SWD samples. In a few of our test samples, the SWD-specific primers even outcompete the internal control primers, resulting in a single band at 263 bp (e.g., fig. 2b, lane 12; fig. 2c, lane 1). When the 263 bp SWD-specific product is present and 1248 bp control product is absent, a single band at 263 bp can be interpreted as a positive test result.

This assay was also performed using crude extract from ethanol preserved larvae, pupae, and adult samples because this is a common collection and storage condition and produced amplification results and conclusions identical to those in reactions using live samples (data not shown). This direct larval tissue PCR method is economical and practical for rapid identification of single larva because the PCR reaction can be assembled in minutes, while the use of extracted genomic DNA typically requires at least an hour as well as additional equipment and reagents. Finally, we verified that our multiplex PCR diagnostic assay also performs well with genomic DNA extracted from all nine *Drosophila* species and observed that the SWD-specific primers only amplified *D. suzukii* DNA (data not shown).

Conclusion

We have designed an accurate PCR diagnostic that can unambiguously differentiate *D. suzukii* from other common *Drosophila* species using the crude homogenate of a single larva as the DNA source without the need for additional steps such as genomic DNA extraction, sequencing, or restriction digestion. The use of the recently sequenced *D. suzukii* genome and a comparative genomic approach facilitated the discovery of the diagnostic marker we presented here. In addition, our genomic analysis generated many other candidate genes that can be used for species-specific diagnostics as well as examination of trait evolution in Drosophilids.

Supplementary material

The supplementary material for this article can be found at http://www.journals.cambridge.org/BER

Acknowledgements

We thank the laboratories of David Begun (UC Davis Department of Evolution and Ecology), Artyom Kopp (UC Davis Department of Evolution and Ecology), and Kelly Hamby (University of Maryland) for sharing *Drosophila* strains. *D. suzukii* from Watsonville, CA, USA. were collected with permission by Garroutte Farms, Inc. (Watsonville, CA). We also thank Claudio loriatti from the Center for Technology Transfer, FEM-IASMA for *D. suzukii* collected from Italy. We thank Ernest K. Lee (AMNH, New York, NY) as well as the editor and reviewers for their valuable input into this manuscript. This project is supported by a Washington Tree Fruit Research Commission grant awarded to JCC and UC Davis Provost's Undergraduate Fellowship and McBeth Memorial Fellowship awarded to KAM.

References

Atallah, J., Teixeira, L., Salazar, R., Zaragoza, G. & Kopp, A. (2014) The making of a pest: the evolution of a fruit-penetrating ovipositor in *Drosophila suzukii* and related species. *Proceedings of the Royal Society B: Biological Sciences* **281**(1781), 20132840.

Behura, S.K. (2006) Molecular marker systems in insects: current trends and future avenues. *Molecular Ecology* **15**, 3087–3113.

Calabria, G., Maca, J., Bachli, G., Serra, L. & Pascual, M. (2012) First records of the potential pest species *Drosophila suzukii* (Diptera: Drosophilidae) in Europe. *Journal of Applied Entomology* **136**, 139–147.

Chiu, J.C., Lee, E.K., Egan, M.G., Sakar, I.N., Coruzzi, G.M., DeSalle, R. (2006) OrthologID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics* **22**(6), 699–707.

Chiu, J.C., Jiang, X., Zhao, L., Hamm, C.A., Cridland, J.M., Saelao, P., Hamby, K.A., Lee, E.K., Kwok, R.S., Zhang, G., Zalom, F.G., Walton, V.M., Begun, D.J. (2013) Genome of *Drosophila suzukii*, the Spotted Wing Drosophila. *G3: Genes, Genomics, Genetics* **3**(12), 2257–2271.

Cini, A., Ioraitti, C. & Anfora, G. (2012) A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. *Bulletin of Insectology* **65**, 149–160.

Dhami, M.K. & Kumarasinghe, L. (2014) A HRM real-time PCR assay for rapid and specific identification of the emerging pest Spotted-Wing Drosophila (*Drosophila suzukii*). *PLoS ONE* **9**(6), e98934.

Gariepy, T.D., Kuhlmann, U., Gillott, C. & Erlandson, M. (2007) Parasitoids, predators and PCR: the use of diagnostic molecular markers in biological control of Arthropods. *Journal of Applied Entomology* **131**(4), 225–240.

Hauser, M. (2011) A historic account of the invasion of *Drosophila suzukii* (Matsumura) (Diptera-Drosophilidae) in the continental United States, with remarks on their identification. *Pest Management Science* **67**, 1352–1357.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270**(1512), 313–321.

Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. (1996) Real time quantitative PCR. *Genome Research* **6**, 986–994.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**(4), 772–780.

Kim, S.S., Tripodi, A.D., Johnson, D.T. & Szalanski, A.L. (2014) Molecular diagnostics of *Drosophila suzukii* (Diptera: Drosophilidae) using PCR-RFLP. *Journal of Economic Entomology* **107**(3), 1292–1294.

Lee, J.C., Bruck, D.J., Curry, H., Edwards, D., Haviland, D.R., Van Steenwyk, R.A., Yorgey, B.M. (2011) The susceptibility of small fruits and cherries to the spotted-wing drosophila, *Drosophila suzukii*. *Pest Management Science* **67**, 1358–1367.

Marygold, S.J., Leyland, P.C., Seal, R.L., Goodman, J.L., Thurmond, J.R., Strelets, V.B., Wilson, R.J. FlyBase consortium. (2013) FlyBase: improvements to the bibliography. *Nucleic Acids Research* **41**(D1), D751–D757.

modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., Washietl, S., Arshinoff, B. I., Ay, F., Meyer, P.E., Robine, N., Washington, N.L., Di Stefano, L., Berezikov, E., Brown, C.D., Candeias, R., Carlson, J.W., Carr, A., Jungreis, I., Marbach, D., Sealfon, R., Tolstorukov, M.Y., Will, S., Alekseyenko, A.A., Artieri, C., Booth, B.W., Brooks, A.N., Dai, Q., Davis, C.A., Duff, M. O., Feng, X., Gorchakov, A.A., Gu, T., Henikoff, J.G., Kapranov, P., Li, R., MacAlpine, H.K., Malone, J., Minoda, A., Nordman, J., Okamura, K., Perry, M., Powell, S.K., Riddle, N.C., Sakai, A., Samsonova, A., Sandler, J.E., Schwartz, Y.B., Sher, N., Spokony, R., Sturgill, D., van Baren, M., Wan, K.H., Yang, L., Yu, C., Feingold, E., Good, P., Guyer, M., Lowdon, R., Ahmad, K., Andrews, J., Berger, B., Brenner, S.E., Brent, M.R., Cherbas, L., Elgin, S.C., Gingeras, T.R., Grossman, R., Hoskins, R.A., Kaufman, T. C., Kent, W., Kuroda, M.I., Orr-Weaver, T., Perrimon, N., Pirrotta, V., Posakony, J.W., Ren, B., Russell, S., Cherbas, P., Graveley, B.R., Lewis, S., Micklem, G., Oliver, B., Park, P.J., Celniker, S.E., Henikoff, S., Karpen, G.H., Lai, E.C.,

MacAlpine, D.M., Stein, L.D., White, K.P., Kellis, M. (2010) Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**(6012), 1787–1797.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.

Stamatakis, A. (2012) RAxML GitHub repository. https://github.com/stamatak/standard-RAxML/.

Walsh, K., Boonham, N., Barker, I. & Collins, D.W. (2005) Development of a sequence-specific real-time PCR to the melon thrips Thrips palmi (Thysan., Thripidae). *Journal of Applied Entomology* **129**, 272–279.

Walsh, D.B., Bolda, M.P., Goodhue, R.E., Dreves, A.J., Lee, J., Bruck, D., Walton, V.M., O'Neal, S.D., Zalom, F.G. (2011) *Drosophila suzukii* (Diptera: Drosophilidae): invasive pest of ripening soft fruit expanding its geographic range and damage potential. *Journal of Integrated Pest Management* **2**, 1–7.

Williams, J.G., Kybelik, A.R. & Livak, K.J. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531–6535.

Wyman, A.R. & White, R. (1980) A highly polymorphic locus in human DNA. *Proceedings for National Academy of Science of the United States of America* **77**(11), 6754–6758.