Utilising samples collected in an existing biodiversity network to identify the presence of potential insect vectors of *Xylella fastidiosa* in the UK

Kirsty J Park[1], Matthew Guy[1], Elisa Fuentes-Montemayor[1], Katherine Lester[2], Stuart A'Hara[3], Joan Cottrell[3]

1. Biological & Environmental Sciences, University of Stirling, Stirling FK9 4LA
2. SASA, Roddinglaw Road, Edinburgh, EH12 9FJ
3. Forest Research, Northern Research Station, Roslin, Midlothian, Scotland, EH25 9SY

SUMMARY

Xylella fastidiosa is a bacterial disease, transmitted between plants by xylem-feeding invertebrates. It has been estimated that, in the UK, approximately 20 xylem-feeding insect species could potentially vector *Xylella,* including those in the family Cicadellidae (leaf hoppers), and the superfamily Aphrophoridae (frog hoppers/spittlebugs). Little is known about the distribution or ecology of species in these groups within the UK, meaning that we lack information about how the presence and density of vectors varies between habitats. For this project, we screened and attempted to morphologically identify existing invertebrate samples for Auchenorrhyncha (with a particular focus on Aphrophoridae), collected using malaise traps from across 78 secondary woodlands within well characterised agricultural landscapes in central Scotland and England in 2015.

In particular, three spittlebugs were morphologically identified to species level within sampled woodlands. These were of particular interest as they are considered to be the most likely vectors to spread Xylella if the disease were to arrive in Scotland. In order of descending abundance these were: *Philaenus spumarius*, *Aphrophora alni* and *Neophilaenus lineatus*. Our study revealed that landscape-level attributes were more important than woodland characteristics in determining spittlebug abundance; spittlebugs were more abundant in woodlands located in landscapes with relatively low percentages of broadleaved woodland and other semi-natural habitats (i.e. those landscapes with high levels of agricultural land-use), and with low hedgerow densities.

We used DNA barcoding on a broad taxonomic range of morphologically identified UK arthropod samples to determine the optimal primer pair required for metabarcoding part of the COI region of mitochondrial DNA for this range of organisms in the UK. We were also able to identify arthropod species from pitfall trap samples collected in 2013, 2014 and 2016, with early indications that the DNA in these samples remain of a good quality and potentially suitable for future metabarcoding projects.

A proof-of-concept study successfully provided metabarcoded sequences from each of a sub-sample of 34 malaise traps and a mock community of previously morphologically identified and barcoded arthropod samples. Two primer pairs were tested on the mock community samples and one of the primer pairs was successful in detecting all three main spittlebug vector species. This primer pair was used in the metabarcoding of the 34 malaise trap multi-species samples. Results from morphological and metabarcoding identification in the 34 malaise trap samples were in almost complete agreement regarding the samples in which the three main spittlebug vectors were present and this is encouraging for the application of the metabarcoding approach to detect these vectors. In addition to the three vector species, the metabarcoding study identified over 1000 arthropod species across the 34 malaise trap samples highlighting the utility of this approach to detect additional potential vectors of Xylella and other plant diseases.

The refinement of the molecular detection methods for potential vectors species will be of value should a suspected *Xylella* outbreak require rapid investigation. Additionally, training and experience in invertebrate ID has increased capacity in the Scottish science community to identify key vectors. Future work in this area should focus on further optimisation of molecular approaches, and exploration of alternative techniques such as Minion that may substantially reduce analytical costs. It is important to note that this study focussed on

woodlands, and little is known about how the abundance of these vectors varies spatially (between and within habitats) and temporally. These are key areas for future work if the spittlebug ecology and distribution is to be better understood.

INTRODUCTION

Xylella fastidiosa is a bacterial disease with many sub-species and strains which is not known to occur in the UK. In mainland Europe, most notably France (Corsica and mainland France) and Italy, there have been several outbreaks of different sub-species which have had significant detrimental impacts on plants both in the wider environment and those grown commercially for olive production. In 2016, *Xylella* was detected in Spain for the first time on cherry trees in a nursery on the Balearic Islands and then in 2017 it was discovered for the first time on mainland Spain on almond trees. Although plant imports into the UK are regulated by the EU, there are nevertheless understandable concerns about the risk of introduction of *Xylella* to the UK and consequential economic and environmental impacts.

Xylella lives in the xylem of plants and therefore insects that feed on the xylem are potential vectors of the bacterium and can spread disease between plants. Of the known outbreaks to date, the *X. fastidiosa* sub-species multiplex detected in Corsica, mainland France, mainland Spain and in nursery stock on the Spanish Balearic Islands is of most concern to the UK. This subspecies is able to survive in cooler climates and can affect a wide range of hosts, including many native broadleaved trees such as oak. *Xylella fastidiosa* is known to be transmitted by insects in the Cicadellidae (leafhoppers) and Aphrophoridae (froghoppers/spittlebugs).

The meadow spittlebug, *Philaenus spumarius*, has been associated with transmission of *Xylella* in olive in Europe, and is the principle insect vector responsible for spreading the disease in Italy and France. However, spittlebugs are not associated with transmission in the Americas, where instead the sharpshooters *Oncometopia fascialis* and *Homalodisca vitripennis* (Cicadellidae) are the principal vectors of *Xylella* in citrus and grapevine, respectively. It is estimated that there are 20 xylem-feeding insect species that could potentially vector *Xylella* in the UK (e.g. Malumphy 2017), in particular species of: *Cicadella*, *Euscelis*, *Evacanthus* and *Graphocephala* (leafhoppers), and *Aphrophora*, *Cercopis* and *Neophilaenus* (froghoppers).

Common and widespread spittlebug species in the UK include *Aphrophora alni*, *Neophilaenus campestris*, *N. lineatus* and *Philaenus spumarius*. *Philaenus spumarius* is the most common, widespread and polyphagous species in the UK, occurring in a wide range of habitats, including those subject to considerable human mediated influence. Remarkably little is known about the distribution or ecology of these species. In the context of *Xylella* it is important to quantify variation in vector presence and density between habitats and across different landscape settings, and to determine how this varies seasonally, as this will influence the risk of disease spread between plants if *Xylella* were to establish in Scotland.

The WrEN project

The WrEN project (http://www.wren-project.com/) is a large-scale, long-term natural experiment examining the effects of past land-use change on current biodiversity (Watts et al.

2016). The project comprises a network of 106 secondary woodlands matching specified criteria within two study areas in Scotland (n = 67 sites) and England (n = 39; Figure 1). Both study areas are dominated (>70%) by agricultural land and represent typical lowland landscapes in the UK. The woodlands range from 0.5 to 32 ha in area, and were planted between 10 and 160 years ago, are surrounded by less than 1% to 17% broadleaved woodland within a 3-km buffer, and are between 7 and 1573 m away from the nearest broadleaved woodland. To date, over 1100 species have been recorded in these secondary woodlands, including vascular plants, lichens, bryophytes, small mammals, bats, birds, spiders, beetles, craneflies, hoverflies and moths. Vegetation structure has been characterised at all sites including information on tree species richness, tree density and size, understorey and canopy cover. In addition, the surrounding landscape has been mapped at a range of spatial scales up to 3 km from each site; % of habitat types and connectivity metrics have been calculated. An additional 27 ancient woodlands within the same landscapes have been surveyed for a range of taxa.

Malaise and pitfall trap samples collected over several years at WrEN sites have been screened for a range of invertebrate taxa, which have been identified to species (see above), but a large number of un-identified specimens remain (e.g. from Diptera, Hymenoptera, Hemiptera). Thus, data already collected at these sites presents a unique opportunity to map the occurrence of potential *Xylella* vectors within agricultural woodlands across Scotland and central England.

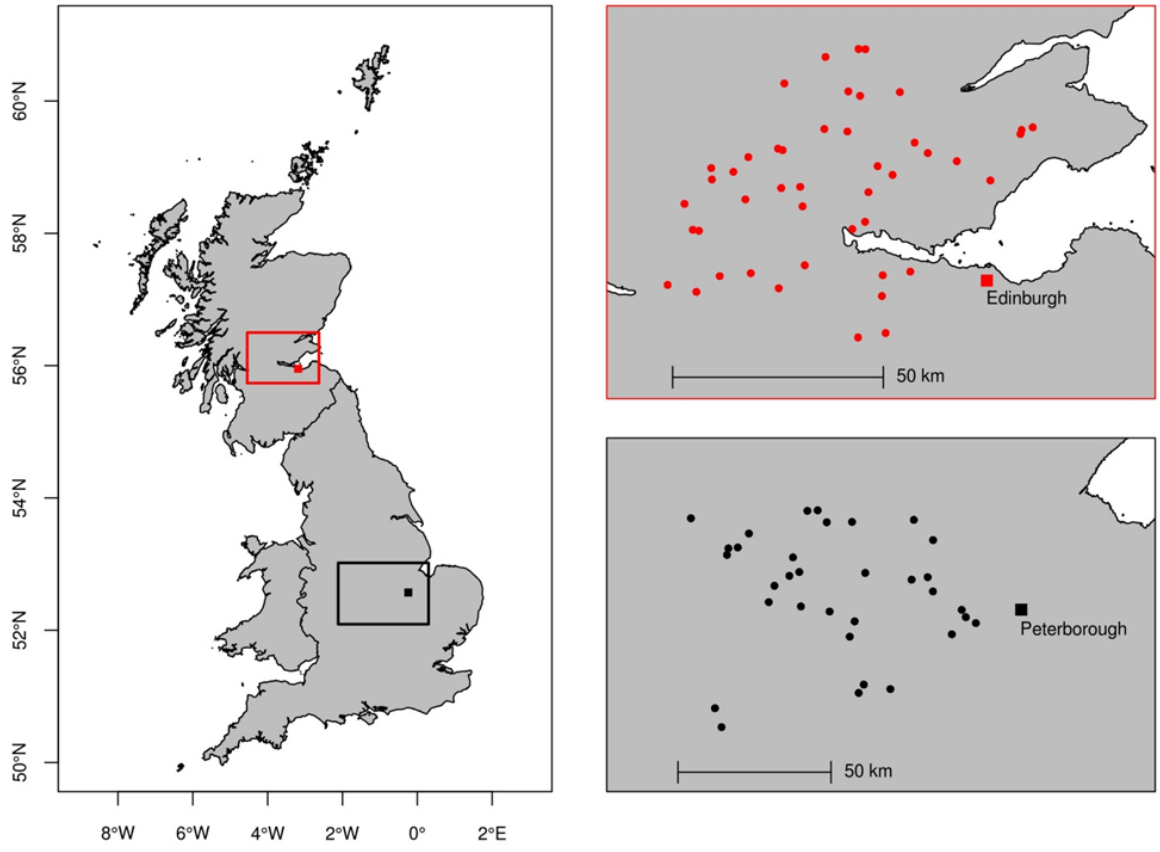


Figure 1. Map of WrEN study sites screened for *Xylella* vectors across central England and central Scotland (From Fuller et al. 2018).

RESEARCH AIMS

In this project, our aims were to:

1. Examine patterns of *Xylella* vector abundance (based on morphological id) in woodland patches within agricultural landscapes, and evaluate the importance of site- and landscape-level attributes on vector abundance;
2. Use DNA barcoding on a broad taxonomic range of morphologically identified UK arthropod samples to a) determine the optimal primer combination and b) identify individuals from a range of species, sampled through pitfall trapping, to assess the quality of DNA in the WrEN samples and the suitability of these samples for subsequent metabarcoding.
3. Conduct a meta-barcoding proof-of-concept study on a subset of malaise trap samples. This will lay the foundation for testing the potential of the metabarcoding approach to offer a fast and efficient way to monitor presence of vector species.

METHODS

1. Invertebrate sampling and identification

As part of the WrEN project, malaise traps were deployed at 78 secondary woodland sites (45 Scottish, 33 English; Figure 1). At each site, a trap was left in place for a sampling period of 7 days each in June, July and August 2015 (these are referred to below as M1 samples); an additional trap was left in each site for one sampling period of 7 days to capture variation across the site (M2 sample). For a detailed description of sampling methods see Fuller et al. 2018.

In addition, pitfall traps were deployed at 86 sites (60 secondary, 26 ancient) for six week periods between June and August in 2013, 2014 and 2016, with some sites surveyed in multiple years. Traps were laid along transects from the edge to the interior of the woodland with 5 – 15 traps set per transect, and a total of 3432 samples collected (unpublished data).

For the project outlined here, a total of 279 malaise trap samples (M1 and M2) from 78 woodland sites were screened for Auchenorrhyncha (leafhoppers, plant hoppers, froghoppers, tree hoppers and cicadas) using morphological methods (Wilson et al. 2015). Specimens were counted and identified to species where possible (or higher taxonomic level otherwise). All spittlebugs (the most likely vectors of *Xylella*) were identified to species.

2. Statistical analysis to identify drivers of spittlebug abundance

In the project funded by the Plant Health Centre (Scotland's Centre of Expertise), the distribution of morphologically identified spittlebugs (superfamily Cercopoidea; referred to as potential *Xylella* vectors) in secondary woodlands was mapped and statistical analyses were conducted to examine the extent to which the abundance of spittlebugs (pooled across survey periods for each site) relates to characteristics of woodland sites and surrounding landscape (aim 1). We used Structural Equation Models (SEMs), a multivariate technique that can be

used to test whether a priori hypothesised direct and indirect causal relationships between variables are supported by observed data and compare relative effect sizes of predictor variables (Lefcheck 2016). Variables in SEMs included 1) site-level woodland characteristics: patch size, age (years since planting), tree species richness, tree density, understorey % cover and the variation (standard deviation) of tree diameter at breast height (used as an index of structural heterogeneity); and 2) landscape-level attributes: percentage of land covered by agriculture, broadleaved woodland and other semi-natural habitat within 1 km of each focal site, hedgerow length within 1 km of each site and distance to nearest broadleaved woodland; 'region' was included in the models to account for potential variability between the two study areas (i.e. England and Scotland). A conceptual model for this analysis is shown in Figure 2. For details of predictor variables and how they were quantified see Watts et al. 2016.

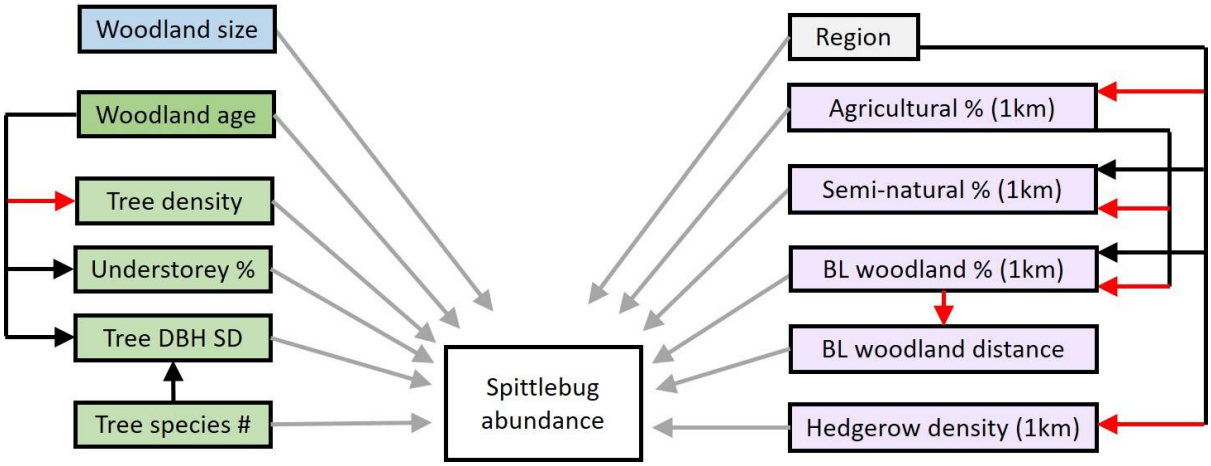


Figure 2. Conceptual model of hypothesised direct and indirect causal relationships between the abundance of *Xylella* vectors and a range of predictor variables: 1) site-level attributes related to habitat quality (green boxes), and patch geometry (blue box), 2) landscape-level attributes (purple boxes), and 3) region (grey box). Arrow colour indicates directionality of hypothesized associations (black = positive; red = negative; grey = variable or unknown). Thus, woodland age could have a *direct* effect on vector abundance or could influence abundance *indirectly* through the effects of e.g. tree density.

3. DNA barcoding

DNA barcoding uses primer pairs designed to amplify a region of the genome, often the 658 base pair C terminal region in the gene encoding mitochondrial oxidase subunit 1 (COI). The DNA sequence of this region is highly variable between species and this property has been harnessed as a means of enabling species identification. This is facilitated by the development of a publicly available database known as the Barcode of Life in which sequence for this genomic region for millions of species across the world is deposited. By searching the database for the sequence generated by the specimen of interest it is possible to arrive at a species identification. The arthropods present in the UK are well represented in the Barcode of Life database and therefore this approach works well for this category of organisms.

i) Reference database of 658bp COI region in 132 arthropod species

SASA amplified the 658 base pair COI region using Folmer et al. (1994) primers for 132 UK arthropodsamples, the majority of which had already been morphologically identified by an arthropod expert (aim 2). These arthropods were collected by SASA and FR from a variety of British sources, including spittlebugs collected from WrEN sites. These included five potential *Xylella* vector species and a further 22 species representing a variety of UK arthropod species in an attempt to represent the taxonomic range of UK arthropods. Of these 132 sequencing attempts, 36 failed to amplify (further work is to be conducted on these samples) but high quality sequence was obtained for the remaining 94 species. Aquatic arthropods were not included and within the timeframe it was not possible to source members of the Orders Thysanoptera (thrips) Sarcoptiformes (ticks), Pscoptera (booklice), Orthoptera (grasshopper), Neuroptera (lacewings), Mallophaga (lice) and Opiliones (harvestmen). However, the most commonly occurring orders were represented. This collection of arthropods for which there was good quality COI sequence available in our database was used to create a 'mock community' that was subjected to metabarcoding along with the 34 samples from the WrEN sites. The mock community was constructed so that it contained 50ng of DNA from each of the 132 component species.

ii) In silico testing of primers prior to metabarcoding

Successful amplification of DNA is dependent on several factors, one of which is how well a primer complements the target sequence. Since the original work of Folmer et al. (1994), multiple primers have been designed to optimise detection of specific target organisms as well as to amplify a shorter region of the CO1 region required for metabarcoding. Possible choices of primer annealing position are provided in Figure 3. Primer sequences are provided in Table 1.

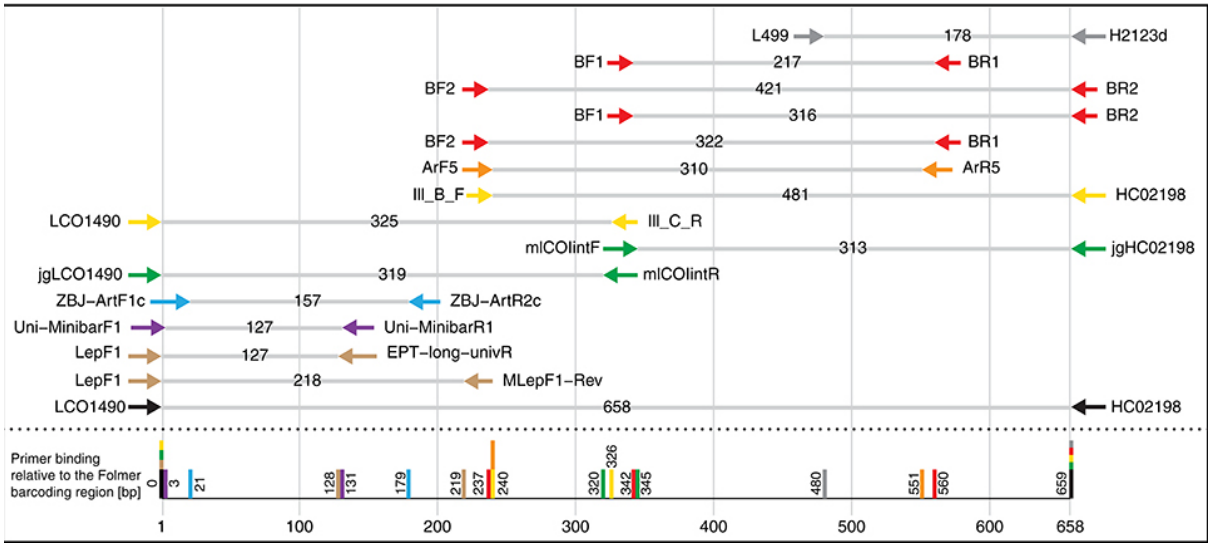


Figure 3. Location of binding sites for primer pairs designed to amplify shorter regions of the 658 bp region amplified by the Folmer et al. 1994 primer pair. Reproduced from Elbrecht & Leese 2017.

Table 1. Primer pairs subjected to *in silico* testing with various taxonomic groups of arthropods; (F=Forward primer, R= Reverse primer)

Primer pair and reference number		F/R	Binding Position	Primer sequence	Reference source
LCO1490	1	F	0	GGTCAACAAATCATAAAGATATTGG	Folmer et al. 1994
HC02198	1	R	659	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. 1994
Uni-MinibarR1	2	R	131	GAAAATCATAATGAAGGCATGAGC	Meusnier et al. 2008
Uni-MinibarF1	2	F	3	TCCACTAATCACAARGATATTGGTAC	Meusnier et al. 2008
ZBJ-ArtF1c	3	F	21	AGATATTGGAACWTTATATTTTATTTTTGG	Zeale et al. 2010
ZBJ-ArtR2c	3	R	179	WACTAATCAATTWCCAAATCCTCC	Zeale et al. 2010
mlCOIintF	4	F	345	GGWACWGGWTGAACWGTWTAYCCYCC	Leray et al. 2013
mlCOIintR	4	R	320	GGRGGRTASACSGTTCASCCSGTSCC	Leray et al. 2013
LepF1	5	F	0	ATTCAACCAATCATAAAGATATTGG	Hebert et al. 2004
EPT-long-univR	5	R	128	AARAAAATYATAAYAAAIGCGTGIAIIGT	Hajibabaei et al. 2011
MLepF1-Rev	5	R	219	CGTGGAAAWGCTATATCWGGTG	Brandon-Mong et al. 2015
lll_C_R	6	R	326	GGIGGRTAIACIGTTCAICC	Shokralla et al. 2015
lll_B_F	6	F	240	CCIGAYATRGCITTYCCICG	Shokralla et al. 2015
BF1	7	F	342	ACWGGWTGRACWGTNTAYCC	herein
BF2	7	F	237	GCHCCHGAYATRGCHTTYCC	herein
BR1	7	R	560	ARYATDGTRATDGCHCCDGC	herein
BR2	7	R	659	TCDGGRTGNCCRAARAAYCA	herein
ArF5	8	F	240	GCICCIGAYATRKCITTYCCICG	Gibson et al. 2014
ArR5	8	R	551	GTRATIGCICCIGCIARIACIGG	Gibson et al. 2014
jgLCO1490	9	F	0	TITCIACIAAYCAYAARGAYATTGG	Geller et al. 2013
jgHCO2198	9	R	659	TAIACYTCIGGRTGICCRAARAAYCA	Geller et al. 2013
L499	10	F	480	ATTAATATACGATCAACAGGAAT	Van Houdt et al. 2010
H2123d	10	R	659	TAWACTTCWGGRTGWCCAAARAATCA	Van Houdt et al. 2010
LCO1490	11	F	0	GGTCAACAAATCATAAAGATATTGG	Folmer et al. 1994
mlCOIintGLR	11	R	5	GGNGGRTANANNGTYCANCCNGYNCC	Doug Wu pers. comm.
jgLCO1490	12	F	0	TITCIACIAAYCAYAARGAYATTGG	Geller et al. 2013
mlCOIintGLR	12	R	5	GGNGGRTANANNGTYCANCCNGYNCC	Doug Wu pers. comm.

In silico tests were carried using the R package Primer Miner developed by Elbrecht et al. (2017) using the following primer pairs:

LCO1490/HCO 2198	Folmer et al. 1994
mlCO1lintF/jgHCO2198	LeRay et al. 2013/Geller et al. 2013
BF1/BR1	Elbrecht & Leese 2017
BF2/BR2	Elbrecht & Leese 2017
BF1/BR2	Elbrecht & Leese 2017
BF2/BR1	Elbrecht & Leese 2017
LCO1490/mlCOIintGLR	Folmer et al 1994/Doug Wu pers.comm.

jgLCO1490/mlCOIintGLR Geller et al. 2013/Doug Wu pers.comm.

We were able to identify the vector sequences emerging from the metabarcoding exercise, using SASA's database of vector DNA sequences. Sequences from all other organisms were identified by comparison with existing publicly available databases such as Genbank and Barcode of Life.

In addition to invertebrates sampled using malaise traps, 20 invertebrates (from a range of taxa including Coleoptera, Diptera, Araneae, Hympenoptera, Collembola) sampled through pitfall trapping in 2013, 2014 and 2016 were tested to evaluate DNA quality (to determine their suitability for future DNA screening after several years in storage).

4. Metabarcoding

Instead of amplifying and sequencing a 658bp part of the COI region from a single organism using barcoding, metabarcoding allows the amplification of a subsection of the 658bp part of the COI region to provide individual DNA sequence for specimens in a mixed species sample consisting of a large number of different species caught in a number of individual traps. Labels attached to the amplification products of each trap allows the sequences to be traced back to the trap in which they were captured. As with barcoding the sequences can be searched for in the Barcode of Life database to provide a list of the species present in each trap.

A total of 34 malaise trap (M2) samples from the Scottish sites (8 taken in June, 12 in July, 14 in August) were subjected to metabarcoding analysis (aim 3; example samples shown in Appendix 1). An additional two samples, consisting of the mock communities referred to in section 3 (i) were amplified using two different pairs of primers.

i) DNA extraction

Extractions were carried out serially, to avoid any possibility of cross-contamination. The workbench was cleaned with DNA-Exitus plus (Anachem) and pipettes irradiated under UV light for 15 minutes between samples. Each sample was prepared by first pouring off the ethanol in which the samples had been stored and allowing the sample to air dry on filter paper in a flow-bench. Depending on the size of the sample and number of large carabids present, this took from 3-12 hours. Each sample was homogenised. Four grams of the homogenised sample was placed in a 50 ml tube and 21.6 mls ATL and 1.2 ml Proteinase K (Qiagen, Germany) were added. The volume of ATL and Proteinase K was adjusted proportionately if the homogenised tissue weighed less than 4g. The mixture was then further homogenised with an Omni 1000 hand held homogeniser. The 50 ml tube was then incubated at 56° C overnight, with gentle hand mixing where possible. DNA was extracted using the Qiagen blood and tissue extraction kit. Six extractions were carried out per sample and pooled in a single tube, representing the final extraction from that sample. To test the reproducibility of the method, two sets of extractions were carried out from the same starting material with three of the samples.

ii) DNA Metabarcoding

The amplified region consisted of part of the 658 base pair C terminal region in the gene encoding COI, as this region has been successfully used to identify arthropods (e.g. Araneae, Coleoptera, Formicidae Lepidoptera; Barrett & Hebert, 2005; Hajibabaei et al., 2006; Ji et al.,

2013). Polymerase Chain Reactions (PCR) were performed using the following primer pair for the COI region of interest:

Primer pair 1

- Forward: LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3')
- Reverse: mlCOIintGLR (5'-GGN GGR TAN ANN GTY CAN CCN GYN CC-3').

An additional primer pair was used on a single sample based on the 'mock community' so that the effectiveness of the two sets of primer pairs could be compared in an identical sample that was subjected to metabarcoding. The primer sequences for this pair are as follows:

Primer pair 2

- Forward: mlCOIintF (5'-GGWACWGGWTGAACWGTWTAYCCYCC-3') and
- Reverse: jgHCO2198 (5'-TAIACYTCIGGRTGICCRAARAAYCA-3').

These two sets of primer pairs amplify a different section of the 658bp COI region (Figure 3). As PCRs using the undiluted pooled DNA did not amplify well, a 1:10 dilution of the pooled DNA was subsequently used for the PCRs.

Six separate PCRs were carried out for each sample. An aliquot was checked on a 1.4% agarose gel PCR product pooled before library construction. Samples were then handed over to Edinburgh Genomics for library preparation, sequencing preparation of OTU (Operational Taxonomic Unit) list and species identification. Non-arthropod species, including Arionidae (slugs) and Lumbricidae (worms), were removed from the dataset. For the initial analysis we set a Genus confidence threshold of >0.9 and >9 reads per OTU for a species to be considered as present.

The final species lists were checked against previous records of species occurrence in Britain using primarily the National Biodiversity Networks Gateway (NBN Gateway, 2015), but also Fauna Europaea (de Jong et al., 2014), Antweb (AntWeb, 2015) and Araneae: Spiders of Europe (Nentwig et al., 2019). Where no previous record was found to species level, occurrence in Britain to Genus level was checked. There were nine cases where the Genus was present in Britain, but not the species; these were accepted by Genus name only. This anomaly could arise because the sequence of the British species is absent from the Barcode of Life database so that the software matches to a non-British species that is taxonomically closest to the sequence of interest present in the Barcode of Life.

RESULTS

1. Patterns of *Xylella* vector occurrence and abundance in woodland patches within agricultural landscapes

During this project, a total of 10,322 Auchenorrhyncha specimens were classified morphologically into 68 species (or higher taxonomic level when species id was not possible). On average, 132 individuals (range 15-687) and 9 species (range 1-21) were detected per site. Three Aphrophoridae, also known as spittlebugs, species were recorded: *P. spumarius* (159 specimens), *A. alni* (139 specimens) and *N. lineatus* (57 specimens); average spittlebug abundance per site was 4.6 specimens (range 0-48). Two other xylem feeder species were

also identified: *Cicadella viridis* (2 specimens) and *Evacanthus acuminatus* (1 specimen). A full list of species of Auchenorrhyncha identified is provided in Appendix 2.

There were no regional differences in the abundance of spittlebugs across woodland patches; however, distribution maps indicated higher relative abundance in more easterly sites in both Scotland and England (Figures 4 & 5).

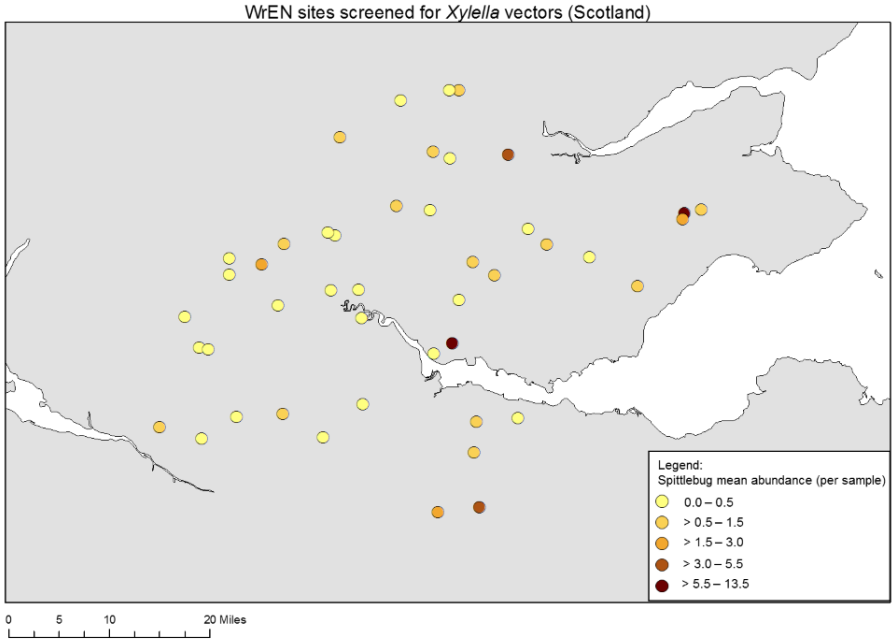


Figure 4. Map of WrEN study sites screened for spittlebugs showing the relative abundance of spittlebugs across central Scotland.

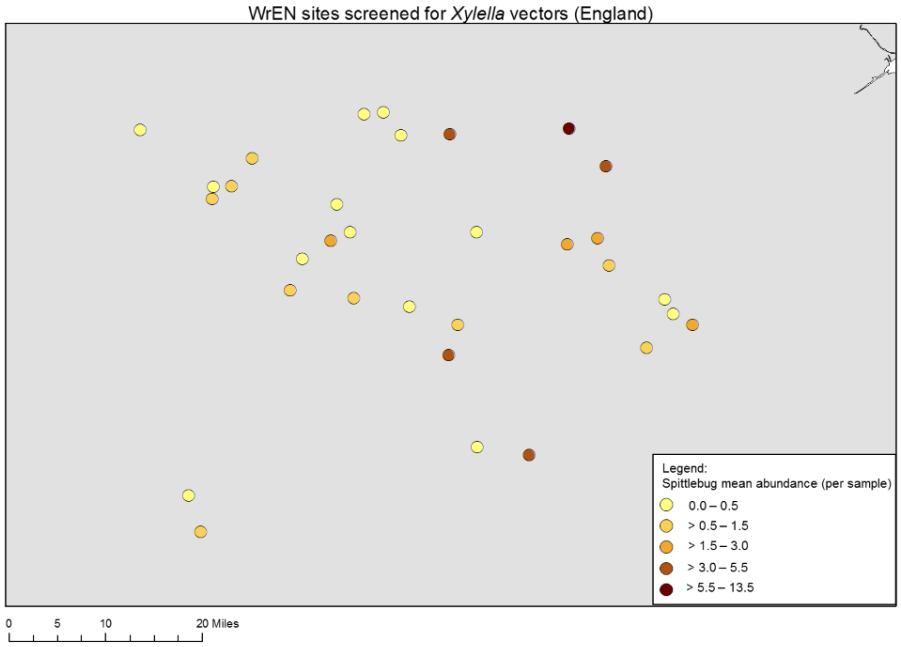


Figure 5. Map of WrEN study sites screened for spittlebugs showing the relative abundance of spittlebugs across central England.

Landscape-level attributes were more important than woodland characteristics in determining spittlebug abundance (Figure 6); spittlebugs were more abundant in woodlands located in landscapes with relatively low percentages of broadleaved woodland and other semi-natural habitats (i.e. a high percentage of agricultural land), and with low hedgerow densities (Figures 6 & 7). There were no direct effects of region or % agricultural land surrounding the focal woodlands. However, the Scottish woodlands in this study have significantly lower amounts of surrounding agriculture than England (and marginally higher amounts of semi-natural habitat), and % agriculture negatively covaries with % semi-natural habitat and % broad-leaved woodland. Of the woodland characteristics, only understorey cover marginally influenced spittlebug abundance (negatively, i.e. higher abundance in woodlands with a relatively open understorey; Figures 6 & 7).

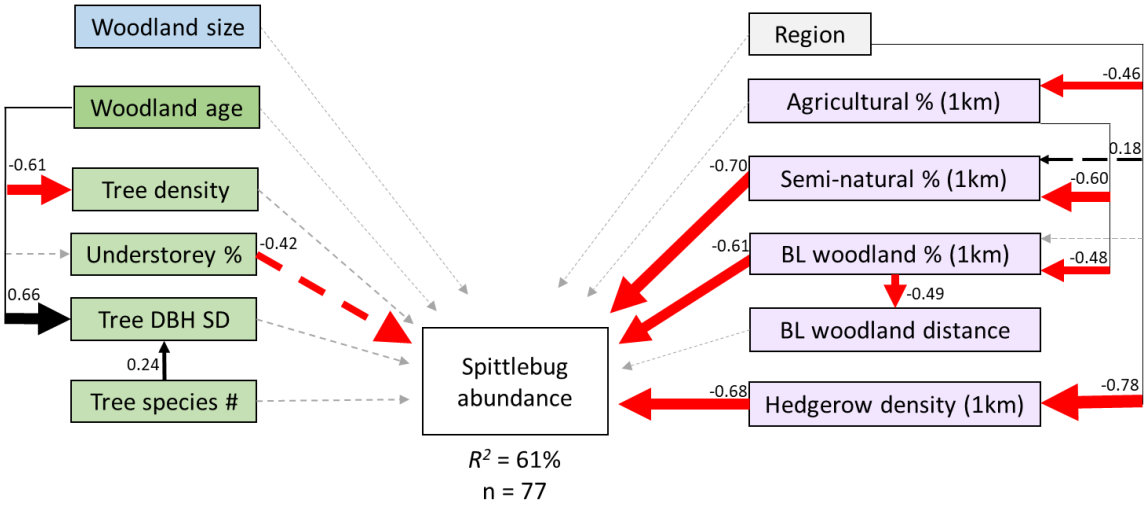


Figure 6. Effects of site- and landscape-level variables on spittlebug abundance in woodlands within agricultural landscapes. Coloured boxes indicate variable types: grey = geographical, green = vegetation structure, blue = patch geometry, purple = landscape. Arrow type and colour indicate statistical significance (solid black/red = significant association, i.e. $P<0.05$; dashed black/red = marginally significant association; i.e. $P<0.1$; dashed grey = non-significant association, i.e. $P>0.1$) and directionality of associations (black = positive; red = negative). Arrow thickness represents relative effect sizes (thicker arrows = larger effect sizes). Effect sizes are shown for all significant associations. Global goodness-of-fit: Fisher's C = 54.28 with P-value = 0.992, on 82 degrees of freedom.

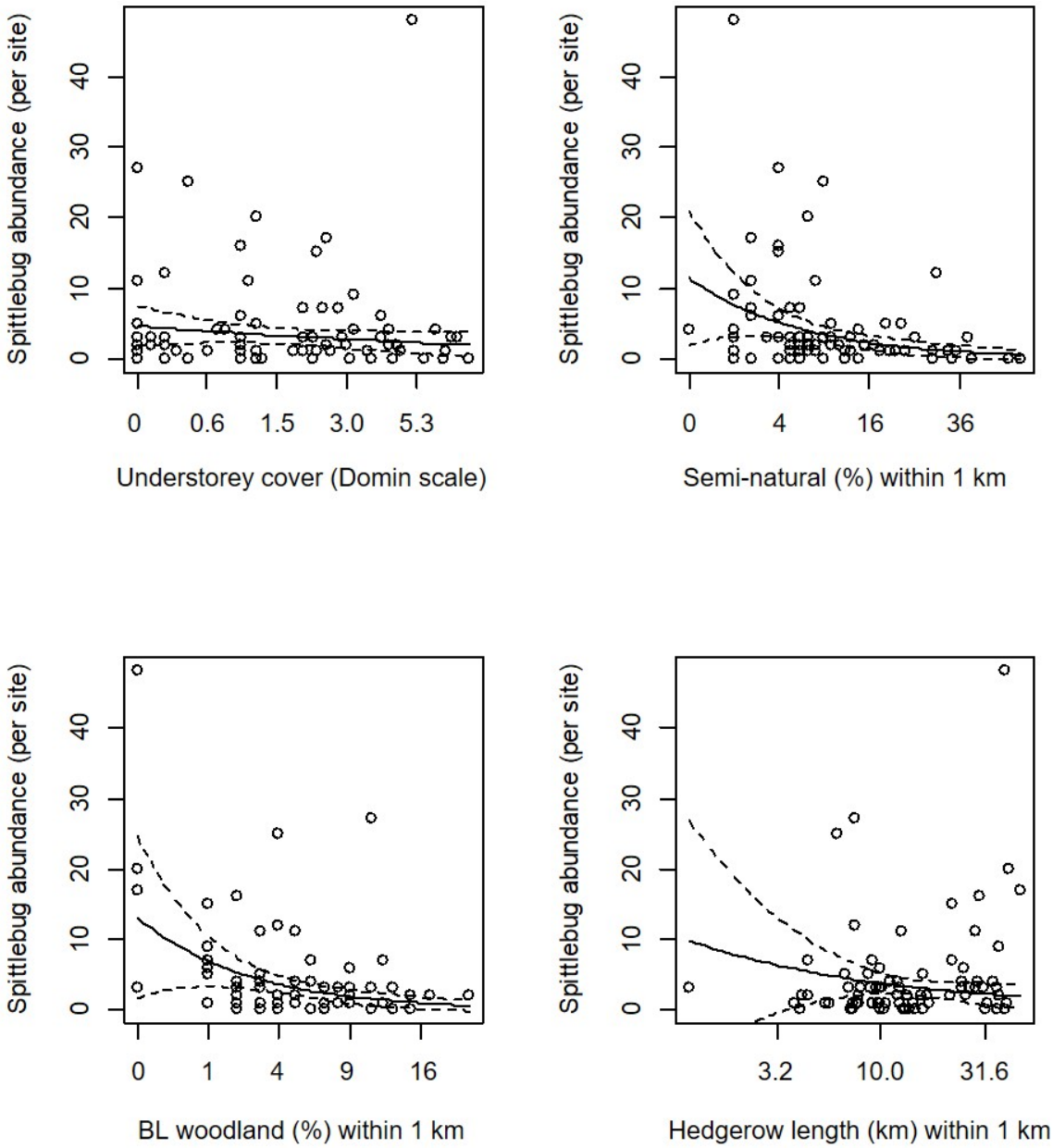


Figure 7. Plots of key predictors for spittlebug abundance in woodlands within agricultural landscapes. Dots represent observed values; solid line represents model predicted values ± 95% CI (dotted lines). Axes have been back-transformed for ease of interpretation (from a square-root transformation for % semi-natural and broadleaf (BL) woodland; log10 transformation for hedgerow length and understorey cover).

2. In silico primer testing and barcoding of pitfall trap samples

Across the taxonomic groups subjected to *in silico* testing, several primer pairs were identified that should successfully amplify parts of the 658 bp COI region of mitochondrial DNA (aim 2; Table 2). The primers were also tested *in vitro* via PCR on a broad range of taxa including Hemiptera, Hymenoptera, Coleoptera, Diptera, Lepidoptera, Araneae, Isopoda, Collembola. The combined result of these assays and the *in silico* testing was used to determine which primers were best taken forward to be used for metabarcoding.

Table 2. Results of *in silico* testing of primer pairs in five classes of UK arthropods (aquatic invertebrates not included). The red cells indicate high likelihood of amplification success. Note that it was not possible to carry out many *in silico* tests for cases in which one of the primers in a pair was located at the ends of the 658bp COI region as the sequence to which the primer was intended to attach was absent in the sequences present in the databanks. These primers are highlighted in blue. *=Primer pair 1, **=Primer pair 2

	Hemiptera ALL	**Opiliones (54)**	**Hymenoptera (20750)**	**Coleoptera (13671)**	**Diptera (24552)**
LCO/HCO	8.03	0.00	10.91	3.94	0.00
mlCO1lintF/jgHCO2198**	51.06	100.00	73.03	67.56	85.42
BF1/BR1	96.91	100.00	88.15	90.58	97.64
BF2/BR2	6.15	No Data	1.28	12.08	8.07
BF1/BR2	96.42	100.00	91.15	92.28	97.35
BF2/BR1	7.41	No Data	3.70	11.95	11.95
LCO1490/mlCOIintGLR*	37.89	0.00	65.49	44.32	0.00
jgLCO1490/mlCOIintGLR	88.11	100.00	97.79	96.77	98.17

The 658bp region of DNA was also successfully amplified using the Folmer et al. (1994) primer pair from invertebrates sampled through pitfall trapping in 2013, 2014 and 2016 (aim 2b). Species tested from pitfall traps included insects from Araneae, Coleoptera, Hymenoptera and Diptera. The DNA in these samples has preserved well and any sequence failures are due to issues with the extraction process and/or primers. Work on optimising the procedure for these samples is on-going.

3. Metabarcoding results

Every malaise trap sample subjected to metabarcoding produced a high number of reads ranging from 218848-786687. The large volume of data will require considerable analysis in order to achieve a full interpretation of the results.

Total Operational Taxonomic Units

We obtained 6729 OTUs across all metabarcoded samples (two mock community and 34 malaise trap samples) of which the majority (6687) were classified as Arthropods. Considering the malaise trap samples alone, the metabarcoding identified a total of 2106 species which was reduced to 1009 when a confidence threshold of 0.90 was imposed. Raising the

confidence threshold to 0.95 reduces the number of species to 772 and resulted in the omission of *P. spumarius* from the species list, so we have adopted a confidence threshold of 0.90 for reporting identified species (Appendix 3).

Mock community

Equal amounts of DNA from a broad taxonomic range of over 100 arthropod individuals was used as the basis of two of the metabarcoded samples; one sample was amplified using Primer pair 1 and the other using Primer pair 2. An attempt was made to morphologically identify all the components to species but a large number could only be identified to a higher taxonomic level. We also attempted to produce a barcode of 658 bp long using the Folmer primer which should include both regions amplified by the two metabarcoding primer pairs (Figure 3). Primer pair 1 identified 53 OTUs to species level in the mock community whereas primer pair 2 was less effective and only identified 33 species, of which 29 were also identified by primer pair 1. Of the total of 56 species identified by either metabarcoding primer pair 1 or 2, we obtained an individual barcode for 36 which when BLASTED against BOLD database agreed with the metabarcoding species id. A further three species detected by metabarcoding were identified morphologically to species level but failed to produce a barcode. The remaining 17 species identified by metabarcoding may be present in the mock community but may have failed to produce a barcode, or were not identified morphologically to species level. Some of the 17 species corresponded to a specimen that was identified morphologically or via barcoding to the same genus but a different species. Further work will enable a more in depth analysis of the mock community results to determine why morphological, barcoding and metabarcoding results do not entirely agree.

Primer pair 1 was successful in allowing the three main Xylella vector species *Philaenus spumarius*, *A. alni* and *N. lineatus* to be detected by metabarcoding of the mock community but primer pair 2 failed to detect any of these three species.

Malaise traps

At the 0.90 confidence (Appendix 3) the species list for the arthropod captured across the 34 malaise traps was dominated by Diptera (355), followed by Hymenoptera (148), Coleoptera (106), Lepidoptera (83) and Hemiptera (81), and smaller numbers of other groups. Of these the three species with the highest number of reads across all samples were *Phaonia valida*, *Helina depuncta* and *Phaonia angelicae*, all Dipteran flies.

Metabarcoding successfully detected the presence of the three main candidate vector species. With two exceptions, results from morphological and metabarcoding identification were in complete agreement regarding the samples in which the three spittlebug vectors were present. *Philaenus spumarius* was detected by morphological identification in ten samples that were subsequently identified by metabarcoding. Additionally there was one sample in which *P. spumarius* was detected by metacoding but not by morphological id. However, this sample contained several nymphs which are notoriously difficult to identify on the basis of morphology and this may explain the cause of the discrepancy. *Neophilaenus lineatus* was detected in the same five samples by both methods. Finally, *A. alni* was detected in three samples morphologically and subsequently identified by metabarcoding in two of these samples.

Members of the Cicadellidae were also detected in the Malaise trap samples subjected to metabarcoding. The Cicadellidae species that were detected by metabarcoding were: *Alebra*

wahlbergi, Balcutha punctate, Cicadella viridis, Empoasca decipiens, Empoasca luda, Eupteryx calcarata, Fagocyba douglasi, Notus sitka, Oncopsis flavicolis, Ribautiana ulmi, Speudotettix subfausculus, Typhlocyba quercus and *Typhlocyba rosae*. No individual Cicadellidae species was universally present across all Malaise trap samples but the most common were *Speudotettix subfusculus*, *Eupteryx calcrata* and *Empoasca luda*. Five of these species were identified morphologically in the Malaise trap samples from Scotland but the morphological list contained a large proportion of samples that could only be identified to genus or family level.

DISCUSSION

Patterns of Xylella vector abundance (aim 1)

The three species of Aphrophoridae or spittlebug morphologically identified within the sampled woodlands are widespread, with the most common, *P. spumarius*, also being the most abundant in our samples. Whilst many spittlebugs are typically associated with grasslands, they can be found in a wide range of habitats, and *P. spumarius* (the meadow spittlebug) is found in most terrestrial habitats (Weaver & King 1954; Yurtsever 2000).

Collectively, in this study, spittlebugs were more abundant in woodlands located in landscapes with relatively low percentages of broadleaved woodland and other semi-natural habitats. It is possible that higher abundance in such sites is a result of spittlebug plant host species being relatively uncommon in woodland and other semi-natural habitats, but this is not something we have information on. It is also important to note that this study was based wholly in predominantly broadleaved woodlands and we cannot compare patterns of spittlebug occurrence and abundance we have uncovered with those of other habitats. We recommend that a wider range of habitat types and landscape settings (e.g. different matrix habitats) should be examined for patterns of vector occurrence and abundance. To better understand movement between, and within habitats (i.e. it is suspected that adult spittlebugs migrate into the tree canopies before returning to the understorey to lay eggs), these surveys need to include both larvae and adults using a variety of sampling techniques (e.g. sweep nets, hand-held suction traps). This would enable stronger characterisation of vector presence in Scottish landscapes.

DNA barcoding on morphologically identified UK arthropods (aim 2)

We identified a primer pair which was successful in enabling a broad taxonomic range of UK arthropod samples (including the three main Xylella vector species) to be detected by metabarcoding. The barcoding analyses undertaken indicated that the DNA quality of arthropods in pitfall samples collected more than five years ago (and not collected or stored for the purposes of genetic analyses) was relatively high and potentially suitable for future metabarcoding studies.

Meta-barcoding proof-of-concept study (aim 3)

The meta-barcoding proof-of-concept study successfully extracted DNA from all malaise traps sampled. The recommended number of reads to aim for per sample in a metabarcoding project is 60,000 and our results far exceeded this with between 218,848 - 786,687 per sample. This offers the opportunity to utilise the purchased number of reads in a

metabarcoding project to be used more economically. For example, by aiming to obtain 60,000 reads per sample we could increase the number of samples analysed by a factor of three. As it is the preparation of the libraries that is the expensive step in metabarcoding, this would double the price of the project but enable three times more samples to be analysed. Care needs to be taken in reducing the number of reads per sample, however, if the focus of the project is to detect the three main candidate vector species as they were present in low numbers and therefore produced a relatively low number of reads per sample.

Our study offers a unique opportunity to compare the relative effectiveness of identifying mixed samples of arthropods by morphological, barcoding and metabarcoding approaches as we have used all three approaches on our mock community samples. More in depth analysis is required to understand how species lists drawn up using these three approaches differ and whether accuracy is compromised by the faster metabarcoding approach.

It proved difficult to compare likely primer success *in silico* across a broad taxonomic range of arthropod species because many of the sequences available in publicly available databases did not contain the primer region. However, the pilot study does demonstrate the importance of primer pair choice in metabarcoding as only one of the primer pairs was effective in detecting the three main candidate *Xylella* vector species in the mock community. This primer pair was also more effective in identifying the other component species of mock community. This primer also detected the three main vector species in the Malaise trap samples. Therefore, Primer pair 1, rather than pair 2, should be used for any future metabarcoding study of arthropods.

Arthropods belonging to the Cicadellidae are also of interest in terms of possible *Xylella* vectors and the metabarcoding study detected 13 of these species across all Malaise trap samples. Only five of these species were morphologically identified in Malaise trap samples from across Scotland and many of these samples were only identified to Genus or family level. Morphological identification is hampered by the fact that many of the larval stages are difficult to identify to species level. One of the advantages of the barcoding and metabarcoding approaches is that they make identification of larval stages much easier provided an accurate species morphological id accompanies any barcoded sequences that are available in publicly available databases. The use of metabarcoding therefore shows great promise for future studies, both for early detection of pest species, as well as detection of rare species of conservation concern.

Metabarcoding detected a total of 1009 individual species across all the Malaise traps. This number is of the expected magnitude as it is similar to the 1128 arthropod species detected in a metabarcoding project involving Malaise trapping over an eight week period in 12 oak, pine or mixed woods in SE England (http://conferences.au.dk/fileadmin/C_Bruce_COI_metabarcoding_of_forest_arthropods_01.pdf). Their finding that 60% of the identified species were Diptera with a lower prevalence of Hymenoptera, Hemiptera, Lepidoptera, Coleoptera, Arachnida and others also agrees with what we found. These findings demonstrate that in addition to application for detection of specific species of interest, metabarcoding has the potential to provide a tool to assess general arthropod diversity in woodlands. This could be useful for assessing the effects of landscape, species composition and management on the biodiversity supported by a range of woodlands,

and of potential 'at risk' species if control measures to manage plant diseases were implemented.

Conclusions and future directions

The presence of a range of xylem-feeding invertebrates has been confirmed in broadleaved and mixed woodlands in a range of settings in central Scotland. Specifically, spittlebugs known to be vectors, or implicated in the transmission of *Xylella* elsewhere were documented. The presence (and abundance) of spittlebugs is related to features of the surrounding landscape, i.e. there is an increased likelihood of the vectors in woodlands located in landscapes with relatively low percentages of broadleaved woodland and other semi-natural habitats. However, this study focussed predominantly on broadleaved woodlands, and little is known about how the abundance of vectors varies spatially (between and within habitats) and temporally. These are key areas for future work if the ecology of these vectors is to be better understood.

Molecular detection methods (barcoding and metabarcoding) have been refined (e.g. improved primers), confirmed against morphological ID and found not to be generating false positives. This will be of value should a suspected outbreak require rapid investigation. Additionally, this project has increased capacity in the Scottish science community to identify key vectors through additions to SASA's database on vector DNA and training in invertebrate ID. Future work in this area should focus on further optimisation of these molecular approaches, and exploration of alternative approaches that may substantially reduce analytical costs.

Acknowledgements

The WrEN project was developed with funding from the Forestry Commission, Forest Research, University of Stirling, Natural England, Department for Environment, Food and Rural Affairs, The National Forest Company, Scottish Natural Heritage and Tarmac. Invertebrate samples were collected and processed by Lauren Fuller, Katja Bitenc, Nadine Royle, Roy Allen and Ian Hayward. We are grateful to all the land owners who allowed access to their woodlands. The work outlined in this report was funded by the Scottish Government's Rural and Environment Science and Analytical Services (RESAS) Division through the Centre of Expertise for Plant Health.

REFERENCES

Barrett RDH & Hebert PDN (2005). Identifying spiders through DNA barcodes. *Canadian Journal of Zoology* 83, 481–491.

Brandon-Mong GJ, Gan HM, Sing KW, Lee PS, Lim PE & Wilson JJ (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin Entomological Research* 105, 717–727.

de Jong Y, Verbeek M, Michelsen V, Bjørn P, Los W, Steeman F, Bailly N, Basire C, Chylarecki P, Stloukal E, Hagedorn G, Wetzel F, Glöckler F, Kroupa A, Korb G, Hoffmann A, Häuser C, Kohlbecker A, Müller A, Güntsch A, Stoev P, Penev L (2014). Fauna Europaea – all European

animal species on the web. *Biodiversity Data Journal* 2, e4034. https://doi.org/10.3897/BDJ.2.e4034

Elbrecht V & Leese F (2017). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science*, doi:10.3359/fens.2017 00011.

Eyre MD (2005). Habitat diversity in the conservation of the grassland Auchenorrhyncha (Homoptera: Cercopidae, Cicadellidae, Cixidae, Delphacidae) of northern Britain. *Journal of Insect Conservation* 9, 309 – 317.

Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R (1994). DNA primers for amplification of mitochondrial c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology & Biotechnology* 3, 294–299.

Fuller L, Fuentes-Montemayor E, Watts K, Macgregor NA, Bitenc, K & Park KJ (2018). Local attributes determine the suitability of woodland creation sites for Diptera. *Journal of Applied Ecology* 55, 1173-1184 Supplementary S1.2018.

Geller J, Meyer C, Parker M & Hawk H (2013). Redesign of PCR primers for mitochondrial cytochrome coxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources* 13, 851–861.

Gibson J, Shokralla S, Porter TM, King I, van Konynenburg S & Janzen DH (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences of the United States of America* 111, 8007–8012.

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ & Whitefield JB (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology* 6, 959–964.

Hebert PDN, Penton EH, Burns JM, Janzen DH & Hallwachs W (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101, 14812–14817.

Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P, Edwards FA, Larsen TH, Hsu WW, Benedick S, Hamer KC, Wilcove DS, Bruce C, Wang X, Levi T, Lott M, Emerson BC & Yu DW (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* 16, 1245–1257.

Lefcheck J (2016). PIECEWISESEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods in Ecology & Evolution* 7, 573 – 579.

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT & Machida RJ (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10, 34.

Liu S, Wang X, Xie L, Tan M, Li Z, & Su X (2016). Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources* 16, 470–479.

Malumphy C (2017). European Conference on *Xylella fastidiosa* 2017: finding answers to a global problem. Mallorca, Spain.

Meusnier I, Singer GA, Landry J-F, Hickey DA, Hebert PD & Hajibabaei M (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9, 214.

Nentwig W, Blick T, Gloor D, Hänggi A & Kropf C (2019). Araneae Version 05.2019 Online at https://www.araneae.nmbe.ch. doi: 10.24436/1

Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH & Hallwachs W (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports* 5, 9687.

Van Houdt JKJ, Breman FC, Virgilio M & De Meyer M (2010). Recovering full DNA barcodes from 613 natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular Ecology Resources* 10, 459–465.

Watts K, Fuentes-Montemayor E, Macgregor NA, Peredo-Alvarez V, Ferryman M, Bellamy C, Brown J & Park KJ (2016). Using historical woodland creation to construct a long-term, large-scale natural experiment: The WrEN project. *Ecology and Evolution*, 6, 3012– 3025.

Weaver CR & King DR (1954). Meadow spittlebug *Philaenus leucophtalmus* (L.). Ohio *Agricultural Experimental Station Research Bulletin* 741, 1-99.

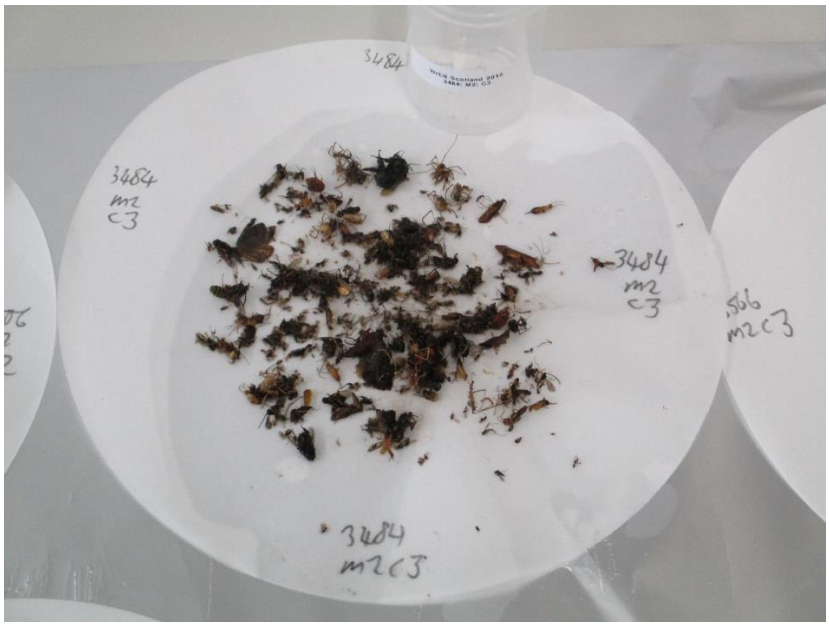
Wilson M, Stewart A, Biederman R, Nickel H & Niedringhaus R (2015). The Planthoppers and Leafhoppers of Britain and Ireland. *Identification keys to all families and genera and all British and Irish species not recorded from Germany*. ISBN: 9783939202066.

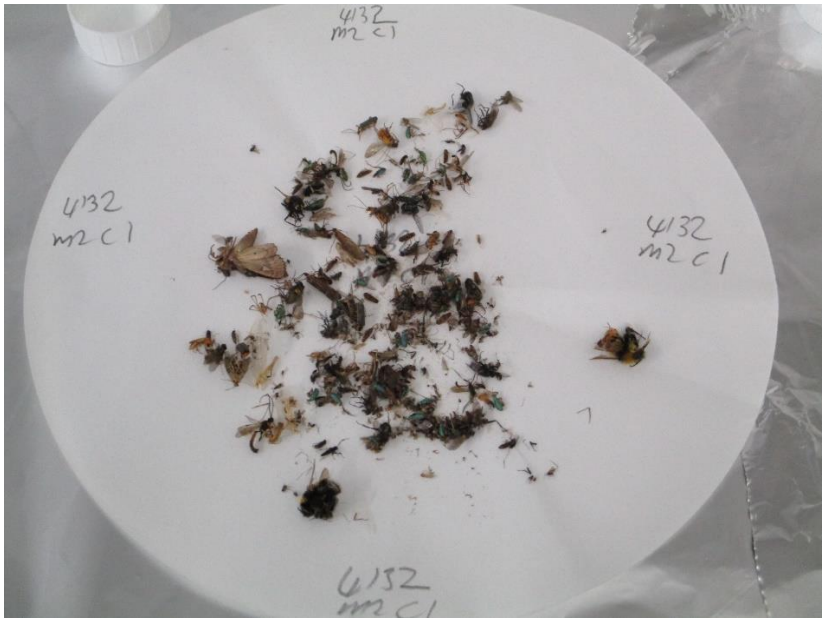
Yurtsever S (2000). On the Polymorphic Meadow Spittlebug, Philaenus spumarius (L.) (Homoptera: Cercopidae). *Turkish Journal of Zoology* 24, 447-459.

Zeale MRK, Butlin RK, Barker GLA, Lees DC & Jones, G. (2010). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources* 11, 236–244.

APPENDICES

Appendix 1. Examples of malaise trap samples, pre-processing.





Appendix 2. List of Auchenorrhyncha species detected at WrEN sites. Spittlebugs (denoted [b]) and other Xylem feeders (denoted [a]) are shown in bold.

Adarrus ocellaris	*Eypteryx stachydearum*
Agallia laevis/ribauti	*Grypotes puncticollis*
Alebra albostriella	*Iassus lanio*
Alebra wahlbergi	*Javesella discolor*
Allygus mixtus	*Javesella dubia*
Allygus modestus	*Javesella pellucida*
Aphrodes makarovi	*Lamprotettix nitidulus*
Aphrophora alni [a, b]	*Linnavuoriana decempunctata*
Arthaldeus pascuellus	*Linnavuoriana sexmaculata*
Balclutha punctata	*Macropsinae*
Cicadella viridis [a]	*Macropsis scotti*
Cicadula frontalis	*Macropsis scutellata*
Cicadula persimilis/saturata	*Macrosapinae nymph*
Cixius nervosus	*Macrosteles sp.*
Cixius simplex	*Macrosteles variatus*
Conomelus anceps	*Macustus grisescens*
Conosanus obsoletus	*Megopthalmus scabripennis*
Delphacinus mesomelas	*Muellerianella fairmairei*
Dicranotropis hamata	***Neophilaenus lineatus*** [a, b]
Dikraneura variata	*Oncopsis alni*
Elymana sulphurella	*Oncopsis flavicollis*
Eupelix cuspidata	*Oncopsis flavicollis/subangulata*
Eupterycyba jucunda	*Oncopsis subangulata*
Eupteryx aurata	*Oncopsis tristis*
Eupteryx cyclops	*Opsius stactogalus*
Eupteryx cyclops/urticae	*Paluda adumbrata*
Eupteryx florida	***Philaenus spumarius*** [a, b]
Eupteryx melissae	*Speudotettix subfusculus*
Eupteryx stachydearum	*Stenocranus minutus*
Eupteryx urticae	*Struebingianella lugubrina*
Eupteryx vittata	*Tachycixius pilosus*
Eurhadina pulchella	*Thamnotettix confinis*
Euscelis incisus	*Thamnotettix dilutior*
Evacanthus acuminatus [a]	*Typhlocyba quercus*

Appendix 3 See separate Excel spreadsheet for a list of arthropods identified through metabarcoding from the WrEN project malaise traps. Only Arthropods with a confidence level of > 0.90 are shown, and are presented in taxonomic order (OTU = Operating Taxonomic Unit), with presence (1) / absence (0) recorded for each site.