

Using citizen science data to assess the population genetic structure of the common yellowjacket wasp, *Vespula vulgaris*

Iona Cunningham-Eurich^{1,2}  | Danai Kontou^{1,3}  | Monika Yordanova^{1,4}  |
Alejandro Maeda-Obregon¹ | Emeline Favreau¹  | Jinliang Wang⁵ |
Adam G. Hart⁶ | Seirian Sumner¹

¹Centre for Biodiversity and Environment Research, University College London, London, UK

²Department of Life Sciences, The Natural History Museum, London, UK

³Department of Plant Sciences, University of Cambridge, Cambridge, UK

⁴Department of Life Sciences, Imperial College London, London, UK

⁵Institute of Zoology, Zoological Society of London, London, UK

⁶Department of Natural and Social Science, University of Gloucestershire, UK

Correspondence

Iona Cunningham-Eurich, Centre for Biodiversity and Environment Research, University College London, London, UK.
Email: iona.cunningham-eurich.16@ucl.ac.uk

Funding information

Royal Entomological Society

Abstract

Monitoring insect genetic diversity and population structure has never been more important to manage the biodiversity crisis. Citizen science has become an increasingly popular tool to gather ecological data affordably across a wide range of spatial and temporal scales. To date, most insect-related citizen science initiatives have focused on occurrence and abundance data. Here, we show that poorly preserved insect samples collected by citizen scientists can yield population genetic information, providing new insights into population connectivity, genetic diversity and dispersal behaviour of little-studied insects. We analysed social wasps collected by participants of the Big Wasp Survey, a citizen science project that aims to map the diversity and distributions of vespine wasps in the UK. Although *Vespula vulgaris* is a notorious invasive species around the world, it remains poorly studied in its native range. We used these data to assess the population genetic structure of the common yellowjacket *V. vulgaris* at different spatial scales. We found a single, panmictic population across the UK with little evidence of population genetic structuring; the only possible limit to gene flow is the Irish sea, resulting in significant differentiation between the Northern Ireland and mainland UK populations. Our results suggest that queens disperse considerable distances from their natal nests to found new nests, resulting in high rates of gene flow and thus little differentiation across the landscape. Citizen science data has made it feasible to perform this study, and we hope that it will encourage future projects to adopt similar practices in insect population monitoring.

KEYWORDS

big wasp survey, citizen science, insects, population genetics, vespines

INTRODUCTION

There is no doubt that anthropogenic action is having wide-spread impacts on biodiversity. However, biodiversity research is often

taxonomically biased towards large and/or charismatic animals. Most insects have received very little attention (Leather, 2018; Rocha-Ortega et al., 2021; Sumner et al., 2018). This is of particular concern as insects constitute a significant proportion of biodiversity, with some estimates exceeding 5.5 million species (Stork, 2018). In addition, insects provide important ecosystem services, including provisioning (e.g., food, pharmaceuticals), regulating (e.g., pollination,

Danai Kontou, Alejandro Maeda-Obregon and Monika Yordanova contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Insect Molecular Biology* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society.

predation), cultural (e.g., tourism) and supporting (e.g., nutrient cycling, food source) services (Cardoso et al., 2011). Although insect populations have been widely reported as declining worldwide (Hallmann et al., 2017; Powney et al., 2019), recent research has shown that some taxa may instead benefit from some forms of anthropogenic action (Bowler et al., 2021; Crossley et al., 2020; Macgregor et al., 2019; Outhwaite et al., 2020). Overall, it is clear that we currently have little understanding of how different insect species are influenced by the complexities of anthropogenic change. There is evidently a need to shift the emphasis of recording and reporting to include insects that have historically received less attention. One of the biggest challenges is finding ways to do this quickly and affordably.

Citizen science has become increasingly popular over the past decade or so as a method to help scientists gather information on biodiversity rapidly and across large spatial scales. Some of the biggest initiatives, such as iNaturalist (www.inaturalist.org) or eBird (www.ebird.org), have contributed millions of biological occurrence records, providing insights into important ecological questions, for instance species distributions (Comont et al., 2012; Zapponi et al., 2017); invasive species monitoring (Brown et al., 2018); effects of climate change and land-use (Whitehorn et al., 2022); and changes in population abundance over time (Hallmann et al., 2017), thus demonstrating the value of citizen scientists to mainstream science. An exciting development of citizen science in recent years has seen members of the public extending their contribution to collecting samples. Freshwater research has been particularly proactive in recruiting volunteers to collect samples, to the extent that citizen scientists now play an integral role in freshwater ecosystem monitoring and research (Metcalf et al., 2022; Thornhill et al., 2019). Furthermore, having citizen scientists willing to provide samples opens the possibility of extracting genetic information. Such data can help understand (and predict) population composition and change (Schwartz et al., 2007), yielding insights into demography and population dynamics, including population structuring, differentiation and rates of gene flow across the landscape (Bohonak, 1999). This can help answer questions relating to geographic barriers to dispersal (e.g., Brunke et al., 2019), effects of land-use change on dispersal and survival (e.g., Dreier et al., 2014), population size (e.g., Furlan et al., 2012) or invasion history (e.g., Ascunce et al., 2011).

To our knowledge, there are currently only a few examples of genetic studies that have used citizen science/volunteer-collected samples in ecological research (Agersnap et al., 2022; de Virgilio et al., 2020; Guindon et al., 2015; Neveceralova et al., 2022; Skrbinišek et al., 2019). However, the reach of these studies remains limited given that volunteers are required to possess specialist skills/knowledge and/or be specifically trained to process the samples. On the other hand, insect-based research has the potential for more inclusive and extensive sampling using citizen-science: insects are small and typically easy to catch in traps. Samples can then be transferred to laboratories for more extensive analyses, including identification to lower taxonomic levels, which often requires expert input (Breeze et al., 2021; Le Féon et al., 2016; Lucky et al., 2014; O'Connor et al., 2019). Some studies focusing on insects have already used these attributes to their

advantage and have coupled them with genetic analyses. These include the UK Pollinator Monitoring Scheme (<https://ukpoms.org.uk/>), which performs DNA barcoding on pan-trap collected samples to assess species diversity (Breeze et al., 2021; Creedy et al., 2020) and the School of Ants project (<https://andrealucky.com/school-of-ants/>), which characterised the population genetic structure of the invasive ant *Tetramorium immigrans* in North America (Zhang et al., 2019).

Vespine wasps (yellowjackets and hornets) are among those insects that remain poorly studied (Sumner et al., 2018) despite increasing evidence of their importance in supporting ecosystem functioning (Brock et al., 2021). Most recent research has focused on populations in non-native regions due to their high success as invasive species around the world with large economic and ecological impacts (Lester & Beggs, 2019). These wasps are eusocial with typically annual colonies (spring to autumn in native regions) that can host tens of thousands of workers which forage from a fixed nest. Sexuials (gynes and males) emerge in the autumn and, after mating, males die; mated gynes on the other hand disperse and overwinter before commencing their own nesting cycle in the spring (Lester & Beggs, 2019). Indeed, vespine wasps are well-known for their dispersal abilities, with data from invasive regions showing that they can spread at rates exceeding tens of kilometres per year (Masciocchi & Corley, 2013; Robinet et al., 2017). Population genetics studies have shown that populations are generally panmictic across continuous landscapes (i.e., unbroken by large geographical barriers such as oceans or seas) (Arca et al., 2015; Chau et al., 2015; Elof et al., 2020; Hoffman et al., 2008; Schmack et al., 2019), suggesting high dispersal rates. It must be noted, however, that this is not always the case, for instance, in *Vespa mandarinia* at the continental scale in its native range (Arca et al., 2015) and *Vespula germanica* in South Eastern Australia (invasive range) across only a few tens of kilometres (although the extent to which this might be an effect of genetic bottlenecks and invasion history is unclear) (Goodisman et al., 2001). However overall, this indicates that whilst workers forage intensively as close to the nest as possible, probably within a few hundred metres (Akre et al., 1975; Greenleaf et al., 2007; Matsuura & Yamane, 1990), gynes are likely to disperse further from their natal nests before founding their own. Thus, the behavioural differences between these two castes may generate different population structuring at different spatial and temporal scales; however, we lack comprehensive tests of these conjectures. Additionally, there are currently no comprehensive studies on the population genetic structure of native populations of vespines in Europe, particularly the UK, which is the source country for multiple invasive populations around the world (Brenton-Rule et al., 2018; Lester & Beggs, 2019; Schmack et al., 2019). An understanding of the genetic structure of their native populations is essential to predict how their ecosystem services will be/are being affected by anthropogenic change (Brock et al., 2021).

Here, we used citizen science-collected samples from the Big Wasp Survey (BWS) (www.bigwaspsurvey.org) to analyse the UK-wide population genetic structure of the common yellowjacket wasp *Vespula vulgaris* using polymorphic microsatellite markers. The BWS is a citizen science project that has run annually in the UK since 2017. It aims to improve our knowledge of social vespine wasps and encourage people

to gain a better understanding of these little-loved insects. Members of the UK's public were invited to hang a simple homemade bottle trap in their gardens for a week in late August and then post the samples to identification hubs where experts could verify species identity. Since 2020, participants have identified their own samples using online identification tools developed for the project (Perry et al., 2021). The project has proved highly successful in engaging with the public: between 2017 and 2021, 3389 unique participants/participant groups set out 7916 traps across the UK, catching over 62,000 wasps. The data have produced reliable species distribution maps that are comparable in quality to those generated from four decades worth of data collected by expert recorders (Sumner et al., 2019). Moreover, the project stores nearly 50,000 physical wasp samples (collected between 2017 and 2019) now preserved in alcohol, which have the potential to provide molecular insights into native vespine wasps across the UK.

Because of the way the samples were collected, the DNA was likely to be quite degraded. Given restriction on budget, our first aim was to determine a cost-effective method to obtain DNA from individual wasps that would be suitable for analyses at multiple microsatellite markers (Aim 1). We tested two DNA extraction methods and then assayed existing microsatellite markers developed for different vespine species for their utility in studying the population genetics of *V. vulgaris*. We then used these data to determine the population structure of *V. vulgaris* in the UK at four different geographical scales: within traps, local (<10 km), regional (South-East of England, 185 km) and national (across the whole of the UK, 850 km) (Aim 2).

Different scales can inform on different temporal trends, including long-term dispersal patterns at larger scales (regional and national) and recent dispersal events at the finer scales (local) (Bluher et al., 2020). In this case, we were interested in testing hypotheses about how the contrasting dispersal patterns of queens and workers might give rise to different patterns of genetic structure (Figure 1). Specifically, we predict that there will be evidence of viscous population structure at a fine (local) spatial scale, given that workers are assumed to forage as close to the nest as possible. Accordingly, at the finer spatial scales (within traps and between nearby traps), we expect to detect siblings (or half-siblings) among foraging workers. At the intermediate spatial scale (<200 km distance), we expect little or no population differentiation given that gynes are assumed to be good dispersers. We expect a similarly low/absence of genetic differentiation at the national scale, unless geographical features (e.g., sea bodies) limit dispersal. This study is (to our knowledge) the first comprehensive population genetic study of one of the most invasive *Vesputa* species (*V. vulgaris*) in its native range and provides much-needed insights into their ecology and dispersal behaviours. Second, it is a rare example of how poorly preserved citizen-science-collected insect samples can be used for genetic analyses.

RESULTS

Sampling

We selected a total of 393 samples collected in two consecutive years of the survey: 2017 and 2018. These were selected using a

hierarchical geographical design at different scales (see Introduction), limited by the locations that had been sampled by BWS participants. At the national scale, 137 wasps from the 2017 samples were selected: 91 individuals were randomly selected, and a further 46 individuals were selected from under-sampled areas (Northern Ireland and Northern England). These samples were divided into six different regional subsets defined by county lines and latitudinal gradient, with an additional separation between East and West England (Figure 1). For the purpose of these analyses, we treated each region as a putative 'population'.

For the finer scale analyses (regional, local and within trap), we selected 256 samples from 2018; samples were selected from this year rather than the previous one as the BWS had resulted in higher participation in 2018, giving us more choice regarding sample selection. To create putative 'populations' for these analyses, we defined 10 clusters of 4–6 traps each, with any two traps within the cluster separated by up to 10 km, and with at least one pair of traps per cluster separated by less than 1 km. These distances were selected based on reported foraging distances of social wasps and other Hymenoptera (Akre et al., 1975; Greenleaf et al., 2007; Matsuura & Yamane, 1990). Any samples that did not belong to a cluster were removed. For the local scale analyses, we treated each trap within each of the 10 clusters as a 'population'. For the regional scale analyses, we treated each of the clusters located in Southern England (across 185 km) as a 'population'; note that one of the clusters was omitted from this analysis as it was located in Edinburgh, Scotland and outside the geographical scope of this analysis, therefore this analysis consisted of nine putative 'populations'.

Aim 1: Determining a cost-effective but reliable method for genotyping degraded citizen science samples

We sampled a total of 393 wasps: 302 were genotyped at up to 15 loci using a modified Chelex protocol (256 from 2018 and 46 from 2017) (henceforth referred to as the Chelex method) and 91 (from 2017) were genotyped at up to 12 loci using a DNeasy® Blood and Tissue kit (QIAGEN) (the DNeasy method). We were unsuccessful in extracting DNA from 10 of the DNeasy samples. Two further DNeasy samples and 80 Chelex samples (20 from 2017 and 60 from 2018) did not amplify at seven or more loci. These samples were removed from subsequent analyses. The final sample sizes were 105 individuals from 2017 and 196 from 2018; of these, 79 were DNeasy samples and 222 were Chelex samples. We tested the influence of methods, days and liquid in trap using a generalised linear model (GLM). Amplification success was significantly influenced by the DNA extraction method ($p = 7.5 \times 10^{-3}$), with the DNeasy method being more efficient than the Chelex method. There was no effect of the liquid (either orange juice or beer) used in the trap nor the number of days the trap was set out for. We did not include year in this analysis given that year and method were strongly associated ($X^2 = 203.26$, $p < 2.2 \times 10^{-16}$).

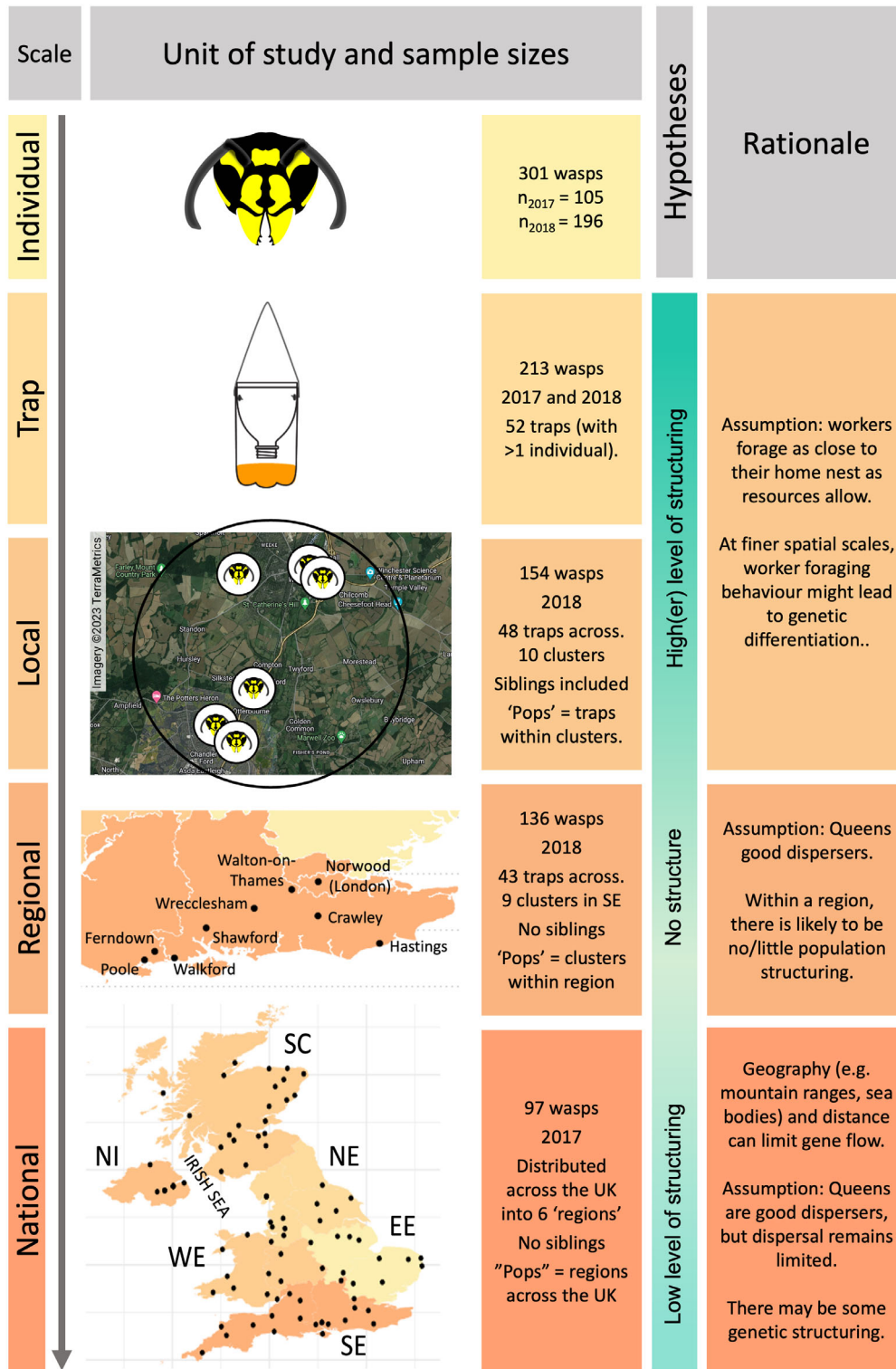


FIGURE 1 Summary of sampling scheme and research questions addressed in Aim 2. Each geographical scale considered in this study is presented here along with information about the samples pertaining to each scale, and specific hypotheses, rationales and predictions. At the national scale, SE = Southern England, EE = Eastern England, WE = Wales and West England, NE = Northern England, SC = Scotland, NI = Northern Ireland. At the local scale, the Hastings cluster is presented as an example (image pulled from Google Maps); each pin corresponds to the location of a trap. Table S4 contains more detailed information relating to sample sizes.

Locus D2-185 showed high genotyping error rates (0.312) and was removed from all further analyses (Table S2). Locus LIST2004 failed to amplify successfully in 62% of Chelex samples. It was

subsequently removed from all 2018 analyses but was retained for the 2017 analyses given that, in combination with DNeasy samples, it amplified successfully in 68% of all 2017 samples.

We estimated sibship between individuals with a primary aim to remove the confounding effects of high relatedness on analyses of genetic structure (Rodríguez-Ramilo & Wang, 2012). In total, 25 pairs of siblings (involving 34 individuals in total) were inferred with a probability of >0.85 (Table S3). One of each sibling pair, including 8 individuals from the 2017 dataset and 10 from the 2018 dataset, was removed from subsequent analyses at the national and regional scales, but were retained for within-trap and local scale analyses.

The total number of alleles (N_A) per locus varied between 6 and 18 in the 2017 dataset and between 4 and 16 in the 2018 dataset (Table 1). In 2017, mean observed heterozygosity (H_O) was 0.586 and mean expected heterozygosity (H_E) was 0.673; in 2018, mean H_O was 0.583 and mean H_E was 0.681. There was no significant difference in N_A between DNA extraction methods ($t = 0.847$, $p = 0.407$, not including LIST2004, VMA4 or D2-185). All loci except LIST2013, LIST2018, LIST2001, VMA6 and Rufa19 in the 2017 dataset and LIST2003,

LIST2001, LIST2011, R1-169 and D3-15 in the 2018 dataset deviated significantly from HWE after Bonferroni correction (Table S2). Locus-specific F_{IS} values ranged between -0.093 and 0.352 (Table 1).

Multiple factors could contribute to these deviations from Hardy-Weinberg Equilibrium (HWE) and positive F_{IS} values. Genotyping errors and null alleles (NAs; Morin et al., 2009; Waples, 2015) could be a possible cause. However, genotyping error rates and NA frequencies were overall low compared with those from non-invasive sampling such as faecal samples (Carlsson, 2008; Dakin & Avise, 2004; Rico et al., 2017), with genotyping error rates ranging between 0 and 0.115 and NA frequency estimates between 0 and 0.155. Another explanation for the deviation from HWE and positive F_{IS} values is inbreeding or the presence of population substructuring, known as a Wahlund effect (Waples, 2015). However, structuring was not detected at the regional or national levels using unsupervised Bayesian structure analysis (see Results: Aim 2).

TABLE 1 Loci characteristics.

Locus	Year	n	N_A	N_E	H_O	H_E	F_{IS}
LIST2007	2017	97	6	1.746	0.278	0.427	0.348
	2018	181	8	1.746	0.365	0.422	0.147
LIST2003	2017	96	18	5.166	0.740	0.806	0.083
	2018	176	17	5.012	0.733	0.800	0.084
LIST2013	2017	88	10	3.919	0.693	0.745	0.069
	2018	161	10	3.580	0.652	0.721	0.095
LIST2018	2017	94	14	7.910	0.766	0.874	0.123
	2018	165	16	7.578	0.697	0.868	0.197
VMA3	2017	90	9	2.852	0.500	0.649	0.230
	2018	155	9	2.940	0.516	0.660	0.218
LIST2004	2017	71	11	3.314	0.521	0.698	0.254
	2018	N/A	N/A	N/A	N/A	N/A	N/A
LIST2001	2017	92	18	9.911	0.826	0.899	0.081
	2018	181	16	8.444	0.785	0.882	0.110
LIST2011	2017	69	6	1.525	0.304	0.344	0.116
	2018	182	4	1.432	0.330	0.302	-0.093
VMA6	2017	83	14	7.448	0.783	0.866	0.095
	2018	143	12	6.201	0.664	0.839	0.208
R1.169	2017	96	18	3.843	0.656	0.740	0.113
	2018	180	16	3.427	0.694	0.708	0.019
LIST2017	2017	96	9	1.346	0.167	0.257	0.352
	2018	178	7	1.327	0.191	0.246	0.224
Rufa19	2017	87	12	4.426	0.793	0.774	-0.025
	2018	132	14	5.138	0.667	0.805	0.172
D3.15	2017	N/A	N/A	N/A	N/A	N/A	N/A
	2018	147	17	4.994	0.721	0.800	0.098
VMA4	2017	N/A	N/A	N/A	N/A	N/A	N/A
	2018	120	15	4.822	0.558	0.793	0.296

Note: In total, the sample sizes are 97 in 2017 and 186 in 2018.

Abbreviations: N_A , number of alleles; N_E , number of effective alleles; H_O , observed heterozygosity; H_E , expected heterozygosity; F_{IS} , inbreeding coefficient.

Aim 2: Population structure of *V. vulgaris* across the UK

We analysed the population genetic structure of *V. vulgaris* at four different geographical scales: within-trap, local (between traps up to 10 km apart), regional (between clusters of traps up to 185 km apart) and national (between traps up to 850 km apart). We expected to find evidence of high population structuring at the fine-scale (Figure 1) because *Vespula* are central-place foragers. Conversely, we expected to find less structuring at the regional and national scales if queens can disperse widely and if dispersal is unrestricted by geography.

Evidence of siblings/half-siblings at finer-spatial scales

Across all samples (105 from 2017 and 196 from 2018), we found evidence for eight pairs of full siblings within traps in the 2017 dataset and 16 pairs of full siblings within traps in the 2018 dataset (Table S3). There was evidence for only one pair of half siblings in traps 3253 and 5597 in the 2018 dataset with a probability of 0.869; however, we chose to retain both of these individuals for further analyses as the traps were located over 135 km apart, and any inferred family relationship was considered unlikely.

Little evidence of population differentiation at the local scale

We analysed the population genetic structure between traps within 10 different clusters with a total sample size of 154 wasps (Table S4 for sample size breakdown). Siblings were retained for this analysis. Each cluster contained between 11 and 21 wasps (mean: 15.4), distributed between three and six traps (mean traps per cluster: 4.7; mean individuals per trap: 3.28). Three clusters indicated significant positive isolation by distance (IBD): the Hastings cluster ($r = 0.268$, $p = 0.015$, $n = 11$), the Crawley cluster ($r = 0.271$, $p = 0.018$, $n = 14$) and the Wrecclesham cluster ($r = 0.268$, $p = 0.030$, $n = 14$) (Table S5). Additionally, high pairwise F_{ST} values were found between traps in four of the 10 clusters: the Norwood cluster ($F_{ST} = 0.159$ between traps 1319 and 2351), the Crawley cluster ($F_{ST} = 0.258$ between traps 2366 and 3760; $F_{ST} = 0.199$ between traps 625 and 3760), the Shawford cluster ($F_{ST} = 0.258$ between traps 1726 and 3253, $F_{ST} = 0.169$ between traps 3253 and 5637; Tables S6–S15). Both results of population differentiation and structuring are weak, given the large, bootstrapped confidence intervals and associated overall low F_{ST} values, respectively. Further sampling would improve our understanding of population structuring.

No evidence of population differentiation at the regional scale

We analysed the population genetic structure of 136 wasps between nine clusters of traps (mean number of individuals per cluster: 15). Siblings were not included in this analysis. We found no evidence for population structure: F_{ST} values were consistently low, ranging between 0 and 0.035 (mean: 0.010; Table 2) and there was no significant IBD (Mantel test: $r = 0.007$, $p = 0.374$; Figure 2a). Our STRUCTURE analysis suggested that the highest ΔK was 119.46 at $K = 2$. However, individual admixture estimates from STRUCTURE analyses showed no visible evidence of any evident structuring for any K from $K = 2$ to $K = 9$ (Figure 2b). Overall,

unsupervised population structure analysis and classical F_{ST} analysis are concordant, suggesting the absence of regional population structuring.

Limits to dispersal at the national scale

We analysed the population genetic structure of 97 wasps between the six putative regions across the UK. This analysis did not include siblings. These analyses indicated low but significantly positive IBD (Mantel test: $r = 0.059$, $p = 0.038$; Figure 2a). Our STRUCTURE analysis yielded the highest ΔK of 23.10 at $K = 2$ (Figure 2b). The analysis at $K = 2$ showed that most of the samples from Northern Ireland clustered together; indeed, F_{ST} values indicated that the Northern Irish samples were slightly, but significantly differentiated from the others, although these values remained low (<0.035 ; Table 3). We repeated the Mantel test without the samples from Northern Ireland to determine the importance of this location in the previous result. IBD without the Northern Irish samples was no longer significantly positive ($r = 0.038$, $p = 0.20$), confirming that these samples were driving the previous result. This suggests that wasps disperse widely across the UK, but that the Irish Sea presents a slight barrier to this. However, this differentiation remains weak and there remains evidence of gene flow between these locations.

DISCUSSION

Vespine wasps have a notoriously rocky relationship with humans (Sumner et al., 2018). They are considered a health risk, highly successful as invasive species and are ubiquitous across the globe. Moreover, most people can recognise them as ‘a wasp’. Despite this, vespine wasps remain poorly studied relative to other insects, like their cousins the bees and ants (Sumner et al., 2018). This means that key biological traits, like population structure, are little studied. This is especially true for most of the vespine wasps (which include the wasps best recognised by people) in their native ranges. We address this data gap by analysing samples collected by citizen scientists to reveal the population genetic structure of the common yellowjacket wasp *V. vulgaris* across the UK at different spatial scales. Our analyses reveal a panmictic population across 850 km, with little evidence for any structuring at finer scales. Our findings provide evidence for two important aspects of the ecology and behaviour of *V. vulgaris*. First, queens disperse considerable distances from their natal nest to found new nests, resulting in high rates of gene flow and thus little differentiation across the landscape. Second, foragers forage close to their nests: worker siblings could be detected within traps but not between closely situated traps. Our study also showcases the potential for using citizen scientist-collected samples for molecular analyses, even when samples are poorly preserved. We anticipate our study will incentivise further engagement of citizen scientists in the urgent quest to enrich our understanding of insect populations, better equipping us in assuring future security of the natural capital these insects provide (Brock et al., 2021).

TABLE 2 Genetic variation between 'populations' at the regional level.

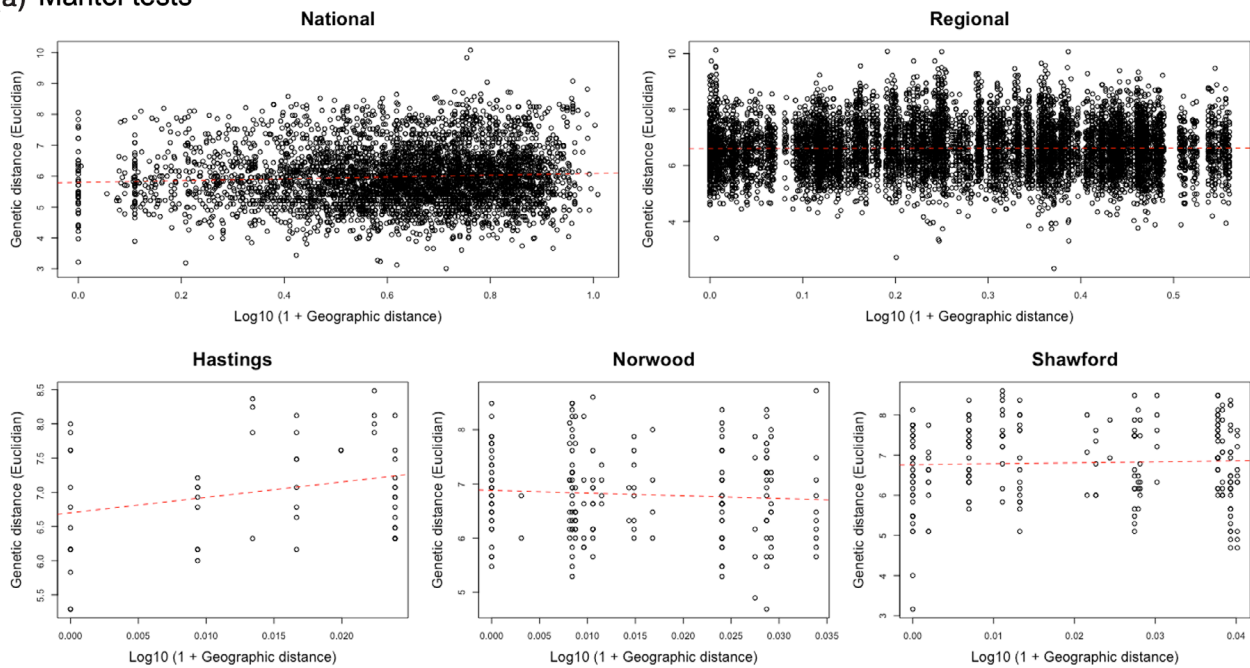
Cluster	Walkford	Hastings	Poole	Norwood	Crawley	Ferndown	Shawford	Wrecclesham	Walton-on-Thames
Walkford	-								
Hastings	0.029 (0.004–0.049)	-							
Poole	0.002 (0.000–0.021)	0.034 (0.011–0.056)	-						
Norwood	0.004 (0.000–0.015)	0.030 (0.000–0.062)	0.001 (0.000–0.013)	-					
Crawley	0.012 (0.000–0.035)	0.033 (0.006–0.065)	0.000 (0.000–0.000)	0.003 (0.000–0.025)	-				
Ferndown	0.003 (0.000–0.016)	0.016 (0.000–0.054)	0.000 (0.000–0.020)	0.007 (0.000–0.037)	0.017 (0.000–0.046)	-			
Shawford	0.016 (0.001–0.034)	0.035 (0.015–0.066)	0.004 (0.000–0.016)	0.006 (0.000–0.018)	0.003 (0.000–0.012)	0.004 (0.000–0.017)	-		
Wrecclesham	0.000 (0.000–0.016)	0.018 (0.000–0.055)	0.000 (0.000–0.008)	0.011 (0.001–0.026)	0.009 (0.000–0.026)	0.001 (0.000–0.021)	0.000 (0.000–0.026)	-	
Walton-on-Thames	0.004 (0.000–0.024)	0.031 (0.000–0.066)	0.001 (0.000–0.015)	0.001 (0.000–0.026)	0.011 (0.000–0.026)	0.010 (0.000–0.029)	0.000 (0.000–0.000)	0.012 (0.000–0.027)	-

Population structure

First, we found that *V. vulgaris* shows little to no population genetic structure at the two broader spatial scales studied here and forms an effectively single panmictic population across the UK. These results provide insights into the dispersal biology of this species. Queens are known to disperse far from their natal nests (based on rates of spread of invasive populations; Arca et al., 2015; Masciocchi & Corley, 2013) to set up colonies the following spring. These high rates of dispersal appear to be driven by human mediation; wasps have repeatedly shown a propensity to expand their geographic range through human-assisted means (Chau et al., 2015; Crosland, 1991; Masciocchi & Corley, 2013; Veldtman et al., 2021). There is no reason to believe that human transport does not also occur across their native ranges, and would contribute to the lack of population structuring found in this study. The extent to which dispersal is sex-biased or not remains to be determined (although see Martínez et al., 2021; Masciocchi et al., 2020), but since we found little genetic structure at the regional scale, and little evidence of inbreeding, it is probable that both sexes are dispersing widely from their natal nests. However, our analyses also revealed a clear barrier to dispersal: there was differentiation between the mainland UK and Northern Ireland samples, demonstrating that water bodies (in this case, the Irish Sea) can act as barriers to dispersal. This is in accordance with previous studies that have analysed gene flow of *Vespula* populations across islands (e.g., invasive *V. pennsylvanica* across Hawaiian islands [Chau et al., 2015] or native *V. germanica* across the English channel [Schmack et al., 2019]). The apparent ease with which these wasps can disperse across the landscape is likely a major contributing factor of the success of invasive wasps around the world (Moller, 1996) and highlights the challenge in containing future introductions. Future studies should concentrate on larger-scale comparisons of genetic diversity and structure between native and invasive *V. vulgaris* populations (as has been done for some other *Vespula* and *Vespa* species as detailed previously) to deepen our understanding of their invasion process.

Second, we detected some evidence of low-level population structuring at finer spatial scales (siblings within traps). This low-level population structuring demonstrates family structuring, likely driven by the behaviour of foragers. If foragers are flying close to their nests (as expected for a central-place forager) then we expected to find siblings within the same trap, and perhaps also in nearby traps. Whilst we did not find any evidence for sibling pairs between traps, we demonstrated the reliability of the markers used in this study to infer sibship. Indeed, all except one inferred sib-pair was within-trap. This may be useful for future studies, as sibship inference can be used to understand other aspects of vespine wasp ecology, including nest densities and foraging distances (see Crowther et al., 2019; Dreier et al., 2014; Knight et al., 2005 for examples in bees). By genotyping multiple individuals per trap, and by estimating foraging distances based on previous work, it would be possible to provide an estimate of nest density surrounding a trap (Carvell et al., 2017). This is useful as wasp and other hymenopteran nests can be notoriously difficult

(a) Mantel tests



(b) Population clustering

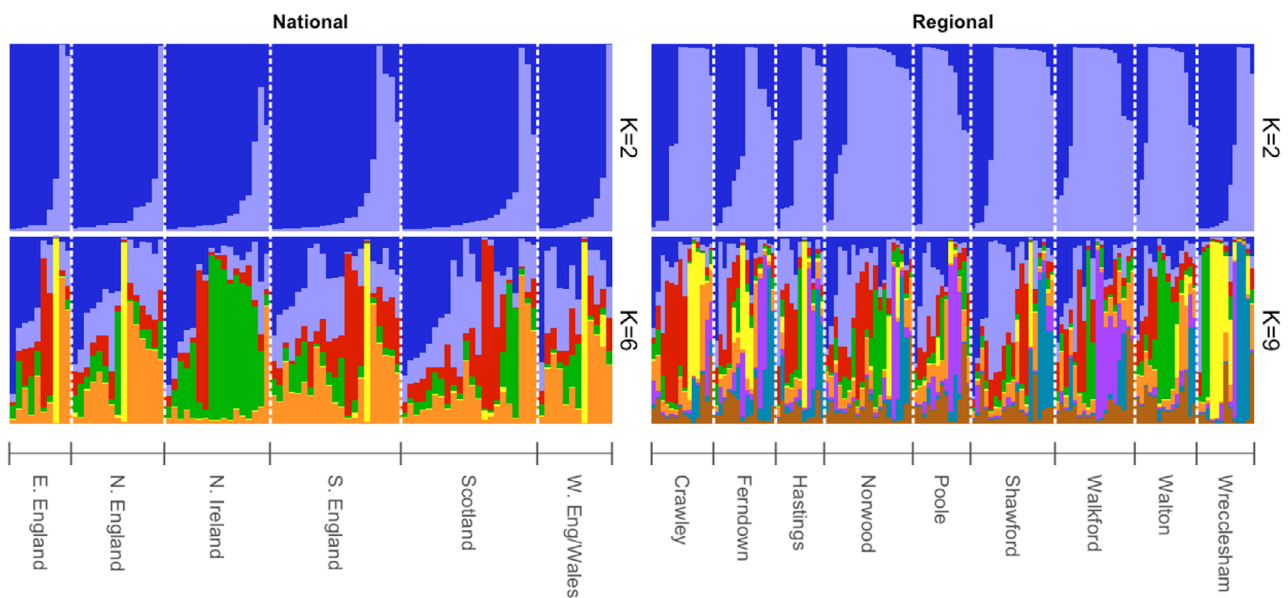


FIGURE 2 (a) Results of Mantel tests at all three geographical scales (*national, regional and selected local scales*); Hastings cluster: Mantel test: $r = 0.268$, $p = 0.015$, $n = 11$; Norwood (London) cluster: $r = -0.067$, $p = 0.703$, $n = 20$; Shawford cluster: $r = 0.037$, $p = 0.242$, $n = 21$.

(b) Population structure analyses (*using STRUCTURE at regional and national scales*). At the national scale: E. England = Eastern England, N. England = Northern England, N. Ireland = Northern Ireland, S. England = Southern England, W. Eng/Wales = Western England and Wales. Presented here at the national scale at $K = 2$ (best ΔK) and for $K = 6$ (number of putative 'populations', here: regions); at the regional scale at $K = 2$ (best ΔK) and $K = 9$ (number of putative 'populations', here: clusters of traps).

and time-consuming to find (Kells & Goulson, 2003; Osborne et al., 2008). Similarly, little is known about vespine (and other flying Hymenoptera) foraging distances because workers are difficult to track. Any evidence for siblings between traps could provide us with insights into worker foraging distances. This study therefore demonstrates the potential scope of citizen science projects like the BWS to

answer ecological and behavioural questions that extend well beyond the initial population monitoring aims.

It is not surprising that genotyping errors and NAs were detected across loci; this is partly due to the low quality of DNA extracted from citizen-collected samples, but is also symptomatic of using microsatellite markers developed for non-target species. NAs and allelic

TABLE 3 Genetic variation at the national level.

Region	NE	SE	SC	EE	WE	NI
NE	–					
SE	0.000 (0.000–0.000)	–				
SC	0.000 (0.000–0.007)	0.004 (0.000–0.013)	–			
EE	0.000 (0.000–0.008)	0.000 (0.000–0.009)	0.002 (0.000–0.010)	–		
WE	0.000 (0.000–0.000)	0.000 (0.000–0.000)	0.000 (0.000–0.000)	0.000 (0.000–0.000)	–	
NI	0.035 (0.015–0.063)	0.034 (0.021–0.049)	0.033 (0.017–0.050)	0.030 (0.011–0.056)	0.035 (0.014–0.062)	–

Note: Values from pairwise F_{ST} tests. Column and row names refer to regions: SE, Southern England; EE, Eastern England; NE, Northern England; WE + Western England + Wales; SC, Scotland; NI, Northern Ireland. Figures in brackets are 95% confidence intervals of F_{ST} obtained by bootstrapping.

dropouts are expected to lead to excess homozygosity and a deficiency in heterozygosity, and thus deviations in genotype frequency from Hardy–Weinberg equilibrium and a positive F_{IS} , as observed in this study. However, these effects are accounted for in our analyses, which take account of genotyping errors. Sibship inference can be vulnerable to mistyping and NAs (Wang, 2004), but we adopted the likelihood method developed to handle noisy data with genotyping error rates much higher than observed in our data (Wang, 2019). Equally, we are confident in our sibship analysis results as all but one sibling pairs were detected within traps. Indeed, F_{ST} and STRUCTURE analyses are both robust to mistyping and NAs, as they work with allele frequencies that are not much affected by genotyping abnormality. This said, the lack of any population structuring of *V. vulgaris* at regional and national levels revealed by our microsatellite data does not preclude the possibility that a weak genetic structure might exist, which might be detected by higher-resolution analyses, for example using many hundred single nucleotide polymorphisms (SNPs), spread across the genome. With the use of dense genomic SNPs, populations of extremely weak differentiations such as $F_{ST} = 0.0007$ can be reliably distinguished by unsupervised structure analysis (Leslie et al., 2015).

Use of citizen science

Samples collected through the BWS allowed us to overcome common sampling limitations relating to finance, time, effort and geographical coverage (Conrad & Hilchey, 2011; Dickinson et al., 2010). In light of the recent Covid-19 pandemic, this is particularly encouraging for future research, demonstrating that samples may continue to be collected despite potential pandemic restrictions, and that this can be used for molecular studies. Additionally, our study tested a key challenge of citizen science data quality. BWS collected wasps had sat in beer or orange juice traps for up to a week, undergone at least two freeze–thaw cycles, and spent several days at room temperature during transit. Yet, we were successful in extracting the DNA from most samples used in this study. We benchmarked two DNA extraction methods: a commercial kit (DNeasy) and a cheaper, less complicated and less time-consuming method (Chelex).

Using prices of the reagents used at the time of study (2018–2019), we calculated that DNA extractions using the DNeasy kit (for 50 samples) cost £3.94/sample (accounting for failed extractions for ~10% of samples), whilst the Chelex method cost £0.07/sample (including Chelex and Proteinase K). The Chelex method was therefore significantly cheaper than DNeasy method. However, the DNeasy method yielded higher quantity and quality DNA, and none of the 81 PCRs required repeating; by contrast, 414 PCRs were necessary to genotype the 196 Chelex individuals used in this study. PCR reagents (based on the price of the Type-it Microsatellite PCR kit [Qiagen] at the time of this study) cost £38.64 per plate and sequencing cost £115 per plate. Accounting for the number of failed PCRs, we calculated that the total cost for DNA extraction, PCR and sequencing cost £12.02 per DNeasy sample and £17.39 per Chelex sample. In terms of labour costs, DNA extraction was less time consuming with the Chelex method than the DNeasy one; however, because of the number of PCR repeats associated with Chelex samples, labour costs with Chelex exceeded those of DNeasy. For future studies wishing to study insect populations at the molecular level with samples collected by citizen scientists using unconventional methods, we suggest that the commercial kit is a more reliable option. However, in line of recent advances which now allow for DNA to be extracted from ancient and museum material (e.g., Rohland & Hofreiter, 2007; Straube et al., 2021), this does not mean that other, lower-cost DNA extraction protocols (including Chelex) may not be further optimised and costs to be further reduced.

We hope that the results of this study will encourage other citizen science-based projects to use their samples for molecular studies, for instance to monitor pollinator genetic diversity and population structures. Pollinator population health is of particular concern, and pollinators can be easily collected, as demonstrated by the UK Pollinator Monitoring Scheme (<https://ukpoms.org.uk/>), a national scheme in the UK that collects pollinators using pan-traps.

CONCLUSIONS

In this study, we showed the feasibility of using citizen science samples to provide a complete assessment of the population genetic structure of *V. vulgaris* across the UK. This is even more noteworthy

since the BWS was not originally designed with a view to performing molecular analyses but was rather intended as a project to gain knowledge relating to vespine wasp distributions and for public engagement. We hope that this study will inspire other insect-related citizen science projects, especially in countries with more restricted budgets, to address ecological and evolutionary questions using molecular tools, and to contribute to insect conservation through a better understanding of their ecological requirements.

EXPERIMENTAL PROCEDURES

Sampling

BWS volunteers set up home-made beer or orange juice-baited bottle traps for 7 days at the end of the summer, when *Vespula* colonies are close to the end of their life cycle, to target late-season workers (Sumner et al., 2019). Insects were removed from the trap after 7 days, washed in tap water and stored dry in household freezers (approx. -20°C) until the samplers were ready to post the samples to experts at the University College London (UCL) for identification. Samples were sent via first-class post as dry specimens, wrapped in tinfoil inside a padded envelope for protection. On arrival at the UCL, samples were relocated to -20°C laboratory grade freezers as soon as possible. For identification, wasps were thawed and examined under a microscope, before being stored in 80% ethanol at room temperature until genetic analyses.

In total, across the UK in 2017, 2942 *V. vulgaris* individuals were collected from 407 traps; and in 2018, 14,804 wasps from 1275 traps. Volunteers provided their post codes with the traps; we transformed these into coordinates using the website UK Grid Reference Finder (www.gridreferencefinder.com). We used these coordinates to calculate the distances between traps to select traps for the population genetic analyses at different geographical scales, as defined in the Results.

DNA extraction

Two conflicting issues were at play in our choice of DNA extraction method. On the one hand, samples collected by the BWS were likely to have been quite degraded, having been stored in warm beer/orange juice for a week, endured 2–4 days at room temperature during posting, and under-gone at least two freeze–thaw cycles, before finally being stored in alcohol. On the other hand, we required a low-cost extraction method that would permit analysis of large sample sizes on the limited budget that is common across citizen science projects (Hecker et al., 2018). Consequently, we tested two different DNA extraction methods: a commercial kit (DNeasy[®] Blood and Tissue kit, ID: 69504, Qiagen) and a less time-consuming and costly (Lienhard & Schäffer, 2019) modified Chelex100 (BioRad) protocol (Gadau, 2009; Moreau, 2014). For both protocols, we used the abdomen, thorax and legs (but not the head and wings) for DNA extraction.

We used the DNeasy kit for a subset of 2017 samples ($n = 91$), and the Chelex method for the others ($n = 46$ samples from 2017, $n = 256$ samples from 2018). The DNeasy method followed the manufacturer's instructions, except for the amount of Proteinase K (20 μL here) and that samples were incubated overnight. The Chelex method consisted of adding 450 μL of 10% Chelex 100 solution to each sample before manually crushing the sample with sterile pestles, then adding 5 μL Proteinase K and incubating the samples at 57°C for 3 h. After incubation, each sample was vortexed and boiled at 95°C for 8–10 min. To separate the supernatant from the wasp particles and the Chelex 100 beads, the samples were centrifuged at $16,873 \times g$ (14,000 rpm) for 15 min, then $\sim 200 \mu\text{L}$ of supernatant was transferred to fresh tubes for PCR. Samples were stored overnight at 4°C or at -20°C for longer periods of time when necessary. We examined the success of DNA extraction using the DNeasy method with a Nanodrop where possible. Samples from both years were stored for similar periods of time and treated in the same way.

Genotyping

We assayed 34 microsatellite markers that had been previously developed for other vespine species (Arca et al., 2012; Daly et al., 2002; Hasegawa & Takahashi, 2002; Thorén et al., 1995) on seven BWS *V. vulgaris* individuals using Qiagen's Multiplex PCR kit, and determined amplification success using gel electrophoresis. Each reaction contained 6.25 μL master mix, 5 μL DNA and 0.21 μL of each forward and reverse primer (with three primer pairs per reaction). The annealing temperature (T_a) was 52°C for all loci. Forward markers were fluorescently tagged with the dyes 6FAM or HEX. Sequencing was performed with an ABI3730xl DNA Analyzer and peaks were scored by two people independently using the Geneious Prime Microsatellite Plugin (v1.4.6). Twenty-eight markers amplified, and of these, 19 were polymorphic. We selected 12 primers for the DNeasy samples (LIST2007, LIST2003, LIST2013, LIST2018, VMA3, LIST2004, LIST2001, LIST2011, VMA6, R1-169, LIST2017 and Rufa19), with the addition of D3-15, VMA4 and D2-182 for the Chelex samples (Table S1 for information on primers and assay results). To compare understand the variables affecting amplification success, we used a GLM that accounted for DNA extraction method (DNeasy or Chelex), liquid in the trap (orange juice or beer), the number of days that the trap was set out for and the trap ID (as a random effects variable). We did not include year in this analysis given that year and method were strongly correlated.

Genetic data analysis

Genotyping error rates were estimated using the method described by Pompanon et al. (2005). We used Micro-Checker v2.2.3 (Van Oosterhout et al., 2004) to screen for allelic dropout. Null allele frequencies, linkage disequilibrium, deviations from Hardy–Weinberg (HWE) and the inbreeding coefficient F_{IS} per locus were calculated

using the R package *Genepop* v1.1.7 (Rousset, 2008) using Weir and Cockerham's method (Weir & Cockerham, 1984). Tests for HWE were corrected for multiple comparisons using the Bonferroni method (Table S2). We calculated number of alleles (N_A), observed heterozygosity (H_O) and expected heterozygosity (H_E) using GenAEx v6.5 (Peakall & Smouse, 2006, 2012).

We estimated sibship within traps and within clusters with two aims: first, to remove the confounding effects of high relatedness on analyses of genetic structure (Rodríguez-Ramilo & Wang, 2012), and second to infer *Vespula* foraging distances. We considered that two individuals were siblings if the probability of inference of either full or half sibship was 0.85 or higher, calculated using COLONY 2.0.6.5 (Jones & Wang, 2010) with a full-likelihood method, medium run lengths and no sibship prior. We performed analyses of population genetic structure (see below) without siblings at the national and regional scales but retained them for the local scale analysis given that our aim was to understand the movements of worker wasps.

Pairwise F_{ST} was calculated according to Nei (1987) between 'populations' at each scale using the R package *adegenet* v2.1.7 (Jombart, 2008) and *hierfstat* v0.5-11 (Goudet, 2005). We considered pairwise F_{ST} to be high if it was >0.15 (Frankham et al., 2002). IBD was assessed using Mantel tests with the package *ade4* v1.7-19 (Bougeard & Dray, 2018; Dray et al., 2007, 2021; Dray & Dufour, 2007). We further analysed genetic structure at the regional and national scales with STRUCTURE v2.3.4 (Pritchard et al., 2000). We used an admixture model with a 10,000 burn-in period and 20,000 Markov Chain Monte Carlo iterations and set with the possible number of populations (K) between one and six for the national scale analysis (number of regions) and between one and nine for the regional analysis (number of clusters). Each analysis was repeated three times for each value of K . We used the R package *pophelper* (Francis, 2017) to implement Evanno's method (Evanno et al., 2005), which determines the most likely number of clusters (ΔK) and to generate graphical outputs. All R analyses were performed with R version 4.1.2 (RC Team, 2022).

AUTHOR CONTRIBUTIONS

Iona Cunningham-Eurich: Conceptualization; methodology; data curation; investigation; validation; formal analysis; writing – original draft; visualization. **Danai Kontou:** Conceptualization; methodology; data curation; investigation; writing – review and editing; formal analysis. **Monika Yordanova:** Methodology; data curation; investigation; formal analysis; writing – review and editing. **Alejandro Maeda-Obregon:** Conceptualization; methodology; data curation; investigation; formal analysis; writing – review and editing. **Emeline Favreau:** Methodology; formal analysis; writing – review and editing. **Jinliang Wang:** Software; supervision; methodology; writing – original draft; formal analysis; writing – review and editing. **Adam Hart G:** Funding acquisition; resources; writing – review and editing; project administration; visualization; conceptualization; investigation. **Seirian Sumner:** Conceptualization; supervision; project administration; funding acquisition; writing – original draft; visualization; resources; investigation; writing – review and editing.

ACKNOWLEDGEMENTS

The Big Wasp Survey is funded through sponsorship and a Goodman Award from the Royal Entomological Society (to Seirian Sumner and Adam G Hart). We thank the several thousand members of the public who have generously contributed their time and effort to sampling the wasps for the Survey, as well as London Natural History Society members and UCL students for sorting. We also thank Adrienne Etard and Gonzalo Albaladejo-Robles for coding assistance and Owen Corbett for figure preparation.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data are made publicly available via online databases; the dataset that support the findings of this study are openly available in figshare at <https://doi.org/10.6084/m9.figshare.21991847.v2>. The R code used in this study is available at <https://zenodo.org/record/7995875>. In addition, the data and the main findings are provided to the wider public at the Big Wasp Survey website www.bigwaspsurvey.org. All members of the public who took part in the Survey are kept informed with the scientific progress made by the project, thanks to their contributions, via annual reports sent out to participants by email.

ORCID

Iona Cunningham-Eurich  <https://orcid.org/0000-0002-7187-1407>

Danai Kontou  <https://orcid.org/0000-0002-0773-9540>

Emeline Favreau  <https://orcid.org/0000-0002-0713-500X>

REFERENCES

- Agersnap, S., Sigsgaard, E.E., Jensen, M.R., Ávila, M.R., Carl, H., Moller, P.R. et al. (2022) A National Scale "BioBlitz" using citizen science and eDNA Metabarcoding for monitoring coastal marine fish. *Frontiers in Marine Science*, 9, 824100. Available from: <https://doi.org/10.3389/fmars.2022.824100>
- Akre, R.D., Hill, W.B., Mac Donald, J.F. & Garnett, W.B. (1975) Foraging distances of *Vespula pensylvanica* workers (Hymenoptera: Vespidae). *Journal of the Kansas Entomological Society*, 48(1), 12–16.
- Arca, M., Capdevielle-Dulac, C., Villemant, C., Mougél, F., Arnold, G. & Silvain, J.F. (2012) Development of microsatellite markers for the yellow-legged Asian hornet, *Vespa velutina*, a major threat for European bees. *Conservation Genetics Resources*, 4(2), 283–286. Available from: <https://doi.org/10.1007/s12686-011-9525-1>
- Arca, M., Mougél, F., Guillemaud, T., Dupas, S., Rome, Q., Perrard, A. et al. (2015) Reconstructing the invasion and the demographic history of the yellow-legged hornet, *Vespa velutina*, in Europe. *Biological Invasions*, 17(8), 2357–2371. Available from: <https://doi.org/10.1007/s10530-015-0880-9>
- Ascunce, M.S., Yang, C.-C., Oakey, J., Calcaterra, L., Wu, W.-J., Shih, C.-J. et al. (2011) Global invasion history of the fire ant *Solenopsis invicta*. *Science*, 331(6020), 1066–1068. Available from: <https://doi.org/10.1126/science.1198734>
- Bluher, S.E., Miller, S.E. & Sheehan, M.J. (2020) Fine-scale population structure but limited genetic differentiation in a cooperatively breeding paper wasp. *Genome Biology and Evolution*, 12(5), 701–714. Available from: <https://doi.org/10.1093/gbe/evaa070>
- Bohonak, A.J. (1999) Dispersal, gene flow, and population structure. *The Quarterly Review of Biology*, 74(1), 21–45.

- Bougeard, S. & Dray, S. (2018) Supervised multiblock analysis in R with the *ade4* package. *Journal of Statistical Software*, 86, 1–17. Available from: <https://doi.org/10.18637/jss.v086.i01>
- Bowler, D.E., Eichenberg, D., Conze, K.-J., Suhling, F., Baumann, K., Benken, T. et al. (2021) Winners and losers over 35 years of dragonfly and damselfly distributional change in Germany. *Diversity and Distributions*, 27(8), 1353–1366. Available from: <https://doi.org/10.1111/ddi.13274>
- Breeze, T.D., Bailey, A., Balcombe, K., Brereton, T., Comont, R., Edwards, M. et al. (2021) Pollinator monitoring more than pays for itself. *Journal of Applied Ecology*, 58(1), 44–57. Available from: <https://doi.org/10.1111/1365-2664.13755>
- Brenton-Rule, E.C., Döbelmann, J., Baty, J.W., Brown, R.L., Dvorak, L., Grangier, J. et al. (2018) The origins of global invasions of the German wasp (*Vespula germanica*) and its infection with four honey bee viruses. *Biological Invasions*, 20(12), 3445–3460. Available from: <https://doi.org/10.1007/s10530-018-1786-0>
- Brock, R.E., Cini, A. & Sumner, S. (2021) Ecosystem services provided by aculeate wasps. *Biological Reviews*, 96(4), 1645–1675. Available from: <https://doi.org/10.1111/brv.12719>
- Brown, P.M.J., Roy, D.B., Harrower, C., Dean, H.J., Rorke, S.L. & Roy, H.E. (2018) Spread of a model invasive alien species, the harlequin ladybird *Harmonia axyridis* in Britain and Ireland. *Scientific Data*, 5(1), 180239. Available from: <https://doi.org/10.1038/sdata.2018.239>
- Brunke, J., Radespiel, U., Russo, I.R., Bruford, M.W. & Goossens, B. (2019) Messing about on the river: the role of geographic barriers in shaping the genetic structure of Bornean small mammals in a fragmented landscape. *Conservation Genetics*, 20(4), 691–704. Available from: <https://doi.org/10.1007/s10592-019-01159-3>
- Cardoso, P., Erwin, T.L., Borges, P.A.V. & New, T.R. (2011) The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, 144(11), 2647–2655. Available from: <https://doi.org/10.1016/j.biocon.2011.07.024>
- Carlsson, J. (2008) Effects of microsatellite null alleles on assignment testing. *Journal of Heredity*, 99(6), 616–623. Available from: <https://doi.org/10.1093/jhered/esn048>
- Carvell, C., Bourke, A.F.G., Dreier, S., Freeman, S.N., Hulmes, S., Jordan, W.C. et al. (2017) Bumblebee family lineage survival is enhanced in high-quality landscapes. *Nature*, 543(7646), 547–549. Available from: <https://doi.org/10.1038/nature21709>
- Chau, L.M., Hanna, C., Jenkins, L.T., Kutner, R.E., Burns, E.A., Kremen, C. et al. (2015) Population genetic structure of the predatory, social wasp *Vespula pensylvanica* in its native and invasive range. *Ecology and Evolution*, 5(23), 5573–5587. Available from: <https://doi.org/10.1002/ece3.1757>
- Comont, R.F., Roy, H.E., Lewis, O.T., Harrington, R., Shortall, C.R. & Purse, B.V. (2012) Using biological traits to explain ladybird distribution patterns. *Journal of Biogeography*, 39(10), 1772–1781. Available from: <https://doi.org/10.1111/j.1365-2699.2012.02734.x>
- Conrad, C.C. & Hilchey, K.G. (2011) A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment*, 176(1–4), 273–291. Available from: <https://doi.org/10.1007/s10661-010-1582-5>
- Creedy, T.J., Norman, H., Tang, C.Q., Chin, K.Q., Andujar, C., Arribas, P. et al. (2020) A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding. *Molecular Ecology Resources*, 20(1), 40–53. Available from: <https://doi.org/10.1111/1755-0998.13056>
- Crosland, M.W.J. (1991) The spread of the social wasp, *Vespula germanica*, in Australia. *New Zealand Journal of Zoology*, 18(4), 375–387. Available from: <https://doi.org/10.1080/03014223.1991.10422843>
- Crossley, M.S., Meier, A.R., Baldwin, E.M., Berry, L.L., Crenshaw, L.C., Hartman, G.L. et al. (2020) No net insect abundance and diversity declines across US long term ecological research sites. *Nature Ecology & Evolution*, 4(10), 1368–1376. Available from: <https://doi.org/10.1038/s41559-020-1269-4>
- Crowther, L.P., Wright, D.J., Richardson, D.S., Carvell, C. & Bourke, A.F.G. (2019) Spatial ecology of a range-expanding bumble bee pollinator. *Ecology and Evolution*, 9(3), 986–997. Available from: <https://doi.org/10.1002/ece3.4722>
- Dakin, E.E. & Avise, J.C. (2004) Microsatellite null alleles in parentage analysis. *Heredity*, 93(5), 504–509. Available from: <https://doi.org/10.1038/sj.hdy.6800545>
- Daly, D., Archer, M.E., Watts, P.C., Speed, M.P., Hughes, M.R., Barker, F.S. et al. (2002) Polymorphic microsatellite loci for eusocial wasps (hymenoptera: Vespidae). *Molecular Ecology Notes*, 2(3), 273–275. Available from: <https://doi.org/10.1046/j.1471-8286.2002.00220.x>
- de Virgilio, M., Cifarelli, S., de Gennaro, P., Garofoli, G. & Degryse, B. (2020) A first attempt of citizen science in the genetic monitoring of a *Posidonia oceanica* meadow in the Italian southern Adriatic Sea. *Journal for Nature Conservation*, 56, 125826. Available from: <https://doi.org/10.1016/j.jnc.2020.125826>
- Dickinson, J.L., Zuckerman, B. & Bonter, D.N. (2010) Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 149–172. Available from: <https://doi.org/10.1146/annurev-ecolsys-102209-144636>
- Dray, S. & Dufour, A.-B. (2007) The *ade4* package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1–20. Available from: <https://doi.org/10.18637/jss.v022.i04>
- Dray, S., Dufour, A.B. & Chessel, D. (2007) The *ade4* package—II: two-table and K-table methods. *R News*, 7(2), 47–52.
- Dray, S., Dufour, A.-B. & Thioulouse, J. (2021) *ade4*: analysis of ecological data: exploratory and euclidean methods in environmental sciences. Available from: <https://CRAN.R-project.org/package=ade4>
- Dreier, S., Redhead, J.W., Warren, I.A., Bourke, A.F., Heard, M.S., Jordan, W.C. et al. (2014) Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape. *Molecular Ecology*, 23(14), 3384–3395. Available from: <https://doi.org/10.1111/mec.12823>
- Eloff, J., Veldtman, R., Bulgarella, M. & Lester, P.J. (2020) Population genetics of the invasive wasp *Vespula germanica* in South Africa. *Insectes Sociaux*, 67(2), 229–238. Available from: <https://doi.org/10.1007/s00040-020-00752-x>
- Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611–2620. Available from: <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Francis, R.M. (2017) Pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17(1), 27–32. Available from: <https://doi.org/10.1111/1755-0998.12509>
- Frankham, R., Ballou, J.D., Briscoe, D.A. & McInnes, K.H. (2002) *Introduction to conservation genetics*. Cambridge University Press: Cambridge University Press.
- Furlan, E., Stoklosa, J., Griffiths, J., Gust, N., Ellis, R., Huggins, R.M. et al. (2012) Small population size and extremely low levels of genetic diversity in Island populations of the platypus, *Ornithorhynchus anatinus*. *Ecology and Evolution*, 2(4), 844–857. Available from: <https://doi.org/10.1002/ece3.195>
- Gadua, J. (2009) DNA isolation from ants. *Cold Spring Harbor Protocols*, 2009(7), pdb.prot5245. Available from: <https://doi.org/10.1101/pdb.prot5245>
- Goodisman, M.A.D., Matthews, R.W. & Crozier, R.H. (2001) Hierarchical genetic structure of the introduced wasp *Vespula germanica* in Australia. *Molecular Ecology*, 10(6), 1423–1432. Available from: <https://doi.org/10.1046/j.1365-294X.2001.01291.x>
- Goudet, J. (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184–186. Available from: <https://doi.org/10.1111/j.1471-8286.2004.00828.x>

- Greenleaf, S.S., Williams, N.M., Winfree, R. & Kremen, C. (2007) Bee foraging ranges and their relationship to body size. *Oecologia*, 153(3), 589–596. Available from: <https://doi.org/10.1007/s00442-007-0752-9>
- Guindon, K., Neidig, C., Tringali, M., Gray, S., King, T., Gardinal, C. et al. (2015) An overview of the tarpon genetic recapture study in Florida—a citizen science success story. *Environmental Biology of Fishes*, 98(11), 2239–2250. Available from: <https://doi.org/10.1007/s10641-015-0440-2>
- Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H. et al. (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS One*, 12(10), e0185809. Available from: <https://doi.org/10.1371/journal.pone.0185809>
- Hasegawa, E. & Takahashi, J. (2002) Microsatellite loci for genetic research in the hornet *Vespa mandarinia* and related species. *Molecular Ecology Notes*, 2(3), 306–308. Available from: <https://doi.org/10.1046/j.1471-8286.2002.00224.x>
- Hecker, S., Mukli, H., Bowser, A., Makuch, Z., Vogel, J. & Bonn, A. (2018) The European citizen science landscape - a snapshot. *Citizen Science - Innovation in Open Science, Society and Policy*. London: UCL Press, pp. 190–200. Available at: <https://www.jstor.org/stable/j.ctv550cf2.20>
- Hoffman, E.A., Kovacs, J.L. & Goodisman, M.A. (2008) Genetic structure and breeding system in a social wasp and its social parasite. *BMC Evolutionary Biology*, 8(1), 239. Available from: <https://doi.org/10.1186/1471-2148-8-239>
- Jombart, T. (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. Available from: <https://doi.org/10.1093/bioinformatics/btn129>
- Jones, O.R. & Wang, J. (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10(3), 551–555. Available from: <https://doi.org/10.1111/j.1755-0998.2009.02787.x>
- Kells, A.R. & Goulson, D. (2003) Preferred nesting sites of bumblebee queens (Hymenoptera: Apidae) in agroecosystems in the UK. *Biological Conservation*, 109(2), 165–174. Available from: [https://doi.org/10.1016/S0006-3207\(02\)00131-3](https://doi.org/10.1016/S0006-3207(02)00131-3)
- Knight, M.E., Martin, A.P., Bishop, S., Osborne, J.L., Hale, R.J., Sanderson, R.A. et al. (2005) An interspecific comparison of foraging range and nest density of four bumblebee (*Bombus*) species. *Molecular Ecology*, 14(6), 1811–1820. Available from: <https://doi.org/10.1111/j.1365-294X.2005.02540.x>
- Le Féon, V., Henry, M., Guilbaud, L., Coiffait-Gombault, C., Dufrière, E., Kolodziejczyk, E. et al. (2016) An expert-assisted citizen science program involving agricultural high schools provides national patterns on bee species assemblages. *Journal of Insect Conservation*, 20(5), 905–918. Available from: <https://doi.org/10.1007/s10841-016-9927-1>
- Leather, S.R. (2018) “Ecological Armageddon”—more evidence for the drastic decline in insect numbers. *Annals of Applied Biology*, 172(1), 1–3. Available from: <https://doi.org/10.1111/aab.12410>
- Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T. et al. (2015) The fine-scale genetic structure of the British population. *Nature*, 519(7543), 309–314. Available from: <https://doi.org/10.1038/nature14230>
- Lester, P.J. & Beggs, J.R. (2019) Invasion success and management strategies for social *Vespula* wasps. *Annual Review of Entomology*, 64(1), 51–71. Available from: <https://doi.org/10.1146/annurev-ento-011118-111812>
- Lienhard, A. & Schäffer, S. (2019) Extracting the invisible: obtaining high quality DNA is a challenging task in small arthropods. *PeerJ*, 7, e6753. Available from: <https://doi.org/10.7717/peerj.6753>
- Lucky, A., Savage, A.M., Nichols, L.M., Castracani, C., Shell, L., Grasso, D.A. et al. (2014) Ecologists, educators, and writers collaborate with the public to assess backyard diversity in the School of Ants Project. *Ecosphere*, 5(7), art78. Available from: <https://doi.org/10.1890/ES13-00364.1>
- Macgregor, C.J., Williams, J.H., Bell, J.R. & Thomas, C.D. (2019) Moth biomass has fluctuated over 50 years in Britain but lacks a clear trend. *Nature Ecology & Evolution*, 3(12), 1645–1649. Available from: <https://doi.org/10.1038/s41559-019-1028-6>
- Martínez, A.S., Rousselot, N., Corley, J.C. & Masciocchi, M. (2021) Nest-departure behaviour of gynes and drones in the invasive yellowjacket *Vespula germanica* (Hymenoptera: Vespidae). *Bulletin of Entomological Research*, 111(2), 174–181. Available from: <https://doi.org/10.1017/S0007485320000462>
- Masciocchi, M., Angeletti, B., Corley, J.C. & Martínez, A.S. (2020) Drone aggregation behavior in the social wasp *Vespula germanica* (Hymenoptera: Vespidae): effect of kinship and density. *Scientific Reports*, 10(1), 7143. Available from: <https://doi.org/10.1038/s41598-020-64232-9>
- Masciocchi, M. & Corley, J. (2013) Distribution, dispersal and spread of the invasive social wasp (*Vespula germanica*) in Argentina. *Austral Ecology*, 38(2), 162–168. Available from: <https://doi.org/10.1111/j.1442-9993.2012.02388.x>
- Matsuura, M. & Yamane, S. (1990) *Biology of the vespine wasps*. Berlin Germany: Springer Verlag. Available at: <https://www.cabi.org/isc/abstract/19910230589>. Accessed 20th September 2022.
- Metcalfe, A.N., Kennedy, T.A., Mendez, G.A. & Muehlbauer, J.D. (2022) Applied citizen science in freshwater research. *WIREs Water*, 9(2), e1578. Available from: <https://doi.org/10.1002/wat2.1578>
- Moller, H. (1996) Lessons for invasion theory from social insects. *Biological Conservation*, 78(1–2), 125–142. Available from: [https://doi.org/10.1016/0006-3207\(96\)00022-5](https://doi.org/10.1016/0006-3207(96)00022-5)
- Moreau, C.S. (2014) A practical guide to DNA extraction, PCR, and gene-based DNA sequencing in insects. *Halteres*, 5(32–42), 11.
- Morin, P.A., Leduc, R.G., Archer, F.L., Martien, K.K., Huebinger, R., Bickham, J.W. et al. (2009) Significant deviations from Hardy–Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources*, 9(2), 498–504. Available from: <https://doi.org/10.1111/j.1755-0998.2008.02502.x>
- Nei, M. (1987) *Molecular evolutionary genetics*. New York: Columbia University Press. Available from: <https://doi.org/10.7312/nei-92038>
- Nevečerálova, P., Carroll, E.L., Steel, D., Vermeulen, E., Elwen, S., Zidek, J. et al. (2022) Population changes in a whale breeding ground revealed by citizen science noninvasive genetics. *Global Ecology and Conservation*, 37, e02141. Available from: <https://doi.org/10.1016/j.gecco.2022.e02141>
- O’Connor, R.S., Kunin, W.E., Garratt, M.P.D., Potts, S.G., Roy, H.E., Andrews, C. et al. (2019) Monitoring insect pollinators and flower visitation: the effectiveness and feasibility of different survey methods. *Methods in Ecology and Evolution*, 10(12), 2129–2140. Available from: <https://doi.org/10.1111/2041-210X.13292>
- Osborne, J.L., Martin, A.P., Shortall, C.R., Todd, A.D., Goulson, D., Knight, M.E. et al. (2008) Quantifying and comparing bumblebee nest densities in gardens and countryside habitats. *Journal of Applied Ecology*, 45(3), 784–792. Available from: <https://doi.org/10.1111/j.1365-2664.2007.01359.x>
- Outhwaite, C.L., Gregory, R.D., Chandler, R.E., Collen, B. & Isaac, N.J.B. (2020) Complex long-term biodiversity change among invertebrates, bryophytes and lichens. *Nature Ecology & Evolution*, 4(3), 384–392. Available from: <https://doi.org/10.1038/s41559-020-1111-z>
- Peakall, R. & Smouse, P.E. (2006) GenAlEx 6: genetic analysis in excel. population genetic software for teaching and research. *Molecular Ecology Notes*, 6(1), 288–295. Available from: <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
- Peakall, R. & Smouse, P.E. (2012) GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28(19), 2537–2539. Available from: <https://doi.org/10.1093/bioinformatics/bts460>

- Perry, J.R., Sumner, S., Thompson, C. & Hart, A.G. (2021) "Citizen identification": online learning supports highly accurate species identification for insect-focussed citizen science. *Insect Conservation and Diversity*, 14(6), 862–867. Available from: <https://doi.org/10.1111/icad.12528>
- Pompanon, F., Bonin, A., Bellemain, E. & Taberlet, P. (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, 6(11), 847–859. Available from: <https://doi.org/10.1038/nrg1707>
- Powney, G.D., Carvell, C., Edwards, M., Morris, R.K.A., Roy, H.E., Woodcock, B.A. et al. (2019) Widespread losses of pollinating insects in Britain. *Nature Communications*, 10(1), 1018. Available from: <https://doi.org/10.1038/s41467-019-08974-9>
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. Available from: <https://doi.org/10.1093/genetics/155.2.945>
- Rico, C., Cuesta, J.A., Drake, P., Macpherson, E., Bernatchez, L. & Marie, A.D. (2017) Null alleles are ubiquitous at microsatellite loci in the wedge clam (*Donax trunculus*). *PeerJ*, 5, e3188. Available from: <https://doi.org/10.7717/peerj.3188>
- Robinet, C., Suppo, C. & Darrouzet, E. (2017) Rapid spread of the invasive yellow-legged hornet in France: the role of human-mediated dispersal and the effects of control measures. *Journal of Applied Ecology*, 54(1), 205–215. Available from: <https://doi.org/10.1111/1365-2664.12724>
- Rocha-Ortega, M., Rodriguez, P. & Córdoba-Aguilar, A. (2021) Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. *Ecological Entomology*, 46(4), 718–728. Available from: <https://doi.org/10.1111/een.13027>
- Rodríguez-Ramilo, S.T. & Wang, J. (2012) The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. *Molecular Ecology Resources*, 12(5), 873–884. Available from: <https://doi.org/10.1111/j.1755-0998.2012.03156.x>
- Rohland, N. & Hofreiter, M. (2007) Ancient DNA extraction from bones and teeth. *Nature Protocols*, 2(7), 1756–1762. Available from: <https://doi.org/10.1038/nprot.2007.247>
- Rousset, F. (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. Available from: <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Schmack, J.M., Brenton-Rule, E.C., Veldtman, R., Wenseleers, T., Beggs, J.R., Lester, P.J. et al. (2019) Lack of genetic structuring, low effective population sizes and major bottlenecks characterise common and German wasps in New Zealand. *Biological Invasions*, 21(10), 3185–3201. Available from: <https://doi.org/10.1007/s10530-019-02039-0>
- Schwartz, M.K., Luikart, G. & Waples, R.S. (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, 22(1), 25–33. Available from: <https://doi.org/10.1016/j.tree.2006.08.009>
- Skrbinšek, T., Luštrik, R., Majič-Skrbinšek, A., Potočnik, H., Kljun, F., Jelenčič, M. et al. (2019) From science to practice: genetic estimate of brown bear population size in Slovenia and how it influenced bear management. *European Journal of Wildlife Research*, 65(2), 29. Available from: <https://doi.org/10.1007/s10344-019-1265-7>
- Stork, N.E. (2018) How many species of insects and other terrestrial arthropods are there on earth? *Annual Review of Entomology*, 63(1), 31–45. Available from: <https://doi.org/10.1146/annurev-ento-020117-043348>
- Straube, N., Lyra, M.L., Pajmans, J.L.A., Preick, M., Basler, N., Penner, J. et al. (2021) Successful application of ancient DNA extraction and library construction protocols to museum wet collection specimens. *Molecular Ecology Resources*, 21(7), 2299–2315. Available from: <https://doi.org/10.1111/1755-0998.13433>
- Sumner, S., Law, G. & Cini, A. (2018) Why we love bees and hate wasps. *Ecological Entomology*, 43(6), 836–845. Available from: <https://doi.org/10.1111/een.12676>
- Sumner, S., Bevan, P., Hart, A.G. & Isaac, N.J.B. (2019) Mapping species distributions in 2 weeks using citizen science. *Insect Conservation and Diversity*, 12(5), 382–388. Available from: <https://doi.org/10.1111/icad.12345>
- Team, R.C. (2022) R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Thorén, P.A., Paxton, R.J. & Estoup, A. (1995) Unusually high frequency of (CT)_n and (GT)_n microsatellite loci in a yellowjacket wasp, *Vespula rufa* (L.) (Hymenoptera: Vespidae). *Insect Molecular Biology*, 4(3), 141–148. Available from: <https://doi.org/10.1111/j.1365-2583.1995.tb00019.x>
- Thornhill, I., Loiseau, S., Clymans, W. & van Noordwijk, C.G.E. (2019) How citizen scientists can enrich freshwater science as contributors, collaborators, and co-creators. *Freshwater Science*, 38(2), 231–235. Available from: <https://doi.org/10.1086/703378>
- Van Oosterhout, C., Hutchinson, W.F.D., Wills, D.P. & Shipley, P. (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3), 535–538. Available from: <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
- Veldtman, R., Daly, D. & Bekker, G.F.H.v.G. (2021) Spatio-environmental analysis of *Vespula germanica* nest records explains slow invasion in South Africa. *Insects*, 12(8), 732. Available from: <https://doi.org/10.3390/insects12080732>
- Wang, J. (2004) Sibship reconstruction from genetic data with typing errors. *Genetics*, 166(4), 1963–1979. Available from: <https://doi.org/10.1534/genetics.166.4.1963>
- Wang, J. (2019) Pedigree reconstruction from poor quality genotype data. *Heredity*, 122(6), 719–728. Available from: <https://doi.org/10.1038/s41437-018-0178-7>
- Waples, R.S. (2015) Testing for Hardy-Weinberg proportions: have we lost the plot? *Journal of Heredity*, 106(1), 1–19. Available from: <https://doi.org/10.1093/jhered/esu062>
- Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. Available from: <https://doi.org/10.2307/2408641>
- Whitehorn, P.R., Seo, B., Comont, R.F., Rounsevell, M. & Brown, C. (2022) The effects of climate and land use on British bumblebees: findings from a decade of citizen-science observations. *Journal of Applied Ecology*, 59(7), 1837–1851. Available from: <https://doi.org/10.1111/1365-2664.14191>
- Zapponi, L., Cini, A., Bardiani, M., Hardersen, S., Maura, M., Maurizi, E. et al. (2017) Citizen science data as an efficient tool for mapping protected saproxylic beetles. *Biological Conservation*, 208, 139–145. Available from: <https://doi.org/10.1016/j.biocon.2016.04.035>
- Zhang, Y.M., Vitone, T.R., Storer, C.G., Payton, A.C., Dunn, R.R., Hulcr, J. et al. (2019) From pavement to population genomics: characterizing a long-established non-native ant in North America through citizen science and ddRADseq. *Frontiers in Ecology and Evolution*, 7, 453. <https://www.frontiersin.org/articles/10.3389/fevo.2019.00453>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1: Supporting Information.

How to cite this article: Cunningham-Eurich, I., Kontou, D., Yordanova, M., Maeda-Obregon, A., Favreau, E., Wang, J. et al. (2023) Using citizen science data to assess the population genetic structure of the common yellowjacket wasp, *Vespula vulgaris*. *Insect Molecular Biology*, 32(6), 634–647. Available from: <https://doi.org/10.1111/imb.12862>