Bucking the trend of pollinator decline: the population genetics of a range expanding bumblebee

Authors: Huml JV[1], Ellis JS[1], Lloyd K[1], Benefer CM[1], Kiernan M[1], Brown MJF[2], Knight ME[1]

Corresponding author: Huml JV[1]

Email: vanessa.huml@plymouth.ac.uk

1. School of Biological and Marine Sciences, University of Plymouth, Plymouth PL4 8AA.

2. Centre for Ecology, Evolution and Behaviour, School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX.

Running title: The population expansion of *B. hypnorum* into the UK.

Keywords: *Bombus hypnorum*; range expansion; pollinators; invasion genetics; ABC population modelling; RAD-seq.

Total word count: 8073

Number of Tables: 5

Number of Figures: 6

ABSTRACT

Recent research has shown drastic reductions in the global diversity and abundance of insects. This is a major concern given the expected cascade effects on ecosystem services, such as pollination. Understanding the patterns and drivers of changes in the distribution and abundance of species in our rapidly changing environment is therefore urgent. Cases of species showing trends that run counter to general population declines, especially when they deliver key ecosystem services, are especially interesting. The tree bumblebee (*Bombus hypnorum*), which belongs to a globally important group of pollinators, has substantially expanded its range in recent years in direct contrast to many other species within this group. Here we reconstructed the likely pattern of colonization of the UK based on RAD-seq population genomic data combined with Bayesian population modelling. This RAD-seq approach also enabled an analysis of genomic regions potentially under selection. We report a complex and dynamic colonization pattern that is most likely ongoing. Current evidence suggests that either a shift in its migration potential, and/or adaptive genomic changes have contributed to the recent range expansion of *B. hypnorum*. Genomic areas of potential adaptive significance included genes involved in regulation of transcription and gene expression, circadian rhythms and innate immunity. Our results are framed within the general context of understanding the factors driving successful population expansions.

INTRODUCTION

Current rates of species extinction are substantially elevated relative to the historical record (Ceballos et al. 2015) and many species are undergoing declines in range and abundance (Hallmann et al. 2017). However, some species remain widespread or are even undergoing range expansion. Such contrasting patterns of population decline and success are especially interesting when they occur among species within close phylogenetic groups (McKinney and Lockwood 1999; Angert et al. 2011; Moran and Alexander 2014). This allows identification of traits that potentially contribute to species' resilience or extinction risk (Purvis 2008; Chichorro et al. 2019) and therefore provides ideal opportunity to gain traction on the factors driving diversity changes (Sax et al. 2007; Moran and Alexander 2014).

A key aspect of understanding these rapid changes in global biodiversity patterns is unravelling the dynamics of colonization. This has historically been achievable only through reconstructing events from observational records. While useful, these are not always available and are typically limited in their capacity to trace historical events with accuracy. However, the progressive accessibility of genome-scale data, and development of population modelling tools (Cornuet et al. 2014; Cabrera and Palsbøll 2017), have offered a powerful approach that is yielding significant insights into the routes and biological signatures of colonizations (e.g. Guzinski et al. 2018). These tools have been successfully used in a range of contexts, such as inference of the evolutionary history of speciation (Momigliano et al. 2017), measuring the success of translocation programmes (Puckett et al. 2014), and investigating meta-population dynamics (Stillfried et al. 2017).

Advances in sequencing technology also facilitate the investigation of the wider genetic effects associated with range expansions. For example, new colonizations are often characterized by a limited number of founders, resulting in a population bottleneck and small initial population sizes (Dlugosch and Parker 2008). However, colonizing species often successfully establish within their new environment despite the predicted loss of genetic diversity and harmful inbreeding effects, posing a genetic paradox (Allendorf and Lundquist 2003; Sax & Brown 2000; Estoup et al. 2016). One potential explanation is that bottlenecks may not lead to a substantial loss of genetic variance in quantitative traits (Lewontin 1965); it may even increase when dominance interactions are considered (Robertson 1952). The loss of genetic diversity during colonization may also be overcome if there are multiple colonization events or high migration rates from the source population, which can eliminate founder effects (Lockwood et al. 2005; Roman and Darling 2007; Dlugosch and Parker 2008).

In parallel with developments in sequencing technologies, there have been theoretical advances in the understanding of range expansions. Facon et al. (2006) proposed a useful framework to consider colonization scenarios, identifying three scenarios that are not mutually exclusive. In the first, 'migration change', a species is limited only by its capacity to migrate to an existing habitat that meets its requirements. Here, colonization may occur if barriers to migration are removed through, for example, human activities. The second scenario, 'environmental change', describes the situation where suitable habitat becomes newly available. The third scenario, 'evolutionary change', assumes genetic changes in the colonizing species that either occur prior to, or during, the colonization, conferring a fitness advantage in the new environment. Successful range expansions following evolutionary

95 change have been linked to adaptations to anthropogenically modified habitats, coined as
96 'Anthropogenically Induced Adaptations to Invade' (Hufbauer et al. 2012).

97

98 Bumblebees (*Bombus* spp.) are a highly pertinent group to investigate factors driving changes
99 in diversity as the group contains taxa with strikingly different population trends. In addition,
100 globally, patterns of decline are predominant, which is an issue of particular concern because
101 of their role as ecologically and economically important pollinators (Goulson et al. 2005;
102 Williams 2005; Williams and Osborne 2009).

103 Patterns of declines with the *Bombus* group are, at least to some degree, phylogenetically
104 structured. For example, members of the subgenus *Thoracobombus* appear to show
105 increased vulnerability to population decline (Cameron et al. 2011; Arbetman et al. 2017), as
106 do several others (Goulson et al. 2008), but those of the subgenus *Pyrobombus*, in contrast,
107 seem to exhibit increased resilience (Arbetman et al. 2017; Richardson et al. 2019). Several
108 *Pyrobombus* species are reported to be increasing in abundance and/or expanding their
109 range, for example*Bombus haematurus* in central Europe (Biella 2020), *Bombus pratorum* and
110 *Bombus monticola* which have colonized Ireland during the last century (Speight 1974;
111 Fitzpatrick et al. 2007), *Bombus bimaculatus, Bombus impatiens, Bombus ternarius,* and
112 *Bombus vagans* in Vermont (Richardson et al. 2019) and *Bombus hypnorum* across the
113 western part of its distribution (Rasmont et al. 2015). These contrasting patterns suggest that
114 this group exhibits traits that make it less vulnerable to threats that are causing declines in
115 the majority of other bumblebees. Investigating the underlying mechanisms for this increased
116 resilience in this group are therefore important to improve understanding of the drivers of

diversity change, and in turn inform conservation efforts, in this important group of pollinators.

A notably successful species among the *Pyrobombus* group is the tree bumblebee, *Bombus hypnorum*, which has recently substantially expanded its range (Goulson and Williams 2001; Prŷs-Jones et al. 2016). *Bombus hypnorum* is one of the most widespread bumblebee species across Europe and Asia (Williams 1991; Goulson and Williams 2001), with a broad palaearctic distribution from Iceland to Japan (Williams 1991; Rasmont et al. 2015). It has been present along the north-western coast of Belgium and France from at least the early 20th century (Rasmont 1988). A recent study showed a lack of genetic structuring of *B. hypnorum* populations in Belgium, indicating a large panmictic population across western parts of Europe (Maebe et al. 2019). Significant differentiation was observed between western European and Baltic populations, suggesting population structure at much larger geographical scales (Maebe et al. 2019). In the last decades, *B. hypnorum* has expanded its range significantly in the western part of its distribution (Rasmont et al. 2015), reaching Iceland in 2010 (Prŷs-Jones et al. 2016) and Ireland in 2017. It was first recorded in the south of England in 2001 (Goulson and Williams 2001) and has since spread rapidly northwards, arriving in Scotland in 2012. It is now one of the most common bumblebee species in the UK, both abundant and widely distributed (BWARS 2019). Given no published data to-date on any negative impacts of this population expansion, we refer to this throughout as a colonization rather than an invasion.

There are a number of characteristics of *B. hypnorum* that may indicate its potential as a successful colonizer. It has a wide distribution across Europe and Asia, where it is found across diverse types of habitats (Goulson and Williams 2001). This reflects its broad niche, including dietary and climatic requirements (BWARS 2019), attributes generally associated with successfully colonizing species (Baker 1965; Willamson and Fitter 1996; Vazquez 2006). It exhibits a facultative bivoltine colony cycle (Edwards and Jenner 2005), which may contribute to an increased rate of population growth facilitating an accelerated spread (Sakai et al. 2001). It also shows facultative polyandrous mating patterns, at least in some parts of it range, which may increase the genetic diversity within colonies (Paxton et al. 2001). Finally, *B. hypnorum* is unique amongst UK *Bombus* species in nesting high above ground (Benton 2006), often in buildings, bird nest boxes and roof structures, which may give it a competitive advantage in a highly urbanized environment (Crowther et al. 2014). However, what has enabled its rapid spread since 2001 remains unclear: these are long-standing characteristics and *B. hypnorum* has nonetheless only very recently colonized the UK. Information on continental populations is incomplete, but available information points to *B. hypnorum* being well established and abundant across western continental Europe for some time (>100 years) prior to its colonization of the UK (Rasmont et al. 1988; Maebe et al. 2019). Although there are intrinsic lag times in the introduction, growth, expansion and detection of colonizing species (Crooks et al 2005), this lag would be particularly prolonged (up to 200 generation since *B. hypnorum* is bivoltine) in this case if the colonization of the UK represents the continuation of a range expansion dating back to over a century ago. Thus, evidence to-date points to a significant change in either the environment, which includes an increase in opportunities for migration, e.g. through human aided transport, or the intrinsic biology of this species that have contributed to its sudden and highly successful range expansion.

Here, we explore the genetic signatures associated with this rapid expansion of *B. hypnorum* into the UK using a RAD-seq population genomic dataset. Our specific goals were to: i) identify whether the UK population was founded from a single event or from multiple and potentially ongoing events; (ii) assess population structure and whether there has been any significant loss of genetic diversity as a consequence of the population expansion; (iii) identify preliminary indicators of any signatures of selection before or during the colonization that may have promoted its success.

MATERIALS AND METHODS

Establishing the spread of Bombus hypnorum *across the UK*

The likely geographical spread of *Bombus hypnorum* across the UK from its first record in 2001 (Goulson and Williams 2001) was reconstructed using the BWARS database (Bees, Wasps & Ants Recording Society 2019), which collates recorded sightings as part of an ongoing dedicated mapping program (Figure 1). This database is reliant on records being sent in by the public, amateur groups and specialists. Although it will inevitably contain some gaps and biases (e.g. towards human population density), all records are verified by specialists within BWARS for quality assurance (van der Wal et al. 2015). While not systematic, it is nevertheless likely to estimate broad patterns reasonably, and has been used previously for such purposes (e.g. Potts et al. 2010; Ollerton et al. 2014).

Tissue samples and RAD library preparation

185 *Bombus hypnorum* was sampled from seven localities representing different 'fronts' of the
186 range expansion of this species (as established from the BWARS database, see also Results).
187 Six localities were sampled in the UK (Cardiff, Hull, London, Newcastle, Plymouth and
188 Southampton), representing different stages of the expansion, and one in France (Le Havre)
189 representing its already well-established distribution in continental Europe (Maebe et al.
190 2019).

191

192 Forty individual samples were collected across a large area (approximately 10 x 10km) at each
193 site using standard population sampling methods for colony-living Hymenopteran species
194 (see e.g. Goulson et al. 2011) where individual samples were collected a minimum of 200m
195 apart to avoid any significant sampling of sisters (belonging to the same nest). Samples were
196 collected between May and July of 2013 and 2014 and stored in 100% ethanol.

197

198 DNA was extracted from the thoracic muscle tissue of all sampled individuals using an
199 ammonium acetate protocol (Nicholls et al. 2000) and quantified on a Qubit 3.0 Fluorometer
200 using a broad range assay (Thermo Fisher Scientific). DNA (500ng) was digested in 20µl
201 volumes after RNase treatment with 40 units of the restriction enzyme XhoI (New England
202 Biolabs) at 37°C for 3 hours with a 20 minute heat deactivation stage at 80°C. Digested DNA
203 was purified using AmpureXP (Beckmann and Coulter, 1.4X ratio of beads to DNA) and
204 quantified on a Qubit 3.0 Fluorometer (Thermo Fisher Scientific). RAD libraries were prepared
205 using the IonXpress Plus gDNA Fragment Library Kit (Life Technologies). Library preparation
206 was carried out using a pooled approach with population-specific barcodes using equimolar
207 concentrations from each individually digested sample for each sampling site. A pooled

sequencing design represents a well established and cost-effective alternative to individual sequencing to reliably obtain genome wide allele frequency data (Futschik and Schlötterer 2010; Gautier et al. 2013; Schlötterer et al. 2014) and has been used in a wide range of systems (e.g. Guo et al. 2015; Kahnt et al. 2018). (We note that a limitation of this approach is the lack of individual genotypes and heterozygosity necessary for analysis such as assignment tests or estimation of inbreeding coefficients (Andrews et al. 2016)). For quality control, library preparation of the Newcastle sample was performed twice independently using the pooled approach. Additionally, both the Newcastle and Le Havre samples were prepared and sequenced using an individual barcoding approach with twenty individual samples for each sampling site (Gautier et al. 2013). A barcoded Ion Torrent adapter A was ligated using 0.1 µM of barcode adaptor, 200 U of T4 DNA ligase (New England Biolabs), 100 mM of ATP and 2µl of NE 4 Buffer in 40µl volumes for 2 hours at 22°C, followed by heat deactivation at 65°C for 20 minutes, either on an individual or pooled basis for each sampling site. Purification was repeated twice after this step (1.2X ratio of beads to DNA) and prior to shearing using Ion Shear Plus Enzyme Mix II (Life Technologies) following the manufacturer's protocol. After further AmpureXP purification (1.4X ratio of beads to DNA) the Ion Torrent adapter P1 (Thermo Fisher Scientific) was ligated in 49µl volumes for 20 minutes at 25°C followed by heat deactivation at 72°C for 5 minutes following the manufacturer's guidelines. After another step of AmpureXP purification (1.2X ratio of beads to DNA), library amplification was achieved through: 5 minutes at 95°C, followed by 18 cycles of a heat denaturation at 95°C for 15 seconds, annealing at 58°C for 15 seconds and extension at 70°C for 1 minute. Pippin Prep (Sage Science) was employed to select for a fragment size range of 159 and 164bp prior to AmpureXP purification (1.5X ratio of beads to DNA) and quantification was then estimated

231 by qPCR. Libraries were run on an Ion Torrent PGM using one 318 chip for each individually
232 barcoded sample or for two pooled samples respectively.

233

234 *Data processing and SNP calling*

235 Raw reads were trimmed from both ends to excise low quality base-calls (average Q-score <
236 15 across 4bp sliding windows) and filtered for a minimum length of 10bp using Trimmomatic-
237 0.36 (Bolger et al. 2014). Filtered reads were aligned to the *Bombus terrestris* genome
238 (Bter_1.0 assembly, Ensembl) using the Burrow-Wheeler Aligner (BWA) and the MEM
239 algorithm (Li 2013), which has been shown to perform best in the presence of indel errors
240 typical of the Ion Torrent (Ziemann 2016). We chose the *Bombus terrestris* genome, and not
241 that of the more closely related *Bombus impatiens*, because linkage group annotations are
242 available for this species. This allowed us to investigate patterns of diversity across genomic
243 regions. We tested if there was evidence for bias caused by this choice through comparison
244 of mapping statistics against the *Bombus impatiens* genome. Aligned reads were only retained
245 if they mapped uniquely to the reference genome and had a minimum mapping quality score
246 of 20 (Cibulskis et al. 2013). Aligned BAM files were sorted and converted into the mpileup
247 format allowing a maximum of 1,000 reads at a given position using Samtools 0.1.5 (Li et al.
248 2009). SNP calling (excluding indels) for individually barcoded samples was performed using
249 VarScan v.2.3.9 (Koboldt et al. 2009) using all sites that had a minimum coverage of 10 in at
250 least 10 individuals. For these individual samples, requirements for SNP acceptance were a
251 minimum of two reads (default) for the minor allele and a p-value of ≤0.05 derived from
252 Fisher's exact test on read counts (coverage) supporting the reference and variant allele
253 respectively (Koboldt et al. 2009; Koboldt et al. 2013).

The mpileup file of pooled samples was converted to the sync file format and filtered for indels using Popoolation2 (Kofler et al. 2011). Because variation in sequencing depth can impair the accuracy of allele frequency estimates derived from pooled samples, only sites with a sequencing coverage of ≥50 and ≤ 500 within any sample, and a maximum coverage of 3000 across all samples, were considered for pooled samples. Note that specifically for the analysis of the distribution of genome wide polymorphism, which does not rely on allele frequency estimates, a minimum within-sample coverage of 10 was applied. For pooled samples, SNPs were called if they had a minimum of 3 reads per allele across all samples and a minimum within-sample allele frequency of 0.01, as commonly used in population genetic studies (e.g. Bruneaux et al. 2013; Fraser et al. 2014). We also applied a threshold of 0.05 for the identification of F_{ST}-outliers to test for consistency (Roesti et al. 2012).

Analysis of population genetic diversity and characterization of population structure

Average expected genome-wide heterozygosity was calculated following Fischer et al. (2017). The proportion of polymorphic sites was calculated across all base pairs covered in all samples. Pseudo-haplotypes were created based on population allele frequencies using a custom Python script and imported into R v.3.4.2 (R Core Team, 2012). Allelic richness was calculated in the R package Hierfstat v.0.04-22 (Goudet 2004). Pairwise population F_{ST} was calculated using Popoolation2 (Kofler et al. 2011) and confidence intervals were created by bootstrapping over loci for 1000 cycles using a custom Python script. A Mann-Whitney-Wilcoxon test was used to test if there is a significant difference in the pairwise F_{ST} values between UK-Le Havre and UK-UK populations (Mann and Whitney 1947). A Mantel test

(Mantel 1967) was implemented in Genepop 4.2 (Raymond and Rousset 1995) to test for correlation between geographic and genetic distance. A principal coordinate analysis was performed in the R package Hierfstat v.0.04-22 (Goudet 2004). This was carried out for SNPs covered in all sample sites, six sample sites (75%) and four sample sites (50%) to assess the effect of missing data on the resolution of population structure.

Testing hypotheses of colonization history

The approximate Bayesian computation software DIYABC v2.0 (Cornuet et al. 2014) was used to assess the relative probability of different colonization scenarios. These were derived using a 'bottom up' hierarchical nested approach to identify the most likely colonization for each UK sample site independently (described in detail below). The highest probability scenario for each individual sampling site was combined into a final consensus model. The advantages of this approach are that it allows the most robust model to be constructed from the data available. No prior assumptions based on the observed BWARS records were required and the history of each sampling site is built independently. It also allowed model construction using the maximum number of available SNPs at each stage. This approach involved the following steps:

Step 1. For each UK sampling site separately, the likelihood of the panmictic Western European population (sampled at Le Havre, France) as a colonization source was tested in DIYABC against a simulated population, where the Western European population and the simulated population share the same common ancestral population (Figure 2). This was carried out to test for the contribution of a potential other source that is more divergent to the Western European population, e.g. from Scandinavia. The possibility of simulating

300 unsampled populations (with shared ancestry to test populations) is an effective feature of
301 the DIYABC software implementation to account for the possible contribution of other genetic
302 lineages that have not been sampled (Estoup and Guillemaud 2010). Simulated populations
303 are constructed using coalescence theory by generating a genealogy based on the defined
304 order of events in a given scenario (Cornuet et al. 2014).

305 Step 2. This next step tested whether each UK sampling site was (i) founded independently
306 from the most likely source identified in step 1; (ii) was founded sequentially via the most
307 likely source in step one and through another UK sampling site (e.g. Le Havre colonized
308 Southampton which colonized London); (iii) was founded by ongoing colonization from the
309 source identified in step one and concurrent colonization by another UK sampling site (Figure
310 2).

311 Step 3: In the case that more than one UK sampling site was more likely to be the source of
312 another UK sample than the continental sample in step 2, these were tested against each
313 other as well as against a dual (concurrent) colonization from each of them. As an example:
314 for the Cardiff sampling site, the most likely scenario derived from step 1 was a colonization
315 from Le Havre. Testing this scenario against the probability of colonization from every other
316 UK sampling site directly or in combination with Le Havre resulted in both Hull and Newcastle
317 being more likely as a source of colonization than Le Havre. In step 3 a colonization from Hull
318 was tested against a colonization from Newcastle and a dual colonization from both sites with
319 the latter showing highest probability.

320

321 Step 4. The most likely scenarios from steps 1- 3 were then combined into a final consensus
322 model that included all sample sites.

All scenarios implemented were tested with or without a bottleneck.

Each individual scenario was tested using a total of 100,000 iterations, following recommended guidelines (Cornuet et al. 2014). For Newcastle (which included a technical replicate), only the "Newcastle 2" replicate was considered (see Results). Model evaluation was based on summary statistics for all biallelic SNP markers covered in the sampling sites considered. Summary statistics consisted of the mean of non-zero values, variance of non-zero values and mean of the complete distribution for genic diversity and pairwise F_{ST} and Nei's distance. Uniform simulation priors were applied to all demographic parameters. Prior constraints on simulation parameters consisted of the definition of the order of events (t3>t2>t1) and an initial effective population size between 10 and 100, where bottlenecks were simulated. Model scenarios were compared using a subset of 1% of all simulations, which matched the observed data best (Cornuet et al. 2014). Using a logistic regression approach, deviations from the summary statistics among the selected subset of simulations were used to predict the probability of a given scenario (Estoup et al. 2012). In order to reduce correlation among explanatory variables, summary statistics were transformed by linear discriminant analysis prior to logistic regression (Estoup et al., 2012).

Identification of genomic regions under selection

A commonly adopted approach to investigate genomic regions that may be under strong selective regimes is to look for F_{ST} outliers in population genomic datasets where several populations or locations have been sampled (e.g. Vandepitte et al. 2014; Lin et al. 2017; Leydet et al. 2018; Theodorou et al. 2018). This is based on the expectation that areas of

reduced or elevated differentiation are subject to balancing and directional selection respectively (e.g. Lotterhos and Whitlock 2014). Other commonly used approaches to detecting selection in population genomic studies follow the expectation that directional selection decreases genetic variation in the genomic region of the selected site and balancing selection increases it (Oleksyk et al. 2010). Levels of genetic diversity can then be compared to the genomic background to infer candidate regions under selection (Hohenlohe et al. 2010, Bruneaux et al. 2013).

Approach 1: F_{ST}-outlier approaches

To identify signatures of selection, we tested for elevated differentiation by first considering all pairwise comparisons between UK sampling sites and secondly for all pairwise UK-Le Havre comparisons. This allows distinction between patterns of differentiation across the UK and/or between UK and a continental population. Two approaches were used: a permutation approach (Bruneaux et al. 2013) and BayeScan (Foll & Gaggiotti 2008).

For the permutation approach, observed F_{ST} -values averaged across all pairwise comparisons of sampling sites were shuffled 1000 times and compared to the observed F_{ST} values averaged across SNP sites within 10kb sliding windows. A 10kb window size was chosen because linkage has been shown to decrease rapidly over these distances in *Bombus* (Sadd et al. 2015). To avoid single SNPs or RAD-tags driving the average across windows, a minimum SNP density of 3 SNPs was required across at least two independent RAD-tags for a window to be included in the analysis (Purfield et al. 2017; Jacobs et al. 2018). P-values were generated as the proportion of permutations being lower/higher than the observed estimates and corrected

for multiple testing using a FDR approach implemented in the qvalue package in R (Storey et al. 2015). For the BayeScan approach (Foll and Gaggiotti 2008), the program was run using default parameters. BayseScan implements a basic regression model to differentiate between locus and population specific effects on the distribution of F_{ST} -values (Foll and Gaggiotti 2008). A likelihood ratio test is then used to assess if the population specific component is sufficient to explain the observed variation (no selection) or if a SNP specific component (selection) is supported (Foll and Gaggiotti 2008). This allows assessment of population specific demographic effects in contrast to the permutation approach, which identifies larger genomic areas with elevated F_{ST} values averaged across all sampling sites considered.

To reduce the rate of false positives (De Mita et al. 2013), genomic regions were only considered as outliers if they were identified using both the permutation and the BayeScan approach.

Approach 2: Distribution of polymorphic sites

Using the approach of Bruneaux et al. (2013), which evaluates the distribution of polymorphism across the genome against a random null-distribution and does not rely on allele frequency estimates, the proportion of polymorphic sites was calculated. This was performed across non-overlapping 10kb sliding windows or across the length of the contig sequence in the case of genomic regions not placed within the assembled reference genome of *B. terrestris*. To generate a null-distribution 1,000 permutations were applied to the data and p-values were generated as the proportion of permutations being lower/higher than the observed estimates (Bruneaux et al. 2013). Only windows that had a minimum coverage of 100bp were considered in the analysis (Cooper et al. 2004). For the analysis of low

polymorphism, the hypergeometric test was used to derive the minimum coverage in base pairs needed across a window to obtain a probability below 5% of not sampling a SNP within a sliding window given our observed SNP density (Lentner, 1972; Fontanillas et al., 2010). In line with other studies, a false discovery rate (FDR) of 10% (q-value < 0.1) was applied (e.g. Krehenwinkel et al. 2015; Rane et al. 2015) as an appropriate balance between the false discovery rate and statistical power where large numbers of tests are involved (van den Oord 2008). Genes that were found within windows of significantly high/low polymorphism were considered for gene ontology analysis.

Gene ontology analysis

Gene ontology (GO) terms for the *B. terrestris* genome were obtained from the Ensembl database. Enrichment tests based on gene count (the 'classic' algorithm, Alexa et al. 2006) were conducted using the R package topGO (Alexa and Rahnenfuhrer 2016). These were assessed using Fisher's exact test and a minimum node size of 10 in order to prune our hierarchy from nodes with the support of less than 10 annotated genes, a frequently applied threshold (e.g. Ahrens et al. 2013; Rademacher et al. 2017). Correction for multiple testing (FDR < 5%, q-value < 0.05) was carried out using the qvalue package in R (Storey et al. 2015). Additionally, the 'weight' and 'elim' algorithms were used, which account for dependencies within the gene ontology hierarchies (Alexa et al. 2006). Here, multiple testing theory does not directly apply as tests are not independent and raw p-values ≤0.05 were considered as significant (Alexa and Rahnenfuhrer 2016). Genes within sliding windows that were identified as F_{ST} outliers or that showed significantly high or low polymorphism were subject to gene ontology analysis to investigate if outliers show significant enrichment for specific biological or molecular functions.

In order to classify SNPs as synonymous or non-synonymous, the annotation for the *B. terrestris* genome was obtained from the Ensembl database, which is assembled across 18 linkage groups.

RESULTS

The range expansion of Bombus hypnorum *into and across the UK*

After the first record of *B. hypnorum* near Southampton in 2001 (Goulson and Williams 2001) more sightings followed across the south of the UK, with records increasing rapidly thereafter (Figure 1). The sharpest increase in records was observed in 2009, where sightings increased more than four-fold compared to the previous year (likely to have been at least partly influenced by an increasing public awareness of the existence of *B. hypnorum* in the UK). The expansion was first in a north-eastward direction with some isolated records as far north as Hull in 2005 and Newcastle in 2007. The expansion then extended to the west with the first sightings of *B. hypnorum* in Cardiff in 2009, and in Plymouth in 2010.

RAD-seq data summary

Ion Torrent sequencing generated 15,513,192 raw reads for all sample pools. After mapping, quality filtering and the application of coverage thresholds 2,469,636bp were covered in total among pools, representing ~1% of the expected genome size. This covers 40% of all expected cut-sites (57,157) of the enzyme used to digest samples (XhoI). Within each sampling site, the range of coverage was 152,225 to 953,719bp (Table 1). After stringent filtering, 12,823 high confidence SNPs were identified in total, ranging from 464 to 3,695 within each sampling site.

Mapping success was on average 6±3% higher against the *B. impatiens* genome but the proportion of raw reads that were uniquely mapped per sampling site was highly correlated (rho 0.94, p<0.001) between the *B. terrestris* and *B. impatiens* genome, indicating minimal bias. We therefore used the *B. terrestris* genome to take advantage of annotated linkage groups, which has been shown to generally increase the power to detect selection across genomic regions compared with single site comparisons (Shafer et al. 2017). Across all genomic sites that were covered in all sampling sites(1,886, including non-polymorphic and polymorphic sites), the proportion of polymorphic markers varied from 0.006 (London) to 0.014 (Le Havre, Table 1), which fall within the range reported from other studies (e.g. Catchen et al. 2013). Deviation in the proportion of polymorphic markers between replicate pools was 0.001.

Allele frequency validation

In line with other studies we tested the consistency of our genotyping approach by comparing replicate allele frequency estimates (Anand et al. 2016, Guo et al. 2016; Dorant et al. 2019). A total of 3,065 SNP sites were covered in both Newcastle replicates and allele frequencies were highly correlated between them (rho 0.95, p <2.2 e^{-16}) with an average difference in allele frequencies of 0.06. SNP allele frequencies within the individually barcoded Newcastle and Le Havre replicates compared to the pooled samples were high (rho 0.87, p <2.2 e^{-16} across 361 SNP sites for Newcastle 1, rho 0.97, p <2.2 e^{-16} across 507 sites for Newcastle 2, and rho 0.97, p <2.2 e^{-16} across 439 sites for Le Havre). The average allele frequency difference was 0.14 for the Newcastle 1 replicate, 0.07 for the Newcastle 2 replicate and 0.06 for the Le Havre pool in comparison to the respective individual data set.

462

463 *Characterization of population structure*

464 Average heterozygosity ranged from 0.12 in London and Hull to 0.2 in Le Havre and Newcastle
465 (Table 1). Allelic richness ranged from 0.11 in London and 1.26 in Le Havre (Table 1). The
466 deviation in average heterozygosity between the replicate pools was 0.02. Principal
467 coordinate analysis on pairwise F_{ST} values clustered the Newcastle replicates closely together
468 with the Le Havre sampling site irrespective of the threshold of missing data that was applied
469 (Figure 3). F_{ST} -values were generally low (Table 2), and no significant isolation by distance
470 was detected using a Mantel test (R^2 = -2.3^{-5}, p=0.87). There was no significant difference in
471 average F_{ST} between within-UK comparisons and UK–Le Havre comparisons (Mann-Whitney-
472 Wilcoxon test, p-value >0.05). The average F_{ST} for within-UK comparisons was 0.023 ± 0.008
473 and 0.016 ± 0.007 for the UK–Le Havre comparison.

474

475 *Evaluation of colonization history*

476 The continental reference sample (Le Havre), which we have assumed to be representative
477 of the panmictic Western European population, was more likely to be the source of
478 colonization than the simulated unsampled population in all cases (Figure 4). Modelling
479 results gave high support for multiple colonization events, indicating a direct colonization
480 from continental Europe into Southampton, London and Newcastle and complex patterns of
481 colonization within the UK (Figure 4, Table 3). A bottleneck was supported for all populations
482 with the exception of Southampton (Figure 4, Table 3). The highest relative probability was
483 observed for the colonization history of the London, Newcastle and Plymouth sites (Table 3).
484 For the Cardiff, Hull and Southampton sites relative probabilities of the second most likely

485 scenario were more similar and confidence intervals were overlapping (Table 3). In the cases
486 of Cardiff and Southampton, the two top scenarios differed in their support for a dual
487 colonization history, whilst for the Hull site the probability of a colonization from London was
488 closely followed by the probability for a direct colonization from the Western continental
489 population (Table 3).

490

491 *Detection of selection*

492 Of all sites covered (2,469,636bp), 136,626bp fell within coding sequences, of which 557 were
493 polymorphic. Of these polymorphisms, 346 were non-synonymous and 211 synonymous. A
494 total of 75 F_{ST}-outliers (66 under directional selection and 9 under balancing selection) were
495 identified by BayeScan. The permutation approach identified 8 windows with significantly
496 elevated F_{ST}-values (6 for within UK comparisons and 2 for UK-Le Havre comparisons) and 11
497 windows with significantly low F_{ST}-values (3 for within UK comparisons and 8 for UK-Le Havre
498 comparisons) (Figure 5). The analysis of patterns of polymorphism across the genome
499 revealed 1,219 windows with significantly elevated polymorphism (Figure 6). Given an
500 observed SNP density of 4.5 per kB, the minimum coverage in base pairs required to obtain a
501 probability below 5% of not detecting any SNP within a sliding window was 645bp, as shown
502 by the hypergeometric test. This resulted in a total of 383 windows exhibiting significantly
503 reduced polymorphism (Figure 6).

504

505 Genes that showed evidence of directional selection from both approaches and all three
506 methods (i.e. F_{ST}-outlier permutation, BayeScan *and* the polymorphism approach) were the
507 protein vestigial (Le Havre-UK comparison), the circadian locomotor output cycles protein

508 (kaput), and one gene important in signal transduction (serine/threonine-protein kinase NLK),
509 both from within-UK comparisons (Table 4). None of the outliers identified by BayeScan to be
510 under balancing selection fell within windows of low F_{ST} in any of the data sets.

511

512 GO analysis

513 GO analysis revealed areas of high polymorphism that were significantly associated with
514 biological processes, including regulation of transcription and gene expression, signaling and
515 developmental processes (Table 5A). Further, areas of high polymorphism were significantly
516 associated with the molecular functions of protein and sequence specific DNA binding, DNA
517 binding transcription factor activity and zinc ion binding (Table 5A). Areas of significantly low
518 polymorphism were associated with the molecular functions of transmembrane signaling
519 receptor activity and extracellular ligand-gated ion channel activity (Table 5B). GO analysis did
520 not reveal any significant associations for molecular functions or biological processes for
521 windows identified as outliers within any of the F_{ST} data sets.

522

523 DISCUSSION

524 The colonization of the UK by *Bombus hypnorum* is an important and interesting counter-
525 example to the reported widespread decline in bumblebees across the northern hemisphere
526 (Goulson et al. 2008). Here we report (i) evidence for multiple entries into the UK (ii) similar
527 levels of genetic diversity in the sampled UK and continental population, despite indications
528 of initial bottlenecks in some sites, and (iii) preliminary evidence of selection in some genomic
529 regions.

530

531 *The population expansion of* B. hypnorum *into and across the UK*

532 For the first time, we report population genomic evidence that supports a multiple
533 colonization of the UK by *B. hypnorum*. Collated records from the BWARS database indicate
534 an initial colonization of the UK in the south, with a subsequent rapid spread north and east
535 and a later spread westwards. Our modelled scenarios generally support the colonization
536 pattern suggested by the BWARS database although some sample sites were founded from
537 multiple sites. There is a high likelihood of migration into multiple sites from continental
538 Europe across the south and north-east, combined with ongoing and rapid migration from
539 neighboring sites already colonized. In bumblebees, queens are the founders of new nests
540 and therefore drive the effective dispersal of the species (Lepais et al. 2010). Our results
541 suggest that both jump-dispersal over a longer distance outside the native range, as well as
542 diffusion dispersal, the gradual dispersion over shorter distances (Pielou 1979), are playing a
543 role in explaining the colonization pattern of *B. hypnorum* queens into the UK. A pattern of
544 multiple founders from both local and continental sources is supported further by patterns of
545 genetic differentiation. For example, the Newcastle sample clustered most closely with Le
546 Havre in a PCoA on pairwise F_{ST} (Figure 3), which again suggests an independent introduction
547 from the Western continental population rather than colonization from within the UK only.

548

549 *Evidence for loss of genetic diversity in colonizing populations*

550 Whether or not colonization is accompanied by a loss of genetic diversity is important in terms
551 of understanding both the colonization event itself, and the potential evolutionary constraints
552 and vulnerability of the newly established populations (Dlugosch and Parker 2008), especially

553 in haplodiploid species where effective population size is reduced (Lester and Selander 1979).
554 Here, we observed generally comparable proportions of polymorphic markers and
555 measurements of heterozygosity within the continental Le Havre and UK locations (Table 1).
556 This is consistent with a previous report that found no evidence for the loss of genetic
557 diversity in a single UK population (Crowther 2017). The levels of genetic diversity observed
558 here in *B. hypnorum* are similar to other long-established bumblebee populations in North
559 America based on SNP data (Lozier 2014). This is despite the fact that DIYABC analysis suggests
560 that bottleneck events have occurred during the establishment of *B.hypnorum* in the UK
561 (Figure 4). However, the likely occurrence of multiple introductions through the presumed
562 successive influx of incoming dispersing queens, the dynamic patterns of migration within
563 colonized UK sites and the rapid expansion will mitigate a loss of genetic diversity (Nei et al.
564 1975; Pannell and Charlesworth 2000; Zenger et al. 2003; Dlugosch and Parker 2008).

565

566 As we sampled only one continental location, the level of variation in genetic diversity across
567 potential source populations is unknown and so could be under-represented in our study.
568 Studies on other bumblebee species that are widespread and abundant across continental
569 Europe (*B. terrestris* and *B. pascuarum*) did not reveal population structuring on a continental
570 scale (Estoup et al 1996, Widmer and Schmid-Hempel 1999). This suggest that across
571 continental Europe there are no significant barriers to gene flow for bumblebee species with
572 ample dispersal abilities, consistent with other studies suggesting that connectivity is
573 generally high for mainland populations of bumblebees (Lozier 2014). It is therefore likely that
574 population structuring is minimal across the coast of north-western continental Europe in *B.*

575 *hypnorum*. Indeed, a very recent study did not find population structure across populations
576 of *B. hypnorum* in Belgium (Maebe et al. 2019).

577

578 Jones and Brown (2014) used an indirect approach of diploid male production as a preliminary
579 assessment of genetic diversity of *B. hypnorum* in the UK. They reported lower than expected
580 genetic diversity within the UK compared to continental European populations, in contrast to
581 our results here. However, that study focused exclusively on samples from the London area,
582 and here we also report this area as having the lowest diversity of all of the locations sampled
583 (both in terms of heterozygosity and in polymorphic sites), raising the possibility that the
584 London site may be less representative of the wider UK population.

585

586 Our finding of no major loss of genetic diversity in a successfully colonizing species is
587 consistent with the majority of cases documented (Roman and Darling 2007; Dlugosch and
588 Parker 2008). However, there are other contrasting examples, such as the colonization of
589 North America by the solitary bee *Lasioglossum leucozonium,* which was likely initialized by
590 the introduction of a singly-mated female (Zayed et al. 2007). Different patterns of loss of
591 genetic diversity through founding events are likely explicable by the severity and length of
592 periods of reduced population size, consistent with established population genetic theory.
593 This suggests that despite initial small population sizes in the UK as indicated by the detection
594 of bottlenecks in the DIYABC analysis, founding populations expanded quickly, which in
595 combination with the likely ongoing independent colonization events, minimized the loss of
596 genetic diversity. This is further supported by a very recent study using microsatellites that
597 reports no significant increases in diploid males in the UK population (Brock et al. in review).

598

599 *Evidence for selection*

600 The limitation in the identification of signatures of selection across the sampled sites using a
601 RAD-seq approach is that only a portion of the genome is screened (here 1%), so many loci
602 under selection are likely to be missed (Tiffin and Ross-Ibarra 2014; Lowry et al. 2017).
603 Further, we used the genome of a different species to infer functionality of genomic regions.
604 Although this is common practice when working with non-model species (Shafer et al. 2017),,
605 the annotation is likely to be incomplete and areas of high divergence to the reference
606 genome may fail to be annotated. The complex demographic history and rapid expansion may
607 also leave genomic signatures similar to those expected under selection (Excoffier and Ray
608 2008; Li et al. 2012). By taking a stringent approach and only highlighting those genes
609 identified from all our approaches (two F_{ST}–outlier approaches *and* genomic regions showing
610 significantly elevated or low polymorphism) three genes were identified as putatively showing
611 signatures of selection. Only one gene, encoding for the protein 'vestigial' and involved in
612 wing formation, showed significantly elevated differentiation between Le Havre and UK sites.
613 Geographic variation in wing shape and size is common in insects (Hoffmann and Shirriffs
614 2002; Kandemir et al. 2009) and the establishment of geographic clines has been documented
615 within short time scales (Gilchrist et al. 2001), which may also be the case here. Differences
616 in wing morphology may also relate to flight and dispersal abilities. Spatial sorting theory
617 predicts that dispersal ability is a trait under strong selection during range expansions
618 (Berthouly-Salazar 2012). This is explained by the expectation that those individuals with the
619 highest dispersal abilities will be spatially assorted at the expanding range front. This creates

620 a positive feedback loop as expansion continues and accelerates the speed at which new
621 areas are colonized (Berthouly-Salazar 2012).

622

623 Among UK-UK site comparisons, two further genes were highlighted as being potentially
624 under strong directional selection: the CLOCK gene, which has an important role in the
625 regulation of circadian rhythms (Darlington et al. 1998) and has been linked with adaptive
626 responses to environmental conditions across a range of taxa, including invertebrates (Tauber
627 and Kyriacou 2005; O'Malley and Banks 2008; Liedvogel et al. 2009) and serine/threonine
628 protein kinase NLK, which is associated with innate immune function and apoptosis (cell
629 death) (Mirkovic et al. 2002; Li et al. 2014). A serine/threonine protein kinase region was also
630 identified as an F_{ST} – outlier by Theodorou et al. (2018) in a recent study using a RAD-seq
631 approach in *Bombus lapidarius,* where it was suggested as a signature of an adaptive response
632 to increasing urbanization. Colonizing *B. hypnorum* suffer higher prevalence of highly virulent
633 parasites than native species (Jones and Brown 2014; Lloyd et al. unpublished data), which
634 might also be explanatory.

635

636 Further, GO analysis identified areas of significantly high polymorphism were associated with
637 biological processes or molecular functions, largely regulating transcription and gene
638 expression (Table 5). This highlights research into the role of genetic variation in regulatory
639 genes in the adaptability of *B. hypnorum* to new locations as a potentially important area for
640 further studies. The GO analysis of genes within genomic regions of low polymorphism
641 revealed a significant association with ion channel and signalling receptor activity (Table 5)

642 and the investigation of the adaptive role of these genomic regions similarly warrants further
643 research across a wider group of *Bombus* species.

644

645 *Why now?*

646 Our study has revealed a probable pattern of multiple colonization events from western
647 continental Europe of *B. hypnorum*, in addition to ongoing gene flow and spread from within
648 established UK sites. We also report evidence of some genomic areas that may show signals
649 of directional selection. Given the likely pattern of colonization, a key question is why this
650 species has been such a successful and successive colonizer now, when it has been present
651 on continental Europe for many decades prior to the turn of the 21st century. Answers to this
652 can only remain speculative but our study gives some potential insight by suggesting a pattern
653 of multiple, possibly ongoing, colonization routes with no evidence of a reduction in genetic
654 diversity. Following the colonization success framework of Facon et al. (2006) our results
655 highlight a possible role of 'migration change' where barriers to dispersal have recently and
656 relatively suddenly opened. Possible changes to anthropogenic transport routes are one
657 potential explanation. Anthropogenic rates of trade and transport are subject to continuous
658 growth (Hulme 2009) and although there have been no obvious significant changes over the
659 period in question, increased opportunities for assisted introduction may have facilitated the
660 colonization of the UK. Shifts in climatic conditions are another potential explanation, which
661 may facilitate dispersing queens reaching the UK. While temperature changes alone are
662 unlikely to be responsible, given that the UK sits well within the current climatic range of *B.*
663 *hypnorum*, shifts in wind patterns (e.g. Hu et al. 2016; Weber et al 2018) are among these
664 plausible scenarios. There is less compelling current evidence for a role of 'environmental

665 change'. While there are noted ecological differences between *B. hypnorum* and other UK
666 species, both in nesting site preferences and also both habitat and foraging associations
667 (Crowther et al. 2014), there is little evidence to date of any marked change in availability of
668 either nesting sites or habitat over the period in question. Neither is there any evidence for
669 'enemy release' occurring, from recent comparative study of *Bombus* parasite communities
670 (Jones and Brown 2014). Our preliminary results highlighting genes already suggested to be
671 involved in adaptive responses to urbanized environments in similar or other insect groups
672 lend some traction to the third 'evolutionary change' scenario as a contributing factor
673 underpinning the remarkable expansion of this species. Such adaptations to
674 anthropogenically modified habitats within the native range, have been put forward as an
675 important feature to promote range expansions (Hufbauer et al. 2012). Finally, biogeographic
676 factors could play a role: with declines of native species on an island (some UK bees have been
677 completely extirpated and many others are declining), it might be expected that novel species
678 might colonize. Further investigation into the potential drivers of this population expansion,
679 particularly focusing on migration and/or evolutionary changes, are likely to yield key insights
680 to our understanding of these recent population changes. Notably, establishing whether
681 these are ubiquitous signals across a much wider taxonomic range, or whether this particular
682 system has been successful because it is an exception to the norm will give important insights
683 into the rapid changes in distribution and abundance of species currently being witnessed.

684

ACKNOWLEDGEMENTS

We are grateful to the University of Plymouth for funding this project and to our colleagues for insightful comments on earlier drafts of this manuscript.

Declarations

Funding
Funding for this project was provided by the University of Plymouth

Consent to participate
Not applicable

Consent for publication
Not applicable

Ethics approval
Not applicable

Declaration of conflicts of interest
The authors declare no conflicts of interest.

Data Accessibility

Raw reads generated in this study will be made available on GenBank (SRA)

Code availability

A custom python script used for variant calling will be deposited on Dryad upon acceptance of the manuscript.

Author Contributions

MEK, JSE and MJFB conceived the project and secured funding; KL carried out the sampling; KL, CMB and MK generated the raw data; JVH conducted the analysis in discussion with MEK and JSE; JVH, MEK and JSE drafted the manuscript, with input from the other authors; MEK led the project.

References

Ahrens, M., Ammerpohl, O., von Schönfels, W., Kolarova, J., Bens, S., Itzel, T., Teufel, A., Herrmann, A., Brosch, M., Hinrichsen, H., Erhart, W., Egberts, J., Sipos, B., Schreiber, S., Häsler, R., Stickel, F., Becker, T., Krawczak, M., Röcken, C., Siebert, R., Schafmayer, C., Hampe, J. (2013) DNA Methylation Analysis in Nonalcoholic Fatty Liver Disease Suggests Distinct Disease-Specific and Remodeling Signatures after Bariatric Surgery. Cell Metabolism, 18, 296–302. https://doi.org/10.1016/j.cmet.2013.07.004

Alexa, A., Rahnenfuhrer, J., 2016. topGO: Enrichment Analysis for Gene Ontology.

Alexa, A., Rahnenführer, J., Lengauer, T. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics, 22, 1600–1607.

Allendorf, F.W., Lundquist, L.L. (2003) Introduction: population biology, evolution, and control of invasive species. Conservation Biology, 17, 24–30.

Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., Corrado, L., Boneschi, F.M., D'Alfonso, S. and De Bellis, G., 2016. Next generation sequencing of pooled samples: guideline for variants' filtering. Scientific reports, 6, p.33735.

Angert, A.L., Crozier, L.G., Rissler, L.J., Gilman, S.E., Tewksbury, J.J., Chunco, A.J. (2011) Do species' traits predict recent shifts at expanding range edges? Ecology Letters, 14, 677–689. https://doi.org/10.1111/j.1461-0248.2011.01620.x

Arbetman, M.P., Gleiser, G., Morales, C.L., Williams, P., Aizen, M.A. (2017) Global decline of bumblebees is phylogenetically structured and inversely related to species range size and pathogen incidence. Proceedings of the Royal Society B, 284, 20170204. https://doi.org/10.1098/rspb.2017.0204

Baker, H.G. (1965) Characteristics and modes of origin of weeds. Characteristics and modes of origin of weeds, 147-72.

Benton, T. (2006) Bumblebees, the natural history & identification of the species found in Britain, Collins, London.

Biella, P., Ćetković, A., Gogala, A., Neumayer, J., Sárospataki, M., Šima, P. and Smetana, V., 2020. North-westward range expansion of the bumblebee Bombus haematurus into Central Europe is associated with warmer winters and niche conservatism. Insect Science.

Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30,* 2114–2120.

Brock, R. E., Crowther, L. P., Wright, D. J., Richardson, D. S., Carvell, C., Taylor, M. I. & Bourke, A. F. G. No severe genetic bottleneck in a rapidly range-expanding bumblebee pollinator. In review

Bruneaux, M., Johnston, S.E., Herczeg, G., Merilä, J., Primmer, C.R. and Vasemägi, A. (2013) Molecular evolutionary and population genomic analysis of the nine‐spined stickleback using a modified restriction‐site‐associated DNA tag approach. *Molecular Ecology, 22*(3), 565-582.

BWARS (Bees, Wasps & Ants Recording Society), 2019. http://www.bwars.com. 30 January 2019.

Cabrera, A.A., Palsbøll, P.J. (2017) Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in diyabc. Molecular Ecology Resources, *17*, e94–e110. https://doi.org/10.1111/1755-0998.12696

Cameron, S.A., Lozier, J.D., Strange, J.P., Koch, J.B., Cordes, N., Solter, L.F. and Griswold, T.L., 2011. Patterns of widespread decline in North American bumble bees. Proceedings of the National Academy of Sciences, 108(2), pp.662-667.

Catchen, J., Bassham, S., Wilson, T., Currey, M., O'Brien, C., Yeates, Q., Cresko, W.A. (2013) The population structure and recent colonization history of Oregon threespine stickleback

determined using restriction-site associated DNA-sequencing. Molecular Ecology, 22, 2864–2883. https://doi.org/10.1111/mec.12330

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M. (2015) Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances, 1(5), e1400253. https://doi.org/10.1126/sciadv.1400253

Chichorro, F., Juslén, A. and Cardoso, P., 2019. A review of the relation between species traits and extinction risk. Biological Conservation, 237, pp.220-229.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., Getz, G. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology, 31, 213–219. https://doi.org/10.1038/nbt.2514

Cooper, G.M., Brudno, M., Stone, E.A., Dubchak, I., Batzoglou, S., Sidow, A. (2004) Characterization of Evolutionary Rates and Constraints in Three Mammalian Genomes. Genome Ressources, 14, 539–548. https://doi.org/10.1101/gr.2034704

Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M., Estoup, A. (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics, *30*, 1187–1189. https://doi.org/10.1093/bioinformatics/btt763

Crowther, L.P. (2017) The tree bumblebee, *Bombus hypnorum*: ecology and genetics of a naturally colonising pollinator. PhD thesis, University of East Anglia, UK.

Crowther, L.P., Hein, P.-L., Bourke, A.F. (2014) Habitat and forage associations of a naturally colonising insect pollinator, the tree bumblebee Bombus hypnorum. Plos One*,* 9, e107568.

Darlington, T.K., Wager-Smith, K., Ceriani, M.F., Staknis, D., Gekakis, N., Steeves, T.D., Weitz, C.J., Takahashi, J.S., Kay, S.A. (1998) Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. Science, 280, 1599–1603.

Dorant, Y., Benestan, L., Rougemont, Q., Normandeau, E., Boyle, B., Rochette, R. and Bernatchez, L., 2019. Comparing Pool-seq, Rapture, and GBS genotyping for inferring weak population structure: The American lobster (Homarus americanus) as a case study. Ecology and evolution, 9(11), pp.6606-6623.

De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., Vigouroux, Y. (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Molecular Ecology, 22, 1383–1399.

Dlugosch, K.M., Parker, I.M. (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Molecular Ecology, 17, 431–449.

Edwards, M., Jenner, M. (2005) Field guide to the bumblebees of Great Britain & Ireland. Ocelli Ltd.

Estoup, A. and Guillemaud, T. (2010) Reconstructing routes of invasion using genetic data: why, how and so what?. Molecular ecology, 19(19), pp.4113-4130.

Estoup, A., Lombaert, E., Marin, J.M., Guillemaud, T., Pudlo, P., Robert, C.P., Cornuet, J.M. (2012) Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. Molecular Ecology resources, 12(5), 846-855.

Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a genetic paradox of biological invasion? Annual Review of Ecology, Evolution, and Systematics, 47, 51-72.

Estoup, A., Solignac, M., Cornuet, J.M., Goudet, J. and Scholl, A., 1996. Genetic differentiation of continental and island populations of Bombus terrestris (Hymenoptera: Apidae) in Europe. Molecular Ecology, 5(1), pp.19-31.

Excoffier, L. and Ray, N. (2008) Surfing during population expansions promotes genetic revolutions and structuration. Trends in ecology & evolution, 23(7), pp.347-351.

Facon, B., Genton, B.J., Shykoff, J., Jarne, P., Estoup, A., David, P., 2006. A general eco-evolutionary framework for understanding bioinvasions. Trends in ecology & evolution, 21, 130–135. https://doi.org/10.1016/j.tree.2005.10.012

Fischer, M.C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K.K., Holderegger, R., Widmer, A. (2017) Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. BMC Genomics, 18, 69. https://doi.org/10.1186/s12864-016-3459-7

Fitzpatrick, Ú., Murray, T.E., Paxton, R.J., Breen, J., Cotton, D., Santorum, V. and Brown, M.J., 2007. Rarity and decline in bumblebees–a test of causes and correlates in the Irish fauna. Biological conservation, 136(2), pp.185-194.

Foll, M., Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics, 180, 977–993. https://doi.org/10.1534/genetics.108.092221

Fontanillas, P., Landry, C.R., Wittkopp, P.J., Russ, C., Gruber, J.D., Nusbaum, C. and Hartl, D.L. (2010) Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. Molecular Ecology, 19, 212-227.

Fraser, B.A., Künstner, A., Reznick, D.N., Dreyer, C. and Weigel, D. (2015) Population genomics of natural and experimental populations of guppies (Poecilia reticulata). Molecular Ecology, 24(2), 389-408.

Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhué, C., Estoup, A. (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. Molecular Ecology, 22, 3766–3779. https://doi.org/10.1111/mec.12360

Gilchrist, G.W., Huey, R.B. and Serra, L. (2001) Rapid evolution of wing size clines in Drosophila subobscura. In Microevolution Rate, Pattern, Process (pp. 273-286). Springer, Dordrecht

Goudet, J. (2004) hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology notes, 5, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Goulson, D., Hanley, M.E., Darvill, B., Ellis, J.S., Knight, M.E. (2005) Causes of rarity in bumblebees. Biological Conservation, 122, 1–8. https://doi.org/10.1016/j.biocon.2004.06.017

Goulson D, Kaden JC, Lepais O, Lye GC, Darvill B (2011) Population structure, dispersal and colonization history of the garden bumblebee Bombus hortorum in the Western Isles of Scotland. Conservation Genetics 12, 867-879.

Goulson, D., Lye, G., Darvill, B., 2008 Decline and conservation of bumblebees. *Annual Review of Entomology, 53*, 191-208.

Goulson, D., Williams, P. (2001) *Bombus hypnorum* (Hymenoptera: Apidae), a new British bumblebee. British Journal of Entomology and Natural History, 14(3), 129-131.

Guo, B., DeFaveri, J., Sotelo, G., Nair, A., Merilä, J. (2015) Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. BMC Biology, 13.

Guo, B., Li, Z. and Merilä, J., 2016. Population genomic evidence for adaptive differentiation in the Baltic Sea herring. Molecular ecology, 25(12), pp.2833-2852.

Guzinski, J., Ballenghien, M., Daguin-Thiebaut, C., Lévêque, L., Viard, F. (2018) Population genomics of the introduced and cultivated Pacific kelp Undaria pinnatifida: Marinas-not farms-drive regional connectivity and establishment in natural rocky reefs. Evolutionary Applications, 11(9), 1582-1597.

Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D., Kroon, H. de (2017) More than 75 percent decline over

27 years in total flying insect biomass in protected areas. Plos one, 12, e0185809. https://doi.org/10.1371/journal.pone.0185809.

Hoffmann, A.A., Shirriffs, J. (2002) Geographic variation for wing shape in Drosophila serrata. Evolution, 56, 1068–1073.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos* Genetics, 6(2): e1000862 https://doi.org/10.1371/journal.pgen.1000862.

Hu G., Lim KS, Horvitz N., Clark SJ, Reynolds DR, Sapir N, Chapman JW (2016) Mass seasonal bioflows of high-flying insect migrants. Science, 354(6319), 1584-1587

Hufbauer, R.A., Facon, B., Ravigné, V., Turgeon, J., Foucaud, J., Lee, C.E., Rey, O., Estoup, A. (2012) Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. Evolutionary Applications, 5, 89–101. https://doi.org/10.1111/j.1752-4571.2011.00211.x

Hulme, P.E., 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of applied ecology, 46(1), pp.10-18.

Jacobs, A., Hughes, M., Robinson, P., Adams, C. and Elmer, K. (2018) The genetic architecture underlying the evolution of a rare piscivorous life history form in brown trout after secondary contact and strong introgression. Genes, 9(6), p.280.

Jones CM., Brown, M.J.F. (2014) Parasites and genetic diversity in an invasive bumblebee. Journal of animal ecology, *83*,1428–1440 https://doi.org/10.1111/1365-2656.12235

Kahnt, B., Theodorou, P., Soro, A., Hollens-Kuhr, H., Kuhlmann, M., Pauw, A. and Paxton, R.J., 2018. Small and genetically highly structured populations in a long-legged bee, Rediviva longimanus, as inferred by pooled RAD-seq. BMC evolutionary biology, 18(1), p.196.

Kandemir, İ., Moradi, M.G., Özden, B., Özkan, A. (2009) Wing geometry as a tool for studying the population structure of dwarf honey bees (Apis florea Fabricius 1876) in Iran. Journal of Apicultural Research, 48, 238–246. https://doi.org/10.3896/IBRA.1.48.4.03

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Ding, L. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics, 25, 2283–2285. https://doi.org/10.1093/bioinformatics/btp373

Koboldt, D.C., Larson, D.E. and Wilson, R.K. (2013) Using VarScan 2 for germline variant calling and somatic mutation detection. Current protocols in bioinformatics, 44(1), pp.15-4.

Kofler, R., Pandey, R.V., Schlötterer, C. (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics, 27, 3435–3436.

Krehenwinkel, H., Rödder, D. and Tautz, D., 2015. Eco-genomic analysis of the poleward range expansion of the wasp spider A rgiope bruennichi shows rapid adaptation and genomic admixture. Global change biology, 21(12), 4320-4332.

Lentner, M. (1972) Elementary applied statistics. Bogden & Quigley.

Lepais, O., Darvill, B.E.N., O'connor, S., Osborne, J.L., Sanderson, R.A., Cussans, J., Goffe, L. and Goulson, D., 2010. Estimation of bumblebee queen dispersal distances using sibship reconstruction method. Molecular Ecology, 19(4), pp.819-831.

Lester, L.J., Selander, R.K. (1979) Population Genetics of Haplodiploid Insects. Genetics, *92*, 1329–1345.

Leydet, K.P., Grupstra, C.G., Coma, R., Ribes, M. and Hellberg, M.E. (2018) Host-targeted RAD-Seq reveals genetic changes in the coral Oculina patagonica associated with range expansion along the Spanish Mediterranean coast. Molecular Ecology, 27(11), 2529-2543.

Lewontin, R.C. (1965) Selection for colonizing ability, in "The Genetics of Colonizing Species"(HG Baker and GL Stebbins, Eds.).

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv preprint, ArXiv13033997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Li, H., Jakobsson, M., Li, S.E.N., Sjödin, P.E.R., Lascoux, M. (2012) Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Molecular Ecology, 21, 28–44.

Li, S.Z., Zhang, H.H., Liang, J.B., Song, Y., Jin, B.X., Xing, N.N., Fan, G.C., Du, R.L. and Zhang, X.D. (2014) Nemo-like kinase (NLK) negatively regulates NF-kappa B activity through disrupting the interaction of TAK1 with IKKβ. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1843(7), 1365-1372.

Liedvogel, M., Szulkin, M., Knowles, S.C., Wood, M.J., Sheldon, B.C. (2009) Phenotypic correlates of Clock gene variation in a wild blue tit population: evidence for a role in seasonal timing of reproduction. Molecular Ecology, 18, 2444–2456.

Lin, Y., Chen, Y., Yi, C., Fong, J.J., Kim, W., Rius, M., Zhan, A. (2017) Genetic signatures of natural selection in a model invasive ascidian. Scientific reports, 7, 44080. https://doi.org/10.1038/srep44080

Lockwood, J.L., Cassey, P., Blackburn, T. (2005) The role of propagule pressure in explaining species invasions. Trends in Ecology & Evolution, 20, 223–228.

Lotterhos, K.E., Whitlock, M.C. (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Molecular Ecology, 23, 2178–2192. https://doi.org/10.1111/mec.12725

Lowry, D.B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K., Antolin, M.F., Storfer, A. (2017) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology resources, 17*,* 142–152. https://doi.org/10.1111/1755-0998.12635

Lozier, J.D. (2014) Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. Molecular Ecology, 23, 788–801.

Maebe, K., Karise, R., Meeus, I., Mänd, M. and Smagghe, G. (2019) Pattern of population structuring between Belgian and Estonian bumblebees. Scientific reports, 9(1), pp.1-8.

Mann, H.B. and Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, pp.50-60.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. Cancer Research, 27, 209–220.

McKinney, M.L., Lockwood, J.L. (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. Trends in Ecology & Evolution, *14,* 450–453.

Mirkovic, I., Charish, K., Gorski, S.M., McKnight, K. and Verheyen, E.M. (2002) Drosophila nemo is an essential gene involved in the regulation of programmed cell death. Mechanisms of development, 119(1), 9-20.

Momigliano, P., Jokinen, H., Fraimout, A., Florin, A.B., Norkko, A. and Merilä, J., 2017. Extraordinarily rapid speciation in a marine fish. Proceedings of the National Academy of Sciences, 114(23), pp.6074-6079.

Moran, E.V., Alexander, J.M. (2014) Evolutionary responses to global change: lessons from invasive species. Ecology Letters, 17, 637–649.

Nei, M., Maruyama, T., Chakraborty, R. (1975) The bottleneck effect and genetic variability in populations. Evolution, 29, 1–10.

Nicholls, J.A., Double, M.C., Rowell, D.M., Magrath, R.D. (2000) The evolution of cooperative and pair breeding in thornbills Acanthiza (Pardalotidae). Journal of Avian Biology, 31, 165–176.

O'Malley, K.G., Banks, M.A. (2008) A latitudinal cline in the Chinook salmon (Oncorhynchus tshawytscha) Clock gene: evidence for selection on PolyQ length variants. Proceedings of the Royal Society of London. Series B: Biological Sciences**,** 275, 2813–2821.

Oleksyk, T.K., Smith, M.W., O'Brien, S.J. (2010) Genome-wide scans for footprints of natural selection. Philosophical Transactions of the Royal Society B: Biological Sciences, 365*,* 185–205.

Ollerton, J., Erenler, H., Edwards, M., Crockett, R. (2014) Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes. Science, 346, 1360–1362.

Pannell, J.R. and Charlesworth, B. (2000) Effects of metapopulation processes on measures of genetic diversity. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 355(1404), pp.1851-1864.

Paxton, R.J., Thorén, P.A., Estoup, A., Tengö, J. (2001) Queen–worker conflict over male production and the sex ratio in a facultatively polyandrous bumblebee, Bombus hypnorum: the consequences of nest usurpation*.* Molecular Ecology, 10, 2489–2498.

Pielou, E. C. (1979) Biogeography. John Wiley & Sons, New York, New York. 351 pp.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., Kunin, W.E. (2010) Global pollinator declines: trends, impacts and drivers. Trends in ecology & evolution, 25, 345–353. https://doi.org/10.1016/j.tree.2010.01.007

Prŷs-Jones, O.E., Kristjánsson, K., Ólafsson, E. (2016) Hitchhiking with the Vikings? The anthropogenic bumblebee fauna of Iceland – past and present. Journal of Natural History, 50, 2895–2916. https://doi.org/10.1080/00222933.2016.1234655

Puckett, E.E., Kristensen, T.V., Wilton, C.M., Lyda, S.B., Noyce, K.V., Holahan, P.M., Leslie Jr, D.M., Beringer, J., Belant, J.L., White Jr, D. and Eggert, L.S., 2014. Influence of drift and admixture on population structure of American black bears (Ursus americanus) in the Central Interior Highlands, USA, 50 years after translocation. Molecular Ecology, 23(10), pp.2414-2427.

Purfield, D.C., McParland, S., Wall, E. and Berry, D.P. (2017) The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One, 12(5), p.e0176780.

Purvis, A., 2008. Phylogenetic approaches to the study of extinction. Annual Review of Ecology, Evolution, and Systematics, 39, pp.301-319.

Rademacher, N., Wrobel, T.J., Rossoni, A.W., Kurz, S., Bräutigam, A., Weber, A.P.M., Eisenhut, M. (2017) Transcriptional response of the extremophile red alga Cyanidioschyzon merolae to changes in CO2 concentrations. Journal of plant physiology, Light driven reactions, 217, 49–56. https://doi.org/10.1016/j.jplph.2017.06.014

Rane, R.V., Rako, L., Kapun, M., Lee, S.F. and Hoffmann, A.A. (2015) Genomic evidence for role of inversion 3 RP of Drosophila melanogaster in facilitating climate change adaptation. Molecular Ecology, 24(10), 2423-2432.

Rasmont, P. (1988) Monographie écologique et zoogéographique des bourdons de France et de Belgique. Faculté des. Sciences Agronomiques de PEtat (Gembioux), *510*.

Rasmont, P., Franzén, M., Lecocq, T., Harpke, A., Roberts, S.P., Biesmeijer, J.C., Castro, L., Cederberg, B., Dvorak, L., Fitzpatrick, Ú. (2015) Climatic risk and distribution atlas of European bumblebees. Pensoft Publishers.

Raymond, M., Rousset, F. (1995) GENEPOP on the Web (Version 3.4).

Richardson, L., McFarland, K., Zahendra, S., Hardy, S. (2019) Bumble bee (*Bombus*) distribution and diversity in Vermont, USA: a century of change. Journal of Insect Conservation, 23(1), 45-62. Doi: 10.1007/s10841-018-0113-5

Robertson, A. (1952) The Effect of Inbreeding on the Variation Due to Recessive Genes. *Genetics, 37,* 189–207.

Roesti, M., Salzburger, W. and Berner, D., 2012. Uninformative polymorphisms bias genome scans for signatures of selection. BMC evolutionary biology, 12(1), p.94.

Roman, J., Darling, J.A. (2007) Paradox lost: genetic diversity and the success of aquatic invasions. Trends in ecology & evolution, 22, 454–464.

Sadd, B.M., Barribeau, S.M., Bloch, G., De Graaf, D.C., Dearden, P., Elsik, C.G., Gadau, J., Grimmelikhuijzen, C.J., Hasselmann, M., Lozier, J.D. (2015) The genomes of two key bumblebee species with primitive eusocial organization. Genome Biolgy, 16*,* 76.

Sakai, A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A., Baughman, S., Cabin, R.J., Cohen, J.E., Ellstrand, N.C. (2001) The population biology of invasive species. Annual review of ecology and systematics, 32, 305–332.

Sax, D. F., Brown, J. H. (2000). The paradox of invasion. Global Ecology and Biogeography, 9, 363-371.

Sax, D.F., Stachowicz, J.J., Brown, J.H., Bruno, J.F., Dawson, M.N., Gaines, S.D., Grosberg, R.K., Hastings, A., Holt, R.D., Mayfield, M.M. (2007) Ecological and evolutionary insights from species invasions. Trends in ecology & evolution, 22*,* 465–471.

Shafer, A.B., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W. and Wolf, J.B. (2017) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution, 8(8), pp.907-917.

Speight, M.C. and MCD, S., 1974. Bombus lapponicus, Parasyrphus lineola and Phaonia exoleta: insects new to Ireland.

Stillfried, M., Fickel, J., Börner, K., Wittstatt, U., Heddergott, M., Ortmann, S., Kramer-Schadt, S. and Frantz, A.C., 2017. Do cities represent sources, sinks or isolated islands for urban wild boar population structure?. Journal of Applied Ecology, 54(1), pp.272-281.

Storey, J.D., Bass, A.J., Dabney, A., Robinson, D. (2015) qvalue: Q-value estimation for false discovery rate control. R package version 2.0. 0. Available Github Comjdstoreyqvalue Accessed April 14, 2017.

Tauber, E., Kyriacou, C.P. (2005) Molecular Evolution and Population Genetics of Circadian Clock Genes, in: Young, M.W. (Ed.), Methods in Enzymology, Circadian Rhythms. Academic Press, 797–817. https://doi.org/10.1016/S0076-6879(05)93042-5

Team, R.C. (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

Theodorou, P., Radzeviciute, R., Kahnt, B., Soro, A., Grosse, I., Paxton, R.J. (2018) Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidariu*s L.). Proceedings of the Royal Society B, 285(1877), 20172806.

Tiffin, P., Ross-Ibarra, J. (2014) Advances and limits of using population genetics to understand local adaptation. Trends in ecology & evolution, 29, 673–680. https://doi.org/10.1016/j.tree.2014.10.004

Vandepitte, K., De Meyer, T., Helsen, K., Van Acker, K., Roldán-Ruiz, I., Mergeay, J., Honnay, O. (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. Molecular Ecology, 23, 2157–2164.

van den Oord, E.J. (2008) Controlling false discoveries in genetic studies. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 147*,* 637–644.

van der Wal, R., Anderson, H., Robinson, A., Sharma, N., Mellish, C., Roberts, S., Darvill, B., Siddharthan, A. (2015) Mapping species distributions: A comparison of skilled naturalist and lay citizen science recording. Ambio, 44*,* 584–600.

Vazquez, D.P. (2006) Exploring the relationship between nichie breadth and invasion success. In Conceptual ecology and invasion biology: reciprocal approaches to nature (307-322). Springer, Dordrecht.

Weber, J., Gotzens, F., Witthaut, D. (2018) Impact of strong climate change on the statistics of wind power generation in Europe. Energy Procedia. 153. 22-28. 10.1016/j.egypro.2018.10.004.

Widmer, A. and Schmid-Hempel, P., 1999. The population genetic structure of a large temperate pollinator species, Bombus pascuorum (Scopoli)(Hymenoptera: Apidae). Molecular Ecology, 8(3), pp.387-398.

Williams, P. (2005) Does specialization explain rarity and decline among British bumblebees? A response to Goulson et al. Biological Conservation, 122, 33–43. https://doi.org/10.1016/j.biocon.2004.06.019

Williams, P.H. (1991) The bumble bees of the Kashmir Himalaya (*Hymenoptera: Apidae, Bombini*). British Museum (Natural History).

Williams, P.H., Osborne, J.L. (2009) Bumblebee vulnerability and conservation world-wide. Apidologie, 40, 367–387.

Zayed, A., Constantin, Ş.A. and Packer, L. (2007) Successful biological invasion despite a severe genetic load. *PLoS One*, *2*(9), p.e868.

Zenger, K.R., Richardson, B.J. and Vachot-Griffin, A.M. (2003) A rapid population expansion retains genetic diversity within European rabbits in Australia. Molecular Ecology, 12(3), pp.789-794.

Ziemann, M. (2016) Accuracy, speed and error tolerance of short DNA sequence aligners. BioRxiv, 053686.

Tables

Table 1: *Summary of RAD-seq data, including basic measures of genetic diversity.*

Population	raw reads	uniquely mapped reads	bp covered[a]	number of SNPs (polymorph within populations)	proportion of polymorphic sites[b]	Allelic richness[b]	Average heterozygosity[c]
Cardiff	1,832,315	1,306,780	655,120	2,139	0.012	1.22	0.15
Le Havre	2,350,459	1,726,971	323,767	2,251	0.014	1.26	0.20
Hull	2,394,796	1,508,923	593,141	2,384	0.008	1.16	0.12
London	1,159,972	619,972	245,067	851	0.006	1.11	0.12
Newcastle 1	2,788,910	1,781,942	953,719	3,695	0.012	1.23	0.18
Newcastle 2	2,466,615	2,012,501	322,409	1,929	0.011	1.21	0.20
Plymouth	714,223	397,903	152,225	464	0.01	1.18	0.13
Southampton	1,805,902	1,225,225	587,552	1,995	0.009	1.17	0.16
All UK populations	13,162,733	8,853,246	2,443,350	11,845	0.048	1.46	0.16
All populations	15,513,192	10,580,217	2,469,636	12,823	0.054	2	0.17

[a] all bps with a coverage 50-500 within pop and <= 3,000 across pops
[b] across sites covered in all pops (1,886 polymorphic and non-polymorphic sites; 101 polymorphic sites)
[c] across all SNPs found in the population

Table 2: *Population pairwise F_{ST} –values (bottom) and confidence intervals generated across 1000 bootstrap cycles (top) across 101 SNPs covered in all populations.*

	Cardiff	**Hull**	**Le Havre**	**London**	**Newcastle 1**	**Newcastle 2**	**Plymouth**	**Southampton**
Cardiff	0.000	0.020,0.066	0.014, 0.039	0.012, 0.040	0.011, 0.040	0.014, 0.044	0.019, 0.056	0.010, 0.038
Hull	0.040	0.000	0.010, 0.029	0.009,0.066	0.012, 0.040	0.010, 0.030	0.010, 0.027	0.009, 0.043
Le Havre	0.026	0.019	0.000	0.006, 0.035	0.006, 0.015	0.003, 0.007	0.010, 0.034	0.007, 0.031
London	0.025	0.035	0.018	0.000	0.007, 0.025	0.006, 0.035	0.009, 0.057	0.006, 0.040
Newcastle 1	0.024	0.025	0.010	0.014	0.000	0.006, 0.015	0.012, 0.036	0.006, 0.025
Newcastle 2	0.028	0.018	0.005	0.018	0.010	0.000	0.009, 0.034	0.007, 0.030
Plymouth	0.036	0.018	0.020	0.032	0.023	0.019	0.000	0.009, 0.027
Southampton	0.023	0.023	0.017	0.018	0.014	0.017	0.017	0.000

Table 3: *relative probabilities and confidence intervals for the top two scenarios in the DIYABC analysis for each test site*

test site	most likely source of colonization	relative probability [CI]	second most likely scenario	relative probability [CI]
London	Le Havre bottlenecked	0.92 [0.75,1.00]	Hull bottlenecked	0.0660 [0.00,0.24]
Cardiff	Combined colonization Hull-Newcastle bottlenecked	0.5811 [0.14,1.00]	Newcastle bottlenecked	0.4189 [0.00,1.00]
Hull	London bottlenecked	0.5463 [0.35,0.74]	Le Havre	0.4537 [0.26,0.65]
Southampton	Le Havre	0.6065 [0.13,1.00]	Combined colonization Le Havre-Hull	0.3917 [0.00,0.86]
Newcastle	Combined colonization Le Havre-Southampton bottlenecked	1.0000 [1.00,1.00]	NA	0.0000 [0.00,0.00]
Plymouth	London bottlenecked	0.9661 [0.92,1.00]	Le Havre	0.0339 [0.00,0.08]

Table 4: *Genes within genomic regions of significantly elevated differentiation for UK-UK and UK-Le Havre comparisons. Genes that were also identified in areas of significantly high polymorphism are highlighted in bold.*

GeneID	description	chromosome	Exon count	comparison	function
100651612	**protein vestigial**	**LG B12**	**6**	**Le Havre – UK**	**wing formation/ wing specific gene expression**
100650446	**circadian locomotor output cycles protein kaput**	**LG B09**	**12**	**UK populations**	**circadian clock**
100651912	**serine/threonine-protein kinase NLK (nemo-like kinase)**	**LG B03**	**11**	**UK populations**	**involved in signal transduction pathways that regulate growth factor response, proliferation, apoptosis and innate immune response**
110119491	uncharacterized LOC110119491	LG B08	2	UK populations	Unknown

Table 5: *Gene ontology enrichment analysis of Molecular Function (MF) and Biological Processes (BP) for all regions showing significantly elevated/reduced polymorphism across all populations. Results are shown where at least two of the three algorithms used ('classic', 'elim' and 'weight') supported significance for regions with significantly elevated (A) and reduced (B) polymorphism. Significant p-values are highlighted in bold.*

	GO.ID	Term	Annotated	Significant	Expected	classic	elim	weight
A)								
BP	GO:0050794	regulation of cellular process	373	118	86	**0.000002**	**0.025080**	**0.022640**
	GO:0050789	regulation of biological process	381	119	87.85	**0.000003**	**0.034850**	0.376710
	GO:0065007	biological regulation	394	121	90.84	**0.000008**	**0.048680**	0.809450
	GO:0007166	cell surface receptor signaling pathway	37	19	8.53	**0.000130**	**0.000130**	**0.000500**
	GO:0010468	regulation of gene expression	151	53	34.82	**0.000200**	0.458540	**0.000200**
	GO:0006355	regulation of transcription, DNA-templated	144	51	33.2	**0.000210**	**0.000210**	0.162160
	GO:0032502	developmental process	27	15	6.23	**0.000220**	0.130260	**0.000220**
	GO:0007154	cell communication	230	73	53.03	**0.000440**	**0.021020**	**0.036350**
	GO:0023052	signaling	229	72	52.8	**0.000680**	**0.029120**	**0.049620**
	GO:0007275	multicellular organism development	21	12	4.84	**0.000700**	**0.000700**	1.000000
	GO:0007165	signal transduction	221	69	50.96	**0.001150**	**0.045140**	0.540420
MF	GO:0005515	protein binding	500	137	112.33	0.001060	**0.001100**	**0.002110**
	GO:0003700	DNA-binding transcription factor activity	94	34	21.12	0.001300	**0.001300**	**0.001300**
	GO:0043565	sequence-specific DNA binding	110	37	24.71	0.003610	**0.003600**	**0.003610**
	GO:0008270	zinc ion binding	96	29	21.57	0.043440	**0.043400**	**0.043440**
B)								
MF								
	GO:0004888	transmembrane signaling receptor activity	37	17	8.28	0.000940	**0.000940**	**0.000940**
	GO:0005230	extracellular ligand-gated ion channel activity	16	8	3.58	0.013090	**0.013090**	**0.013090**

Figure 1: *Reconstruction of the UK* Bombus hypnorum *colonization from BWARS records of year-wise sightings 2004-2012.*

Figure 2: *schematic representation of steps 1 and 2 of the hierarchical DIYABC analysis done for each UK population independently*

Figure 3: *Principal coordinate analysis on population pairwise F_{ST} values for A) all SNPs covered in all populations (101); B) all SNPs (2166) covered in at least six populations (75%) and C) all SNPs (3475) covered in at least four populations (50%); sample sites are abbreviated as follows: Le Havre (H), Southampton (S), London (L), Hull (Hu), Newcastle (N), Cardiff (C) and Plymouth (P)*

Figure 4: *A) Geographic representation of the DIYABC consensus model B) Schematic consensus model, derived by combining the scenarios with highest likelihood for each UK population. Three independent colonization events from the continental reference population, Le Havre (H), one to Southampton (S) one to London and one to Newcastle (N) are apparent. Further, founders from close UK populations (London (L), Hull (Hu), Southampton (S) and Newcastle (N)) are also involved in the establishment of populations. Note that, whilst the succession of events implies a certain timeline, time is not explicitly evaluated here and not represented by the length of connecting lines.*
**representing the continental European reference population*

Figure 5: *Distribution of average F_{ST} -values across linkage groups across all pairwise comparisons within the UK (points) and between the UK and Le Havre (crosses) respectively. Genomic regions showing significantly high F_{ST} -values based on permutation tests are highlighted in red and regions with significantly low F_{ST} -values are highlighted in blue.*

Figure 6: *Distribution of polymorphism across linkage groups. Genomic regions showing significantly elevated polymorphism based on permutation tests are highlighted in red and areas of significantly reduced polymorphism in blue. Note the different y-axis scale in the NW linkage group.*

685