Received: 16 May 2022 | Revised: 28 August 2023 | Accepted: 7 September 2023

DOI: 10.1111/mec.17138

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Rapid evolution of novel biotic interactions in the UK Brown Argus butterfly uses genomic variation from across its geographical range

Maaike de Jong[1] | Alexandra Jansen van Rensburg[1,2] | Samuel Whiteford[3] | Carl J. Yung[3] | Mark Beaumont[1] | Chris Jiggins[4] | Jon Bridle[1,2]

[1]School of Biological Sciences, University of Bristol, Bristol, UK

[2]Department of Genetics, Evolution and Environment, University College London, London, UK

[3]Institute of Integrative Biology, University of Liverpool, Liverpool, UK

[4]Department of Genetics, University of Cambridge, Cambridge, UK

Correspondence
Jon Bridle, Department of Genetics, Evolution and Environment, University College London, UK.
Email: j.bridle@ucl.ac.uk

Present address
Maaike de Jong, Netherlands eScience Center, Amsterdam, The Netherlands
Alexandra Jansen van Rensburg and Jon Bridle, Department of Genetics, Evolution and Environment, University College London, London, UK

Funding information
Marie Curie Intra-European Fellowship, Grant/Award Number: 332138; Swiss National Science Foundation, Grant/Award Number: P2ZHP2_178363

Handling Editor: Christian Schlötterer

Abstract

Understanding the rate and extent to which populations can adapt to novel environments at their ecological margins is fundamental to predicting the persistence of biological communities during ongoing and rapid global change. Recent range expansion in response to climate change in the UK butterfly *Aricia agestis* is associated with the evolution of novel interactions with a larval food plant, and the loss of its ability to use an ancestral host species. Using ddRAD analysis of 61,210 variable SNPs from 261 females from throughout the UK range of this species, we identify genomic regions at multiple chromosomes that are associated with evolutionary responses, and their association with demographic history and ecological variation. Gene flow appears widespread throughout the range, despite the apparently fragmented nature of the habitats used by this species. Patterns of haplotype variation between selected and neutral genomic regions suggest that evolution associated with climate adaptation is polygenic, resulting from the independent spread of alleles throughout the established range of this species, rather than the colonization of pre-adapted genotypes from coastal populations. These data suggest that rapid responses to climate change do not depend on the availability of pre-adapted genotypes. Instead, the evolution of novel forms of biotic interaction in *A. agestis* has occurred during range expansion, through the assembly of novel genotypes from alleles from multiple localities.

KEYWORDS
adaptation, climate change, contemporary evolution, ecological genetics, population genetics - empirical, species interactions

1 | INTRODUCTION

Predicting population and community persistence in the face of a changing and more variable climate remains an urgent priority for biologists (Bridle & van Rensburg, 2020). A critical unknown is when and to what extent evolutionary responses will buffer the effects of climate change on ecological communities, or will allow species to shift their ranges to track changes in suitable climate (Angert

Maaike de Jong and Alexandra Jansen van Rensburg contributed equally to this study.

et al., 2020; Bridle & Hoffmann, 2022; Hoffmann & Sgrò, 2011). Of particular interest is how specialist biotic interactions between species (e.g. parasites and hosts, herbivores and food plants) will limit the availability of suitable habitat, so preventing range shifts due to abiotic change (Chen et al., 2011), as well as how these interactions will change as species encounter novel environmental regimes (Bridle & Hoffmann, 2022; Nadeau et al., 2017; O'Brien et al., 2022). Addressing these issues is relevant not only to understanding responses to climate change but also to the colonization of novel habitats, and to the persistence of functioning ecological communities in the face of local and global biodiversity loss (Bridle & Hoffmann, 2022).

Population genetic models of adaptation at ecological margins predict that steep or patchy ecological gradients prevent species from tracking changing climate, leading to local (and eventually global) extinction (Bridle et al., 2019; Polechová & Barton, 2015). Other models of evolution at range margins predict that adaptive shifts are more likely if populations have been exposed to variable environments in time and space, because this maintains genetic variation across the species' geographical range (Kopp & Matuszewski, 2014). In support of this, empirical studies suggest that almost 75% species with specialist biotic interactions have failed to shift their ranges to match suitable climate (Hill et al., 2011; Parmesan, 2006). By contrast many generalist species have shifted their distributions to track available habitat.

A key question is whether rapid evolution at ecological margins depends on genotypes that are already present elsewhere in a species' geographical range. For example, under global climate warming, alleles at the equatorial part of a species' range may support adaptation at poleward margins, provided gene flow is sufficiently widespread. For example, such widespread gene flow among marine populations may explain repeated (and convergent) radiations of sticklebacks into freshwater lakes from their marine ancestors (Jones et al., 2012). If such pre-existing ('standing') allelic variation is necessary for rapid evolution, population persistence will depend either on maintaining large populations throughout the species' geographical range, or on translocations from appropriate environments where dispersal among populations is limited (Bridle et al., 2009; Hoffmann & Sgrò, 2018). By contrast, if rapid evolution depends on novel mutations in situ, a population's current size will be a better predictor of its adaptive potential, rather than its connections with other populations, or its historical size.

Many Lepidoptera species are associated with particular host plant species, particularly at the larval stage, due to plant defence against herbivory, and the particular microclimates that host plants provide as oviposition sites (Jaenike, 1990; Stewart et al., 2021). Such specialist biotic interactions slow or prevent range expansion into areas that may be climatically suitable, but where the preferred host plant is rare. In the UK, more than 90% of Lepidoptera species that are habitat specialists have contracted their ranges (Warren et al., 2001). Limits to habitat availability caused by specialist host plant interactions are also associated with lags in climate responses (Chen et al., 2011), with habitat availability explaining 25% of the variation in range expansion rates, even accounting for other factors (Platts et al., 2019). Such effects of habitat availability suggest that range shifts in specialists in response to climate change depend on the evolution of novel biotic interactions at ecological margins. Habitat specialists that have tracked changing climates therefore, provide an exceptional opportunity to understand the evolutionary responses demanded by ecological gradients where they are made locally steep by particular biotic interactions (Bridle et al., 2019; Bridle & Hoffmann, 2022; O'Brien et al., 2022).

The Brown Argus butterfly, *Aricia agestis* (Polyommatinae: Lycaenidae), is a habitat specialist which has approximately doubled its geographic range in the UK since 1970–1982 (Asher et al., 2001) in response to climate change (Bodsworth, 2002; Pateman et al., 2012; Thomas et al., 2001). In mainland Europe, annual plants from the family *Geraniaceae* (*Geranium* and *Erodium*) are its predominant larval hosts (Asher et al., 2001; Tolman, 1997), although host plant use in southern Europe is less well known, partly due to the *Aricia* being found as many ecotypes and subspecies in this area. For most of the 20th century in the UK, however, *A. agestis* only used *Geraniaceae* as a host plant in a few long-established *A. agestis* populations found in southern and eastern coastal sand dunes (Heath et al., 1984). Elsewhere, *A. agestis* was limited to habitat where the rockrose host plant (Cistaceae; *Helianthemum nummularium*) is abundant. Rockrose grows only on chalk or limestone soils, and its perennial growth form probably provides a more predictable larval microclimate and food supply for *A. agestis*, regardless of annual temperature variation (Stewart et al., 2022). By contrast, *Geranium* annuals are four to seven times more sparsely distributed across the landscape and are more affected by seasonal climate, performing well during wet and warm springs and summers, but becoming rare and lower quality when springs are dry and summers are hot (Pateman et al., 2012; Stewart et al., 2021, 2022).

These data suggest that for most of the 20th century, *A. agestis* could persist in the UK on *Geraniaceae* host plants only where south-facing sand dunes provided locally elevated but relatively moist microclimates for rapid larval growth. However, increasing spring and summer temperatures since the 1990s have made non-coastal *Geranium* populations available for sustained occupation by *A. agestis* for the first time, leading to rapid range expansion northwards, into areas where rockrose host plants are rare (Pateman et al., 2012; Figure 1).

Studies of female oviposition preference and genotyping at AFLP markers demonstrate that range expansion into higher latitude *Geranium* sites has been associated with evolutionary change (Buckley et al., 2012). Newly colonised areas show an increased preference of mothers to oviposit on the widespread *Geranium molle* (Bridle et al., 2014; Thomas et al., 2001), as well as reduced variation in female oviposition preference, increased dispersal ability and less consistent preference for the locally most common host plant species (Bridle et al., 2014). In addition, field transplants of individual females onto patches of naturally-growing host plants demonstrate that, although females from rockrose-dominated (established) sites in the south of the range oviposit on both rockrose and *Geranium* plants, females from *Geranium*-dominated (newly colonised) sites north of the range will only oviposit on *Geranium* plants (Buckley &

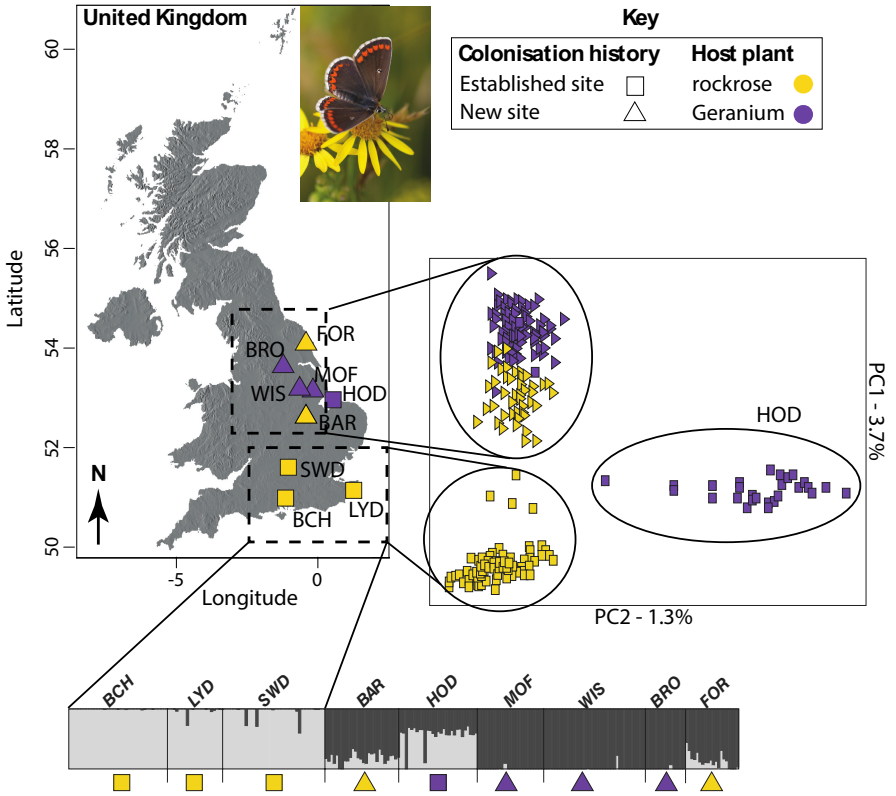

FIGURE 1 Population structure of *Aricia agestis* across its range in the UK. Site information and locations are given in Table 1. Colonisation history and the dominant host plant at each site is represented here with shapes and colours. Note that the plot of genetic variance on a PCA has been rotated to reflect the latitudinal gradient. PC1 explained 3.6 % of the genetic variance and separates the northern and southern populations. PC2 explained 1.3 % of the variation and separates HOD from the rest of the samples. The bottom panel shows the proportion assignment of each individual to one of two genetic clusters as estimated by fast Structure. Each vertical bar represents an individual, with populations ordered from South to North. Photo credit: Callum Macgregor.

TABLE 1 Sample sites and locations.

Site	Population	*N*	Lat	Long	Colonisation history	Host plant	H_{exp} (SD)	F_{ST}	
Beacon Hill	BCH	38 (39)	50.998	−1.136	Established	Cistaceae	0.178 (0.16)	*Overall*	0.031
Lydden	LYD	20 (23)	51.154	1.269	Established	Cistaceae	0.176 (0.17)	*Est* versus *New*	0.026
Swyncombe	SWD	38 (40)	51.618	−1.039	Established	Cistaceae	0.178 (0.16)	*Cis* versus *Ger*	0.018
Barnack	BAR	28 (30)	52.629	−0.414	New	Cistaceae	0.177 (0.17)	*Within Est*[a]	0.016
Holme Dunes	HOD	29 (34)	52.973	0.543	Established	*Geraniaceae*	0.178 (0.16)	*Within New*	0.014
Moor Farm	MOF	25 (26)	53.156	−0.169	New	*Geraniaceae*	0.173 (0.17)		
Whisby	WIS	38 (45)	53.190	−0.640	New	*Geraniaceae*	0.173 (0.17)		
Brockadale	BRO	15 (19)	53.645	−1.219	New	*Geraniaceae*	0.171 (0.18)		
Fordon	FOR	20 (20)	54.162	−0.394	New	Cistaceae	0.174 (0.18)		

Note: Characteristics of the nine *Aricia agestis* sites sampled across the UK, ordered from south to north. The number of individuals (*N*) included in the final analysed data set is shown, with the number of sequenced individuals in brackets. Sites were defined as established (occupied since 1970–1982) or new (occupied since 1995–1997), and as *Geraniaceae* or *Cistaceae* based on the dominant host plant at that site. We report expected heterozygosity (H_{exp}; see Figure S2) as the mean and standard deviation (SD) estimated across all neutral SNPs. Nei's F_{ST} was estimated for between all nine populations (*Overall*), and between the colonisation history (*Est* vs. *New*) and host plant (*Cis* vs. *Ger*) groups, as well as between populations within the established (*within Est*) and new sites (*within New*). HOD was excluded from the *within Est* comparison given its geographic distance from the established South.

[a]Excludes HOD.

Bridle, 2014). This suggests that range expansion driven by climate adaptation in *A. agestis* has been associated with a narrowing of oviposition preference (albeit to exploit a more widely distributed host plant), involving the loss of its ability to use its ancestral UK host plants in the newer parts of its range.

The example of the UK Brown Argus suggests that a climate-driven range shift by a habitat specialist has required an evolutionary change in species' interactions. In this case, the evolution of host plant use has effectively smoothed a steep and patchy ecological gradient at the species' poleward margin, allowing access to regions that have recently become climatically suitable, even though they lack the host plants typically used by long-established populations, especially in hot and dry years (O'Brien et al., 2022). This system therefore provides an opportunity to understand the ecology and genetics of rapid

adaptation, and to assess the likelihood and likely context of such adaptation in other species and circumstances, especially where species depend on particular interactions with other species. In *A. agestis*, a key question is whether its expansion involved the colonization of novel areas by females from coastal UK populations that already used only *Geraniaceae* food plants. Alternatively, did the observed shift in species' interactions involve the creation of novel genotypes in situ, either from new mutations arising locally or from selection on standing genetic variation already found in southern UK populations that use both rockrose and *Geraniaceae* species (Buckley & Bridle, 2014)?

In this paper, we use genome-wide SNP markers to assess the distribution of genetic variation across the UK and to test the genetic basis of adaptation at the newly colonised sites. Specifically, we: (1) test for reduced genomic variation associated with range expansion, suggesting selective sweeps at the range margin for specialization on *Geranium* host plants; (2) identify regions of the genome under selection, and their genomic distribution and likely function; and (3) determine whether evolution during range expansion has occurred through colonization of existing genotypes that use only *Geranium* host plants, or through the assembly of novel genotypes from alleles sourced from throughout the geographical range.

2 | MATERIALS AND METHODS

2.1 | Study system and sample collection

Nine *Aricia agestis* populations (276 individuals, 19–45 sampled per site) were sampled in the summers of 2013 and 2014 across most of their latitudinal distribution in the UK. Populations were chosen to include long established (present since 1970–1982) and newly colonised sites (since 1995–1997; Thomas et al., 2001), and sites were classified as either dominated by *Geraniaceae* (which includes *G. molle*, *G. dissectum* and *Erodium cicutarium*) or Cistaceae (rockrose *Helianthemum nummularium*; Table 1). Sampling effort was focussed (even in the relatively lower density sites of central and northern England), so that at least 15 individuals were genotyped per site, based on past estimates of F_{ST} that suggested high levels of gene flow among sites in the UK, and low levels of systematic variation in genetic variation within sites, as confirmed by the genomic analyses below.

2.2 | Generating genome-wide markers for population genetics

DNA was extracted from the head and half of the thorax each of 276 individuals using a Qiagen DNeasy Blood and Tissue kit. DNA was eluted in EB buffer and quantified using Qubit 2.0 fluorometer with the DNA BR assay kit (Life Technologies).

Genome-wide markers were generated using a modified double-digest restriction-associated DNA (ddRAD) protocol (Peterson et al., 2012). Briefly, genomic DNA of each individual was digested with *PstI* and *EcoRI* restriction enzymes. In total, 276 individuals from nine populations were sequenced across six ddRAD libraries. Individuals from a single population were sequenced in four to six different libraries to avoid confounding population structure with differences in library preparation and sequencing between libraries. Each library comprised 48 individuals uniquely identified using a 6-bp DNA barcode. Libraries were sequenced on an Illumina HiSeq 2000 instrument to generate 150-bp paired-end sequences from insert sizes of 300–450 bp.

2.3 | Bioinformatic analysis

Raw data were demultiplexed based on individual barcodes using ipyRAD v. 0.7.28 (Eaton, 2014) and adapters were removed using Trimmomatic v. 0.36 (Bolger et al., 2014). A long-read-based reference genome has been assembled by the Sanger Institute and is available under accession number GCA_905147365.1 from the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov). To call variants, we first mapped the demultiplexed data to the reference genome using BWA-mem (Li, 2013). Next, we called variants using mapped read pairs with a PHRED-scaled mapping quality higher than 20. We used the SAMtools v.1.8 (Li et al., 2009) *mpileup* and BCFtools v. 1.8 (Li, 2011) *call* functions to call variants simultaneously across all individuals. Raw individual variant calls were output as a VCF and filtered before downstream analyses using VCFtools v.0.017 (Danecek et al., 2011), BCFtools v. 1.8 and PLINK 2.0 (Purcell et al., 2007).

The following filters were applied: (1) We retained only variants called with a genotype likelihood (QUAL) PHRED score of more than 20. (2) A minimum mean depth filter of 6× was applied to reduce spurious heterozygote calls. We chose this threshold because it is the minimum amount of data needed to call heterozygote loci. However, the final data set comprised loci and individuals with much higher mean depth (see Figure S1), with a mean of 235× and a median of 215×, a figure that was quite consistent across sites. (3) A maximum mean depth of the mean plus twice the standard deviation (646×) was applied to remove duplicate loci. (4) Kinship coefficients were estimated between individuals within each population using the KING method (Manichaikul et al., 2010). Individuals were removed to exclude any second-order or higher relatives ($\varphi > .05$). (5) Individuals with a genotyping rate of less than 60% were removed. (6) To include loci that were evenly genotyped across all populations, we excluded loci genotyped in less than 50% of individuals within each population. (7) We allowed a global minimum minor allele frequency (MAF) of 1%. The final data set comprised 251 individuals genotyped at 61,210 variants.

2.3.1 | Changes in genomic variation associated with range expansion

Population structure and genetic variation associated with range expansion

We assessed population structure and genetic diversity across the sampled range using three complementary methods. Firstly, we visually examined the extent of genetic differentiation between

populations using a principal component analysis (PCA) using the R package PCAdapt (Luu et al., 2016). Secondly, we estimated the most likely number of genetic clusters with fastStructure (Raj et al., 2014). fastStructure uses variational Bayesian inference to approximate the log-marginal likelihood of the data. This approach is attractive for large genomic data sets because it is very rapid compared to a traditional Bayesian approach implemented in Structure (Pritchard et al., 2000). We estimated ancestry proportions for up to 10 clusters ($K=2$–10) using a simple prior on the model. Thirdly, we estimated individual co-ancestry based on a pairwise comparison between samples using a Markov chain Monte Carlo (MCMC) coalescence model implemented in fineRADstructure (Malinsky et al., 2018). fineRADstructure estimates coancestry between individuals by comparing haplotypes across all individuals to estimate the nearest neighbour for each locus. Coancestry is divided equally between all individuals with the same haplotype or between individuals that are the nearest neighbour of a rare haplotype. In this way, rare haplotypes and their nearest neighbours receive a higher coancestry weighting than more prevalent haplotypes. Given that rare mutations are on average expected to be of more recent origin than haplotypes that occur at higher frequencies in the population, fineRADstructure is able to estimate recent coancestry in the data set. No population prior was specified for the analysis. We ran the analysis with a burn-in of 100,000 iterations, followed by 100,000 MCMC steps. RADpainter (Malinsky et al., 2018) was used to infer the coancestry matrix and assign individuals to populations using default parameters.

To determine whether genetic divergence between populations increases on average with geographic distance, we estimated population pairwise F_{ST} (Nei, 1973) from a subset of unlinked loci using the R package *adegenet* (Jombart et al., 2008). Unlinked loci were obtained by keeping only a single SNP per ddRAD-tag using the *vcftools --thin 600* filter. We estimated Pearson's correlation coefficient between genetic distance and log-transformed geographic distance, difference in dominant host plant and difference in colonization history. Significance was tested with a Mantel test in the *vegan* package (Oksanen et al., 2015) in R.

Testing for genome-wide signals of specialization in host plant use at the range edge

We tested for genome-wide signals of specialisation on the most prevalent host plant at each site by determining how much genetic variance can be explained by host plant prevalence and site colonisation history. The initial model (basic model) determined how much of the genetic variance could be explained by all the variables combined using a redundancy analysis (RDA) as implemented in the vegan package (Oksanen et al., 2015) in R. The full basic model was GeneticData [MAF matrix] ~ latitude + longitude + host plant + colonisation history. Next, we determined the best model to explain variance in the genetic data by removing one non-significant variable at a time based on an automatic permutation of the model, and selecting the best variables in each case based on Akaike information criterion (AIC). This was implemented using the *ordistep* function from *vegan* in R. We then used two partial RDA analyses to estimate the variance explained by host plant or colonization history when latitude and longitude are kept constant.

Previous data on UK *Aricia agestis* suggest that populations expanding at the range margin lost genetic diversity due to the spread of genotypes laying only on *Geraniaceae* hosts rather than on *Geraniaceae* and rockrose hosts (Bridle et al., 2014; Buckley & Bridle, 2014), as well as (potentially) due to selective sweeps of novel mutations or existing alleles that come under positive selection during expansion. We tested for a signal of a genome-wide reduction in genetic variation associated with range expansion by comparing gene diversity between sites dominated by the different host plants or with different colonization histories. Expected heterozygosity was calculated using the *basic. stats* function in the R package *hierfstat* (Goudet, 2005). We tested whether these estimates differed significantly between established and new populations, as well as between *Geraniaceae* and Cistaceae dominated sites using a Kruskal–Wallis rank sum test implemented in R. We tested the robustness of these pairwise comparisons using two methods: (1) a jackknife approach, where we repeated the analysis while removing each population in turn and (2) we randomised the colonisation history and host plant variables and repeated the Kruskal–Wallis rank sum tests.

2.3.2 | Identifying genomic regions under selection and their geographical distribution

Identifying genomic regions under selection across the UK

We used the F_{ST} outlier approach implemented in BayeScan v2.1 (Foll & Gaggiotti, 2008) to identify adaptive genotypes associated with (1) different host plants and (2) site colonisation history. For the host plant preference analysis, we estimated F_{ST} between all individuals from 'Geranium' pooled together and all individuals from 'rockrose' populations pooled together. Similarly, the colonisation history analysis was based on the pool of 'new' populations compared with the pool of 'established' populations. This model decomposes F_{ST} between populations into a population (beta) and locus (alpha) component. A locus is deemed to be under selection if the alpha component is needed to explain the diversity at that locus. BayeScan implements a reverse-jump MCMC to explore the parameter space with and without the alpha parameter included and uses this to estimate a posterior probability associated with each model, and therefore tests the statistical significance of outliers that are detected. To identify and exclude any loci associated with a single population, we repeated the analysis with each population removed in turn. In this way, we confirm the robustness of a candidate locus by identifying it in multiple analyses (Bonin et al., 2006). Loci that were identified with a false discovery rate of 0.05 in the full analysis and at least two-thirds of the jackknife analyses were included in the final set of outlier loci.

Identifying the function of outliers

We annotated the outlier loci with snpEff (Cingolani et al., 2012) using a database built from the reference genome and annotation available on NCBI (GCA_905147365.1). We used the 'closest' function to identify the closest gene to each outlier locus. The gene name and biological process were obtained by searching the UniProtKB database (The UniProt Consortium, 2021) for the protein identified in each case.

2.3.3 | Determining the geographical and colonisation history of evolutionary responses

Haplotype networks of outlier loci associated with host plant variation

To determine if alleles associated with adaptation to different host plants spread independently or as genotypes during population expansion, we constructed haplotype networks for each of the loci by identifying and phasing all outlier ddRAD tags independently. That is, each 300–450 bp sequence was phased independently of the other 300–450 bp ddRADtags.

The approach we used was as follows:

1. We subset the mapped reads (bam files) by extracting the outlier variants flanked by 300 bp on either side to include all variants within the RAD tag. Since the maximum insert size was 450 bp, this approach ensured we extracted the entire RADtag. As no sequencing data is available beyond the RADtag for any particular individual, the maximum length of the extracted sequence was 450 bp.
2. We retained sequences with at least two variant sites.
3. We phased each locus independently with WhatsHap (Patterson et al., 2015), which uses sequence data from each individual's bam files to inform phasing. This phases the loci within each RADtag assuming complete linkage between loci within the 300–450 bp. As the RADtags were phased independently of each other, they were considered unlinked.
4. All unphased sites were removed using custom bash code (see GitHub project page) and we retained only phased variants for each RADtag.
5. We converted the haplotype sequences into fasta format and constructed haplotype networks for each locus using `haploNet()` with default settings in the Pegas package (Paradis, 2010) in R. This uses a variant of the parsimony-based algorithm in Templeton et al. (1992) to construct the networks. Code used in this analysis is presented on the project GitHub repository.

Reconstructing colonisation history using a coalescent approach

The evolution of *A. agestis* to use only *Geraniaceae* at newly colonised sites could have occurred through the arrival of pre-adapted genotypes that already specialised on these host plants, or through evolution in situ at the range edge. The established north-eastern population (HOD) is dominated by *Geraniaceae* and its geographic proximity to newly colonised sites makes it a potential local source for such pre-adapted genotypes. To test this idea, and to determine the most likely origin of colonists at the new sites, we constructed six demographic models scenarios using the coalescent simulator fastSimCoal2 (Excoffier et al., 2013). We compared three models of demographic history with two different migration scenarios applied to each: Model (1) has the established southern populations (South) as the source; Model (2) has HOD as the source; Model (3) involves secondary contact between HOD and South followed by the colonisation of the new sites (Figure 2). We applied two migration scenarios to each model for a total of six models: (a) a full migration matrix of asymmetric gene flow and (b) a complete absence of gene flow after divergence. Source populations were assumed to have a constant size throughout the simulation given how recently colonisation has occurred. We removed all non-neutral loci identified by our tests for selection and excluded the Z-chromosome from our analysis.

The SFS approach is a composite likelihood method that assumes all sites are unlinked. To accurately estimate site frequency spectra, fastSimCoal simulates the genealogy at unlinked loci (we used the default number of 10,000). An assumption of the method is that

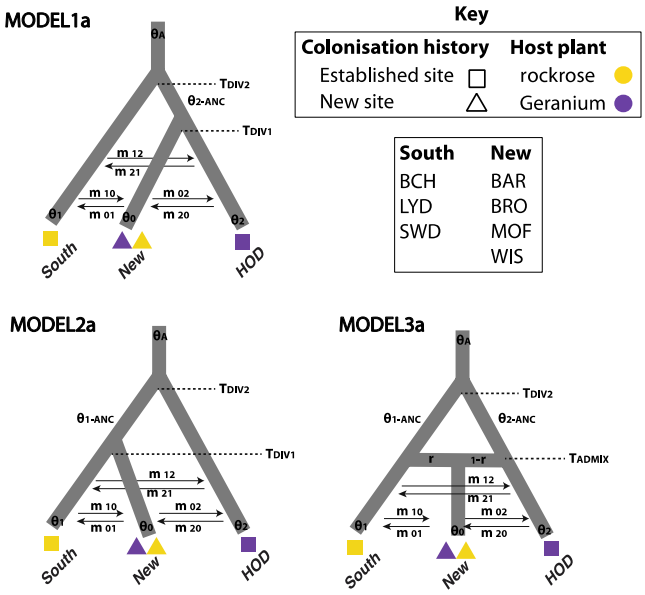

FIGURE 2 Demographic models tested with fastSimCoal2. Three demographic models were compared using FastSimCoal2: Model (1) Colonisation of the new sites exclusively from HOD followed by gene flow between all populations, Model (2) colonisation of new sites exclusively from the South followed by gene flow between all populations, Model (3) secondary contact and subsequent colonisation of the new sites by admixed populations from HOD and the South. All models were tested with two different migration matrices: (a) asymmetric gene flow after divergence, (b) no gene flow. Parameters included in the model were mutation-scaled effective population size (θ), migration rates per generation (m), the time of divergence (T_{DIV}) or admixture (T_{ADM}) between populations and the proportion of the source population transferred to the sink population (r).

all sites are based on the same sample size genotyped across all individuals with no missing data allowed. Missingness is a characteristic of reduced representation libraries because loci sequenced in different libraries do not overlap exactly. However, if we minimise the missingness in the data set, too few loci remain to accurately estimate the site frequency spectrum. At the same time, including sites with a high proportion of missingness can bias the estimated site frequency spectrum because this assumes that all sites have the same sample size. Therefore, to increase the number of loci in the final data set, we used downsampled data within populations to reduce missingness. Our final downsampled data set comprised of 9735 SNPs and was used only for fastSimCoal analyses. The downsampling approach was described and successfully used by the developers of fastSimCoal (Bagley et al., 2017; Pfeifer et al., 2018). Scripts to downsample the data and construct the minor allele frequency spectrum were obtained from Vitor Sousa, University of Lisbon (https://github.com/vsousa/EG_cE3c/tree/master/CustomScripts/Fastsimcoal_VCFtoSFS).

The effective population size was fixed for 'South' in the model so that all parameters could be estimated relative to this value. The effective population size (N_e) was calculated from the mutation rate (μ) and nucleotide diversity (π). Nucleotide diversity was calculated across all variant and invariant sites in windows of 1 kb using vcftools (--window-pi function; $\pi = 0.001$). We assume a mutation rate of 2.9×10^{-9} per base per haploid genome per generation based on the only direct estimate of Lepidoptera mutation rates (*H. melpomone*, Keightley et al., 2014). Given the model is based on the site frequency spectrum, recombination is irrelevant and was designated as '0'. The median effective population size was calculated as $N_e = (\pi/4\mu)$; South = 86,207 (range 898–712,069). For parameter estimation, we assume a generation time of 0.5 years, because *A. agestis* in the UK are bivoltine.

Given the uncertainty associated with these parameters, and the small site frequency spectrum compared with the number of parameters estimated (9–15), we expect wide and overlapping confidence intervals and uncertainty in the parameter estimates. This meant that we used the test to determine the relative likelihood of the different demographic models, rather than the absolute values of the estimated parameters. We ran 100 independent simulations of each model in fastSimCoal2. Each run comprised 100,000 coalescent simulations and 40 expectation maximisation cycles. All parameters and priors are documented in Table S4. To compare the models directly, we rescaled AIC as the difference between the lowest AIC and the AIC for each competing models. This is shown in the results as ΔAIC, where the best model has a ΔAIC of 0 (Table 4). The point estimates of each demographic parameter for the best supported model were obtained from the run with the highest composite maximum likelihood score.

Confidence intervals (CI) were estimated for the parameters by simulating 100 site frequency spectra using the maximum likelihood point estimates for the best run. Parameter estimates were re-estimated using 100 independent simulations of the model for each of the simulated site frequency spectra The lower and upper CI bounds were obtained from the lowest and highest composite maximum likelihood estimates obtained from the 100 iterations.

3 | RESULTS

After applying filters, the final data sets comprised 251 individuals from nine populations ($n = 15$–38; Table 1) genotyped at 61,210 variants.

3.1 | Changes in genomic variation associated with range expansion

3.1.1 | Population structure is associated with latitude

Our results suggest that *A. agestis* populations are largely structured latitudinally, following the likely colonisation route northwards, with the exception of FOR (see below). Overall gene flow was high ($F_{ST} = 0.031$ between all sites; Table 1, Table S1), but genetic distance increased significantly with geographic distance (Mantel's $r = .78$, $p = .001$). Similarly, genetic divergence was low ($F_{ST} = 0.026$) but significant between established and new sites (Mantel's $r = .45$, $p = .04$), with estimates similar to previous results based on 409 AFLP markers ($F_{ST} = 0.025$; Buckley et al., 2012). By contrast, genetic divergence was not significant between sites dominated by different host plants ($F_{ST} = 0.018$; Mantel's $r = .29$, $p = .07$).

Our PCA analysis showed that the most genetic variance was explained by the divergence between northern and southern populations (PCA1 = 3.7%; Figure 1), with FOR more closely related to the southern populations than the other northern populations, and HOD more differentiated than the others. Similarly, fastStructure results supported two genetic clusters ($K = 2$) that also correspond to the northern and southern populations (Figure 1). Both analyses showed that HOD (the only established population where *Geraniaceae* is the most prevalent host plant) is differentiated from the rest of the UK sites. In the PCA, the divergence between HOD and the rest of the populations explains 1.3% of the total genomic variance (PC2). Results from fastStructure also reveal that HOD and BAR (a newly established site) are genetically intermediate to the northern and southern clusters (Figure 1).

Recent coancestry as estimated with fineRADstructure revealed a more complex relationship between populations. Two well-supported groups were recovered that correspond to the northern and southern populations. However, HOD clustered with the three established southern populations (Figure 3). This is in agreement with the fastStructure analysis that shows an ~75% match between HOD and the southern cluster. Similarly, HOD is located closer to the southern than the northern populations along PC1 which otherwise matches the expected differentiation based on the geographic distance between the populations (Figure 1). Substructure within

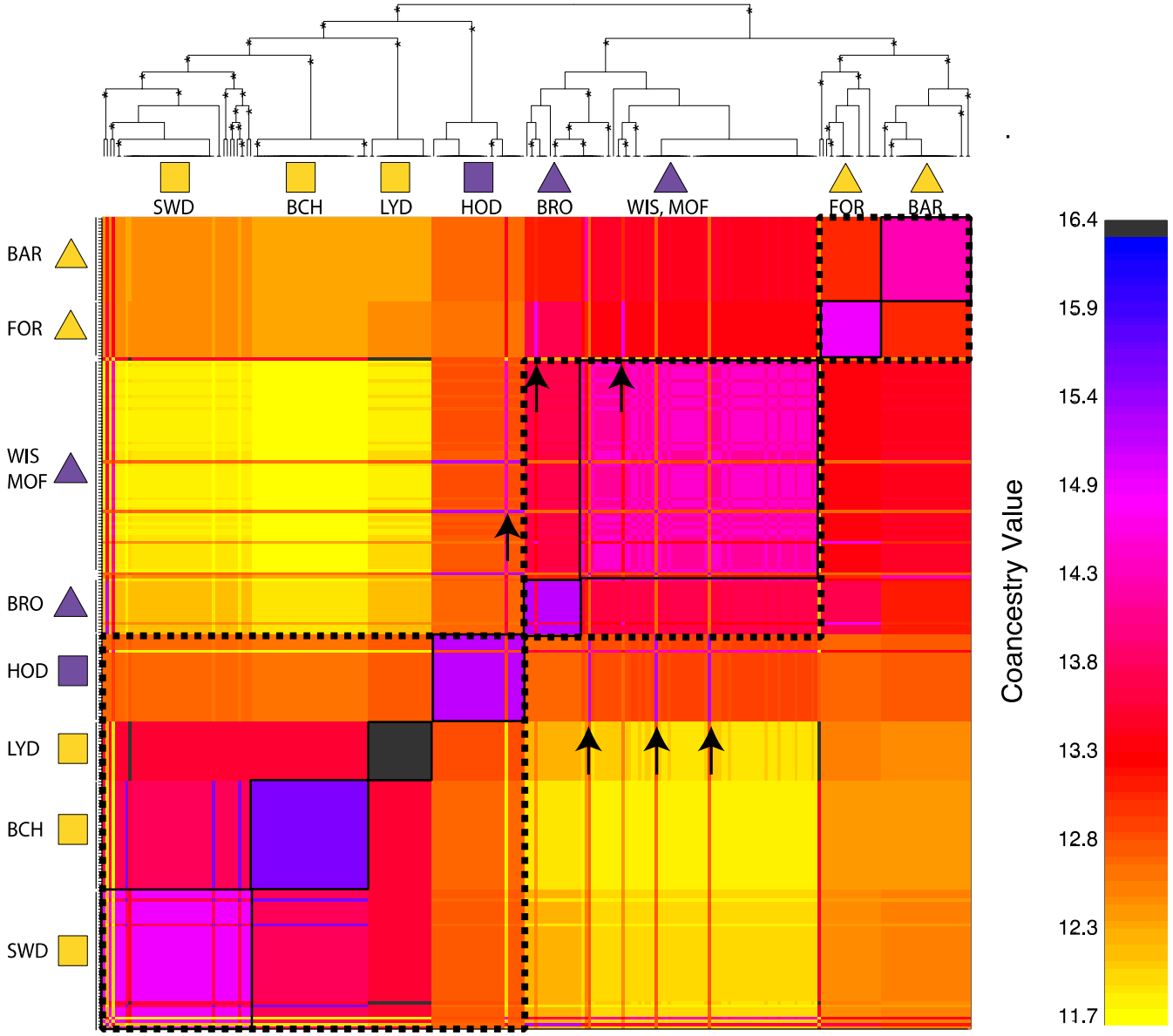

FIGURE 3 fineRADstructure of haplotype variation. Individual coancestry matrix estimated with fineRADstructure and clustered by population using RADpainter. The level of co-ancestry is indicated with colour (high = black/blue; low = yellow) as shown by the colour bar to the right. The maximum a posteriori (MAP) tree shows the inferred ancestry of each individual. Posterior probability branch support above .85 is indicated with an asterisk. The inferred co-ancestry groups largely correspond to geographic populations, although there is evidence of haplotypes moving from the new rockrose sites (BAR and FOR) to the new *Geranium* sites (WIS, MOF and BRO), as well between the new *Geranium* sites and HOD. These are indicated with arrows. Hierarchical structure in the data is shown with solid and dashed lines around grouped populations.

the northern cluster supported differentiation between the newly established *Geraniaceae*-dominated sites and the two sites where rockrose are the most prevalent species. The rockrose-dominated sites (FOR and BAR) are geographically well separated, and the *Geraniaceae* sites (BRO, WIS, MOF) are found between them. This suggests that the new sites were colonised from the established rockrose sites, rather than the established *Geraniaceae* site to the east (HOD). There is also evidence of colonisation through infilling from neighbouring populations, with at least five individuals in the new *Geraniaceae* sites resembling the neighbouring new rockrose and established *Geraniaceae* haplotypes (Figure 3).

3.1.2 | Genomic variation is slightly reduced in newly colonised regions

Our results show that there is genome-wide differentiation between new and established sites, and between *Geraniaceae*- and Cistaceae-dominated sites, but that the variance in the data cannot be significantly explained by either of these variables, independent of spatial variables (Table 2). The basic model explained a large and significant proportion of variance in the genomic data (basic model: 66%, $p = .001$). The best model retained latitude and colonisation history as significant variables. When latitude and longitude were kept constant in the

TABLE 2 Redundancy analysis.

Basic model				
Genetic variation	**Partitioned variance**	**Proportion constrained**	R^2_{adj}	***p***
Total variance	325.10			
Full model (constrained variance)	214.15	1.00	.66	**.001**
Host plant only (Host Plant | ColHist + Geog)	32.20	0.10	.02	.39
Colonisation history only (ColHist | Geog + HostPlant)	43.99	0.14	.08	.15
Geog only (Lat + Long | HostPlant + ColHist)	93.73	0.29	.16	.06

Note: Partitioning of genetic variance in each geographic transect using RDA and partial RDA analyses. The column 'Partitioned variance' shows the total variance of the genetic data (total variance), the proportion of variance that could be explained by the full RDA model which includes HostPlant, colonisation history, longitude and latitude as explanatory variables (basic model) and the proportion of total variance explained by the partial RDA in each case. The column 'proportion constrained' shows the variation explained by each model relative to the total explainable variance. The fit (R^2_{adj}) and significance (*p*) of each model are shown, and with significant *p*-values shown in bold.

partial redundancy analysis, a large proportion of the constrained variance was explained by either host plant (10%) or colonisation history (14%). However, neither host plant nor colonisation history was a significant variable in the unconstrained basic model.

Genetic diversity was marginally but significantly lower in newly colonised sites when compared with established sites (Table 1, Figure S2; median H_s: new = 0.12; established = 0.13), and in *Geraniaceae*-dominated sites compared with Cistaceae sites (median H_s: *Geraniaceae* = 0.12; Cistaceae = 0.13). Randomisation of the host plant or colonisation history variable resulted in non-significant differences. Allelic diversity was significantly different in all cases ($p < 2.2e\text{-}16$), and was still significant when the variables were randomised ($p = .026$).

3.2 | Identifying genomic regions under selection and their distribution

3.2.1 | Identification and putative origin of loci under selection

The F_{ST} outlier analysis conducted in Bayescan identified 12 loci (137 SNPs; 0.22% of total variants) associated with host plant preference, and 19 loci (239 SNPs; 0.39%) associated with colonisation history (Figure S3). Two loci each on the Z-chromosome and chromosome 9 were identified as outliers in both 'colonization history' and 'host plant use' data sets (Table S2). Of the 25 candidate loci that could be assigned an annotation with snpEff, 21 were located in the intron of a gene, and the remaining four were in intergenic regions 1287–71,461 bp from the closest gene. Any functional information available for these genes is recorded in Table S2.

The candidate loci were distributed across the genome, occurring on 8 (host plant) and 11 chromosomes (colonisation history) (Figure 4). The outlier locus with the highest difference in allele frequencies in both cases is a locus on chromosome 9 with pairwise $F_{ST} > 0.4$. This locus was identified as being under strong selection both in the host plant choice and colonisation history comparison.

3.3 | Determining the geographical and colonisation history of evolutionary responses

3.3.1 | Haplotype networks of outlier loci associated with host plant variation

Haplotype networks of the outlier loci associated with host plant use (Figure 5; Figure S4) show that the haplotypes do not contribute to phenotypic variation in host plant preference. They therefore provide no evidence that host plant preference is determined by only a few loci, as would be suggested were the haplotypes to cluster by phenotype (Kautt et al., 2020; Van Belleghem et al., 2018). This suggests that selection on variation in host plant preference acts across many loci with high levels of recombination among them. We also found no evidence that specialisation on *Geraniaceae* occurred through the rapid influx of pre-adapted alleles (or haplotypes) from HOD, given that no selected haplotype currently found in new *Geraniaceae* sites was derived from the established coastal *Geraniaceae* site (HOD). Instead, we found that the *Geraniaceae*-preferring haplotypes in newly colonised areas were similar to common haplotypes that occur in both the South and HOD, or only in the South (Table S3). In addition, haplotype networks of neutral loci were indistinguishable from the loci associated with selection, which suggests that adaptive and neutral alleles spread to the new *Geraniaceae* populations from both the established populations in the South (that can use both *Geranium* and rockrose) and from coastal populations using *Geraniaceae*, across sufficient numbers of generations for recombination to occur between them.

3.3.2 | Reconstructing the colonisation history using a coalescent approach

The best-supported colonisation history model (Model 3a) specified that new sites were colonised by an admixture of the established north-eastern population (HOD) and the established populations in the South (Table 3; Figure 2). Point estimates of the model

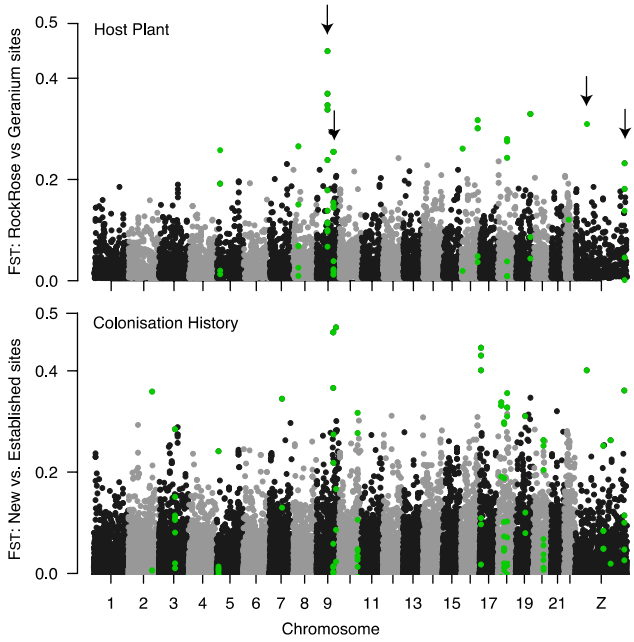

FIGURE 4 Manhattan plot of F_{ST} outliers across the *Aricia agestis* genome. The per locus distribution of F_{ST} across the genome between populations defined by (a) host plant preference, and (b) colonisation history. Outlier loci are coloured green. Four outlier loci, indicated by arrows, were associated with both host plant choice and colonisation history.

parameters suggests a slightly higher proportion of ancestry from the South (56%) than from HOD (44%) (Table 4), although this should be interpreted with caution based on the wide confidence intervals surrounding our estimates.

4 | DISCUSSION

4.1 | Rapid adaptation occurs by formation of genotypes from alleles across the range

The population genomic analyses presented here suggest that the rapid evolution of host plant use at the expanding range edge in *A. agestis* is associated with selection on genomic variation found across the species' range, facilitated by high levels of gene flow between populations. Our analyses of haplotype data (Figure 5) and fastSimCoal2 simulations best support a scenario where climate adaptation occurred through evolution in situ during range expansion, rather than through colonisation by pre-adapted genotypes from coastal populations that already specialise on *Geraniaceae* host plants. This finding is supported by the fact that we find little evidence for a reduction in genome-wide genetic variance at newly colonised sites, despite field experiments that demonstrate that female oviposition preference has narrowed in the new habitats (Buckley & Bridle, 2014). Buckley and Bridle (2014) also provide evidence for the reduced fitness of long-established *Geraniaceae*-using populations on the Norfolk coast, when transplanted to newly-colonised *Geranium* sites in Lincolnshire, suggesting that established forms of *Geraniaceae* use involve different traits to host plant adaptation associated with climate-driven range expansion.

FineRADstructure plots, and the lack of structure in the haplotype networks both in the adaptive and neutral loci, suggest ongoing gene flow from source habitats, or subsequent gene flow following an initial bottleneck associated with colonisation. Range expansion in *A. agestis* seems to occur by the infilling of new habitats rather than a mass northward expansion, with historical and ongoing high connectivity between neighbouring (source) populations. The lack of a strong signal in population size associated with range expansion could also be explained by a polygenic genomic architecture of the trait, with several small effect loci associated with the host plant shift, meaning that selective sweeps during adaptation to novel conditions have had little effect on effective population size (see below).

4.2 | Shifts in host plant use associated with climate-driven range expansion

Although *A. agestis* females from the core range typically prefer rockrose in field host choice experiments (Bridle et al., 2014), they can also oviposit on *Geraniaceae*. However, the expansion into northern England has been associated with the loss of rockrose use, certainly for Lincolnshire populations (Buckley & Bridle, 2014), as well as reduced correlations between local host preference profile and the dominant host plant (Bridle et al., 2014), suggesting an increase in the spatial scale of adaptation associated with range expansion. Presumably therefore, some cost to maintaining preference for both host plants is responsible for the loss of rockrose laying preference in habitats where only *Geraniaceae* is present. Geographic variation in host plant choice has been found in several polyphagous butterfly species (e.g. Hanski et al., 2002; Nygren et al., 2006; Stålhandske et al., 2016), and these are driven by trade-offs such as differences in host plant defence chemicals, host plant morphology and life history (e.g. phenology), or in the regional and local abundance of host plants. Importantly, southern populations of Brown Argus continue to use both plant species, especially in warm and wet years when *Geranium* becomes more common at the margins of rockrose habitat (Stewart et al., 2022). Climatic variation among years and their effects on larval success on host plants may therefore maintain the ability of females to use both species of host plant, especially if *Geranium* host plants provide more productive larval hosts than rockrose in years when summers are relatively wet (Stewart et al., 2021, 2022). By contrast, the use of rockrose at these sites may be maintained by very dry years, when eggs laid on these plants probably show lower mortality than those laid on *Geranium* plants (Stewart et al., 2022).

Rockrose-dominated sites predominate at the current northern range margin, which could be a barrier to further expansion given the loss of rockrose use at new sites documented by Buckley

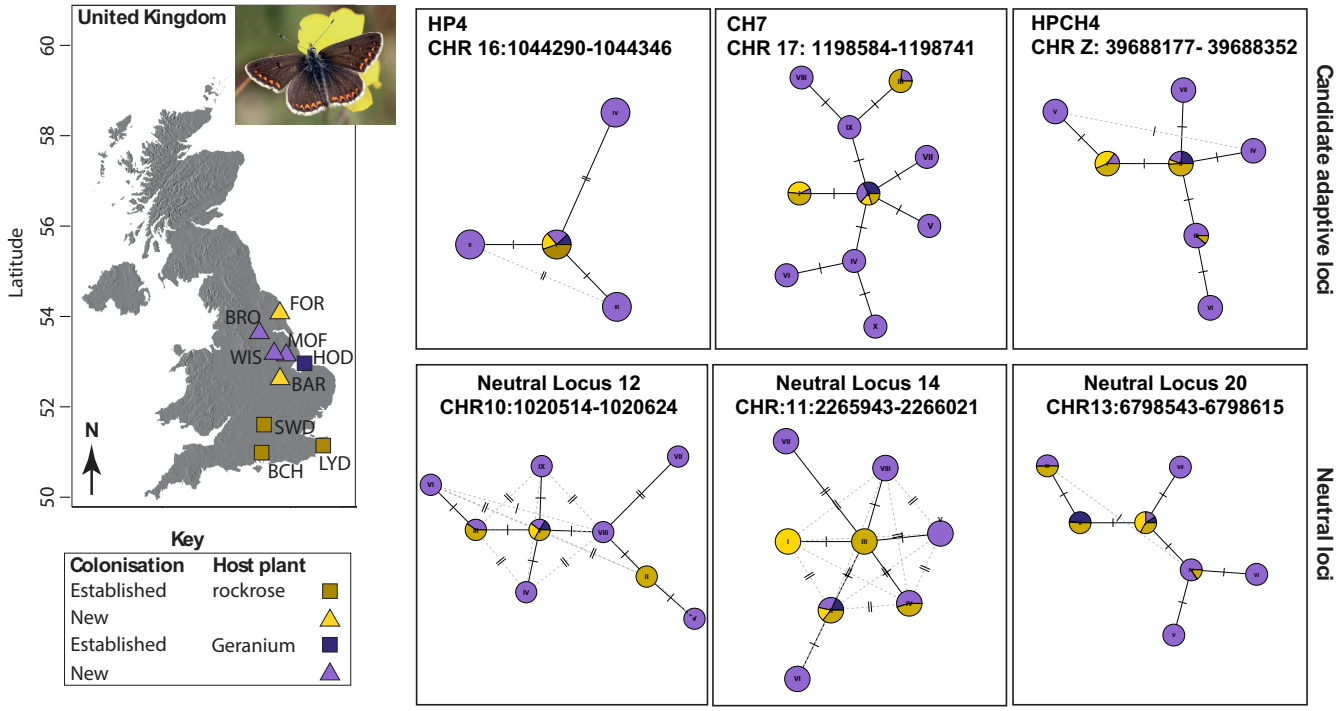

FIGURE 5 Haplotype networks suggest multiple and distant sources of selected alleles. Haplotype networks of a representative subset of the host plant (HP) outliers, colonisation history (CH) outliers and neutral loci. The remaining loci showed a similar overall pattern and are shown in Figure S4. A genomic location of each haplotype is shown in Table S2, and haplotype frequencies in Table S3. The bars represent nucleotide differences between haplotypes, and alternative parsimonious connections are shown by dotted lines. Light yellow and light purple represent the newly established populations, while dark yellow represents the established South and dark purple represents HOD, the potential source of *Geraniaceae* adaptive haplotypes. The star-like configuration of haplotypes is indicative of a recent expansion, where haplotypes found exclusively in the new sites radiate from more common haplotypes found in the established range. If the adaptive haplotypes were introduced from HOD, we would expect the light purple haplotypes to radiate from dark purple haplotypes. Instead, the adaptive haplotypes originate either from the established South or from both the established South and from HOD. The Brown Argus image in Figure 5 is sourced from Wikipedia.

TABLE 3 Comparison of demographic models.

Model	$\log_{10}$L	*k*	AIC	ΔAIC	wi
1a	−19,917.8	15	91,752.86	45.41	0.00
1b	−19,944.37	9	91,863.20	155.75	0.00
2a	−20,007.05	15	92,163.88	456.43	0.00
2b	−20,019.58	9	92,209.59	502.14	0.00
3a	**−19,907.07**	**17**	**91,707.45**	**0.00**	**1.00**
3b	−19,933.18	11	91,815.67	108.22	0.00

Note: Three demographic models (Figure 2) considering (a) asymmetric migration between populations and (b) no migration after population divergence were simulated using FastSimCoal2. We report the log likelihood ($\log_{10}$L), number of parameters estimated (*k*), Akaike information criterion (AIC), rescaled AIC (ΔAIC) and weighting (wi) for each model. The best-supported model is highlighted in bold.

and Bridle (2014). However, two sites included here constitute new rockrose sites; Fordon is the northernmost site at the extreme northern range margin, and Barnack is the southernmost newly colonised site. There is some evidence for rapid local adaptation (10–12 generations) from *Geranium* to rockrose preference at Barnack (Bridle et al., 2014; Buckley & Bridle, 2014), driven either by local adaptation or by re-colonisation from rockrose favouring individuals from further south. The latter is perhaps more likely, given high levels of gene flow found across the entire range (F_{ST} = 0.031), and the genetic similarity of new rockrose sites despite the geographic distance between them (Figures 1 and 3). An alternative explanation for the rockrose-dominant site at the northern range edge (FOR) is that it is an established *A. agestis* site which had not previously been detected. However, our intial assumption that Fordon is a newly established *A. agestis* site is supported by our genetic evidence: FOR clusters with the newly colonized sites in the fastStructure and fineRADstructure analyses and the PCA.

4.3 | The genomic basis of adaptation associated with range expansion

Using ddRAD markers allows us to investigate the genetic basis of adaptation associated with climate-driven range expansion at many loci of known genomic location, an important extension to previous AFLP data (Buckley et al., 2012). In the first instance, it is remarkable that so few loci (and most with relatively low F_{ST} levels for the

TABLE 4 Demographic parameters estimated from the best-supported coalescent model: Model 3a.

Parameter	Estimate	Lower bound	Upper bound	Category
ANCSIZE	544,351	30,800	1,119,249	Population size
NNEW	15,562	5560	44,915	Population size
NSOUTH	86,206[a]			Population size
NHOD	16,659	8691	65,541	Population size
TADMIX	9511	3680	13,034	Time
TPLUSDIV	18,372	4564	19,874	Time (complex parameter)
MIG01	2.31E-04	5.71E-10	6.56E-04	Migration
MIG10	1.57E-08	3.82E-12	7.31E-05	Migration
MIG02	4.78E-07	3.67E-11	3.30E-04	Migration
MIG20	2.35E-05	3.51E-10	4.44E-04	Migration
MIG12	1.45E-04	3.30E-11	7.67E-05	Migration
MIG21	8.97E-05	1.19E-11	3.22E-04	Migration
MUTRATE	3.53E-09	3.26E-09	5.47E-09	Mutation rate
RESIZE1	5.54	1.92	15.50	Relative change in population size
RESIZE2	1.07	0.56	6.27	Relative change in population size
MIGTOM	0.56	0.11	0.95	Migration
RESIZE3	6.31	0.36	12.98	Relative change in population size
TDIV1	27,883	11,227	29,246	Time

Note: Point estimates were obtained from the run with the highest composite maximum likelihood. Upper and lower bounds were estimated from 100 bootstrap replicates of the model using these point estimates to construct the simulated site frequency spectrum. The parameter names are defined Figure 2 (model 3a), and priors are shown in the Table S4. Results for all the models are shown in Table S5.

[a]NSOUTH population size was fixed, and all other population sizes were estimated relative to NSOUTH.

outliers) are associated with the host plant shift, given that rockrose and *Geraniaceae* species belong to different orders within the Rosid clade of angiosperms. It could be that the shift in host plant is not as phenotypically demanding as we imagine, with potentially substantial overlap in the butterfly adult and larval traits needed to sustain both interactions. Instead, the difference in microclimate between host plants could be the main factor driving evolutionary change, demanding adaptation to cope with different pathogen abundances associated with each host, or shifts in egg production or composition to alter desiccation or thermal resistance (see e.g. Stewart et al., 2021).

Alternatively, genomic variation already present and affecting host plant use in the established populations could mean that only small changes in allele frequency across many loci are needed to specialise on *Geranium* plants during expansion (see e.g. Pritchard et al., 2010), making such evolution difficult to detect using genomic methods. *Geraniaceae* species are the usual hosts of *A. agestis* throughout (warmer) continental Europe, and the evolution of alleles that allowed the use of rockrose in UK populations after post-glacial colonisation may have been critical to *A. agestis*' persistence in the UK, with alleles for *Geranium* use being ancestral. However, the analyses we present here suggest that any such ancestral *Geranium*-use alleles have been combined with alleles from long-established (rockrose-use) populations, and (possibly) more recent mutations, to form novel genotypes associated with recent climate-driven range expansion. Certainly, *A. agestis* is known to lay on *Geraniaceae* in the established part of the UK range (Bridle et al., 2014; Buckley & Bridle, 2014), and individual females lay on both species in oviposition experiments where both plant species are available (MA de Jong and JR Bridle, unpublished results). In addition, laboratory rearing of eggs (i.e. in ideal conditions) is more successful on *Geranium* host plants than on rockrose (Bodsworth, 2002), which seem to present a less nutritious host for larvae, but may become a critical resource during hot, dry summers (Stewart et al., 2022). We note also that we know nothing about local adaptation among populations of these host plants, or any evolutionary responses that may have occurred in response to the invasion of Brown Argus into their communities. Alleles determining use of a given host plant in Brown Argus are therefore likely to vary across its geographical range, possibly in response to plant local adaptation, and in relation to local microclimate (Stewart et al., 2021), as has been suggested in the Glanville Fritillary in Aaland (de Jong et al., 2014). Such local adaptation by plants to prevent herbivory is likely to demand the evolution of novel alleles for host plant use, even for plant species that are familiar to butterfly species elsewhere in their range.

Two of the loci associated with host plant choice were located on the X-chromosome (Figure 4), which corresponds with

evidence that the X-chromosome is important in Lepidopteran speciation driven by host plant differences (Janz, 2019; Prowell, 1998; Sperling, 1994). Within species variation in female oviposition choice has also associated with loci on the X-chromosome (e.g. in the comma butterfly, Nygren et al., 2006). Theoretically, genes on the X-chromosome could evolve faster than genes on autosomes, because recessive alleles will be available for selection to act on in the heterogametic sex (Charlesworth et al., 1987). A rapid change in female oviposition preference would therefore be most successful for genes located on the X-chromosome. However, a multispecies comparison of genes associated with Lepidoptera host plant preference found loci were distributed throughout the genome, with a core set of genes located on the autosomes found across all butterfly/plant pairs (Nallu et al., 2018). However, a BLAST search of our loci found no overlap between our outlier loci and these candidate genes.

The four loci associated with both colonisation history and host plant prevalence are particularly interesting because individuals colonising new sites during range expansions are often high dispersers, and increased individual movement (and extended searching) is also associated with *Geranium* host plant use (Bodsworth, 2002). In addition, Bridle et al. (2014) provide evidence for the evolution of increased dispersal ability in *A. agestis* in newly colonised sites in the UK. However, a BLAST search of the candidate regions associated with dominant host plant or colonisation history did not reveal any known regions associated with butterfly flight or movement (Table S2).

4.4 | Understanding the evolution of biotic interactions under climate change

This study provides further evidence that evolutionary responses have been necessary for climatic shifts in the Brown Argus butterfly in the UK, and that such evolution has occurred *in situ*, and has involved shifts in polygenic traits, requiring the creation and establishment of novel genotypes during the range expansion, rather than the colonisation of pre-existing genotypes from elsewhere in the range. The Brown Argus example is highly instructive, because rapid evolutionary responses are likely to be necessary for climate adaptation in many populations and communities that are characterised by specialist biotic interactions, and look likely to depend on the maintenance of gene flow and high levels of genetic variation across a species' geographical range (Bridle & Hoffmann, 2022; Hoffmann et al., 2021). It is also likely that such rapid adaptation will be required to colonise novel habitats, as well as to persist in geographical space through evolutionary rescue. However, increased environmental unpredictability and variation are key to shaping life history and behavioural strategies during adaptation to novel climates (Hoffmann & Bridle, 2021), and may reduce the genetic variation available for later adaptation just when it is most needed (see O'Brien et al., 2022). The Brown Argus may be an example of this, given evolution to allow rapid range expansion in space seems to have reduced its capacity to cope with increasingly unpredictable future conditions across years and in coming decades, by favouring specialisation on a host plant that is only productive during clement years (Stewart et al., 2022). In this way, and due to its rapid evolutionary responses, UK Brown Argus populations may be characterised by greater fluctuations in population size, and loss of genetic variation in coming decades, at least until selection favours the re-establishment of rockrose use in regions where this host plant is present.

AUTHOR CONTRIBUTIONS

MdJ and JRB devised the study, secured EC/FPT/Marie Curie Intra-European Fellowship and NBAF funding, and conducted the sampling. MdJ conducted molecular lab work for the population genomics data set, and data interpretation, and helped write the manuscript. AJvR designed and conducted the bioinformatic and genomic and ecological analyses and interpretation, and helped write the manuscript. SW designed and conducted bioinformatic analysis and created the draft genome assembly. CJY conducted the molecular lab work for the draft genome assembly. JRB hosted MdJ and assisted with study and sampling design, fieldwork, data analysis and interpretation, and wrote the manuscript, with contributions from all authors. MB assisted with genomic analysis and data interpretation and validation. CJ co-hosted MdJ and assisted with sampling design, genomic analysis and data interpretation.

ACKNOWLEDGEMENTS

MdJ was funded by a Marie Curie Intra-European Fellowship (CLIMADAPT Grant agreement ID: 332138). AJvR was funded by a Swiss National Science Foundation Early Postdoc Mobility Fellowship (P2ZHP2_178363). The generation of genomic data was supported by a grant to JB and MdJ from the Biomolecular Analysis Facility (NBAF) of the UK's Natural Environment Research Council (NERC). We thank Roger Butlin and Chris Thomas for useful discussions and advice on sampling and data analysis, and to Natural England, the National Trust and individual landowners, for permission to collect.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT

Sequence data are available on NCBI under accession number PRJNA740142. Data and scripts used for each analysis are described on the archived GitHub project page: https://github.com/alexjvr1/BrownArgus_PopGenMS_MolEcol.

ORCID

Jon Bridle https://orcid.org/0000-0002-5999-0307

REFERENCES

Angert, A. L., Bontrager, M. G., & Ågren, J. (2020). What do we really know about adaptation at range edges? *Annual Review of Ecology, Evolution, and Systematics*, *51*, 341–361.

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, S., & Jeffcoate, G. (2001). *The millennium atlas of butterflies in Britain and Ireland*. Oxford University Press.

Bagley, R. K., Sousa, V. C., Niemiller, M. L., & Linnen, C. R. (2017). History, geography and host use shape genomewide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, *26*, 1022–1044.

Bodsworth, E. (2002). *Dispersal and behaviour of butterflies in response to their habitat*. University of Leeds.

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

Bonin, A., Taberlet, P., Miaud, C., & Pompanon, F. (2006). Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, *23*, 773–783.

Bridle, J., & Hoffmann, A. A. (2022). Understanding the biology of species' ranges: When and how does evolution change the rules of ecological engagement? *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, *377*, 20210027.

Bridle, J., & van Rensburg, A. (2020). Discovering the limits of ecological resilience. *Science*, *367*, 626–627.

Bridle, J. R., Buckley, J., Bodsworth, E. J., & Thomas, C. D. (2014). Evolution on the move: Specialization on widespread resources associated with rapid range expansion in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20131800.

Bridle, J. R., & Hoffmann, A. A. (2022). Understanding the biology of species' ranges: When and how does evolution change the rules of ecological engagement? *Phil Trans Roy Soc.*, *377*, 202100272.

Bridle, J. R., Kawata, M., & Butlin, R. K. (2019). Local adaptation stops where ecological gradients steepen or are interrupted. *Evolutionary Applications*, *12*, 1449–1462.

Bridle, J. R., Polechova, J., & Vines, T. H. (2009). Patterns of biodiversity and limits to adaptation in time and space. In R. Butlin, J. R. Bridle, & D. Schluter (Eds.), *Speciation and patterns of biodiversity* (pp. 77–101). Cambridge University Press.

Buckley, J., & Bridle, J. R. (2014). Loss of adaptive variation during evolutionary responses to climate change. *Ecology Letters*, *17*, 1316–1325.

Buckley, J., Butlin, R. K., & Bridle, J. R. (2012). Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology*, *21*, 267–280.

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, *130*, 113–146.

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, *333*, 1024–1026.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*, 80–92.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

de Jong, M. A., Wong, S. C., Lehtonen, R., & Hanski, I. (2014). Cytochrome P450 gene CYP337 and heritability of fitness traits in the Glanville fritillary butterfly. *Molecular ecology*, *23*(8), 1994–2005. https://doi.org/10.1111/mec.12697

Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*, 1844–1849.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*, e1003905.

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.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*, 184–186.

Hanski, I., Breuker, C. J., Schöps, K., Setchfield, R., & Nieminen, M. (2002). Population history and life history influence the migration rate of female Glanville fritillary butterflies. *Oikos*, *98*, 87–97.

Heath, J., Pollard, E., & Thomas, J. A. (1984). *Atlas of butterflies in Britain and Ireland*. Viking.

Hill, J. K., Griffiths, H. M., & Thomas, C. D. (2011). Climate change and evolutionary adaptations at species' range margins. *Annual Review of Entomology*, *56*, 143–159.

Hoffmann, A. A., & Bridle, J. (2021). The dangers of irreversibility in an age of increased uncertainty: Revisiting plasticity in invertebrates. *Oikos*, *2022*, e08715.

Hoffmann, A. A., Miller, A. D., & Weeks, A. R. (2021). Genetic mixing for population management: From genetic rescue to provenancing. *Evolutionary Applications*, *14*, 634–652.

Hoffmann, A. A., & Sgrò, C. M. (2011). Climate change and evolutionary adaptation. *Nature*, *470*, 479–485.

Hoffmann, A. A., & Sgrò, C. M. (2018). Comparative studies of critical physiological limits and vulnerability to environmental extremes in small ectotherms: How much environmental control is needed? *Integrative Zoology*, *13*, 355–371.

Jaenike, J. (1990). Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics*, *21*, 243–273.

Janz, N. (2019). Sex linkage of host plant use in butterflies. In C. L. Boggs, W. B. Watt, & P. R. Ehrlich (Eds.), *Butterflies: Ecology and evolution taking flight* (pp. 229–240). University of Chicago Press.

Jombart, T., Ahmed, I., Calboli, F., Cori, A., Reiners, T. E., Solymos, P., & Jombart, M. T. (2008). Package 'adegenet'. *Bioinformatics Application Note*, *24*, 1403–1405.

Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., Absher, D. M., Myers, R. M., Reimchen, T. E., Deagle, B. E., Schluter, D., & Kingsley, D. M. (2012). A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology*, *22*, 83–90.

Kautt, A. F., Kratochwil, C. F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Härer, A., Hulsey, C. D., Franchini, P., Pippel, M., Myers, E. W., & Meyer, A. (2020). Contrasting signatures of genomic divergence during sympatric speciation. *Nature*, *588*, 106–111.

Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., Davey, J. W., & Jiggins, C. D. (2014). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, *32*, 239–243.

Kopp, M., & Matuszewski, S. (2014). Rapid evolution of quantitative traits: Theoretical perspectives. *Evolutionary Applications*, *7*, 169–191.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*, 2987–2993.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *bioRxiv*, 1–3.

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*, 2078–2079.

Luu, K., Bazin, E., & Blum, M. G. B. (2016). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *33*, 67–77.

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population inference from RADseq data. *Molecular Biology and Evolution*, *35*, 1284–1290.

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*, 2867–2873.

Nadeau, C. P., Urban, M. C., & Bridle, J. R. (2017). Climates past, present, and yet-to-come shape climate change vulnerabilities. *Trends in Ecology & Evolution*, *32*, 786–800.

Nallu, S., Hill, J. A., Don, K., Sahagun, C., Zhang, W., Meslin, C., Snell-Rood, E., Clark, N. L., Morehouse, N. I., Bergelson, J., Wheat, C. W., & Kronforst, M. R. (2018). The molecular genetic basis of herbivory between butterflies and their host plants. *Nature Ecology & Evolution*, *2*, 1418–1427.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*, 3321–3323.

Nygren, G. H., Nylin, S., & Stefanescu, C. (2006). Genetics of host plant use and life history in the comma butterfly across Europe: Varying modes of inheritance as a potential reproductive barrier. *Journal of Evolutionary Biology*, *19*, 1882–1893.

O'Brien, E. O., Walter, G. M., & Bridle, J. (2022). Environmental variation and biotic interactions limit adaptation at ecological margins: Lessons from rainforest *Drosophila* and European butterflies. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, *377*, 20210017.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2015). *Vegan: Community ecology package*. R package vegan, vers. 2.2-1.

Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*, 419–420.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 637–669.

Pateman, R. M., Hill, J. K., Roy, D. B., Fox, R., & Thomas, C. D. (2012). Temperature-dependent alterations in host use drive rapid range expansion in a butterfly. *Science*, *336*, 1028–1030.

Patterson, M., Marschall, T., Pisanti, N., van Iersel, L., Stougie, L., Klau, G. W., & Schönhuth, A. (2015). WhatsHap: Weighted haplotype assembly for future-generation sequencing reads. *Journal of Computational Biology*, *22*, 498–509.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*, e37135.

Pfeifer, S. P., Laurent, S., Sousa, V. C., Linnen, C. R., Foll, M., Excoffier, L., Hoekstra, H. E., & Jensen, J. D. (2018). The evolutionary history of Nebraska deer mice: Local adaptation in the face of strong gene flow. *Molecular Biology and Evolution*, *35*, 792–806.

Platts, P. J., Mason, S. C., Palmer, G., Hill, J. K., Oliver, T. H., Powney, G. D., Fox, R., & Thomas, C. D. (2019). Habitat availability explains variation in climate-driven range shifts across multiple taxonomic groups. *Scientific Reports*, *9*, 15039.

Polechová, J., & Barton, N. H. (2015). Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 6401–6406.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*, R208–R215.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Prowell, D. P. (1998). *Endless forms: Species and speciation*.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589.

Sperling, F. A. H. (1994). Sex-linked genes and species differences in Lepidoptera. *The Canadian Entomologist*, *126*, 807–818.

Stålhandske, S., Olofsson, M., Gotthard, K., Ehrlén, J., Wiklund, C., & Leimar, O. (2016). Phenological matching rather than genetic variation in host preference underlies geographical variation in host plants used by orange tip butterflies. *Biological Journal of the Linnean Society*, *119*, 1060–1067.

Stewart, J. E., Maclean, I. M. D., Edney, A. J., Bridle, J., & Wilson, R. J. (2021). Microclimate and resource quality determine resource use in a range-expanding herbivore. *Biology Letters*, *17*, 20210175.

Stewart, J. E., Maclean, I. M. D., Trujillo, G., Bridle, J., & Wilson, R. J. (2022). Climate-driven variation in biotic interactions provides a narrow and variable window of opportunity for an insect herbivore at its ecological margin. Philosophical transactions of the Royal Society of London. *Series B, Biological sciences*, *377*(1848), 20210021. https://doi.org/10.1098/rstb.2021.0021

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, *132*(2), 619–633.

The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*, D480–D489.

Thomas, C. D., Bodsworth, E. J., Wilson, R. J., Simmons, A. D., Davies, Z. G., Musche, M., & Conradt, L. (2001). Ecological and evolutionary processes at expanding range margins. *Nature*, *411*, 577–581.

Tolman, T. (1997). *Butterflies of Britain and Europe*. Harpercollins Pub Limited.

Van Belleghem, S. M., Vangestel, C., De Wolf, K., De Corte, Z., Möst, M., Rastas, P., De Meester, L., & Hendrickx, F. (2018). Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genetics*, *14*, e1007796.

Warren, M. S., Hill, J. K., Thomas, J. A., Asher, J., Fox, R., Huntley, B., Roy, D. B., Telfer, M. G., Jeffcoate, S., Harding, P., Jeffcoate, G., Willis, S. G., Greatorex-Davies, J. N., Moss, D., & Thomas, C. D. (2001). Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, *414*, 65–69.

SUPPORTING INFORMATION

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

How to cite this article: de Jong, M., van Rensburg, A. J., Whiteford, S., Yung, C. J., Beaumont, M., Jiggins, C., & Bridle, J. (2023). Rapid evolution of novel biotic interactions in the UK Brown Argus butterfly uses genomic variation from across its geographical range. *Molecular Ecology*, *00*, 1–15. https://doi.org/10.1111/mec.17138