RESEARCH ARTICLE

Deep structure, long-distance migration and admixture in the colour polymorphic land snail *Cepaea nemoralis*

Daniel Ramos-Gonzalez[1] [iD] | Suzanne V. Saenko[2,3] [iD] | Angus Davison[1] [iD]

[1]School of Life Sciences, University of Nottingham, Nottingham, UK

[2]Evolutionary Ecology, Naturalis Biodiversity Center, Leiden, The Netherlands

[3]Institute of Biology Leiden, Leiden University, Leiden, The Netherlands

Correspondence
Angus Davison, School of Life Sciences, University Park, University of Nottingham, NG7 2RD, UK.
Email: angus.davison@nottingham.ac.uk

Abstract

Although snails of the genus *Cepaea* have historically been important in studying colour polymorphism, an ongoing issue is that there is a lack of knowledge of the underlying genetics of the polymorphism, as well as an absence of genomic data to put findings in context. We, therefore, used phylogenomic methods to begin to investigate the post-glacial history of *Cepaea nemoralis*, with a long-term aim to understand the roles that selection and drift have in determining both European-wide and local patterns of colour polymorphism. By combining prior and new mitochondrial DNA data from over 1500 individuals with ddRAD genomic data from representative individuals across Europe, we show that patterns of differentiation are primarily due to multiple deeply diverged populations of snails. Minimally, there is a widespread Central European population and additional diverged groups in Northern Spain, the Pyrenees, as well as likely Italy and South Eastern Europe. The genomic analysis showed that the present-day snails in Ireland and possibly some other locations are likely descendants of admixture between snails from the Pyrenees and the Central European group, an observation that is consistent with prior inferences from mitochondrial DNA alone. The interpretation is that *C. nemoralis* may have arrived in Ireland via long-distance migration from the Pyrenean region, subsequently admixing with arrivals from elsewhere. This work, therefore, provides a baseline expectation for future studies on the genetics of the colour polymorphism, as well as providing a comparator for similar species.

KEYWORDS
Cepaea, colour, polymorphism, snail

1 | INTRODUCTION

The study of animal colour has been a key in providing evidence for some of the central tenets of biology, especially with respect to genetics and evolution. For example, early work on the inheritance of colour traits contributed to an understanding of Mendelian genetics

(Staples-Browne, 1908; Wheldale, 1907). Subsequent studies on the distribution and predation of colour morphs shaped our understanding of how natural selection may operate in wild populations, with perhaps some of the most important insights coming from moths and snails (Cook, 2003; Jones et al., 1977). In the 21st century, DNA sequencing technologies have been used to identify the underlying

genes that determine variation in colour morphs within species and their evolutionary history (Wellenreuther et al., 2014). These same methods have also enabled the subdiscipline of phylogeography to move away from the use of one or a few markers to studies that involve thousands of genetic markers (Dussex et al., 2020; Kotlik et al., 2018; Lucena-Perez et al., 2020). Such studies have revealed how connectivity between populations, or lack of, shapes, the genetic structure of species, and critically, how regions of the genome respond differently depending upon the nature of selection and the genetic architecture of a particular colour trait (e.g. Pardo-Diaz et al., 2012; Poelstra et al., 2014).

Historically, some of the most important animals in studying colour polymorphism have been the two snail species, *Cepaea nemoralis* and *C. hortensis*. One benefit of working with these species is that variation in colour and banding (Figure 1) shows straightforward inheritance (Cook, 2017; Jones et al., 1977; Richards et al., 2013), being controlled by a series of at least nine loci, of which five make

a single 'supergene' containing tightly linked colour, banding and other loci. In these two species, we now have some understanding of the pigments and shell proteome (Affenzeller et al., 2020; Mann & Jackson, 2014; Williams, 2017) and some knowledge underlying the development of the banding pattern (Jackson et al., 2021) and have begun to use genomic methods to try to map and identify the genes involved (Kerkvliet et al., 2017; Richards et al., 2013).

However, although ongoing studies on snails in general continue to provide evidence for the relative role of various forms of natural selection and random drift in promoting and maintaining variation, snails (and molluscs in general) are poorly represented in phylogeographic or population genomic studies. For example, most studies on terrestrial snails have been limited to using just a few genes, even those published recently (de Weerd et al., 2016; Ebbs et al., 2018; Neiber et al., 2017). In our own work on *C. nemoralis*, we sampled 880 individuals from 111 locations, but used only one genetic marker, a fragment of the mitochondrial gene cytochrome c oxidase
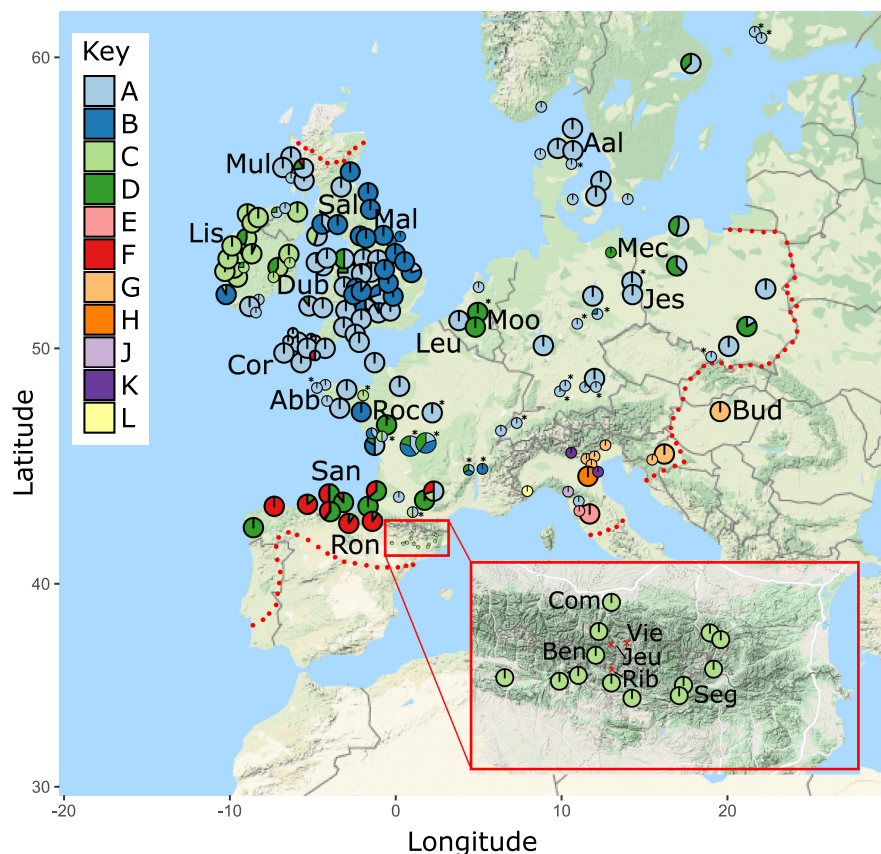

FIGURE 1 Top: European map summarizing the results of the mtDNA study on *C. nemoralis*, as well as the locations and abbreviation code for the samples used for the ddRAD study. Pies are coloured according to mtDNA lineage (see Figure S1); larger pies indicate $n > 5$ per site. Limits of the present-day distribution of *C. nemoralis* shown by the red dotted line. Data from Layton et al. (2019) indicated by asterisk. Bottom: Some of the *C. nemoralis* colour and banding morphs, from left: (1) yellow unbanded, white lip; (2) brown mid-banded, white lip; (3) pink unbanded, normal lip; (4) yellow five-banded, white lip; (5) brown unbanded, normal lip; (6) pink five-banded, normal lip; (7) yellow three-banded (10305), normal lip. Image credit: Angus Davison

subunit 1 (Grindon & Davison, 2013). Otherwise, there are just a few notable exceptions (Haponski et al., 2019; Hirano et al., 2019; Miura et al., 2020; Neiber & Hausdorf, 2015b; Westram et al., 2018), with a few studies in Europe (Koch et al., 2017; Razkin et al., 2016; Xu & Hausdorf, 2021), and a single genome re-sequencing study (Chueca et al., 2021a). This is a problem because understanding colour polymorphism requires a contextual understanding of the history of populations in general, and more precisely, a gene-by-gene account of the impact that drift and selection have had upon the genome.

One issue for *Cepaea* snails is that Jones et al. (1977) questioned whether understanding the polymorphism is 'a problem with too many solutions?', which led to difficulties arguing the case for these snails as a study organism, or molluscan 'model' (Davison & Neiman, 2021). Actually, the misunderstood intention of Jones et al. (1977) was to emphasize the perfect case study provided by *Cepaea*. Simple explanations for phenotypic variation, including colour, are likely an exception, so they were making the point that it is important to study organisms for which polymorphism may be explained by a variety of processes, precisely because they are more representative. Given that present-day genomic technologies should allow us to uncover the relative contributions of each of these processes in making contemporary diversity, this aim should now be realizable. Unfortunately, a second general issue is that extracting DNA from snails is sometimes problematic (Adema, 2021), and snails in general have large and repetitive genomes, which are difficult to assemble. Advances in the understanding of the colour polymorphisms of snails have, therefore, tended to lag behind other species, especially invertebrates with relatively small genomes (Davison & Neiman, 2021).

In this study, we take advantage of a first draft genome sequence of *C. nemoralis* (∼3.5 Gb, 28 537 contigs with N50 of 333 kb; Saenko et al., 2021) and restriction site-associated DNA sequencing methods (ddRAD) to understand the phylogeographic and population genomic structure of *C. nemoralis*. By combining the genomic data with mtDNA sequences generated from three prior studies (Davison, 2000; Grindon & Davison, 2013; Layton et al., 2019) and new sequences from Italy, the proximate aim is to begin to understand the post-glacial history of the species. The ultimate aim is to provide context for subsequent studies on the evolutionary origins of the supergene and the relative roles that natural selection and drift may play in the establishment and loss of colour polymorphism in local populations.

2 | MATERIALS AND METHODS

2.1 | Sample collection

In previous work, we and others collected *C. nemoralis* from across Europe, euthanizing snails by freezing at −80°C upon arrival at the University of Nottingham (Davison, 2000; Grindon & Davison, 2013). In this study, we strategically sampled from this prior collection, including individuals from Germany, Belgium, France,

Hungary, Denmark, Spain, UK and Ireland. We also used snails from a 2017/2018 collection (Ramos Gonzalez & Davison, 2021) from three valleys in the Pyrenees, Valle de Vielha, Valle de Jeu and Valle Noguera Ribagorzana (subsequently "Vie", "Jue" and "Rib").

For the mitochondrial DNA analysis, we combined the data from our own prior studies (Davison, 2000, Grindon & Davison, 2013) with data from another recently published work (Layton et al., 2019), the latter of particular benefit because the study had good coverage of sites in central France. In addition, we also added new data from samples from Italy, useful because the peninsula was a potential refugium during previous glaciations but poorly sampled in the original studies.

For the genomic analyses, individuals were selected to ensure a wide geographic spread and to include representatives of different colour morphs, as well as the major mitochondrial lineages reported in the prior studies (Figure 1, Table 1).

2.2 | DNA methods

High-molecular-weight genomic DNA was extracted from frozen snail foot tissue, using methods similar to those described previously (Gonzalez et al., 2019; Richards et al., 2013). In brief, tissue slices were incubated at 65°C in lysis buffer (10 mM Tris, 0.1 M EDTA, 0.5% SDS), with proteinase K (0.2 mg/ml), with occasional inversion. RNase A (80 μg/ml) was added, and then, each tube was incubated for one more hour. The mixture was cooled to room temperature and then extracted using the standard phenol/chloroform/iso-amyl ethanol (25:24:1) method, with ethanol precipitation. DNA quality was visually verified by agarose gel electrophoresis and then quantified and further checked using a NanoDrop 2000c spectrophotometer and a Qubit 2.0 Fluorometer. ddRAD libraries were generated by SNPsaurus (https://www.snpsaurus.com/), using the method of Russello et al. (2015) and 75 ng of input DNA. Using a service provided by SNPsaurus at the University of Oregon (https://www.snpsaurus.com/), one lane of a HiSeq4000 was then used to generate 150 base pair sequence reads, which were then demultiplexed and stored as compressed fastq files. In addition, a ∼600 base pair (b.p.) fragment of the mitochondrial gene COI was amplified by the polymerase chain reaction (PCR) for some individuals, using either the Folmer et al. (1994) primers, those of Gittenberger et al. (2004) or a combination of the two, as described previously (Grindon & Davison, 2013). The PCR product was then sequenced with forward primers (and reverse, if necessary) using BigDye™ Terminator version 3.0 Cycle Sequencing Kit.

2.3 | Analyses of mitochondrial DNA variation

The randomized axelerated maximum likelihood method, as implemented in RAxML 8.2.12 (Stamatakis, 2014), was used with a partitioned data set (one for the 1st/2nd codon positions and another for the 3rd codon position) and produced a mitochondrial phylogeny,

TABLE 1 Summary showing the locations and phenotypes of the individual samples used for the analyses of genomic variation, including mitochondrial haplotype (Grindon & Davison, 2013); "group" reflects the clustering returned using Gaussian finite mixture modelling

Location			ID	Grindon ID	Collector	Date	Latitude	Longitude	mtDNA	Group	Colour	Banding
Belgium	Leuven	Leu	40	Kesscn5	Thierry Backeljau	6/29/2005	50.883467	4.722461	D39	8	Pink	00300
			45	Kesscn8		6/29/2005			D8	8	Yellow	00000
	Moortsele	Moo	48	Ooscn2a	Thierry Backeljau	7/27/2005	50.958811	3.779839	A27	10	Yellow	12345
Denmark	Aalborg	Aal	46	Aal2cn5	Adele Grindon	7/20/2006	57.051833	9.872750	A51	10	Yellow	12345
			47	Aal2cn7		7/20/2006			A22	10	Pink	12345
England	Cornwall	Cor	26	Coascn3	Elizabeth Bailes	6/18/2005	50.254919	−5.043106	A87	10	Yellow	00000
			66	Coascn5		6/18/2005			A87	10	Yellow	00000
	Malham	Mal	53	Malcn3	Adele Grindon	8/10/2005	54.119917	−2.180333	B8	10	Yellow	12345
			54	Malcn5		8/10/2005			B8	10	Yellow	00000
	St. Bees	Sal	52	Saltcn9	AD	5/2/2007	54.489256	−3.589572	B3	10	Pink	00300
France	Comminges	Com	63	Romancn5	Adele Grindon	4/24/2007	43.027983	0.722267	C59	5	Yellow	00345
			64	Romancn7		4/24/2007			C56	5	Yellow	00000
	La Roche	Roc	31	Berncn2	Adele Grindon	4/23/2007	47.511767	−2.285667	B23	7	Pink	12345
	Pont L'Abbe	Abb	28	Abbecn1	Adele Grindon	4/24/2007	47.883950	−4.182567	A58	10	Yellow	00300
			29	Abbecn3		4/24/2007			A85	10	Yellow	00300
Germany	Jessern	Jes	49	Jesscn1	Malgorzata Ozgo	5/23/2005	52.016667	14.183333	A11	10	Pink	00000
			51	Jesscn9		5/23/2005			A40	10	Pink	00300
	Mecklenburg	Mec	50	Robcn2	Michael Zettler	5/29/2005	53.581667	12.985000	D33	4	Pink	00300
Hungary	Budapest	Bud	42	Hungcn4a	Feher Zoltan	9/13/2005	47.425808	19.448261	G11	1	Yellow	12345
			43	Hungcn5a		9/13/2005			G13	1	Yellow	00300
			44	Hungcn1a		9/13/2005			G12	1	Yellow	00300
Ireland	Dublin	Dub	38	Dublincn4	AD	5/3/2006	53.318475	−6.251375	C15	7	Pink	12345
			39	Dublincn2		5/3/2006			C15	7	Yellow	00345
	Lisdoonvarna	Lis	36	Liscn1	AD	4/7/2007	53.026317	−9.290167	B27	7	Yellow	12345
			37	Liscn2		4/7/2007			C8	7	Pink	12345
Scotland	Mull	Mul	27	Mullcn4	Max Cooper	6/13/2007	56.578339	−6.283422	A47	10	Pink	12345

(Continues)

TABLE 1 (Continued)

Location			ID	Grindon ID	Collector	Date	Latitude	Longitude	mtDNA	Group	Colour	Banding
Spain	Benasque	Ben	61	Bencn3	Adele Grindon	4/16/2007	42.599367	0.527000	C19	9	Yellow	12345
			62	Bencn1		4/16/2007			C17	9	Yellow	00000
	Jueu	Jue	34	DRG106_15	DRG	6/29/2018	42.677900	0.705500	-	9	Yellow	00000
			56	DRG103_5		6/29/2018			-	9	Pink	00000
	Ribagorzana	Rib	57	DRG77_14b	DRG	10/12/2017	42.479200	0.714000	-		Yellow	00000
			58	DRG124_16		6/29/2018			-		Pink	00000
	Roncesvalles	Ron	41	Ron2cn3	Adele Grindon	4/14/2007	43.009300	-1.319067	F41	3	Pink	12345
	Santander	San	60	Santacn8	AD	10/21/2006	43.414400	-3.741300	D52	2	Yellow	00000
			65	Santacn9		10/21/2006			F25	2	Yellow	00000
			30	Playacn9	Adele Grindon	4/12/2007	43.487067	-3.792683	D49	2	Yellow	00000
			35	Playacn10		4/12/2007			D47	2	Yellow	00000
			55	Playacn7		4/12/2007			F24	2	Yellow	00000
	Segre	Seg	59	Serge2cn1a	Adele Grindon	5/23/2006	42.359817	1.479550	C1	6	Yellow	12345
	Vielha	Vie	32	DRG17_9	DRG	10/12/2017	42.690500	0.874200	-	9	Yellow	12345
			33	DRG17_12		10/12/2017			-	9	Pink	00000

with rapid bootstrap resampling (100 replicates). The phylogeny was then used to place individual haplotypes into lineages, as previously described (Grindon & Davison, 2013).

2.4 | Quality checking, variant calling and filtering

FASTQC v0.11.9 software was used to check the quality of the reads, and then, adapters were removed using TRIMMOMATIC 0.39 (Bolger et al., 2014). The fastq files were then aligned to the draft genome (Saenko et al., 2021) using the Burrows-Wheeler Alignment method (Li et al., 2009) in BWA v0.7.17-r1188, using default settings and marking low-quality alignments (−M). SAMTOOLS v1.11 was then used to sort and compress the sam files to bam files, then BCFTOOLS v1.10 (Danecek & McCarthy, 2017) was used to call variants. VCFTOOLS (Danecek et al., 2011) was initially used to explore the unfiltered data, to guide subsequent filtering thresholds (e.g. see https://speciationgenomics.github.io/filtering_vcfs/ or the Davison github repository). A filtered data set of bi-allelic loci was then created, retaining those sites where there was a missing call in <25% of individuals, a mean sequencing depth across all individuals of between 0.25 and 25 reads (removes false positives and avoids repetitive regions), a minimum quality score of 20 and minor allele frequency (MAF) of 0.1, the latter based on an inspection of the distribution of allele frequencies. PCR duplication was checked using a script 'checkHetsIndvVCF.sh', originally written by David Marques and edited by Joana Meier (https://github.com/alphaneer/scripts-1/blob/master/checkHetsIndvVCF.sh; Meier et al., 2018).

2.5 | Analyses of genomic variation

We used principal component analysis (Patterson et al., 2006), whole-genome phylogeny and admixture tests to analyse the relationships between the individuals.

First, as others have done (Bryc et al., 2010; Decker et al., 2014; Marques et al., 2022), principal components analysis (PCA) and Gaussian finite mixture modelling methods were used as an independent means of exploring the broad phylogenomic relationship between individuals, independent of models of sequence evolution. For the PCA, but not other analyses, the filtered variant calling file was first pruned using PLINK v1.9 (Purcell et al., 2007), then linkage-pruned sites were used to generate eigenvalues and vectors. The PCA output was plotted in R version 3.4.1 (R Core Team, 2020) using TIDYVERSE v1.3.0 (Wickham et al., 2019). To identify possible clusters within the data, Gaussian finite mixture modelling was applied to the PCA values using MCLUST 5.4.6 (Scrucca et al., 2016), assuming between 1 and 20 possible clusters. The best-fitting model was then determined using the Bayesian Information Criteria (BIC), with significant differences established by using a bootstrap likelihood ratio test, assuming varying orientations and homogeneity of variance.

To understand the distribution of genomic variation by geography, pairwise values of Weir & Cockerham's F_{st} between populations

were estimated using VCFTOOLS (Danecek et al., 2011). To achieve this, genome-wide F_{st} was used directly by plotting against geographic distance or transformed into a dissimilarity matrix, using the VEGAN 2.5-6 package in R (Jari Oksanen et al., 2018) and again plotted against geographic distance. Significance was tested using a Mantel test based on Spearman's rank correlation and 9999 permutations, again using VEGAN 2.5-6.

In the future—when whole-genome analyses are possible and the colour and banding loci have been finely mapped—we aim to understand how gene flow impacts upon different regions of the genome, for example, using a window-based approach. As a starting point, we estimated F_{st} by contig and also estimated Mantel's r statistic for the F_{st} for each contig and all pairwise comparisons. To achieve this, the same analyses as above were carried out on each of the individual contigs, again testing significance using a Mantel test.

To further understand the phylogenomic relationship between individuals, the concatenated biallelic SNP ddRAD data were used to generate a phylogeny, using RAXML v8.2.9 with the GTRGAMMA model and 100 bootstrap replicates (Stamatakis, 2014).

The program ADMIXTURE v1.3.0 was used to determine the best number of clusters (K), by computing the fivefold cross-validation error (Alexander & Lange, 2011), with the aim of then using TREEMIX 1.13 (Pickrell & Pritchard, 2012) to investigate possible migration events.

To support these analyses, D statistics were used to explore evidence of introgression between Pyrenean and central European snails, using ADMIXTOOLS version 6.0 (Patterson et al., 2012) and ADMIXR version 0.7.1 (Petr et al., 2019). Leave-one-out cross-validation was used to determine the best-supported value, with significance evaluated using block jackknife and the Z-score.

For the D statistics, the underlying principle (Ravinet & Meier, 2020) is that if W, X, Y, Z represent four taxa with a relationship (([W,X,]Y),Z), with no hybridization, then the number of discordant SNPs grouping W and Y together (BABA patterns) is expected to be roughly the same as the number of SNPs grouping X and Y together (ABBA). Therefore, we set out to test whether the Irish samples and others show evidence of introgression from focal populations; groupings and analyses were based on prior PCA, admixture analyses and RAxML phylogenies. Using Hungarian samples as an outgroup (Z), we compared individual central European populations (W; England: Cornwall, St. Bees, Malham; Germany: Jessern; Belgium: Moortsele; Denmark: Aalborg) and individual central Pyrenean populations (Y; Vielha, Jueu, Benasque, Segre and Comminges) to the populations of interest (X; Ireland: Lisdoonvarna, Dublin; France: La Roche; Scotland: Mull).

3 | RESULTS

3.1 | Mitochondrial DNA

Here, we combined data from two prior studies (Davison, 2000; Grindon & Davison, 2013) with a new study (Layton et al., 2019). We also added 17 individuals from a further nine locations in Italy, making a combined total of 1521 individuals from 180 locations across Europe (Table S1). The revised mtDNA phylogeny (Figure S1; rooted using the sister taxon *C. hortensis*) shows the same pattern as before (Figure 1), with the new finding that Italy has several lineages that are not found anywhere else, except in Croatia and Hungary. The majority of Europe (Figure 1) has snails with mtDNA lineages A, B, C, D and F, with the other lineages relatively rare and/or range restricted.

3.2 | Restriction site-associated DNA sequencing

ddRAD libraries were prepared from 41 snails and 22 locations across Europe (Table 1). The locations used commonly have mtDNA lineages A (Aal, Abb, Cor, Jes, Leu, Mul), B (Lis, Mal, Roc, Sal), C (Ben, Com, Dub, Jeu, Lis, Rib, Seg, Vie), D (Dub, Mec, Moo, Mul, Ron, San), F (Ron, Sand), G (Bud).

After discarding two samples due to a low number of reads (S57 and S58 both from Ribagorzana valley in Spain; Table S2), 39 individuals were put through to the main analysis. An average of 95% of reads mapped to the reference genome (Table S2). After filtering, the whole-genome data set was reduced to 8689 bi-allelic loci contained on 2323 genome contigs, making a total length of about 990 Mb (out of 28 538 contigs in the reference genome of ~3.5 Gb; Table S3). There was a mean of 3.74 SNPs per contig and, as expected, a correlation between contig length and number of SNPs ($r = 0.11$, $p < 0.001$; Figure S2). There was no evidence for PCR duplication problems.

3.3 | Different genomic evolutionary history of the colour and geographic variation

A principal components analysis of the whole-genome data showed that individuals of *C. nemoralis* broadly cluster according to geography (Figure 2) with PC1, PC2 and PC3 together explaining 37% of the variance (Table S4). Samples from Hungary (Budapest, $n = 3$), North-West Spain (Santander, $n = 5$) and the West Pyrenees (Roncesvalles, $n = 1$) were separated from the other samples on PC1 (Figure 2a; PC1 = 15%). PC3 separated the Hungarian samples from the North-West Iberian samples, with the others intermediate (Figure 2b; PC3 = 10%). PC2 separated the populations found in the Central Pyrenees (Vielha, Jueu & Benasque, n = 6; Comminges, n = 2; Segre, n = 1) from samples found elsewhere (Figure 2c; PC2 = 12%). Of the other samples, those from Budapest ($n = 3$) and from Ireland (Dublin, $n = 2$; Lisdoonvarna, $n = 2$) and the West of France (Roche, $n = 1$) were placed near to the central Pyrenean samples, the closest of which was an individual from Segre in the Eastern Pyrenees.

Gaussian finite mixture modelling returned 10 clusters as the best-fitting model (EII, spherical, equal volume; BIC 349.23 and ICL 349.22; $p < 0.001$ compared with the next best model); Figures 2–4 are coloured according to these groupings. Figure 2 shows the clusters in relation to the principal component analysis; Figure 3 shows the geographic distribution of three clusters.
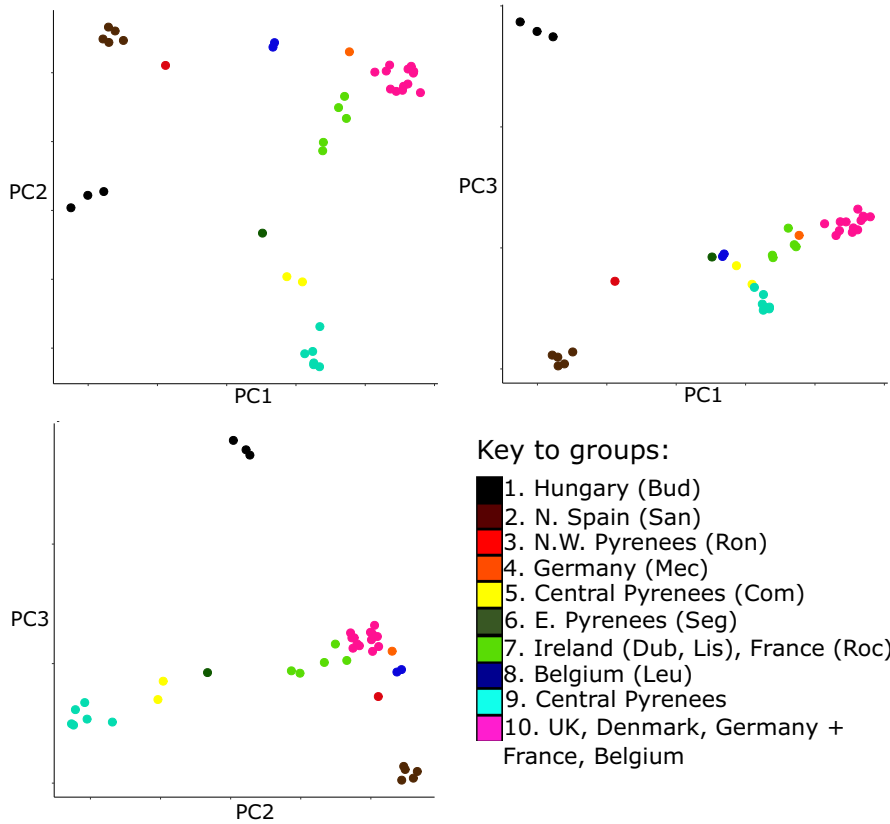
FIGURE 2 First three axes of principal component analysis of the whole-genome data of individual *Cepaea nemoralis* snails using 8689 biallelic loci. Individuals are coloured according to the clusters defined by Gaussian finite mixture modelling. Groups 1 to 6 and 8 are made up of snails from single sample sites; group 7 samples from Ireland (Lis, dub) and NW France (roc); group 9 samples from the Central Pyrenees (ben, Jue, vie, rib), with group 10 all remaining samples (UK: Mul, Sal, mal, Cor; France: Abb; Belgium: Moo, Denmark: Aal; Germany: Jes)
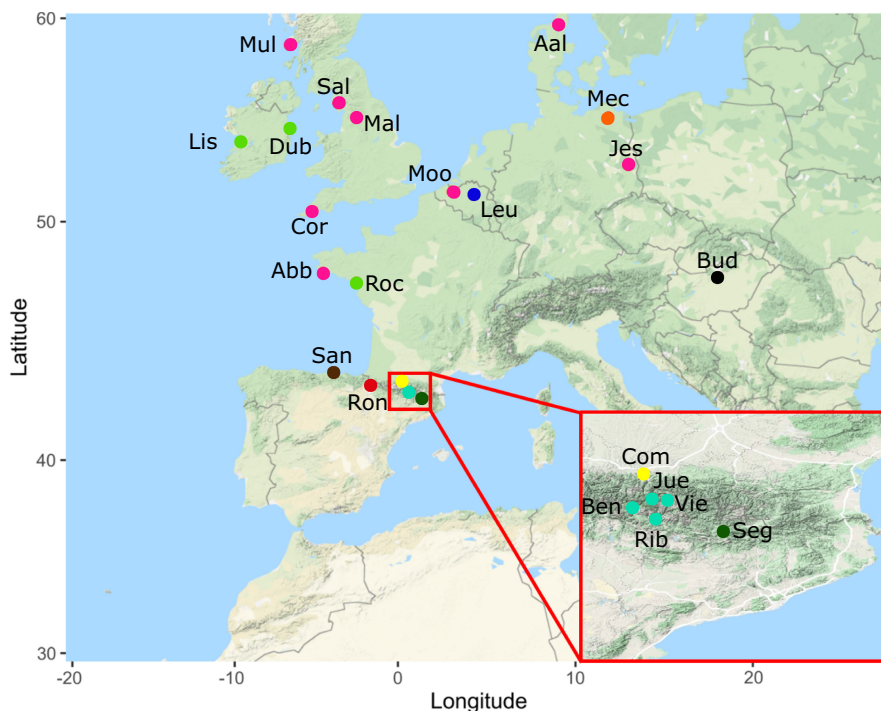
Key to groups:
1. Hungary (Bud)
2. N. Spain (San)
3. N.W. Pyrenees (Ron)
4. Germany (Mec)
5. Central Pyrenees (Com)
6. E. Pyrenees (Seg)
7. Ireland (Dub, Lis), France (Roc)
8. Belgium (Leu)
9. Central Pyrenees
10. UK, Denmark, Germany + France, Belgium

FIGURE 3 European map showing the locations and abbreviation code for the samples used in this study, with locations coloured according to the groups identified by a cluster analysis of the genomic data

A RAxML phylogeny (Figure 4) based on the whole-genome data mainly corroborated the principal components analysis in recovering a Northern Spanish and West Pyrenean group (Santander and Roncesvalles), a Hungarian group, a central European group (French, German, Danish and British) and multiple lineages in the central and Eastern Pyrenees. There were some populations/individuals that did not fall in this general grouping, including snails from Leuven in

Belgium (*n* = 2) and Mecklenburg in Germany (*n* = 1). Although the phylogeny shows some well-resolved groups, the deeper structure is not well supported.

Across all pairwise population comparisons, the mean genome-wide F_{st} was 0.14, but with a wide range of individual values (0.00 to 0.44; SD = 0.10; Table 2). The frequency distribution revealed around three groups, with peaks of F_{st} at approximately 0.05–0.015,

FIGURE 4 Mid-point rooted RAxML phylogeny, based on an analysis of 8689 biallelic loci. Bootstrap support >70% is shown. Colours are in accordance with the clusters defined by Gaussian finite mixture modelling
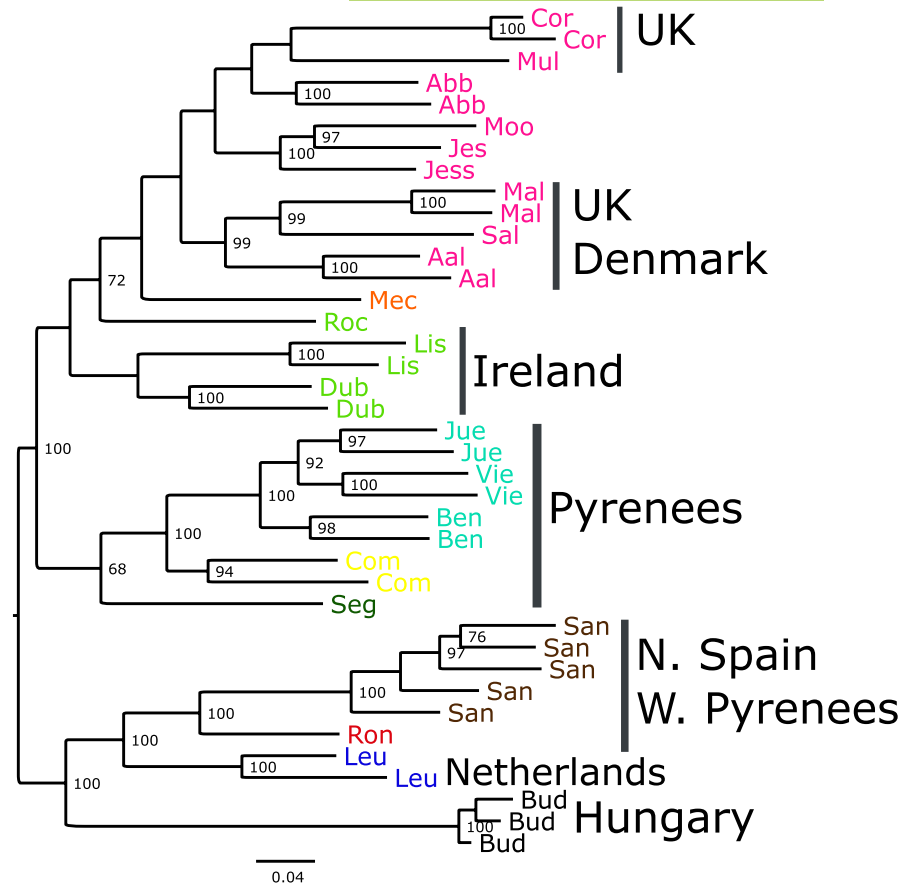
0.2 and 0.3 (Figure 5a). The highest structure was between samples from Budapest, Hungary and elsewhere ($F_{st} > \sim 0.3$), with the next-most divergent samples those from Santander, Spain and elsewhere ($F_{st} \sim 0.2$). There was a shallow, but significant relationship of F_{st} by geography (Figure 6), irrespective of the precise method used.

In individual contigs, there was considerable variation of F_{st}, showing a mean of 0.23 across the whole genome (range 0.03 to 0.51; SD = 0.09; Figure 5b). To further understand geographic and genomic variation, we estimated Mantel's r statistic for the F_{st} for each contig and all pairwise comparisons of populations. Although the mean r was greater than zero (Figure 5c; mean = 0.021), only ~1/3 of the genome contigs showed a significant association with geography.

Admixture analyses were used to test the hypothesis that some populations are descended from introgression between Pyrenean and other European snails (Figure 7). For both $K = 2$ and $K = 3$, individuals from Ireland, especially the West coast (Lisdoonvarna) showed mixed ancestry between the Pyrenean snails and elsewhere. Samples from the East coast of Ireland (Dublin), and possibly also the North West coast of France (La Roche) and the Scottish island of Mull, may also show some evidence of mixed ancestry, albeit to a less degree (Figure 7).

D-statistical analyses were used to further explore the patterns of admixture (Table S5). Specifically, we set out to test whether the Irish samples (Lisdoonvarna, Dublin) and some others (e.g. La Roche, Mull) show evidence of introgression from other populations. In general, there was no evidence for an excess of allele sharing (Table S5, Z-score between −3 and 3) and therefore, no evidence for introgression since their common ancestry. The main exception was a significant enrichment of shared sites between Segre in the Eastern Pyrenees and Lisdoonvarna in Ireland, irrespective of whichever central European site used (Z-score around −5 to −6). This is consistent with these West Irish samples being partly derived from snails in the Eastern Pyrenees (Table S5). There was also limited evidence that the Dublin population may be partly derived from Pyrenean snails and the same for samples from La Roche (Z-score < −3 in two comparisons each). These results are, therefore, consistent with Irish and possibly some French populations being derived from an introgression event between Pyrenean and central European populations of snails. In the future, it should be possible to better test for evidence of introgression, using analyses such as treeMix (Pickrell & Pritchard, 2012), but unfortunately, a relatively high degree of missing data precluded this analysis here.

4 | DISCUSSION

Investigating the phylogeography of *Cepaea nemoralis*–and the relative degree of gene flow between populations—should be an important step towards understanding the relative roles that natural selection and drift have to play in the evolutionary history of the colour and banding polymorphism. In this study, we used a newly

available draft genome (Saenko et al., 2021) and the ddRAD method, supplemented with a wide geographic survey of mtDNA data, with a primary aim to begin to understand the post-glacial history of the species, but also to provide context for future studies on the evolutionary origins of the supergene, and the roles that selection and drift have in determining the local colour polymorphism. Of course, these methods are now standard in many animal groups, but the study is relatively novel for snails, partly because molluscs in general lag behind (Davison & Neiman, 2021), with, for example, only a few other land snail species having a genome assembly (Chueca et al., 2021b; Guo et al., 2019; Liu et al., 2021). The reality is that although phylogeographic and genomic studies are relatively common in other invertebrates, there are remarkably few similar studies in snails or even the wider group of molluscs.

Broadly, analysis of the whole-genome data expands and adds nuance to the mtDNA survey. The analysis shows that European-wide patterns of differentiation are primarily due to multiple deeply divergent groups of snail. Geographically one group is dominant, being found over much of Central Europe (Figures 1 and 2). There are also likely multiple Italian populations, some of which may have colonized to the limits of the eastward distribution of the species (e.g. in Croatia and newly colonized Budapest). Finally, there are multiple lineages in Iberia, roughly centred along the Northern coast and Cantabrian mountains, with separate populations in regions of the Pyrenees. In the future, further sampling, especially in potential

refugia in the Balkans and Italy, may reveal other divergent groups, some of which may have very localized distributions.

The results corroborate previous studies in that populations showed a high degree of genetic structuring, sometimes over short geographic distances (Davison & Clarke, 2000). For example, the maximal genomic distance was between snails in Hungary and the rest of Europe ($F_{st} > 0.3$; Table 2; Figure 5). In comparison, samples from the UK, Denmark, Germany, France and Belgium were relatively homogenous, with an F_{st} around 0.05 to 0.15. Previously, Ochman et al. (1983) showed using allozymes that the Pyrenees contain three deeply differentiated sets of populations, or 'molecular area effects', for which the geographic patterns did not correlate with the shell ground colour and banding frequencies. The genomic evidence also supports these findings, showing deep divergence between geographically close populations on the Atlantic coast of Northern Spain (e.g. Santander), and the Central Pyrenees (e.g. Vielha) with possible hybrid populations in both the Western and Eastern Pyrenees (e.g. Roncesvalles in West and Segre in East).

4.1 | Origins of the Irish *Cepaea nemoralis*

One of the key characteristics of most population genetic studies of land snails is that they show high population structure, and deeply diverged mitochondrial lineages, the latter frequently considerably

TABLE 2 Mean F_{st} across all loci; values are coloured according to whether they are between 0.2 and 0.3, or >0.30

		Leu	Moo	Aal	Cor	Mal	Sal	Roc	Abb	Jes	Mec
Belgium	Leuven										
	Moortsele	0.02									
Denmark	Aalborg	0.11	0.06								
England	Cornwall	0.20	0.06	0.22							
	Malham	0.18	0.04	0.13	0.26						
	St. Bees	0.06	0.08	0.02	0.11	0.01					
France	La Roche	0.04	0.11	0.01	0.06	0.04	0.09				
	Pont L'Abbe	0.10	0.09	0.03	0.03	0.13	0.03	0.07			
Germany	Jessern	0.06	0.11	0.02	0.15	0.11	0.06	0.05	0.02		
	Mecklenburg	0.01	0.36	0.03	0.09	0.05	0.10	0.11	0.05	0.24	
Hungary	Budapest	0.33	0.30	0.40	0.44	0.42	0.33	0.30	0.35	0.37	0.31
Ireland	Dublin	0.10	0.03	0.08	0.18	0.13	0.05	0.07	0.04	0.05	0.03
	Lisdoonvarna	0.15	0.05	0.16	0.23	0.19	0.06	0.01	0.14	0.12	0.06
Scotland	Mull	0.14	0.11	0.07	0.06	0.04	0.17	0.14	0.09	0.10	0.13
Spain	Benasque	0.16	0.01	0.15	0.24	0.24	0.06	0.01	0.08	0.09	0.03
	Commitges	0.12	0.00	0.12	0.18	0.17	0.03	0.02	0.07	0.08	0.02
	Jueu	0.17	0.07	0.17	0.24	0.21	0.09	0.05	0.14	0.12	0.08
	Roncesvalles	0.03	0.19	0.12	0.20	0.16	0.22	0.17	0.10	0.26	0.16
	Santander_1	0.17	0.20	0.23	0.25	0.23	0.21	0.18	0.18	0.21	0.17
	Santander_2	0.17	0.20	0.24	0.25	0.24	0.21	0.19	0.19	0.21	0.17
	Segre	0.06	0.02	0.07	0.13	0.10	0.01	0.04	0.03	0.03	0.01
	Vielha	0.16	0.06	0.15	0.23	0.21	0.08	0.04	0.12	0.11	0.06

<5% *within* species (Chiba, 1999; Thomaz et al., 1996), for which a variety of explanations have been put forward (Davison, 2006; Parmakelis et al., 2013). In *Cepaea*, one of the most striking findings from previous work was that some Irish *C. nemoralis* are most likely derived from a post-glacial colonization event from the Pyrenees (Grindon & Davison, 2013), in part because of identical mitochondrial haplotypes in the two locations.

Here, the genomic data provide further evidence for the long-distance colonization, but also show that the present-day snails in Ireland are likely descendants of admixture between Pyrenean snails and individuals from the population that colonized most of Central Europe. The snails from the West coast of Ireland show the greatest proportion of Pyrenean ancestry (Figure 7; Table S5), consistent with the observation that they also have some of the same shell characters, including large size and shells with a pale or white lip colour (Cameron, 1969; Cook & Peake, 1960, 1962). The analyses may also indicate, albeit a lesser signal, of this same admixture in snails from Dublin, Ireland, and the North-West of France (La Roche; Figure 7; Table S5). As there is also another mitochondrial DNA lineage present in Ireland (D), which may derive from Northern Spanish or West Pyrenean snails (directly or indirectly), then Ireland may have been colonized on at least two and perhaps three or more occasions.

Similar patterns of mixed ancestry have been reported for other species (especially those involving the 'Celtic fringe'; Brace et al., 2016; Kotlik et al., 2018; McDevitt et al., 2011; Searle et al., 2009). Altogether, these new results contribute to the ongoing debate regarding the uncertain origins of the Irish biota (e.g. McDevitt et al., 2020).

Although not proven, we believe that the most likely scenario is that *C. nemoralis* snails arrived directly in Ireland from the Pyrenean region, intentionally or accidentally assisted by humans. For the moment, the precise geographic origin remains unknown—identical mtDNA haplotypes are shared between multiple sites in the Pyrenees and Ireland, with mtDNA F_{st} values not significantly different from zero (Grindon & Davison, 2013). In the genomic study here, a single site (Segre) in the Eastern Pyrenees falls most closely with the Irish snails. However, this population in itself could be a hybrid population, with the actual Irish ancestor elsewhere.

It is reasonable to suppose that snails of the same lineage could also have been distributed along the western fringes of France. Of use here, an inspection of the data in Layton et al. (2019) shows that snails of mtDNA lineage C are found in central and North Western France (see their Figure 1; Deux-Sevres, Paimpont), whereas previously lineage C was reported from the Pyrenean region and Ireland only (Grindon & Davison, 2013). This finding is, therefore, consistent with the genomic evidence (especially the admixture analyses) that indicates a possible hybrid origin of snails in North-West France (La Roche; Table S5). However, the specific mitochondrial C type haplotype found in France has not yet been found in Ireland. It is,

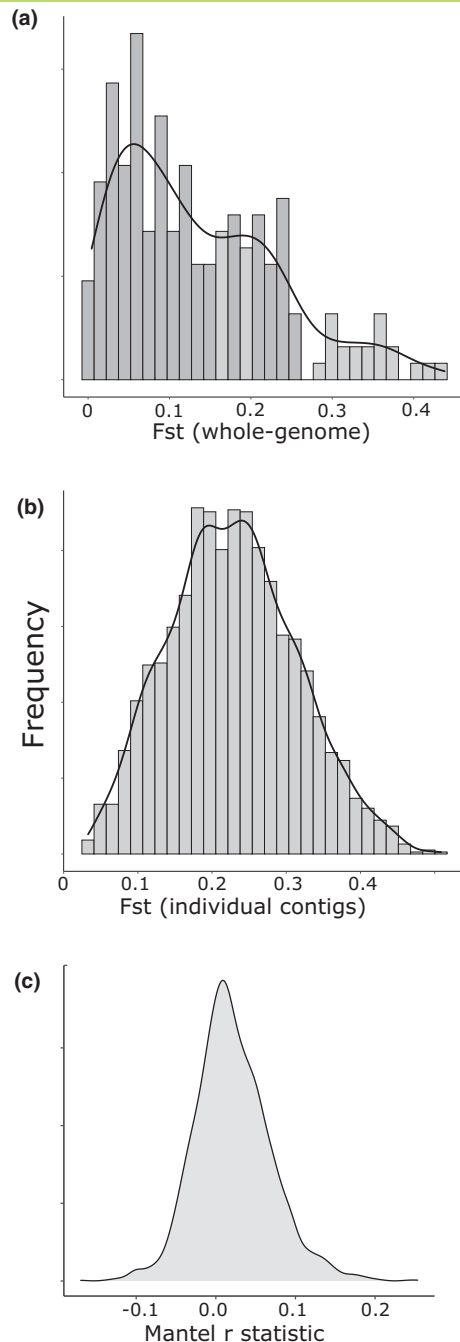| Bud	Dub	Lis	Mul	Ben	Com	Jue	Ron	San_1	San_2	Seg	Vie
0.37											
0.38	0.02										
0.30	0.09	0.08									
0.37	0.10	0.15	0.08								
0.31	0.04	0.09	0.04	0.02							
0.35	0.11	0.15	0.05	0.06	0.01						
0.29	0.22	0.28	0.17	0.11	0.07	0.14					
0.29	0.20	0.21	0.21	0.22	0.18	0.24	0.05				
0.28	0.19	0.22	0.20	0.22	0.20	0.24	0.05	0.01			
0.25	0.03	0.00	0.04	0.10	0.07	0.03	0.03	0.19	0.19		
0.34	0.09	0.14	0.04	0.05	0.03	0.05	0.11	0.24	0.24	0.09	

FIGURE 5 Histograms showing how genetic differentiation varies between individuals and between loci. (a) Estimates of F_{st} between individuals based on whole-genome data. (b) Estimates of F_{st} between individuals for each of the 2323 genome contigs (c) estimated mantel r statistic, derived by testing the correlation of F_{st} for each of the 2323 genome contigs against geographic distance, using Euclidean Haversine distances

therefore, feasible that these populations in Ireland and France are the result of separate admixture events.

In terms of dating, fossil material indicates that *Cepaea nemoralis* may have colonized Britain first, perhaps 1000 years before Ireland, the latter of which was at least 8000 years ago (Newlands Cross, Co. Dublin: 7600+/−500 BP Cartronmacmanus, Co. Mayo:

8207+/−165; Kerney et al., 1980; Preece et al., 1986; Speller, 2006). Assuming that these relative dates are correct, then a new arrival of *C. nemoralis* in Ireland would have quickly become established in the vacant habitat. In comparison, in Britain, new arrivals would have had to compete with more numerous existing snails and so perhaps only became established on the fringes of the distribution. Populations have since hybridized and introgressed. In perhaps a similar manner, founding populations of mice were likely transported by humans (Jones et al., 2013). Moreover, *C. nemoralis* are arguably 'synanthropic', liking disturbed habitats but also being consumed as food source, especially in Iberia, since at least 11 000 years before the present (Gutiérrez Zugasti, 2011), and they are commonly found as an item of food refuse in Mediterranean archaeological sites just prior to the onset of agricultural economies (although not in Ireland – Eva Laurie, pers. comm.). Recently founded populations of *C. nemoralis* also illustrate the ease with which they may be introduced by humans, especially in disturbed habitats, including in much of Poland (Ożgo et al., 2019), and parts of Sweden (Cameron & von Proschwitz, 2020), the Czech Republic (Peltanová et al., 2012), Hungary (Budapest; Szili-Kovács et al., 2012) and Russia (Moscow; Egorov & Sverlova, 2021).

4.2 | Colour and genomic geographic variation

One of the characteristics of the genus *Cepaea* is that all, or nearly all, populations show some degree of colour or banding polymorphism (Cook, 2017; Jones et al., 1977). Nonetheless, large-scale and local surveys of the geographical distribution of shell ground colour have shown clear geographical patterns (Cameron & Cook, 2012; Cook, 2014; Davison et al., 2019; Ożgo & Schilthuizen, 2012; Ramos Gonzalez & Davison, 2021; Silvertown et al., 2011; Worthington et al., 2012). Future studies that aim to test trends in shell variation may, therefore, wish to take account of this knowledge in analyses. For the moment, the whole-genome data enable a more nuanced understanding, for example, showing that large scale differentiation due to isolation by distance is mainly driven by about a third of the genome (Figure 3c).

To what extent is the lack of geographic differentiation for most loci due to neutral processes such as drift (enabled by gene flow), or else because of a selective sweep? Unfortunately, the relatively low contiguity of the genome assembly and the relatively shallow ddRAD sampling means that the present data set lacks power for more sophisticated analysis, such as a genome scan. It is, therefore, premature to understand the relative contribution of chance events, gene flow or selection in explaining patterns at individual loci. In the future, we hope that it should be possible to identify the genes involved in determining the polymorphism, in which case it would be fruitful to compare associations of colour alleles with geography. For the moment, it is our impression that the yellows, pinks and even browns of many Pyrenean (and Irish) populations are somewhat different—more 'exuberant' as Franks and Oxford (2009) might put it—compared with other populations

FIGURE 6 Correlations of F_{st} between individuals versus geographic distance using the ddRAD data, a) shows the F_{st} (averaged across all loci) versus geographic distance. b) Shows the same data but transformed to represent a Euclidian genetic distance and geographic Haversine distance
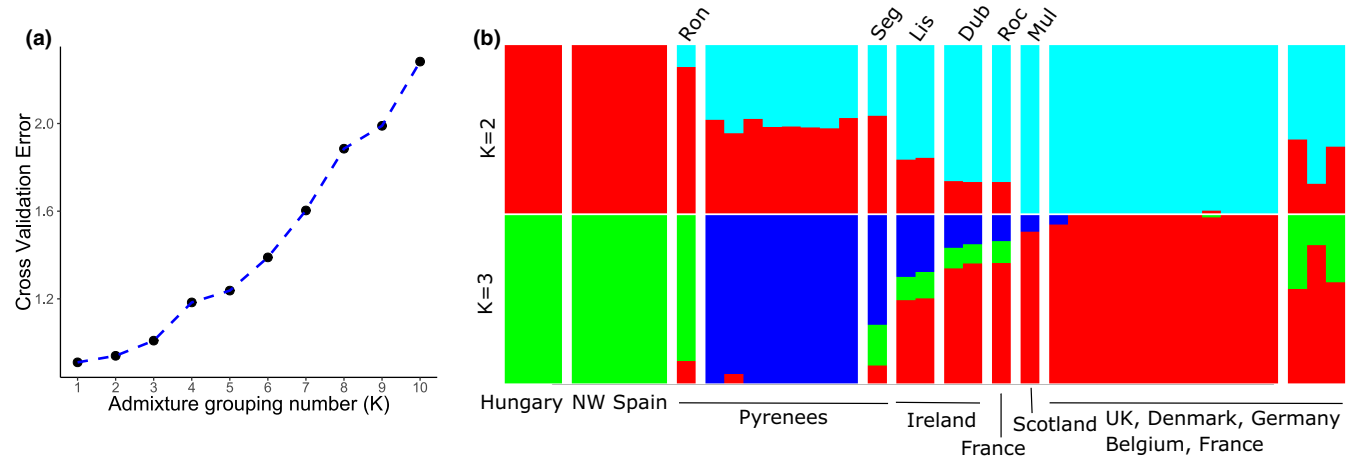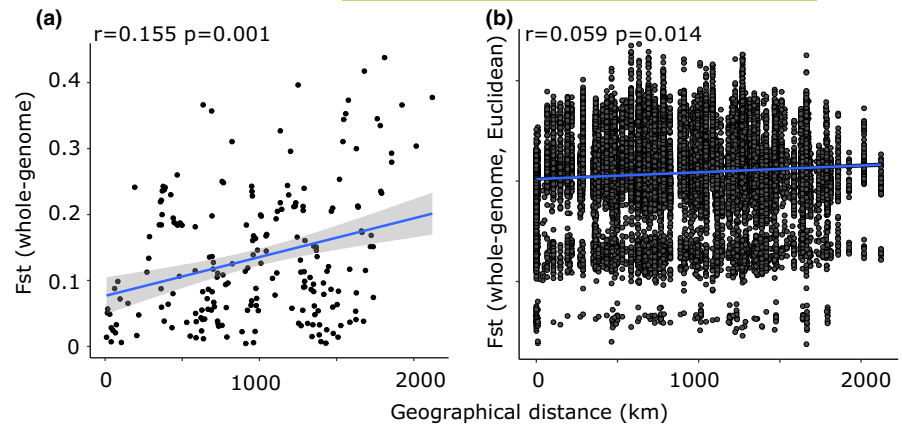

FIGURE 7 Population clustering using ADMIXTURE. a) Output of cross-validation assay. b) Population clustering using K = 2 and K = 3. Each vertical bar represents an individual, with the proportion of each colour representing the inferred ancestry

(Figure 1b). Although only speculation, a possible explanation is that the alleles for colour in the separate populations have also diverged, which should be evident in any phylogenies based on colour-linked loci.

In some other species, such as *Heliconius* butterflies, phylogenies based on colour-linked loci show strong associations with the wing colour. This is especially the case when markers are tightly linked and occur irrespective of the species (e.g. Pardo-Diaz et al., 2012). As our previous work showed that colour is 'indiscrete' or continuous, and not three separate colours in *Cepaea* (Davison et al., 2019), further sequence analysis of colour alleles will inform how selection has impacted upon the genome and more specifically, the nature of this selection.

4.3 | The evolutionary history and future of *Cepaea*

A recent revision of the taxonomy means the genus *Cepaea* now contains only two species, *C. nemoralis* and *C. hortensis*, with *C. sylvatica* and *C. vindobonesis* removed to another genus, *Caucasotachea* (Kajtoch et al., 2017). As the known sister taxa to *Cepaea* are Iberian (Neiber et al., 2016, 2021; Neiber & Hausdorf, 2015a), then a guess

is that the ancestor of *Cepaea* itself may have originated in Iberia, subsequently colonizing the rest of Europe. Unfortunately, the genomic data presented here do not easily inform on the origination of the genus or the evolution of the polymorphism. The mitochondrial DNA data on their own are derived from the short 'barcoding' region, with potential outgroups to *Cepaea* rather distant. In consequence, although the phylogeny (Figure S1) indicates that lineages from Italy may be basal, this positioning is dependent upon an outgroup which may not be appropriate.

In comparison, the data presented here are more informative of relatively recent events. Certainly, there are multiple divergent lineages in Iberia (mtDNA and genome), which could equally imply the origination of the species there, but caution is required because there are also divergent populations and lineages in other likely refugia, especially Italy/Croatia (mtDNA) and Hungary (ddRAD—population recently established and so must be derived from elsewhere?). Also, although this and the previous work (Grindon & Davison, 2013) show that much of mainland Europe was likely colonized by snails with mtDNA lineages A and B, the actual refugial location(s) are not clear. Further fine sampling is required in Italy, the Balkans, Alps, Mediterranean coast, western coast of France and the valleys of the Pyrenees. For the moment,

on the basis of the evidence provided here, and the distribution of samples, it seems plausible that the central European population is derived from an as yet unsampled refugium in the Italian or alpine region, or perhaps adjacent Balkans—which would make the pattern most similar to that of the *Chorthippus parallelus* paradigm suggested by Hewitt (2004).

In the future, we believe that a higher level assembly of the *C. nemoralis* genome should enable subsequent studies on the evolutionary origins of the supergene, and the relative roles that natural selection, recombination and drift may play in the establishment and loss of colour polymorphism in species as a whole and, more specifically, local populations. The same phylogeographic and population genomic questions should also be addressed in the sister species *C. hortensis*, which has a more northerly distribution; many records for *C. hortensis* in the Iberian Peninsula are arguably erroneous, instead of being pale-lipped *C. nemoralis* (Puente et al., 1998, Ramos Gonzalez & Davison, 2021; see also records in e.g. iNaturalist). This species also colonized North America, likely more than 6000 years ago (Pearce et al., 2010), and is long present in Iceland, and perhaps once present in Greenland (Johnson, 1906). It is not known how this species colonized these locations, but population genomics would reveal whether it was from one or several sources and whether a 'stepping-stone' was involved (e.g. Europe – Iceland – North America).

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The Illumina ddRAD reads for each individual are available via the NCBI Bioproject accession PRJNA730503. New mtDNA haplotypes have Genbank accession numbers ON556571-ON556582. The mitochondrial alignment, ddRAD vcf files and phylip file used for the analyses are available in the Dryad database https://doi.org/10.5061/dryad.brv15dvc5. Some simple scripts are available on https://github.com/angusdavison.

ORCID

Daniel Ramos-Gonzalez 🅳 https://orcid.org/0000-0003-4497-6374
Suzanne V. Saenko 🅳 https://orcid.org/0000-0002-9019-6590
Angus Davison 🅳 https://orcid.org/0000-0003-2554-8585

REFERENCES

Adema, C. M. (2021). Sticky problems; extraction of nucleic acids from molluscs. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 376, 20200162.

Affenzeller, S., Wolkenstein, K., Frauendorf, H., & Jackson, D. J. (2020). Challenging the concept that eumelanin is the polymorphic brown banded pigment in *Cepaea nemoralis*. *Scientific Reports*, 10, 2442.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12, 246.

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

Brace, S., Ruddy, M., Miller, R., Schreve, D. C., Stewart, J. R., & Barnes, I. (2016). The colonization history of British water vole (*Arvicola amphibius* [Linnaeus, 1758]): Origins and development of the Celtic fringe. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 283, 20160130.

Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C. D., & Ostrer, H. (2010). Genomewide patterns of population structure and admixture among Hispanic/Latino populations. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 8954–8961.

Cameron, R., & von Proschwitz, T. (2020). *Cepaea nemoralis* (L.) on Öland, Sweden: Recent invasion and unexpected variation. *Folia Malacologica*, 28, 303–310.

Cameron, R. A. D. (1969). The distribution and variation of *Cepaea nemoralis* L. near Slievecarron, county Clare and county Galway, Eire. *Proceedings of the Malacological Society of London*, 38, 439–450.

Cameron, R. A. D., & Cook, L. M. (2012). Habitat and the shell polymorphism of *Cepaea nemoralis* (L.): Interrogating the evolution Megalab database. *Journal of Molluscan Studies*, 78, 179–184.

Chiba, S. (1999). Accelerated evolution of land snails *Mandarina* in the oceanic Bonin Islands: Evidence from mitochondrial DNA sequences. *Evolution*, 53, 460–471.

Chueca, L., Schell, T., & Pfenninger, M. (2021a). Whole genome resequencing data to infer historical demography and speciation processes in land snails: The study of two *Candidula* sister species. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 376, 20200156.

Chueca, L. J., Schell, T., & Pfenninger, M. (2021b). *De novo* genome assembly of the land snail *Candidula unifasciata* (Mollusca: Gastropoda). *bioRxiv*, 2021.01.23.427926. [PREPRINT]

Cook, L. M. (2003). The rise and fall of the *carbonaria* form of the peppered moth. *Quarterly Review of Biology*, 78, 399–417.

Cook, L. M. (2014). Morph frequency in British *Cepaea nemoralis*: What has changed in half a century? *Journal of Molluscan Studies*, 80, 43–46.

Cook, L. M. (2017). Reflections on molluscan shell polymorphisms. *Biological Journal of the Linnean Society*, 121, 717–730.

Cook, L. M., & Peake, J. F. (1960). A study of some populations of *Cepaea nemoralis* L. from the Dartry Mountains, co. Sligo, Ireland. *Proceedings of the Malacological Society of London*, 34, 1–11.

Cook, L. M., & Peake, J. F. (1962). Populations of *Cepaea nemoralis* from sand dunes of the Mullaghmore Peninsula, co. Sligo, Ireland; with a comparison with those from Annacoona, Darty Mts, co. Sligo. *Proceedings of the Malacological Society of London*, 35, 7–13.

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., & Group, GPA. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics*, *33*, 2037–2039.

Davison, A. (2000). An east-west distribution of divergent mitochondrial haplotypes in British populations of the land snail, *Cepaea nemoralis* (Pulmonata). *Biological Journal of the Linnean Society*, *70*, 697–706.

Davison, A. (2006). The ovotestis: An underdeveloped organ of evolution. *BioEssays*, *28*, 642–650.

Davison, A., & Clarke, B. (2000). History or current selection? A molecular analysis of 'area effects' in the land snail *Cepaea nemoralis*. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *267*, 1399–1405.

Davison, A., Jackson, H. J., Murphy, E. W., & Reader, T. (2019). Discrete or indiscrete? Redefining the colour polymorphism of the land snail *Cepaea nemoralis*. *Heredity*, *123*, 162–175.

Davison, A., & Neiman, M. (2021). Mobilising molluscan models and genomes in biology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *376*, 20200163.

de Weerd, D. R. U., Robinson, D. G., & Rosenberg, G. (2016). Evolutionary and biogeographical history of the land snail family Urocoptidae (Gastropoda: Pulmonata) across the Caribbean region. *Journal of Biogeography*, *43*, 763–777.

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., de Almeida, C., Regitano, L., Yildiz, M. A., Heaton, M. P., Liu, W.-S., Lei, C.-Z., Reecy, J. M., Saif-Ur-Rehman, M., … Taylor, J. F. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics*, *10*, e1004254.

Dussex, N., Alberti, F., Heino, M. T., Olsen, R. A., van der Valk, T., Ryman, N., Laikre, L., Ahlgren, H., Askeyev, I. V., Askeyev, O. V., Shaymuratova, D. N., Askeyev, A. O., Doppes, D., Friedrich, R., Lindauer, S., Rosendahl, W., Aspi, J., Hofreiter, M., Liden, K., … Diez-del-Molino, D. (2020). Moose genomes reveal past glacial demography and the origin of modern lineages. *BMC Genomics*, *21*, 854.

Ebbs, E. T., Loker, E. S., & Brant, S. V. (2018). Phylogeography and genetics of the globally invasive snail *Physa acuta* Draparnaud 1805, and its potential to serve as an intermediate host to larval digenetic trematodes. *BMC Evolutionary Biology*, *18*, 103.

Egorov, R., & Sverlova, N. G. (2021). Shell colour and banding polymorphism in *Cepaea nemoralis* (Gastropoda, Pulmonata, Helicidae) from the Moscow region. *Ruthenica*, *31*, 27–38.

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

Franks, D. W., & Oxford, G. S. (2009). The evolution of exuberant visible polymorphisms. *Evolution*, *63*, 2697–2706.

Gittenberger, E., Piel, W. H., & Groenenberg, D. S. J. (2004). The Pleistocene glaciations and the evolutionary history of the polytypic snail species *Arianta arbustorum* (Gastropoda, Pulmonata, Helicidae). *Molecular Phylogenetics and Evolution*, *30*, 64–73.

Gonzalez, D. R., Aramendia, A. C., & Davison, A. (2019). Recombination within the *Cepaea nemoralis* supergene is confounded by incomplete penetrance and epistasis. *Heredity*, *123*, 153–161.

Grindon, A. J., & Davison, A. (2013). Irish *Cepaea nemoralis* land snails have a cryptic Franco-Iberian origin that is most easily explained by the movements of Mesolithic humans. *PLoS One*, *8*, e65792.

Guo, Y., Zhang, Y., Liu, Q., Huang, Y., Mao, G., Yue, Z., Abe, E. M., Li, J., Wu, Z., Li, S., Zhou, X., Hu, W., & Xiao, N. (2019). A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *GigaScience*, *8*, giz124.

Gutiérrez Zugasti, F. I. (2011). Early Holocene land snail exploitation in northern Spain: The case of La Fragua cave. *Environmental Archaeology*, *16*, 36–48.

Haponski, A. E., Lee, T., & Foighil, D. O. (2019). Deconstructing an infamous extinction crisis: Survival of *Partula* species on Moorea and Tahiti. *Evolutionary Applications*, *12*, 1017–1033.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*, 183–195.

Hirano, T., Saito, T., Tsunamoto, Y., Koseki, J., Ye, B., Do, V. T., Miura, O., Suyama, Y., & Chiba, S. (2019). Enigmatic incongruence between mtDNA and nDNA revealed by multi-locus phylogenomic analyses in freshwater snails. *Scientific Reports*, *9*, 6223.

Jackson, H. J., Larsson, J., & Davison, A. (2021). Quantitative measures and 3D shell models reveal interactions between bands and their position on growing snail shells. *Ecology and Evolution*, *11*, 6634–6648.

Jari Oksanen, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R., Simpson, G. L., Solymos, P., & Stevens, M. H. H. (2018). *Vegan: Community ecology package*. R package version 2.

Johnson, C. W. (1906). On the distribution of *Helix hortensis* Mueller, in North America. *The Nautilus*, *20*, 73–80.

Jones, E. P., Eager, H. M., Gabriel, S. I., Johannesdottir, F., & Searle, J. B. (2013). Genetic tracking of mice and other bioproxies to infer human history. *Trends in Genetics*, *29*, 298–308.

Jones, J. S., Leith, B. H., & Rawlings, P. (1977). Polymorphism in *Cepaea*: A problem with too many solutions? *Annual Review of Ecology and Systematics*, *8*, 109–143.

Kajtoch, Ł., Davison, A., Grindon, A., Deli, T., Sramkó, G., Gwardjan, M., Kramarenko, S., Mierzwa-Szymkowiak, D., Ruta, R., Ścibior, R., Tóth, J. P., Wade, C., Kolasa, M., Egorov, R. V., & Fehér, Z. (2017). Reconstructed historical distribution and phylogeography unravels non-steppic origin of *Caucasotachea vindobonensis* (Gastropoda: Helicidae). *Organisms Diversity & Evolution.*, *17*, 679–692.

Kerkvliet, J., de Boer, T., Schilthuizen, M., & Kraaijeveld, K. (2017). Candidate genes for shell colour polymorphism in *Cepaea nemoralis*. *PeerJ*, *5*, e3715.

Kerney, M. P., Preece, R. C., & Turner, C. (1980). Molluscan and plant biostratigraphy of some late Devensian and Flandrian deposits in Kent. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *291*, 1–43.

Koch, E. L., Neiber, M. T., Walther, F., & Hausdorf, B. (2017). High gene flow despite opposite chirality in hybrid zones between enantiomorphic door-snails. *Molecular Ecology*, *26*, 3998–4012.

Kotlik, P., Markova, S., Konczal, M., Babik, W., & Searle, J. B. (2018). Genomics of end-Pleistocene population replacement in a small mammal. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *285*, 20172624.

Layton, K. K. S., Warne, C. P. K., Nicolai, A., Ansart, A., & deWaard, J. R. (2019). Molecular evidence for multiple introductions of the banded grove snail (*Cepaea nemoralis*) in North America. *Canadian Journal of Zoology*, *97*, 392–398.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

Liu, C. H., Ren, Y. W., Li, Z. Y., Hu, Q., Yin, L. J., Wang, H. C., Qiao, X., Zhang, Y., Xing, L. S., Xi, Y., Jiang, F., Wang, S., Huang, C., Liu, B., Liu, H. W., Wan, F. H., Qian, W. Q., & Fan, W. (2021). Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Molecular Ecology Resources*, *21*, 478–494.

Lucena-Perez, M., Marmesat, E., Kleinman-Ruiz, D., Martínez-Cruz, B., Wecek, K., Saveljev, A. P., Seryodkin, I. V., Okhlopkov, I., Dvornikov,

M. G., Ozolins, J., Galsandorj, N., Paunovic, M., Ratkiewicz, M., Schmidt, K., & Godoy, J. A. (2020). Genomic patterns in the widespread Eurasian lynx shaped by late quaternary climatic fluctuations and anthropogenic impacts. *Molecular Ecology*, 29, 812–828.

Mann, K., & Jackson, D. J. (2014). Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genomics*, 15, 249.

Marques, A. J. D., Hanson, J. O., Camacho-Sanchez, M., Martinez-Solano, I., Moritz, C., Tarroso, P., Velo-Anton, G., Verissimo, A., & Carvalho, S. B. (2022). Range-wide genomic scans and tests for selection identify non-neutral spatial patterns of genetic variation in a non-model amphibian species (*Pelobates cultripes*). *Conservation Genetics.*, 23, 387–400.

McDevitt, A. D., Coscia, I., Browett, S. S., Ruiz-González, A., Statham, M. J., Ruczyńska, I., Roberts, L., Stojak, J., Frantz, A. C., Norén, K., Ågren, E. O., Learmount, J., Basto, M., Fernandes, C., Stuart, P., Tosh, D. G., Sindicic, M., Andreanszky, T., Isomursu, M., ... Wójcik, J. M. (2020). Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox *Vulpes vulpes*, in Europe. *bioRxiv*, 2020.02.21.954966. [PREPRINT]

McDevitt, A. D., Vega, R., Rambau, R. V., Yannic, G., Herman, J. S., Hayden, T. J., & Searle, J. B. (2011). Colonization of Ireland: Revisiting 'the pygmy shrew syndrome' using mitochondrial, Y chromosomal and microsatellite markers. *Heredity*, 107, 548–557.

Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of parallel ecological speciation in Lake Victoria cichlids. *Molecular Biology and Evolution*, 35, 1489–1506.

Miura, O., Urabe, M., Mori, H., & Chiba, S. (2020). Ancient drainage networks mediated a large-scale genetic introgression in the east Asian freshwater snails. *Ecology and Evolution*, 10, 8186–8196.

Neiber, M. T., Chueca, L. J., Caro, A., Teixeira, D., Schlegel, K. A., Gómez-Moliner, B. J., Walther, F., Glaubrecht, M., & Hausdorf, B. (2021). Incorporating palaeogeography into ancestral area estimation can explain the disjunct distribution of land snails in Macaronesia and the Balearic Islands (Helicidae: Allognathini). *Molecular Phylogenetics and Evolution*, 162, 107196.

Neiber, M. T., & Hausdorf, B. (2015a). Molecular phylogeny reveals the polyphyly of the snail genus *Cepaea* (Gastropoda: Helicidae). *Molecular Phylogenetics and Evolution*, 93, 143–149.

Neiber, M. T., & Hausdorf, B. (2015b). Phylogeography of the land snail genus *Circassina* (Gastropoda: Hygromiidae) implies multiple Pleistocene refugia in the western Caucasus region. *Molecular Phylogenetics and Evolution*, 93, 129–142.

Neiber, M. T., Razkin, O., & Hausdorf, B. (2017). Molecular phylogeny and biogeography of the land snail family Hygromiidae (Gastropoda: Helicoidea). *Molecular Phylogenetics and Evolution*, 111, 169–184.

Neiber, M. T., Sagorny, C., & Hausdorf, B. (2016). Increasing the number of molecular markers resolves the phylogenetic relationship of '*Cepaea*' *vindobonensis* (Pfeiffer 1828) with *Caucasotachea* Boettger 1909 (Gastropoda: Pulmonata: Helicidae). *Journal of Zoological Systematics and Evolutionary Research*, 54, 40–45.

Ochman, H., Jones, J. S., & Selander, R. K. (1983). Molecular area effects in *Cepaea*. *Proceedings of the National Academy of Sciences of the United States of America*, 80, 4189–4193.

Ożgo, M., Cameron, R. A. D., Horsák, M., Pokryszko, B., Chudaś, M., Cichy, A., Kaczmarek, S., Kobak, J., Marzec, M., Mierzwa-Szymkowiak, D., Parzonko, D., Pyka, G., Rosin, Z., Skawina, A., Soroka, M., Sulikowska-Drozd, A., Surowiec, T., Szymanek, M., Templin, J., ... Żołądek, J. (2019). *Cepaea nemoralis* (Gastropoda: Pulmonata) in Poland: Patterns of variation in a range-expanding species. *Biological Journal of the Linnean Society*, 127, 1–11.

Ożgo, M., & Schilthuizen, M. (2012). Evolutionary change in *Cepaea nemoralis* shell colour over 43 years. *Global Change Biology*, 18, 74–81.

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O., & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, 8, e1002752.

Parmakelis, A., Kotsakiozi, P., & Rand, D. (2013). Animal mitochondria, positive selection and cyto-nuclear coevolution: Insights from pulmonates. *PLoS One*, 8, e61970.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192, 1065–1093.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2, e190.

Pearce, T. A., Olori, J. C., & Kemezis, K. W. (2010). Land snails from St. Elzéar cave, Gaspé peninsula, Quebec: Antiquity of *Cepaea hortensis* in North America. *Annals of Carnegie Museum*, 76, 65–78.

Peltanová, A., Dvořák, L., & Juřičková, L. (2012). The spread of non-native *Cepaea nemoralis* and *Monacha cartusiana* (Gastropoda: Pulmonata) in The Czech Republic with comments on other land snail immigrants. *Biologia*, 67, 384–389.

Petr, M., Vernot, B., & Kelso, J. (2019). Admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics*, 35, 3194–3195.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8, e1002967.

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M. G., & Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344, 1410–1414.

Preece, R. C., Coxon, P., & Robinson, J. E. (1986). New biostratigraphic evidence of the post-glacial colonization of Ireland and for mesolithic forest disturbance. *Journal of Biogeography*, 13, 487–509.

Puente, A. I., Altonaga, K., Prieto, C. E., & Rallo, A. (1998). Delimitation of biogeographical areas in the Iberian Peninsula on the basis of Helicoidea species (Pulmonata: Stylommatophora). *Global Ecology and Biogeography Letters*, 7, 97–113.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., 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.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramos Gonzalez, D., & Davison, A. (2021). Qualitative and quantitative methods show stability in patterns of *Cepaea nemoralis* shell polymorphism in the Pyrenees over five decades. *Ecology and Evolution*, 11, 6167–6183.

Ravinet, M., & Meier, J. I. (2020). *Speciation & population genomics: A how-to-guide*. Vol. 2022. https://speciationgenomics.github.io/

Razkin, O., Sonet, G., Breugelmans, K., Madeira, M. J., Gomez-Moliner, B. J., & Backeljau, T. (2016). Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex Pyramidula: The power of RADseq data. *Molecular Phylogenetics and Evolution*, 101, 267–278.

Richards, P. M., Liu, M. M., Lowe, N., Davey, J. W., Blaxter, M. L., & Davison, A. (2013). RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology*, 22, 3077–3089.

Russello, M. A., Waterhouse, M. D., Etter, P. D., & Johnson, E. A. (2015). From promise to practice: Pairing non-invasive sampling with genomics in conservation. *PeerJ*, 3, e1106.

Saenko, S. V., Groenenberg, D. S. J., Davison, A., & Schilthuizen, M. (2021). The draft genome sequence of the grove snail *Cepaea nemoralis*. *G3 (Bethesda)*, 11, jkaa071.

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8, 205–233.

Searle, J. B., Kotlik, P., Rambau, R. V., Markova, S., Herman, J. S., & McDevitt, A. D. (2009). The Celtic fringe of Britain: Insights from small mammal phylogeography. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *276*, 4287–4294.

Silvertown, J., Cook, L., Cameron, R., Dodd, M., McConway, K., Worthington, J., Skelton, P., Anton, C., Bossdorf, O., Baur, B., Schilthuizen, M., Fontaine, B., Sattmann, H., Bertorelle, G., Correia, M., Oliveira, C., Pokryszko, B., Ozgo, M., Stalazas, A., ... Juan, X. (2011). Citizen science reveals unexpected continental scale evolutionary change in a model organism. *PLoS One*, *6*, e18927.

Speller, G. R. (2006). Molluscan biostratigraphy, stable isotope analyses and dating of Irish Holocene tufas. In *Zoology*. University of Cambridge. PhD thesis.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313.

Staples-Browne, R. (1908). On the inheritance of colour in domestic pigeons, with special reference to reversion. *Proceedings of the Zoological Society of London*, *78*, 67–104.

Szili-Kovács, T., Majoros, G., & Csontos, P. (2012). *Cepaea nemoralis* (Linnaeus, 1758) emergence associated with ornamental planting in Budapest.[Dísznövény-telepítéssel összefüggésbe hozható Cepaea nemoralis (Linnaeus, 1758) felbukkanások Budapesten]. *Malakológiai Tájékoztató/Malacological Newsletter*, *30*, 39–44.

Thomaz, D., Guiller, A., & Clarke, B. (1996). Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *263*, 363–368.

Wellenreuther, M., Svensson, E. I., & Hansson, B. (2014). Sexual selection and genetic colour polymorphisms in animals. *Molecular Ecology*, *23*, 5398–5414.

Westram, A. M., Rafajlovic, M., Chaube, P., Faria, R., Larsson, T., Panova, M., Ravinet, M., Blomberg, A., Mehlig, B., Johannesson, K., & Butlin, R. (2018). Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters*, *2*, 297–309.

Wheldale, M. (1907). The inheritance of flower colour in *Antirrhinum majus*. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *79*, 288–305.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*, 1686.

Williams, S. T. (2017). Molluscan shell colour. *Biological Reviews (Cambridge)*, *92*, 1039–1058.

Worthington, J. P., Silvertown, J., Cook, L., Cameron, R., Dodd, M., Greenwood, R. M., McConway, K., & Skelton, P. (2012). Evolution MegaLab: A case study in citizen science methods. *Methods in Ecology and Evolution*, *3*, 303–309.

Xu, J., & Hausdorf, B. (2021). Repeated hybridization increased diversity in the door snail complex *Charpentieria itala* in the southern Alps. *Molecular Phylogenetics and Evolution*, *155*, 106982.

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: Ramos-Gonzalez, D., Saenko, S. V., & Davison, A. (2022). Deep structure, long-distance migration and admixture in the colour polymorphic land snail *Cepaea nemoralis*. *Journal of Evolutionary Biology*, *35*, 1110–1125. https://doi.org/10.1111/jeb.14060