Received: 29 January 2020 | Revised: 21 August 2020 | Accepted: 26 August 2020
DOI: 10.1111/mec.15630

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Rapid homoploid hybrid speciation in British gardens: The origin of Oxford ragwort (*Senecio squalidus*)

Bruno Nevado[1] | **Stephen A. Harris**[1] | **Mark A. Beaumont**[2] | **Simon J. Hiscock**[1,3]

[1]Department of Plant Sciences, University of Oxford, Oxford, UK
[2]School of Biological Sciences, University of Bristol, Bristol, UK
[3]Oxford Botanic Garden and Arboretum, Oxford, UK

Correspondence
Bruno Nevado, Department of Plant Sciences, University of Oxford. South Parks Road, OX1 3RB Oxford, UK.
Email: bruno.nevado@plants.ox.ac.uk

Present address
Bruno Nevado, CE3C – Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

Funding information
University of Oxford, Grant/Award Number: 162/091

Abstract

Hybridisation can lead to homoploid hybrid speciation, i.e., the origin of new species without change in chromosome number between parents and offspring. Central to homoploid hybrid speciation is the role of hybridisation in the establishment of reproductive isolation between the hybrid and the parental species in the early stages of speciation, when typically all species occur at least partly in sympatry. In this work we analyse genome-wide polymorphism data obtained by transcriptome sequencing of the British hybrid species Oxford ragwort (*Senecio squalidus*, Asteraceae), its two Italian parental species (*S. aethnensis* and *S. chrysanthemifolius)* and their naturally occurring hybrids on Mt Etna (Italy). We show that Oxford ragwort most likely originated from de novo hybridisation between its two Italian parental species whilst they were in cultivation in British gardens at the turn of the 18th century. Reproductive isolation between the new hybrid species and its parental species probably resulted from inheritance of genetic incompatibilities between the two parental species and subsequent ecological segregation – both of which have been shown in previous studies. Our results imply that *S. squalidus* meets the most stringent criteria set forth to identify homoploid hybrid speciation, and call attention to the creative role of hybridisation in responding to novel environmental conditions.

KEYWORDS
ABC analysis, demographic inference, homoploid hybrid speciation, hybridisation, Oxford ragwort, RNA sequencing

1 | INTRODUCTION

Hybridisation between species is a well-accepted phenomenon that can impact both adaptation and speciation (Abbott et al., 2013; Arnold, 1997). Potential outcomes of hybridisation include homoploid hybrid speciation, i.e., the origin of a new species without a change in chromosome number. Although putative cases have been described in both animals (Mavarez et al., 2006) and plants (Gross & Rieseberg, 2005; Rieseberg et al., 2003), this form of speciation is thought to be rare in nature (Schumer et al., 2014; Yakimowski & Rieseberg, 2014). Central to homoploid hybrid speciation is the establishment of reproductive isolation between the hybrid and the parental species in the early stages of speciation, when typically all three species occur at least partly in sympatry. Here, hybridisation can promote reproductive isolation in at least two ways: chromosomal rearrangements and genetic incompatibilities between species can cause post-zygotic reproductive isolation (Lai et al., 2005; Rieseberg et al., 1995, 1999); and new ecological or reproductive traits in the hybrid species can cause prezygotic reproductive isolation (Gross & Rieseberg, 2005; Mavarez et al., 2006). Establishing conclusively that homoploid hybrid speciation has occurred requires showing that at least one of these barriers evolved as a direct result

FIGURE 1 Herbaria specimens from the beginning of the 18th century showcasing the range of variation of *Senecio* specimens growing in (a) the gardens of Oxford, and (b–d) Badminton. (a) Individuals present in the Oxford Botanic Garden at this time resemble typical *Senecio chrysanthemifolius* ("From Mr Jacob Bobart of Oxford"; OXF; Dubois-5439). (b) Typical *Senecio squalidus* individuals were already present in Badminton ("One of Dr Sherards Badminton Specimens"; OXF; Dubois-5437). (c, d) Other individuals growing in Badminton encompassed the range of variation from *S. chrysanthemifolius*-like (c: "From the Duchess of Beauforts Garden at Badminton"; OXF; Dubois-5438) to *S. aethnensis*-like (d: "ex semine a P. Cupani misso, succravit in Horto Badminton" [paraphrase: "seed from Cupani, it grew in the garden at Badminton"]; OXF; Sherard-5318). Labels for each specimen contain also William Sherard's listing of the polynomials associated with it

of hybridisation, which has been achieved in only a handful of studies (Schumer et al., 2014).

Senecio squalidus L. (Oxford ragwort) is a short-lived, perennial herb commonly found along railway lines, motorway verges and waste-ground in urban areas throughout Britain. This species is derived from material introduced to Britain from Italy in the late 17th century (Abbott et al., 2000; Harris, 2002; James & Abbott, 2005), and subsequently grown more or less continuously at the Oxford Botanic Garden (Abbott et al., 2009). By the second half of the 18th century the species was growing on old college walls outside the Oxford Botanic Garden (Kent, 1956, 1960), and over the next two centuries it spread throughout the UK, aided by the establishment of the railway system, roads and ultimately motorways (Abbott et al., 2009; Harris, 2002).

The origin of *S. squalidus* from interspecific crosses between two Italian species that form a hybrid zone on Mt Etna, Sicily (*Senecio aethnensis* Jan *ex* DC. and *Senecio chrysanthemifolius* DC.), is well supported by both morphological (Abbott et al., 2000, 2009; Brennan et al., 2012) and genetic data (Abbott et al., 2000; James & Abbott, 2005). Notwithstanding this, the source of hybrid material introduced to the UK and the early evolutionary history of *S. squalidus* remains elusive (Abbott et al., 2009; Harris, 2002). The prevailing view is that this species arose from hybrid material collected from the hybrid zone on Mt Etna and introduced into the Oxford Botanic Garden in the late 17th century (Abbott et al., 2000; Druce, 1927; James & Abbott, 2005; Kent, 1956). Indeed, George Druce's (1927) description of *S. squalidus* was used by Crisp (1972) to propose that the *S. squalidus* introduced to Britain came from a natural hybrid zone between *S. aethnensis* and *S. chrysanthemifolius* on Etna. Druce's evidence for a pre-eighteenth century Sicilian introduction of *S. squalidus* to the Oxford Botanic Garden was based on a description and illustration in Morison (1699) and herbarium specimens apparently collected by Jacob Bobart. However, the description and illustration are copied from Boccone (1674), with the herbarium specimens unlikely to be as old as the late seventeenth century (Harris, 2002). Moreover, morphological (Figure 1a) and distributional evidence are consistent with Boccone's plant being *S. chrysanthemifolius*. An alternative scenario posits that *S. squalidus* originated from recent hybridisation while in cultivation in the UK. Based on early-eighteenth-century herbarium specimens, and known contacts among gardens in Sicily and England, specimens spanning the range of variation from *S. aethnensis* to *S. chrysanthemifolius* were growing at Badminton (Harris, 2002). The most likely source of the Badminton plants was William Sherard, via Francisco Cupani at the Hortus Catholicus, Palermo (Harris, 2002). Catalogues of the Hortus Catholicus (Cupani, 1694, 1696) show that polynomial names for taxa matching descriptions of *S. aethnensis*, *S. chrysanthemifolius* and *S. squalidus* were being grown in Badminton (Figure 1b–d), and hybrid material from their crosses was probably introduced to the Oxford Botanic Garden at this time (Harris, 2002).

Revealing the early evolutionary history of *S. squalidus* is central to understanding the role of hybridisation in its origin and the process of homoploid hybrid speciation more generally. If *S. squalidus* resulted from introduction of hybrid material from Mt Etna into the UK, it follows that reproductive isolation was due to human-mediated geographic isolation between hybrid and parental species. However, if *S. squalidus* originated from more recent hybridisation between *S. chrysanthemifolius* and *S. aethnensis* in the UK, it follows that hybrids and parental species may have been growing together in the early stages of divergence at Badminton and Oxford. In this case, reproductive isolation must have evolved by means other than simple spatial (allopatric) isolation, possibly driven by ecological differences (ecological homoploid hybrid speciation) or sorting of genetic incompatibilities (recombinational homoploid hybrid speciation) between parental and hybrid species (Grant, 1981; Rieseberg, 1997; Yakimowski & Rieseberg, 2014).

Here, we use transcriptome data from multiple individuals of *S. squalidus* collected from throughout its distribution range in the UK, together with transcriptome data from its two parental species and their naturally occurring hybrids sampled on Mt Etna, to explore both scenarios for the homoploid hybrid origin of *S. squalidus* and to investigate the genome-wide effects of homoploid hybrid speciation.

2 | MATERIALS AND METHODS

2.1 | Data collection

We collected plant cuttings from 28 individuals of *S. squalidus* sampled from throughout the UK (Table S1), and grew them in glasshouses in Oxford, UK in a mixture of soil and perlite (4:1) with natural sunlight supplemented with artificial light to maintain a 16:8 hr light-dark period. For *S. aethnensis* and *S. chrysanthemifolius* we used previously published data (Chapman et al., 2013), together with six new samples of each species grown in glasshouses in Catania, Italy (Table S1). Individuals from the naturally-occurring hybrid zone on Mt Etna ($n = 14$) were grown in the glasshouse in Oxford from achenes collected in the field. To maximise the number of genes analysed plants were allowed to flower and different tissues (inflorescence, stems and young leaves) were flash-frozen in liquid nitrogen and kept in −80°C until RNA extraction.

We extracted total RNA from each sample using the RNeasy Plant extraction kit (Qiagen) following manufacturer's protocol and including the additional DNA digestion step with DNase. Sample quality and total RNA concentration were assessed with 1% agarose gel, Nanodrop and Qubit. High-throughput sequencing libraries were prepared at the genomics facility of the Wellcome Centre for Human Genetics (Oxford, UK). RNA extracts were enriched for mature poly-A transcripts by annealing to oligo-(dT) magnetic beads, and Illumina Truseq libraries prepared, size-selected and multiplexed. Sequencing was performed on the Illumina HiSeq platform with a minimum of 50 million reads obtained per sample (Table S1). We quality-checked raw read data using FASTQC v0.11 (Andrews, 2010) and removed adaptors and low quality bases (Phred score below 20) with TRIM_GALORE v0.4 (Krueger, 2015). We chose to use RNAseq data in this study because this is a cost-effective approach to obtain genome-wide polymorphism data, and it furthermore allowed us to increase the sample sizes of the parental species by reanalysing previously published RNAseq data (Chapman et al., 2013).

We assembled a reference transcriptome for *S. squalidus* using TRINITY v2.6 (Haas et al., 2013) with default settings and including an initial trimming step to remove low quality bases with TRIMMOMATIC (Haas et al., 2013). We calculated expression values for each sample using RSEM v1.3 (Haas et al., 2013), and filtered transcripts with a very low normalised expression level in all samples (Transcripts Per Million, TPM < 2). We identified and removed vector contaminants by blasting all transcripts against the univec NCBI database, and other contaminants by blasting against NCBIs nonredundant protein database (we excluded hits to nonviridiplantae organisms that were longer than 20 aminoacids and with identity percentage above 98%). We identified putative coding sequences within each transcript de novo using TRANSDECODER v3.0 (Haas et al., 2013) and annotated each transcript using the TRINOTATE pipeline (Bryant et al., 2017). Chloroplast genes were identified and removed from subsequent analysis by blasting against the complete chloroplast genome of *Jacobaea vulgaris* Gaertn. (Doorduin et al., 2011), with identity threshold of 95 and minimum length of 50 bp.

We mapped trimmed reads from each sample to the newly assembled *S. squalidus* transcriptome reference using BWA v0.7 (Li & Durbin, 2009) with the mem algorithm, marked duplicated reads with PICARD v1.1 (available from http://broadinstitute.github.io/picard) and performed indel realignment with GATK v3.4 (McKenna et al., 2010). We obtained genotype calls using the SAMTOOLS package v1.2 (Li et al., 2009) for positions covered by at least eight reads after removing reads with mapping quality below 20 (−q 20) and bases with base quality below 20 (−Q 20). We included in the output homozygous-reference regions with a minimum depth of eight reads (−g 8), and excluded single-nucleotide polymorphisms (SNPs) within 3 bp of an indel (−g3). We further filtered out SNPs with quality below 15 and heterozygous SNPs with fewer than two reads supporting each allele, and converted resulting VCF files into fasta format using VCF2FAS (available from https://github.com/brunonevado/vcf2fas). For subsequent analyses, for each gene we used only the isoform with less missing data across all samples, and excluded genes with more than 80% missing data overall.

2.2 | Population structure, polymorphism and linkage disequilibrium

To characterise genetic structure among species, we randomly selected one SNP within 4-fold degenerate sites for each gene, and identified the number of clusters of the data set that maximised the marginal likelihood with FASTSTRUCTURE v1.0 (Raj et al., 2014) within the range $K = [1, 6]$. We further visualised genetic structure by performing a principal component analysis (PCA) based on the correlation matrix (after centering and scaling the allele frequency data) with the R package ADEGENET v2.1 (Jombart & Ahmed, 2011). For *S. squalidus*, we tested these genome-wide randomly selected SNPs for excess heterozygosity using the global Hardy-Weinberg test implemented in GENEPOP v. 4.7 (Rousset, 2008). Only SNPs without missing data were considered for these analyses.

For each gene with a single best open reading frame over 300 bp long (identified by TRANSDECODER) we calculated polymorphism (Watterson's theta estimator) neutrality tests (Tajima's *D*) and divergence (F_{ST}) using MSTATSPOP v0.1 (available from https://github.com/CRAGENOMICA/mstatspop). For each gene we excluded individuals with fewer than 200 bp scored, after which we removed genes with more than 50% missing data and with fewer than 10 individuals sequenced. For each gene, we calculated the same measures of polymorphism and divergence considering only of 4-fold degenerate sites obtained with GET4FOLDSITES (available from https://github.com/brunonevado/get4foldSites), after excluding genes with fewer than 100 four-fold degenerate sites. For estimates of divergence between *S. squalidus* and each parental species and hybrid population, we only considered genes with at least 10 individuals sequenced in each population. To estimate linkage disequilibrium (LD) for each species we first phased SNPs within each gene using FASTPHASE v1.4 (Scheet & Stephens, 2006) with default values. Genotypes with a posterior probability below 0.9 were coded as missing and genes shorter than 1,000 bp were excluded. We then performed chi-square tests of independence between all pairs of SNPs (within each gene) with at least 20 haplotypes scored. Results were binned according to distance between SNPs, and the proportion of significant chi-square tests ($p < .001$) counted within each bin. To account for the effect of sample size on power to detect LD between SNPs, for each gene we randomly selected 16 *S. squalidus* individuals for this analysis (i.e., the same number of samples we analysed for each parental species). We note that the distance between SNPs used in this study was calculated along transcripts, as performed in previous work (Chapman et al., 2013). This means that the physical distance along the genome will be larger for many of the SNP pairs considered. However, the comparative framework employed herein is robust to this issue insofar as there are no consistent differences in the distance between SNP pairs on the different taxa—a bias which is not present in the data set (Figure S1).

2.3 | Approximate Bayesian Computation analysis

To understand the early origin of *S. squalidus* we investigated two alternative evolutionary scenarios (described in the main text). We simulated 75,000 data sets for each scenario using the coalescent simulation software MS (Hudson, 2002). Each simulation consisted of 50 loci of 10,000 bp each, and the summary statistics used (Table S2) were calculated for each simulation as averages across all 50 loci. For each data set, the recombination rate between adjacent sites was randomly selected from a uniform distribution $U(10^{-10}, 10^{-8})$, and recombination was allowed to occur at any site along a locus. The resulting data was analysed with two alternative Approximate Bayesian Computation (ABC) approaches: a regression-based approach using neural networks as implemented in the R package ABC (Csilléry et al., 2012); and the random forests approach implemented in the R package ABCRF (Pudlo et al., 2016).

The parameters used for simulation of each scenario (and the priors used) are denoted in Table 1 and are based on current knowledge of this system. In particular, we set a prior on divergence time between the two parental *Senecio* species of 100,000–250,000 generations ago (Osborne et al., 2013), and allowed bidirectional gene flow between species mediated by the hybrid zone (modelled as a third population) as inferred previously (Osborne et al., 2013). Effective population size of *S. aethnensis* was set to 100,000–500,000 individuals (Osborne et al., 2013), and that of *S. chrysanthemifolius* and of the hybrid zone population were set to between half and 1.5× that of *S. aethnensis* based on their similar levels of polymorphism. Given the large area currently occupied by *S. squalidus* compared to its parental species, we set a prior on effective population size of this species of 1× to 1,000× that observed in *S. aethnensis*. For the evolutionary history of *S. squalidus*, we set three important times according to historical records (Harris, 2002): introduction of material to the UK (*T1*), naturalisation in the UK (*T2*) and beginning of exponential expansion as *S. squalidus* started to spread across the UK (*T3*). We assumed a generation time of 1 year as *S. squalidus* typically flowers within

TABLE 1 Demographic history parameters and priors used in simulations

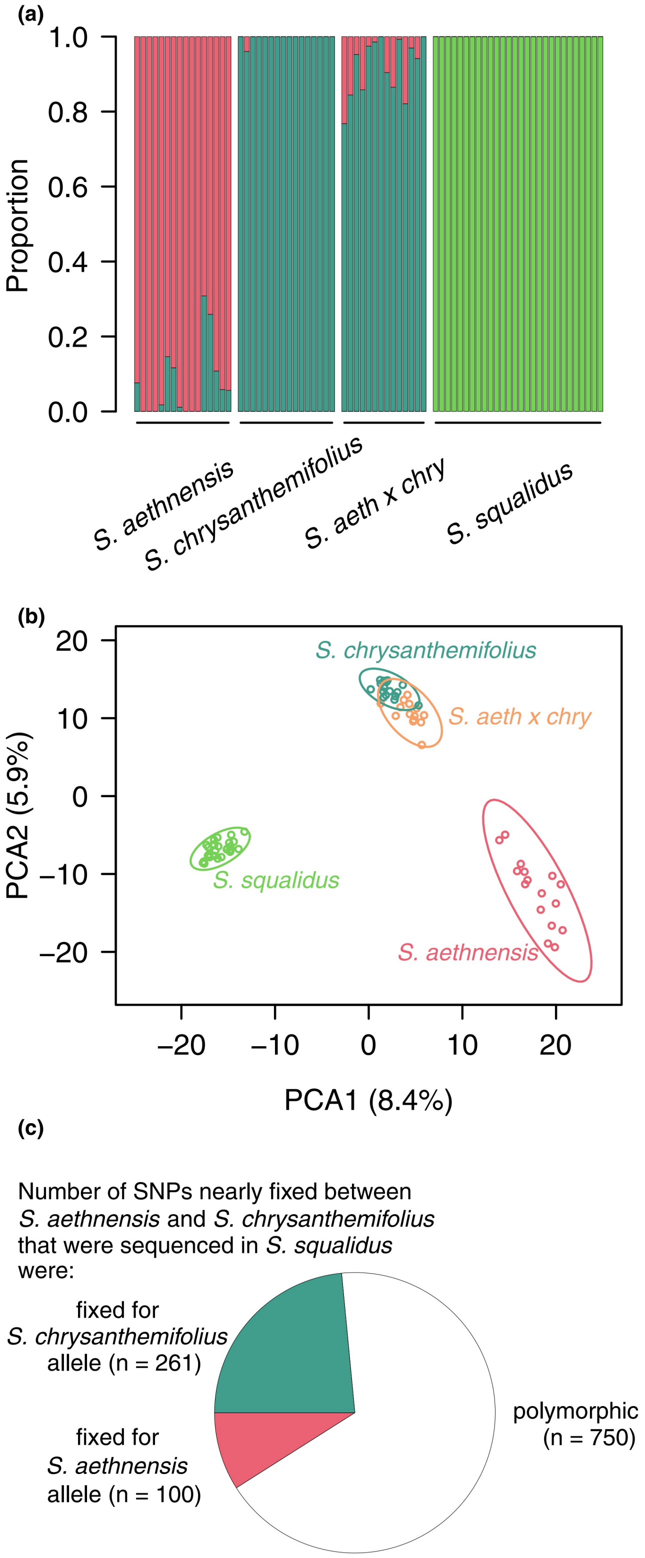
Parameter	Description	Prior	Posterior*
MEtna1	Migration rate parameter between *Senecio chrysanthemifolius* and the hybrid population (symmetric in each direction)	Uniform (0.1, 0.9)	0.68 (0.16–0.89)
MEtna2	Migration rate parameter between *Senecio aethnensis* and the hybrid population (symmetric in each direction)	Uniform (0.05, 0.9)	0.75 (0.13–0.89)
Torigin	Time of split between *S. aethnensis* and *S. chrysanthemifolius* (thousands of generations)	Uniform (100, 250)	178 (116–243)
Na	Population size of *S. aethnensis* (thousands of individuals)	Uniform (150, 500)	202 (151–356)
Nc	Population size of *S. chrysanthemifolius* (relative to *Na*)	Uniform (0.5, 1.5)	0.52 (0.50–0.66)
Nh	Population size of hybrid population (relative to *Na*)	Uniform (0.5, 1.5)	1.41 (1.09–1.49)
T1	Time of introduction of seeds to UK (generations)	Uniform (300, 400)	333 (301–388)
T2	Time of naturalisation of *Senecio squalidus*	Uniform (200, 300)	257 (206–297)
T3	Beginning of exponential expansion of *S. squalidus*	Uniform [100, 200]	144 (102–198)
P	Proportion of *S. squalidus* genepool coming from *S. chrysanthemifolius*	Uniform [0.5, 0.9]	0.59 (0.51–0.75)
Ns	Population size of *S. squalidus* (relative to *Na*)	Uniform [1, 1,000]	573 (57–969)
Nnat	Population size of *S. squalidus* during naturalisation period (relative to *Nb*)	Uniform [10, 100]	61 (13–98)
Nb	Population size of *S. squalidus* at moment of origin	Uniform [1, 200]	155 (77–197)

*Posterior shows median and 95% posterior distribution.

its first year, and set *T1* to 300–400 generations ago (approximate introduction of material to the UK at around the turn of the 18th century), *T2* to 200–300 generations (following reports of *S. squalidus* growing outside the Oxford Botanic Garden by the mid-18th century) and *T3* to 100–200 generations ago (following spread of *S. squalidus* along the railway lines starting in the mid 19th century). For Model 1, the introduction of hybrid material into the UK at time *T1* was modelled by randomly selecting a number of individuals (*Nb*, between 1 and 200 individuals) from the hybrid zone population to give rise to *S. squalidus*. For Model 2, the individuals (*Nb*, 1–200 individuals) that gave rise to *S. squalidus* were instead randomly sampled from either *S. chrysanthemifolius* (with a probability, *p*, between .5 and .9) or *S. aethenesis* (with probability 1 − *P*). Population size for the period immediately after *T1* was modelled as constant (and equal to *Nb*), for the period after *T2* was modelled as constant but larger than *Nb* (*Nnat*, between 10 and 100× *Nb*), and for the period after *T3* was modelled as exponentially increasing (exponential growth factor calculated for each simulation depending on assumed population size of *S. squalidus* at *T1* and at present).

For the regression-based ABC approach we used a reduced set of six summary statistics to avoid the "curse of dimensionality". The curse of dimensionality in the context of ABC refers to the inability to simulate data sets that closely match the observed data as the number of summary statistics increases. Ideally, we would like to match the data in as many ways as possible, by using a large number of different summary statistics to describe it. However, the more summary statistics we use the less likely we are to generate data sets that are similar to the observed data, so there is a bias/variance trade-off. Thus, for this first ABC approach we selected only those summary statistics that convey the information we sought a priori to be relevant to distinguish the two scenarios of *S. squalidus* origin: the amount of divergence to each potential parental population, the magnitude of bottleneck suffered while colonising the UK, the relative loss of genetic diversity and the relative increase in linkage disequilibrium in *S. squalidus*. The six summary statistic selected were thus: the F_{ST} between *S. squalidus* and each of the three potential parental sources ($F_{ST.ac}$, $F_{ST.cs}$, $F_{ST.hs}$); the Tajima's *D* observed in *S. squalidus* (D_{squa}); the relative reduction of genetic diversity in *S. squalidus* compared to the three potential parental populations (θ_{squa}/mean [θ_{aeth}, θ_{chry}, θ_{hybr}]); and the observed linkage disequilibrium in *S. squalidus* relative to the average LD in the other three taxa ($LD1_{squa}$/mean [$LD1_{aeth}$, $LD1_{chry}$, $LD1_{hybr}$]). To evaluate whether this approach is able to distinguish the two alternative models we computed misclassification errors with the function cv4postpr of the ABC package, using

FIGURE 2 Genome-wide analyses confirm the origin of *Senecio squalidus* from hybridisation between *Senecio aethnensis* and *Senecio chrysanthemifolius*. (a) Clustering and (b) principal component analysis based on 2,081 non-coding SNPs distributed across the genome show that *S. squalidus* is genetically distinct from both parental species and individuals from the naturally-occurring hybrid zone on Mt Etna (denoted *S. aeth* × *chry*). (c) Analysis of SNPs showing nearly fixed differences between both parental species reveals that *S. squalidus* is polymorphic in the majority of SNPs (*n* = 750) and is fixed for *S. chrysanthemifolius* alleles more than twice as often as for *S. aethnensis* alleles



100 simulated data sets from each model. To infer which model was best supported by the data, we calculated the approximate posterior model probability of each model with the function postpr, and the approximate Bayes factors as the ratio of model probabilities of each model. Parameter estimates were obtained for the best supported model using the function abc, with parameter values logit transformed and using the prior for each parameter as bounds. For all calculations we used the neural network ABC approach with a tolerance of 0.01 as implemented in ABC package.

To test whether the best supported model provided a good fit to the observed data we performed posterior predictive checks, by randomly sampling with replacement 1,000 multivariate parameters from the posterior distribution of each model, simulating new summary statistics using these parameters, and plotting the resulting distributions both separately (for each summary statistic) and jointly in a PCA in the space of all summary statistics. Here, a good model fit is indicated when the observed values lay within the distributions obtained in the a posteriori simulations.

For the random forests ABC approach, no a priori selection of summary statistics was performed. We built a machine learning classifier containing 1,000 trees with the function abcrf, using all summary statistics denoted in Table S2 complemented by the single axis of a linear discriminant analysis (LDA) of the two models as an additional statistic. We then used this classifier to predict the best fitting model with the function predict.abcrf. To test whether the models simulated provide a good fit to the observed data we projected the observed summary statistics on the single axis of the LDA, together with the simulated data sets for each model—a good fit is indicated when the observed data is in close vicinity of the simulated data sets.

3 | RESULTS

3.1 | Population structure, polymorphism and linkage disequilibrium

To characterise the genome-wide effects of hybrid speciation we used transcriptome data from multiple individuals of *S. squalidus* collected from throughout the UK (*n* = 28), together with transcriptome data from its two parental species (*n* = 16), and naturally-occurring interspecific hybrids sampled on Mt Etna (*n* = 14).

Senecio squalidus was resolved as a well-separated cluster from all other groups in both Structure and PCA analyses (Figure 2a,b). Conversely, the hybrid individuals collected on Mt Etna showed variable degrees of admixture between the two parental species, and based on the admixture proportions none appeared to be a first-generation hybrid (Figure 2a). This is in agreement with previous results using different genetic markers (James & Abbott, 2005; Wong et al., 2020) and suggests that F1 hybrids on Mt Etna are rare. The stronger clustering of all hybrid individuals with *S. chrysanthemifolius* accords with the relatively low elevation where hybrid seeds were collected for this study (1,353–1,530 m; Table S1) which is closer to the distributional limits of typical *S. chrysanthemifolius* (below 900 m) than *S. aethnensis* (above 2,500 m; Abbott & Brennan, 2014), and is more likely to reflect the geographic origin of the putative hybrid material introduced to the Oxford Botanic Garden in the late-17th century (Harris, 2002). Furthermore, some individuals of *S. aethnensis* sequenced previously (Chapman et al., 2013) and reanalysed here were collected at around 2,000 m elevation, and exhibit some admixture with *S. chrysanthemifolius* (Figure 2a).

We identified 1,121 SNPs where the two parental species were fixed or nearly fixed (frequency of most common allele >90%) for different alleles. Of 1,111 such SNPs that were sequenced in *S. squalidus*, we found that 67.5% (750 SNPs) were polymorphic, while 23.5% ($n = 261$) were fixed or nearly fixed for the *S. chrysanthemifolius* allele and 9% ($n = 100$) for the *S. aethnensis* allele (Figure 2c). We found no evidence for genome-wide excess heterozygosity in *S. squalidus* (global Hardy-Weinberg test, $p = 1$) and only six out of 970 SNPs tested for excess heterozygosity in this species showed a significant result ($p < .05$; only four of these remained significant after multiple test correction).

Analysis of polymorphism revealed that genetic diversity is on average 25% lower in *S. squalidus* (diversity on entire coding sequences, $\theta_{\text{squa}} = 0.0034$; on synonymous sites only, $\theta s_{\text{squa}} = 0.008$) compared to the two parental species ($\theta_{\text{aeth}} = 0.0049$ and $\theta s_{\text{aeth}} = 0.0119$; $\theta s_{\text{chry}} = 0.0039$ and $\theta s_{\text{chry}} = 0.0094$) and their naturally-occurring hybrids ($\theta_{\text{hybr}} = 0.0047$ and $\theta s_{\text{hybr}} = 0.0110$; Figure 3a; Table S3). Tajima's *D* estimated across all genes (Figure 3b; Table S3) was higher in *S. squalidus* ($D_{\text{squa}} = 0.77$ and $Ds_{\text{squa}} = 0.79$) than in the two parental species ($D_{\text{aeth}} = 0.18$ and $Ds_{\text{aeth}} = 0.22$; $D_{\text{chry}} = 0.14$ and $Ds_{\text{chry}} = 0.20$) or their hybrid population ($D_{\text{hybr}} = 0.20$ and $Ds_{\text{hybr}} = 0.27$). Estimated F_{ST} values between *S. squalidus* and the Italian taxa were always higher than between any pair of Italian taxa (Table S3). The degree of linkage disequilibrium observed was also higher in *S. squalidus* compared to either parental species or their hybrid population, with ca. 25% of SNPs within 1 kb of each other showing significant LD in *S. squalidus* (Figure 3c).

3.2 | ABC analysis

To reveal the early evolutionary history of *S. squalidus*, we performed ABC analyses of the two alternative scenarios for the homoploid hybrid origin of this species. This approach relies on simulating data sets under each model, computing summary statistics for each simulated data set, and comparing these summary statistics to the observed values—better fitting models should result in summary statistics more similar to the ones observed in the empirical data (Beaumont et al., 2002). We considered two models that differed only in the

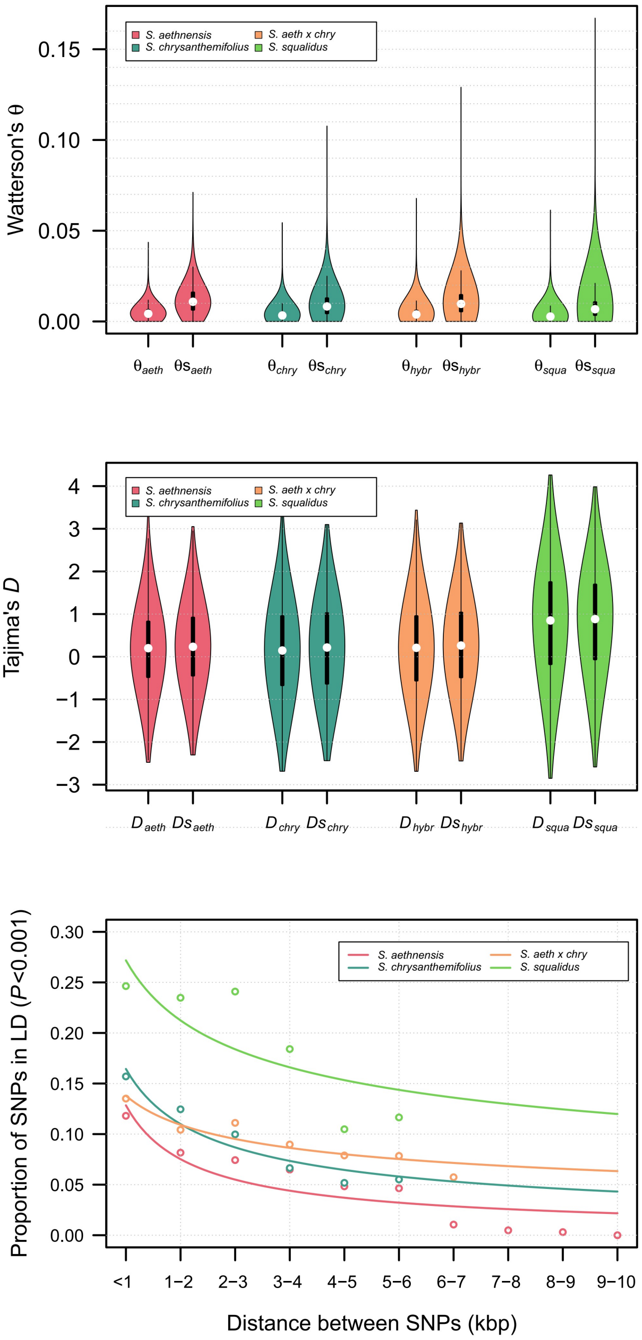


FIGURE 3 Hybrid speciation and population bottleneck in the origin of *Senecio squalidus* affected genome-wide patterns of polymorphism. (a) Analysis of genetic polymorphism shows lower average genetic diversity, (b) higher Tajima's *D* values, and (c) a larger proportion of SNPs in linkage disequilibrium in *S. squalidus* compared to parental species and their naturally-occurring hybrids. For genetic diversity and Tajima's *D* estimates, values are shown when using all sites (θ, D) or only synonymous sites (θ_s, D_s) for each species and the naturally-occurring hybrids from Mt Etna (*S. aeth* × *chry* in all panels). For each violin plot in panels a and b white dots represent median; black vertical rectangles represent the interquartile range (IQR); and black vertical lines denote the range of the most extreme data points that are within 1.5 × IQR below lower quartile and 1.5 × IQR above upper quartile

very early evolutionary history of *S. squalidus*. Under Model 1, *S. squalidus* originated from introduction of hybrid material from the naturally-occurring hybrid zone on Mt Etna into the UK, while under Model 2 *S. squalidus* originated from a recent hybridisation event between the two parental species in the UK (Figure 4a,b). We found that we are able to distinguish the two models with high accuracy (cross-validation test, all pseudo-observed data sets correctly identified), and retrieved strong support for Model 2 compared to Model 1 (approximate posterior probabilities of 0.945 and 0.055 respectively; approximate Bayes factor of 17.11). An independent random

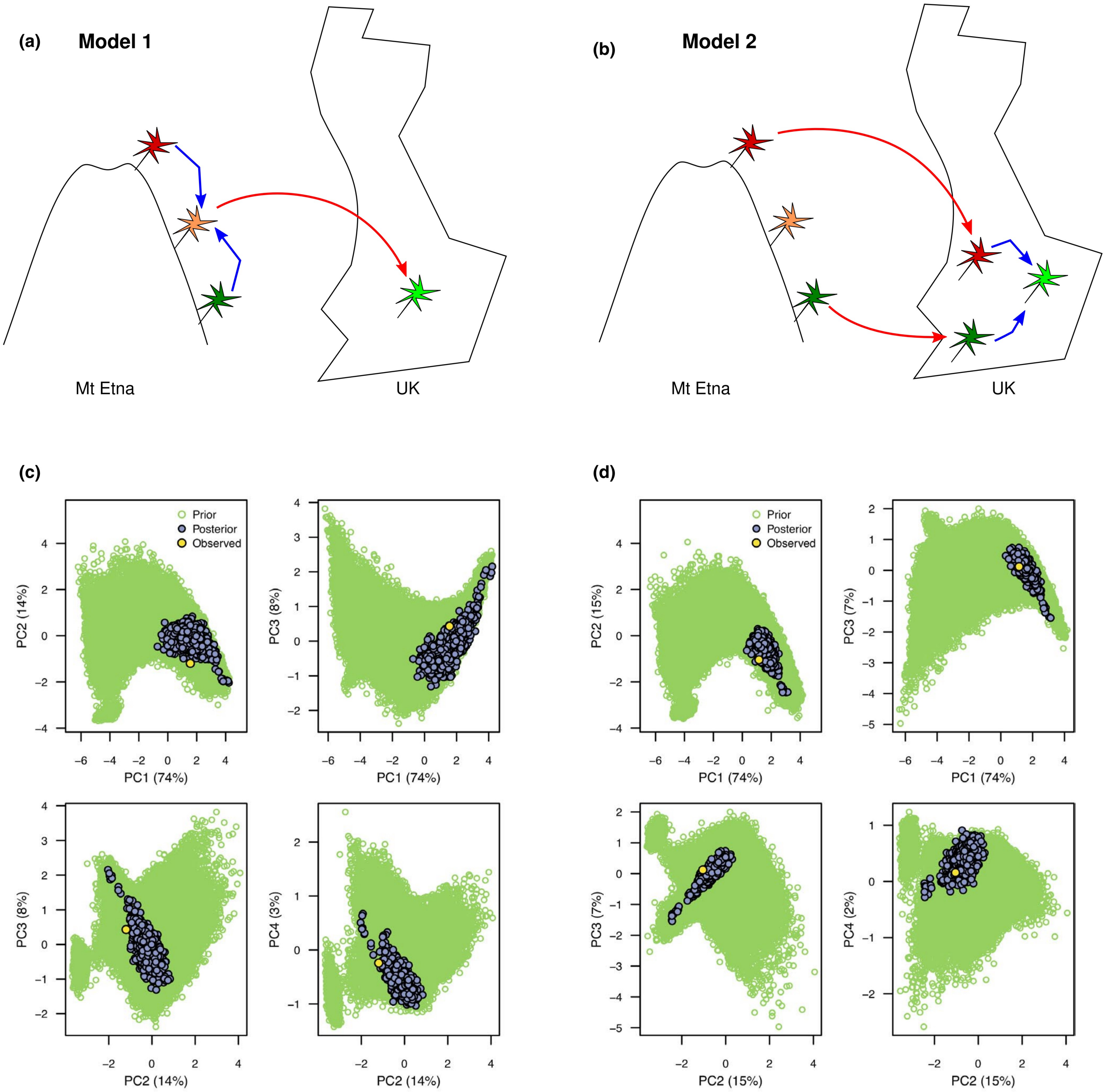


FIGURE 4 Approximate Bayesian Computation analysis supports a recent hybridisation event in the origin of *Senecio squalidus*. (a, b) The two models differ in whether *S. squalidus* originated from introduction of hybrid individuals from the naturally-occurring hybrid zone on Mt Etna (Model 1, panel a) or from recent hybridisation in the UK between introduced specimens of both parental species (Model 2, panel b). Dashed red lines represent human-mediated dispersal between Mt Etna and the UK, and solid blue lines represent hybridisation between *S. aethnensis* and *Senecio chrysanthemifolius*. (c, d) Principal component analysis in the space of summary statistic used in the ABC neural network regression approach, showing simulated data sets from the prior (open green circles) and from the posterior predictive distributions (filled blue circles), and the observed data set (filled yellow circle). The observed values fall within the posterior predictive distribution on all four PCA axes (accounting for 98% of variation) for Model 2 (d) but not for Model 1 (c)

forests ABC approach using all 18 summary statistics calculated (Table S2) also supported Model 2 over Model 1 (approximate posterior probability of Model 2 of .998; approximate Bayes factor of 499; Figure S2). To further confirm that Model 2 provides a good fit to the data, we performed posterior predictive checks for each model using 1,000 multivariate parameters from the posterior distribution of each model. We found that the distributions of the summary statistics calculated from these a posteriori-simulated data sets encompassed the values observed in the empirical data in all cases (Figure S3), and a PCA on the space of all summary statistics confirmed the better fit of Model 2 to the data (Figure 4c,d). Posterior estimates for all demographic parameters are given in Table 1, but these typically had wide confidence intervals and should be interpreted with caution.

4 | DISCUSSION

Using transcriptome data from multiple individuals of *S. squalidus*, its two parental species and individuals from their naturally occurring hybrid zone, we have characterized the genome-wide effect of homoploid hybrid speciation in *S. squalidus*, and inferred the most likely scenario for the origin of this species. The results of the analysis of genetic data obtained in this study must be interpreted in view of the extensive and detailed historical records of the origin and spread of *S. squalidus* throughout the UK (reviewed in Abbott et al., 2009; Harris, 2002). Of particular importance, it is well established that seeds of *Senecio* from Mt Etna were introduced to gardens in the UK at the end of the 17th century, and that *S. squalidus* originated from material descending from these seeds that escaped cultivation in the Oxford Botanic Garden in the 18th century (Abbott et al., 2009; Harris, 2002; James & Abbott, 2005). Furthermore, because neither parental species has ever been recorded in the wild anywhere in the UK (Sell et al., 2006), hybridisation between *S. squalidus* and either parental species cannot have occurred after *S. squalidus* escaped from the Oxford Botanic Garden. This allows us to dismiss alternative scenarios for genetic admixture in *S. squalidus*, and in particular that introgression of genetic material with either parental species occurred after the origin of *S. squalidus*—a scenario that is usually difficult to exclude in other systems (Schumer et al., 2014).

4.1 | Genome wide effect of hybrid speciation in *S. squalidus*

Senecio squalidus forms a genetically divergent group from both parental species and individuals from their naturally occurring hybrid zone, as inferred from both PCA and Structure analyses (Figure 2). This is a different pattern than has been observed in other homoploid hybrid species (e.g., Elgvin et al., 2017). Because both of these analyses use allele frequencies to infer clustering patterns, the clear separation between *S. squalidus* and all other groups probably reflects the strong drift experienced by *S. squalidus* during its origin and in the early stages of its colonisation of the UK, and the lack of gene flow between *S. squalidus* and either parental species after *S. squalidus* escaped cultivation in the Oxford Botanic Garden.

The majority of SNPs showing fixed differences between the two parental species are polymorphic in *S. squalidus*, confirming the very recent hybrid origin of this species. We also confirm a greater contribution of genetic material from *S. chrysanthemifolius* than *S. aethnensis* as reported previously (James & Abbott, 2005). Although this could result from preferential fixation of *S. chrysanthemifolius* alleles, it may also reflect more frequent backcrosses with this species, in line with the co-occurrence of hybrid specimens and *S. chrysanthemifolius* in the Oxford Botanic Garden in the early-18th century (Harris, 2002; see also next section) prior to *S. squalidus* escape and colonisation of the UK.

Despite the strong bottleneck experienced by *S. squalidus*, genetic diversity in this species is relatively high, and on average only 25% lower compared to either parental species (Figure 3). This is in line with a previous estimate of 32% reduction in microsatellite diversity in this species (Brennan et al., 2013). *Senecio squalidus* also exhibits a wider range of genetic diversity values, with 13% of genes exhibiting higher diversity than observed in either parental species (Figure 3a), possibly resulting from retention of alleles from both species at these genes. This, together with the self-incompatibility system in *S. squalidus* that promotes maintenance of genetic diversity (Abbott & Forbes, 1993; Hiscock, 2000), may explain why genetic diversity in *S. squalidus* remained relatively high despite the strong and recent bottleneck experienced by this species. The decrease in genetic diversity and increase in Tajima's *D* observed in *S. squalidus* are expected signatures of a recent population bottleneck, while the increase in linkage disequilibrium may result from a recent bottleneck, recent hybridisation between the two parental species, or both.

4.2 | Origin of *S. squalidus* from recent hybridisation in British gardens

Colonisation of the UK by *S. squalidus* after escaping cultivation at the Oxford Botanic Garden has been extensively recorded by British botanists to give an extremely accurate chronological record (reviewed in Abbott et al., 2009), but the early evolutionary history of this species remains elusive (Abbott et al., 2009; Harris, 2002). We investigated two alternative models for the origin of *S. squalidus*: introduction of individuals from the naturally occurring hybrid zone on Mt Etna (Model 1); and de novo hybridisation between the two parental species while in cultivation in British Gardens (Model 2). The main difference between the two models is the age of the hybridisation event giving rise to *S. squalidus*. Under Model 1, introduction of individuals from the naturally occurring hybrid zone on Mt Etna would imply a more ancient hybrid origin, as the hybrid zone on Mt Etna is relatively long (ranging from 900 m to 2,500 m elevation) and stable such that first-generation hybrids are rare. Instead, hybrid individuals along the hybrid zone show different degrees of ancestry and backcrossing with each parental species depending on elevation (James & Abbott, 2005; Wong et al., 2020). Conversely, Model 2 assumes a very recent hybridisation event between *S. aethnensis* and *S. chrysanthemifolius* as giving origin to *S. squalidus*. We note that first generation hybrids may occur sporadically on Mt Etna, driven by long-distance dispersal of pollen by pollinators. However, these events are rare (S. Hiscock, G. Walter and J. Bridle, personal observations) so it is unlikely these individuals would make up any significant proportion of the seeds collected on Mt Etna and introduced into the UK at the turn of the 18th century.

We find that a model of very recent hybridisation in the origin of *S. squalidus* (Model 2) is strongly supported. Two different methods for estimating posterior probabilities give values of .945 and .998 in favour of this model, based on neural network regression and random forest classification (respectively). The contrasting model, where established hybrid material from Mt Etna was introduced to the UK at the turn of the 18th century (Model 1) is only very weakly supported—our analysis suggests that the probability of obtaining the observed summary statistics is 17 to 500 times higher under Model 2 compared to Model 1. Model choice using ABC can be affected by the choice of summary statistics, which may lead to insufficient information retained to confidently distinguish between competing models (Robert et al., 2011). However, cross-validation analysis revealed that even when using the restricted set of six summary statistics with the neural network regression approach all pseudo-observed data sets are correctly assigned to their respective models. Posterior predictive checks further confirm the good fit of Model 2 to the data: when considered separately, all observed summary statistics fall within the 95% posterior distributions (Figure S3); and when considered jointly, observed values fall within the posterior distribution on all projections of a PCA on the space of all summary statistics (Figure 4d). As mentioned above, "pure" populations of *S. aethnensis* and *S. chrysanthemifolius* do not usually meet—"pure" *S. aethnensis* individuals are restricted to high elevation regions above 2,500 m, while "pure" *S. chrysanthemifolius* individuals occur at elevations below 900 m (Abbott & Brennan, 2014)—and it is highly unlikely that the seeds introduced into the UK would come from a first generation cross between the two species. Instead, recent hybridisation between the two species most likely occurred in the UK, in the gardens at Badminton where individuals from both species were in cultivation at the end of the 17th century (Figure 1b-d; Harris, 2002). Hybrid material from these crosses was most likely introduced to the Oxford Botanic Garden in the early 18th century, where *S. chrysanthemifolius* was also cultivated at this time (Figure 1a; Harris, 2002). This implies that the newly formed hybrid individuals were growing in close proximity with at least one of the parental species before escaping cultivation and spreading throughout the UK (as *S. squalidus*).

This now raises the question of what mechanism(s) caused reproductive isolation in this system. While we do not address this question in this study, recent research supports the existence of genetic incompatibilities between both *S. aethnensis* and *S. chrysanthemifolius* (Brennan et al., 2014; Chapman et al., 2016), and between these two species and *S. squalidus* (Brennan et al., 2019). Moreover, *S. squalidus* appears to have inherited genetic incompatibilities from both parental species (Brennan et al., 2019), providing a possible mechanism for an initial degree of reproductive isolation following homoploid hybrid speciation (but see Schumer & Brandvain, 2016). At the same time, no wild populations of either parental species have ever been reported in the UK (Sell et al., 2006); and common garden experiments in the UK have shown that *S. squalidus* outperforms both parental species, their naturally occurring hybrids from Mt Etna, and newly-synthesised hybrids (Ross, 2010). This suggests a strong degree of ecological divergence between *S. squalidus* and both parental species, which furthermore appears to be more strongly associated with differences in soil composition rather than climatic variables (Ross, 2010). Analysis of differential gene expression between *S. squalidus* and its parental species (Hegarty et al., 2009) identified over 200 genes with transgressive expression values in *S. squalidus* (i.e., gene expression values in *S. squalidus* significantly higher or lower than both parental species), possibly underpinning the transgressive physiological adaptations that allowed the hybrid species to colonise an environment where neither parental species can live. Thus, hybridisation is likely to have caused a degree of reproductive isolation from the onset (due to post-zygotic mechanisms), while *S. squalidus*' subsequent escape from cultivation and spread through the UK provided additional (prezygotic) isolation in the form of habitat and ecological segregation.

Given that *S. squalidus* originated from recent interspecific crosses in the UK—as opposed to introduction of already stabilised hybrid material from Mt Etna—our results have important implications for understanding homoploid hybrid speciation. First, we can infer that establishment of a stable hybrid lineage occurred rapidly. The original interspecific crosses occurred at the turn of the 18th century (1690–1710), whilst individuals sent to Carolus Linnaeus in the 1730s were formally described as *S. squalidus* in 1753, with the first unambiguously dated specimen from Oxford being collected in 1792 (Harris, 2002). Thus a period of 30–100 generations elapsed during formation of the new species, *S. squalidus*, which is in line with estimates obtained in *Helianthus* (Rieseberg, 2000; Rieseberg et al., 1996) and some, but not all, simulation studies (Ungerer et al., 1998). Second, we can infer that ecological divergence occurred rapidly following hybridisation, possibly as early as the mid-18th century, when *S. squalidus* was naturalised in old walls outside the Oxford Botanic Garden (Harris, 2002), i.e., adaptation to the novel habitat occurred within ca. 50 generations of hybridisation. Third, we can infer that both pre- and post-zygotic reproductive isolation barriers evolved during, or rapidly after, hybridisation, in line with results in *Helianthus* (Rieseberg, 2006), and suggesting that homoploid hybrid speciation may require interactions between both types of reproductive isolation barriers to produce a new hybrid species.

In conclusion, we found that, by the turn of the 18th century, the homoploid hybrid species *S. squalidus* had arisen from hybridisation between its two parental species while in cultivation in two British gardens. This is a remarkable example of the creative role of hybridisation in the face of environmental change—two species, each unable to survive in a novel habitat, hybridized and formed a new species that not only survived but was able to spread rapidly and colonize the novel habitat. In line with the well-documented cases of homoploid hybrid speciation in North American *Helianthus* (reviewed in Rieseberg, 2006), speciation of *S. squalidus* appears to have resulted from both genomic incompatibilities between hybrid and parental species, and transgressive phenotypes created by hybridisation that allowed colonisation of novel habitats. However, compared to the hybrid *Helianthus* species, *S. squalidus* originated much more recently, and its invasion into a novel habitat has been extremely well documented (reviewed in Abbott et al., 2009), providing a window into the very earliest stages of homoploid hybrid speciation—the period during which this process is most different from other forms of speciation (Nolte & Tautz, 2010). Ongoing genome assembly of this species, together with experimental crosses between *S. squalidus* and its parental species (Brennan et al., 2019) will lead to a better understanding of the genomic architecture of reproductive isolation in this system, and allow pinpointing of genomic regions and evolutionary mechanisms responsible for rapid ecological divergence between *S. squalidus* and its parental species.

ACKNOWLEDGEMENTS

This work was supported by a “start-up” grant from the University of Oxford to S.J.H., small research grants from Christ Church, Oxford (to S.J.H.) and a John Fell Fund grant from the University of Oxford (JFF grant 162/091 to B.N.). We thank the staff at the Wellcome Trust Centre (Oxford) for sequencing and computational support, and acknowledge the use of the University of Oxford Advanced Research Computing facility in carrying out this work (http://dx.doi.org/10.5281/zenodo.22558). We thank Adrian Brennan, Richard Abbott and Gregory Walter for providing samples; Edgar Wong, Ying-Qi Lou and Tim Wroe for help with greenhouse work; and Dmitry Filatov for discussions in the early stages of this work. We thank Varvara Fazalova for comments and discussions on various versions of this manuscript.

AUTHOR CONTRIBUTIONS

B.N., and S.J.H. conceived this study. B.N. designed the work, collected and analysed the data and wrote the manuscript with input from S.J.H., S.H., and M.B. All authors have read and agree with the manuscript.

DATA AVAILABILITY STATEMENT

Raw sequence data collected for this study is available from GenBank's Short Read Archive (BioProject ID PRJNA549571), and filtered vcf files from the DRYAD repository (https://doi.org/10.5061/dryad.9ghx3fff3).

ORCID

Bruno Nevado https://orcid.org/0000-0002-9765-2907

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., & Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott, R. J., & Brennan, A. C. (2014). Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1648), 20130346. https://doi.org/10.1098/rstb.2013.0346

Abbott, R. J., Brennan, A. C., James, J. K., Forbes, D. G., Hegarty, M. J., & Hiscock, S. J. (2009). Recent hybrid origin and invasion of the British Isles by a self-incompatible species, Oxford ragwort (*Senecio squalidus* L., Asteraceae). *Biological Invasions*, *11*(5), 1145–1158. https://doi.org/10.1007/s10530-008-9382-3

Abbott, R. J., & Forbes, D. G. (1993). Outcrossing rate and self-incompatibility in the colonizing species *Senecio squalidus*. *Heredity*, *71*(2), 155–159. https://doi.org/10.1038/hdy.1993.119

Abbott, R., James, J. K., Irwin, J., & Comes, H. (2000). Hybrid origin of the Oxford Ragwort, *Senecio squalidus* L. *Watsonia*, *23*, 123–138.

Andrews, S. (2010). *FastqQC: A quality control tool for high throughput sequence data*. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035.

Boccone, P. (1674). *Icones & descriptiones rariorum plantarum Siciliae, Melitae, Galliae, & Italiae. Quarum unaquaeque proprio charactere signata, ab uliis ejusdem classis facile distinguitur.* Theatro Sheldoniano.

Brennan, A. C., Barker, D., Hiscock, S. J., & Abbott, R. J. (2012). Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: A study of*Senecio squalidus* (Asteraceae). *Heredity*, *108*(2), 87–95. https://doi.org/10.1038/hdy.2011.46

Brennan, A. C., Harris, S. A., & Hiscock, S. J. (2013). The population genetics of sporophytic self-incompatibility in three hybridizing *Senecio* (Asteraceae) species with contrasting population histories. *Evolution*, *67*(5), 1347–1367. https://doi.org/10.1111/evo.12033

Brennan, A. C., Hiscock, S. J., & Abbott, R. J. (2014). Interspecific crossing and genetic mapping reveal intrinsic genomic incompatibility between two *Senecio* species that form a hybrid zone on Mount Etna, Sicily. *Heredity*, *113*, 195–204. https://doi.org/10.1038/hdy.2014.14

Brennan, A. C., Hiscock, S. J., & Abbott, R. J. (2019). Completing the hybridization triangle: The inheritance of genetic incompatibilities during homoploid hybrid speciation in ragworts (*Senecio*). *AoB Plants*, *11*(1), ply078. https://doi.org/10.1093/aobpla/ply078

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Peshkin, L., Tabin, C. J., & Whited, J. L. (2017). A tissue-mapped Axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Reports*, *18*(3), 762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Chapman, M. A., Hiscock, S. J., & Filatov, D. A. (2013). Genomic divergence during speciation driven by adaptation to altitude. *Molecular Biology and Evolution*, *30*(12), 2553–2567. https://doi.org/10.1093/molbev/mst168

Chapman, M. A., Hiscock, S. J., & Filatov, D. A. (2016). The genomic bases of morphological divergence and reproductive isolation driven by ecological speciation in *Senecio* (Asteraceae). *Journal of Evolutionary Biology*, *29*(1), 98–113. https://doi.org/10.1111/jeb.12765

Crisp, P. (1972). *Cytotaxonomic studies in the section Annui of Senecio*. Ph.D. thesis, University of London.

Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, *3*(3), 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x

Cupani, F. (1694). *Syllabus plantarum siciliae nuper detectarum (quantum vires suppetierunt) à P. F. Francisco Cupani, etc*. John Adam.

Cupani, F. (1696). *Hortus Catholicus*. Franciscum Benzi.

Doorduin, L., Gravendeel, B., Lammers, Y., Ariyurek, Y., Chin, A. W. T., & Vrieling, K. (2011). The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Research*, *18*(2), 93–105. https://doi.org/10.1093/dnares/dsr002

Druce, G. C. (1927). *The flora of Oxfordshire* (2nd ed.). Clarendon Press.

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., Ravinet, M., Jensen, H., & Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*(6), e1602996. https://doi.org/10.1126/sciadv.1602996

Grant, V. (1981). *Plant speciation* (2nd ed.). Columbia University Press.

Gross, B. L., & Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, *96*(3), 241–252. https://doi.org/10.1093/jhered/esi026

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., & Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: Reference generation and analysis with Trinity. *Nature Protocols*, *8*(8). https://doi.org/10.1038/nprot.2013.1084

Harris, S. (2002). Introduction of Oxford ragwort, *Senecio squalidus* L. (Asteraceae), to the United Kingdom. *Watsonia*, *24*, 31–43.

Hegarty, M. J., Barker, G. L., Brennan, A. C., Edwards, K. J., Abbott, R. J., & Hiscock, S. J. (2009). Extreme changes to gene expression associated with homoploid hybrid speciation. *Molecular Ecology*, *18*(5), 877–889. https://doi.org/10.1111/j.1365-294X.2008.04054.x

Hiscock, S. J. (2000). Self-incompatibility in *Senecio squalidus* L. (Asteraceae). *Annals of Botany*, *85*, 181–190. https://doi.org/10.1006/anbo.1999.1058

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338. https://doi.org/10.1093/bioinformatics/18.2.337

James, J. K., & Abbott, R. J. (2005). Recent, allopatric, homoploid hybrid speciation: The origin ofSenecio squalidus (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution*, *59*(12), 2533–2547. https://doi.org/10.1111/j.0014-3820.2005.tb00967.x

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Kent, D. H. (1956). *Senecio squalidus* L. in the British Isles. 1. Early records (to 1877). *Proceedings of the Botanical Society of the British Isles*, *2*, 115–118.

Kent, D. H. (1960). *Senecio squalidus* L. in the British Isles. 2. The spread from Oxford (1879–1939). *Proceedings of the Botanical Society of the British Isles*, *3*, 375–379.

Krueger, F. (2015). *Trim galore*. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Lai, Z., Nakazato, T., Salmaso, M., Burke, J. M., Tang, S., Knapp, S. J., & Rieseberg, L. H. (2005). Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics*, *171*(1), 291303. https://doi.org/10.1534/genetics.105.042242

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., & Subgroup, G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Mavarez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in *Heliconius* butterflies. *Nature*, *441*(7095), 868–871. https://doi.org/10.1038/nature04738

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Morison, R. (1699). *Plantarum Historiae Universalis Oxoniensis. Pars Tertia seu Herbarum Distributio Nova, per Tabulas Cognationis & Affiniatis ex Libro naturae observata & detecta*. Theatro Sheldoniano.

Nolte, A. W., & Tautz, D. (2010). Understanding the onset of hybrid speciation. *Trends in Genetics*, *26*(2), 54–58. https://doi.org/10.1016/j.tig.2009.12.001

Osborne, O. G., Batstone, T. E., Hiscock, S. J., & Filatov, D. A. (2013). Rapid speciation with gene flow following the formation of Mt. Etna. *Genome Biology and Evolution*, *5*(9), 1704–1715. https://doi.org/10.1093/gbe/evt127

Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866. https://doi.org/10.1093/bioinformatics/btv684

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, *28*, 359–389. https://doi.org/10.1146/annurev.ecolsys.28.1.359

Rieseberg, L. H. (2000). Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution*, *54*(3), 859–865. https://doi.org/10.1111/j.0014-3820.2000.tb00086.x

Rieseberg, L. H. (2006). Hybrid speciation in wild sunflowers. *Annals of the Missouri Botanical Garden*, *93*(1), 34–48. https://doi.org/10.3417/0026-6493(2006)93[34:Hsiws]2.0.Co;2

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., & Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, *301*(5637), 1211–1216. https://doi.org/10.1126/science.1086949

Rieseberg, L. H., Sinervo, B., Linder, C. R., Ungerer, M. C., & Arias, D. M. (1996). Role of gene interactions in hybrid speciation: Evidence from ancient and experimental hybrids. *Science*, *272*(5262), 741–745.

Rieseberg, L. H., Van Fossen, C., & Desrochers, A. M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*, *375*(6529), 313–316. https://doi.org/10.1038/375313a0

Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, *152*(2), 713–727.

Robert, C. P., Cornuet, J. M., Marin, J. M., & Pillai, N. S. (2011). Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), 15112–15117. https://doi.org/10.1073/pnas.1102900108

Ross, R. I. C. (2010). *Local adaptation and adaptive divergence in a hybrid species complex in* Senecio. Doctor of Philosophy PhD, University of Oxford.

Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Scheet, P., & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, *78*(4), 629–644. https://doi.org/10.1086/502802

Schumer, M., & Brandvain, Y. (2016). Determining epistatic selection in admixed populations. *Molecular Ecology*, *25*, 2577–2591. https://doi.org/10.1111/mec.13641

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution*, *68*(6), 1553–1560. https://doi.org/10.1111/evo.12399

Sell, P., Murrell, G., & Walters, S. M. (2006). *Flora of Great Britain and Ireland: Volume 4*, *Campanulaceae – Asteraceae*. Cambridge University Press.

Ungerer, M. C., Baird, S. J., Pan, J., & Rieseberg, L. H. (1998). Rapid hybrid speciation in wild sunflowers. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 11757–11762. https://doi.org/10.1073/pnas.95.20.11757

Wong, E. L. Y., Nevado, B., Osborne, O. G., Papadopulos, A. S. T., Bridle, J. R., Hiscock, S. J., & Filatov, D. A. (2020). Strong divergent selection at multiple loci in two closely related species of ragworts adapted to high and low elevations on Mount Etna. *Molecular Ecology*, *29*(2), 394–412. https://doi.org/10.1111/mec.15319

Yakimowski, S. B., & Rieseberg, L. H. (2014). The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany*, *101*(8), 1247–1258. https://doi.org/10.3732/ajb.1400201

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Nevado B, Harris SA, Beaumont MA, Hiscock SJ. Rapid homoploid hybrid speciation in British gardens: The origin of Oxford ragwort (*Senecio squalidus*). *Mol Ecol*2020;29:4221–4233. https://doi.org/10.1111/mec.15630