

Chapter 3 - The genetic landscape of hybridisation in UK flora

Max R. Brown Peter M. Hollingsworth Laura Forrest Michelle Hart
Laura Jones Col Ford Natasha de Vere Alex D. Twyford

Abstract

Hybridisation has a profound impact on the evolution of plants, with consequences including the generation of novel phenotypes (Lexer et al. 2003), introgression of adaptive alleles (Chapman and Abbott 2010), and the origin of new species (Ainouche et al. 2009), as well as the blurring of species boundaries (Bardy et al. 2011), and the extinction of rare taxa. While natural hybridisation is common and widespread both geographically and phylogenetically (Beddows and Rose 2018; Whitney et al. 2010; Ellstrand, Whitkus, and Rieseberg 1996), the ecological factors and species traits that promote or prevent hybridisation in natural plant communities are poorly understood. Many important traits related to hybridisation, such as parental genetic distance and ploidy level differences have not been systematically quantified across diverse species in a flora, while estimates for phylogenetic effects of hybridisation in plant communities are highly variable (Beddows and Rose 2018; Mitchell et al. 2019; Whitney et al. 2010). Here, we combine use phylogenetic mixed effect models to combine estimates of hybridisation from extensive field observations (Stace, Preston, and Pearman 2015) with DNA barcoding data and ecological attributes for over 1,000 native species of flowering plant. Our results quantify the influence of different predictors, and the genetic distances over which hybridisation is most likely. We also quantify the importance of phylogenetic relatedness and parental ploidy differences in shaping the likelihood of hybrid formation. Life history of parental species does not impact hybrid formation despite perennial life history being pervasive in both parental species and hybrids. Although the effect of genetic distance requires careful interpretation as low parental divergence may either allow hybrids to form or be a consequence of genetic homogenisation, taken together with ploidy and phylogenetic effects, genetic factors are key predictors of hybridisation across diverse taxa in the UK flora.

Main

Natural hybridisation plays an important role in plant evolution by facilitating adaptation and promoting species survival, generating novel variation, or alternatively in some cases leading to a reduction in fitness and even causing extinction (Chapman and Abbott 2010; Becker et al. 2013; Rieseberg, Archer, and Wayne 1999; Rhymer and Simberloff 1996). Focused studies on plant evolutionary model systems such as *Helianthus* (Lexer et al. 2003), *Senecio* (Abbott et al. 2009), and *Tragopogon* (Novak, Soltis, and Soltis 1991) have revealed how hybridisation may contribute to range expansion, invasiveness and phenotypic and genomic novelty. These focused genetic studies have selected species and study systems for their noteworthy hybrid outcomes, and it is currently unclear how these results generalise to natural plant communities and entire floristic assemblages. A relatively small body of work has looked more broadly at hybridisation between diverse species in floras, with the aim of understanding the likelihood of hybridisation in the context of species attributes, ecology, and phylogeny (Whitney et al. 2010; Beddows and Rose 2018; Mitchell et al. 2019; Ellstrand, Whitkus, and Rieseberg 1996; Marques et al. 2018). A limitation of these studies has been the reliance on phylogenies where the tips represent higher taxonomic units such as genera, families, or orders (Beddows and Rose 2018; Mitchell et al. 2019; Whitney et al. 2010). Hybridisation is an outcome of interactions between species, and a resolved species-level phylogeny is required to calculate the genetic distances of hybridising taxa and to

accurately estimate the phylogenetic signal of hybridisation.

DNA barcoding is proving increasingly useful for the genetic characterisation of species assemblages and complex communities (Gomez-Rodriguez et al. 2015; Young et al., n.d.). By amplifying a small and standardised set of target loci, DNA barcoding is inherently scalable to deployment across large sample sets and is well-suited to comparative analyses of diverse taxa. As DNA data is generated at pace and at scale it is becoming feasible to generate comprehensive DNA barcoding datasets for all plant species in a country. This provides the opportunity for the integration with ecological data collected at a national level to test major ecological and evolutionary questions at broad geographic scales.

In this study we characterise the genetic landscape of natural hybridisation across flowering plant species in the UK. The UK flora is an ideal study system for investigating hybridisation as it contains ~1400 species and is therefore manageable for genetic characterisation, but has sufficient diversity to include ~20% of all known angiosperm plant families and numerous hybrid combinations (Stace, Preston, and Pearman 2015). There is also a hybrid flora — a unique resource describing all known vascular plant inter-specific hybrids present across the British Isles, of which 616 represent hybridisation between native UK flowering plant species or archaeophytes (introduced pre-1500 or potentially natives) (Stace, Preston, and Pearman 2015; Stace 1975; Preston and Pearman 2015). The hybrid flora also summarises their ecology, distribution, cytology and parentage (Stace, Preston, and Pearman 2015). We use the Barcode UK dataset (Jones et al. In Prep.), a new DNA barcoding resource that includes a three locus DNA barcode of *rbcL*, *matK*, and ITS2 for the native and archaeophyte seed plants of the UK. We integrate this DNA barcode data with information on hybridisation from the hybrid flora, as well as other species level traits and ecological information. We assess the determinants of hybridisation across the flora in the context of parental species range overlap, genus size, life history (annual vs perennial), ploidy differences, parental genetic distance and phylogeny.

Out of the 6117 possible unique pairwise congeneric combinations between 1100 species suitable for analysis in the UK flora (see Methods), only 7.8% produce hybrids. From 244 genera containing multiple species, 96 contain hybrids, and the 480 recorded hybrids are disproportionately concentrated in just five genera, with 45.8% of hybrids found in *Euphrasia* ($n = 62$), *Carex* ($n = 50$), *Rosa* ($n = 40$), *Epilobium* ($n = 35$), and *Salix* ($n = 33$). Some genera have many species that prolifically hybridise (e.g. *Euphrasia* has 4 species that are each parents to 14+ hybrids), while others include widespread species that disproportionately contribute to the number of hybrids (e.g. *Rumex crispus*, involved in 12 hybrid combinations in our data). We then explored whether the number of hybrids is a simple function of genus size using phylogenetic mixed models. Genus size is predicted to affect hybridisation as larger genera contain more possible pairwise combinations of species, and thus greater opportunity for hybridising with congeners (Johnson 2018). Our models show that the probability congeneric species hybridising is independent of genus size (pMCMC = 0.92, Supplementary Table 1). Although hybridisation tends to occur in species rich genera, not all species rich genera form hybrids (see also Preston and Pearman (2015)). For example, genera such as *Trifolium* (19 species; usually self-incompatible) and *Alchemilla* (12; all apomictic) are relatively species rich but form no naturally occurring hybrids, whilst *Veronica* (15) and *Galium* (13) form only one naturally occurring hybrid each. This shows that the number of hybrid taxa deviates from a simple model determined by the number of congeneric taxa, with other factors such as phylogenetic relatedness, life history, genetic distance, and ploidy level (explored below) interacting to determine hybridity.

We then investigate hybrid formation in the context of a newly generated 3-locus phylogeny of the British Flora from the Barcode UK data (see Methods). Hybridisation propensity, here defined as the number of hybrid combinations a species produces weighted by genus size, is highly uneven across the plant phylogeny (Figure 1). The phylogenetic signal of hybridisation is high at 0.62 (0.32-0.77 CI – 95% Credible Intervals) meaning that closely related lineages are likely to have similar levels of hybridisation (Supplementary Table 2). Inspection of the species level Best Linear Unbiased Predictors (BLUPs; see Methods) from the phylogenetic model shows that the monocot Asparagales clade including orchids (Orchidaceae, Iridaceae, Asparagaceae

and Amaryllidaceae) are most likely to hybridise after accounting for other model factors, whilst legumes (Fabaceae) are the least likely (Supplementary Figure 1). These results are consistent with previous work; many legumes, such as clovers (*Trifolium*) and related genera (e.g. peas, *Lathyrus*) are known to hybridise very little due to strong between species incompatibilities (Evans 1962), while orchids have been shown to hybridise rampantly, even across ploidy levels (De Hert et al. 2012). Overall, the pattern of hybridisation is highly heterogeneous across the phylogeny for the British flora, with phylogenetic position a good predictor of hybrid formation.

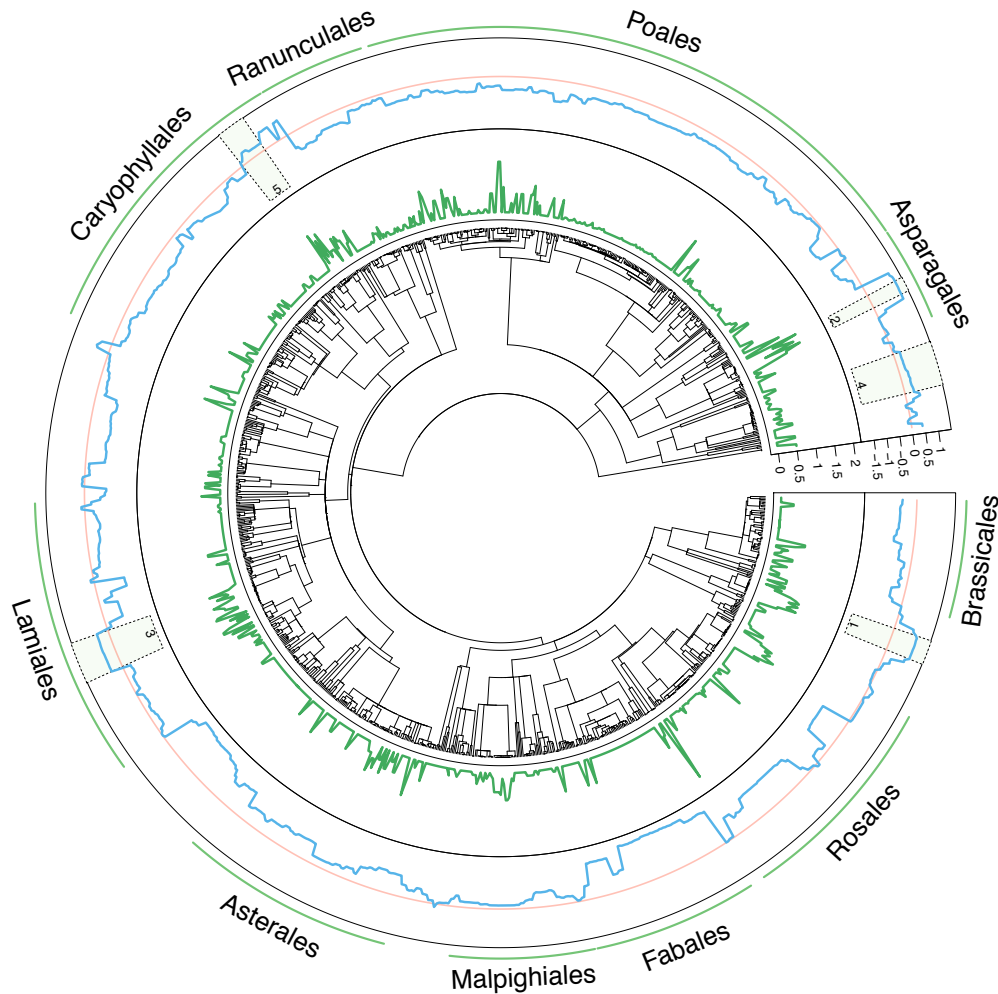


Figure 1: Distribution of hybrids across the phylogeny of the British flora. Innermost-ring shows phylogenetic relationships of 1098 British native species from an alignment of ITS2, matK and rbcL. Phylogenetic reconstructions used maximum likelihood implemented in IQ-TREE. The middle ring (green line) shows species-level hybrid propensity weighted by size of genus. The outer ring shows the probit scale posterior mean of the probability of a particular species hybridising (blue line). The zero line is represented in pale red and positive probit values indicate higher probabilities of hybridisation. The figure is annotated with the five genera with highest probabilities of hybridisation, given variation in model fixed effects, indicated from the sum of the species level posterior means from the phylogenetic model (1. *Epilobium*, 2. *Euphrasia*, 3. *Dactylorhiza*, 4. *Potamogeton*, and 5. *Rumex*), and the 10 largest plant orders around the outside.

Previous studies have explored the relationship between life history and hybridisation in plant species (Beddows and Rose 2018; Mitchell et al. 2019; Stace 1975; Ellstrand, Whitkus, and Rieseberg 1996; Preston and Pearman 2015). It has been documented that most hybridising species are perennial, at least in temperate floras (e.g. 97% in the Michigan flora Beddows and Rose (2018), 68% in our data). Out of all potential congeneric species pairs of the same life history in the British flora however, perennial only parental combinations form proportionally fewer hybrids (7.7%, $n = 4725$) than annual only parental combinations (15.6%, $n = 588$). Perennial plant species are thought to participate in hybridisation events more frequently because (a) perennials tend to outcross more than annuals, which tend to be more highly selfing (Morgan, 2001) and (b) perennial plants are longer lived and therefore more gametes are produced over a longer period of time (Ellstrand, Whitkus, and Rieseberg 1996). Despite the high frequency of perennial parent plant species we find no evidence for a significant effect of life history on the probability of hybrid formation ($\text{Chi}^2 = 4.04$, $\text{df}=2$, $P=0.13$; Supplementary Table 1). Life history may still be important in the persistence of hybrid lineages (Ellstrand and Schierenbeck 2000), however our results show that the probability of forming a hybrid is independent of parental species life history.

Increasing overlap in parental species distribution is expected to increase the probability of hybridisation due to the greater opportunity for crossing and therefore hybrid formation. Whilst this is an important variable when considering hybridisation, few studies have critically looked at range overlap with respect to hybridisation in multi-species systems (Mitchell et al. 2019; Preston and Pearman 2015). We leverage accurate information on the distribution of British plant species to infer the influence of species range overlap measured as a count of the number of 10x10km grid (hectad) overlaps. As expected, mean overlap in congeneric parental species distribution is higher for pairs of species known to give rise to hybrids (739 hectads ± 27 SE, Standard Error), compared to those that have not been recorded to successfully hybridise (353 hectads ($=10 \times 10 \text{km}^2$) ± 6 SE). Our models predict significantly higher probabilities of hybrid formation when there is larger overlap in parental species distribution (pMCMC < 0.001 , Supplementary Table 1). Although significant, the variability in the effect of parental distribution overlap is very low (posterior SD: 0.0001) compared to that of genetic distance between parental species, which is five orders of magnitude more variable (posterior SD: 3.77; Supplementary Figure 2). This suggests that range overlap and correlated attributes such as species abundance may be secondary to intrinsic genetic factors in determining hybridisation at a broad-spatial scale (Brown 1984), though they may be key factors at a local scale (Heinze 2011). In sum, there are opportunities for hybrids to occur even in areas where closely related parental species overlap little or not at all, with hybrid presence potentially affected by historical range overlap, long distance cross-pollination when parental species have coincident phenologies, or independent dispersal of hybrids (Lamont et al. 2003; Preston and Pearman 2015).

Ploidy level variation is frequent, both within and between species in the same genus (Husband, Baldwin, and Suda 2013). Hybrids formed from parental species of differing ploidy level (cross ploidy hybrids) appear in the British flora, with 131 detected (38% of hybridising species pairs with ploidy information Stace, Preston, and Pearman (2015)), but their relative importance has not yet been investigated across a flora. Hybrid formation is expected to decrease when the parental species have contrasting ploidy levels due to endosperm imbalance in the fertilised embryo (Tate, Soltis, and Soltis 2005). We used ploidy data for 684 species across the British Flora and determined for each pairwise comparison of species whether they were of the same or different ploidy. Our model shows that parental species with the same ploidy are 35% more likely to form hybrids than parents of differing ploidy levels, when fixed at mean overlap in geographical distribution, mean branch length between species pairs, and accounting for phylogenetic effects (pMCMC < 0.001 , Supplementary Table 3; Supplementary Figures 3 and 4). Cross ploidy hybridisation has been reported in many plant genera and has led to the generation of new species (Elkington 1984) and introgression of genes affecting fitness (Chapman and Abbott 2010), highlighting the importance of rare hybridisation between ploidy levels.

We then investigated the impact of parental genetic distance on hybrid formation across taxa in the flora. The likelihood of hybrid formation is expected to decrease with parental genetic distance due to a greater number

of genetic incompatibilities (Edmands 2002). However, low genetic distances may also be a consequence of genetic homogenisation from hybridisation. We observe a ten-fold variation in mean congeneric ITS distance (see Methods for definition) across the 35 genera containing hybrids with more than five taxa, from low mean pairwise distance in *Agrostis*, *Cochlearia* and *Rosa*, to high distance in *Geranium*, *Juncus* and *Saxifraga*. Overall, hybridising congeneric species showed a significantly lower pairwise genetic distance (mean ITS distance = 0.097, SE = 0.004) than non-hybridising congeneric species pairs (mean ITS distance = 0.215, SE = 0.001, Wilcoxon Test, $P < 0.001$; Figure 2). In our tree based phylogenetic models, the probability of forming a hybrid strongly decreases as branch length between parental species increases (pMCMC < 0.001 , Supplementary Table 1 and Supplementary Figure 5) and shows a greater standardised effect size than both pairwise overlap in distribution and size of genus. The stronger effect of genetic distance is evident from the joint probability distribution of hybridisation and geographical distance (Supplementary Figure 1), which are predicted by theory to be correlated (Felsenstein 1976). For species which hybridise, the highest average parental genetic distance is seen in the genus *Saxifraga* (ITS2 distance: 0.28), with other divergent hybridising taxa seen in the genera *Poa* (0.22), *Cardamine* (0.19), *Potamogeton* (0.17) and *Fumaria* (0.17), showing rare examples of hybrid formation between divergent taxa. In each of these five genera, hybrids that do form tend to be sterile and therefore introgression and genetic homogenisation are unlikely (Hegde et al. 2006). This contrasts with genera characterised by low mean parental genetic distance, such as *Salicornia* (ITS2 distance: 0.00), *Prunus* (0.00), *Rosa* (0.02), *Epipactis* (0.05), and *Atriplex* (0.05), where hybrids that form tend to be fertile (Stace, Preston, and Pearman 2015; Martin and Mendelson 2018). Overall these results show parental genetic divergence is a good predictor of hybrid formation, but that low parental genetic divergence may either be a cause of hybrid formation, or a consequence of genetic homogenisation. In cases where hybridisation does occur, there may be long lasting evolutionary consequences such as phenotypic novelty (Stelkens and Seehausen 2009) and polyploid hybrid speciation, especially when there is high parental divergence (Chapman and Burke 2007).

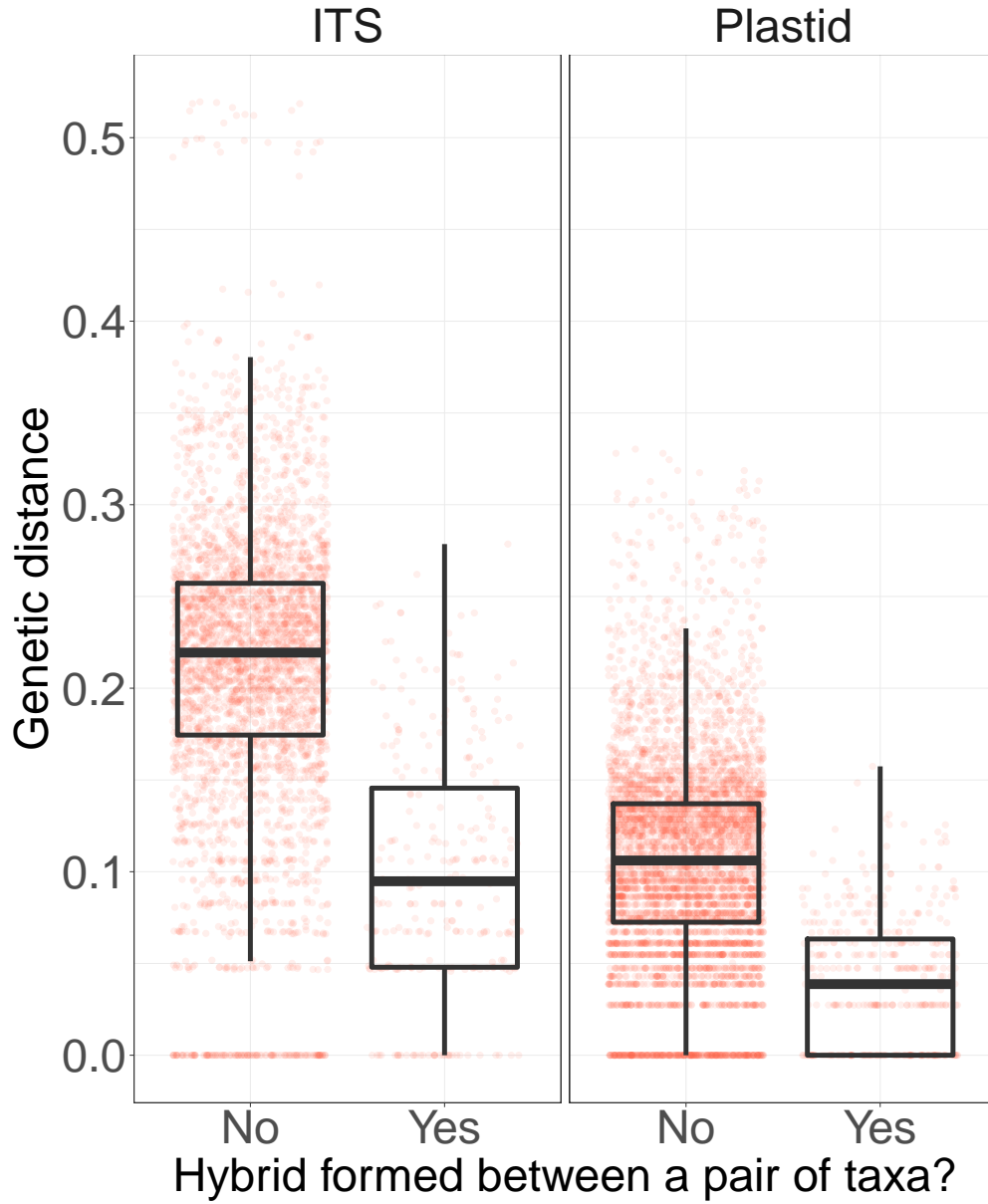


Figure 2: Hybrid formation in the context of genetic divergence in the British flora. Jittered points represent genetic distances between pairs of congeneric taxa, grouped by whether a pair of taxa produced a recorded hybrid or not. Both ITS and plastid loci are shown.

This study is the first to integrate a species level phylogeny, species traits, and ecological information for an entire flora, to understand the relative effects of genetic and ecological factors that affect hybridisation in plants. We report a high phylogenetic signal of hybridisation in flowering plants, supporting some (Whitney et al. 2010) but not all previous studies (Mitchell et al. 2019; Beddows and Rose 2018), and highlighting the importance of taking species relationships into account. Phylogenetic effects are complex, and may

be attributed to multiple unmeasured traits of parental species, such as habitat preferences, chromosomal stability, or mating systems (Ramsey, Bradshaw, and Schemske 2003; Bittencourt 2019; Brys et al. 2014). Genetic distance between parental species emerges as the strongest predictor of hybridisation, emphasising the importance of this variable known to be important in pre and postzygotic isolation (Edmands 2002; Moyle, Olson, and Tiffin 2004). Causality is not absolute however, as low genetic distance between parental species could also be driven by genetic homogenisation reducing sequence variability. Recently diverged species within genera such as *Euphrasia*, *Sorbus* and *Epipactis* (Ennos et al. 2012) are examples of genera with closely related species which hybridise extensively and where plastid and ITS introgression is likely. Hybridisation is also affected by ploidy level of parental species, with parents of differing ploidy less likely to form hybrids. Whilst ploidy differences are expected to lower the probability of hybridisation due to prezygotic barriers such as meiotic irregularities (Tate, Soltis, and Soltis 2005), the probability of forming a hybrid in this situation is still well above zero. Ploidy level can therefore be considered a leaky barrier to hybridisation (Abbott and Lowe 2004). Overall, genetic factors at all levels have a profound impact on hybridisation.

Other, non-genetic, factors have been reported to play an important role in the prezygotic isolation of plant species and therefore in hybrid formation (Widmer, Lexer, and Cozzolino 2009). Factors such as genus size and life history are expected to predict hybridisation (Whitney et al. 2010), however we find no evidence of this. In the case of genus size we attribute the lack of a clear cut effect to phylogenetic signal and genetic distance between species outweighing increased opportunity for hybridisation due to more congeners alone. Previous work has emphasised the importance of the perennial life history in hybridisation (Preston and Pearman 2015; Mitchell et al. 2019; Beddows and Rose 2018; Ellstrand, Whitkus, and Rieseberg 1996). We detect no effect of life history in the probability of hybrid formation, and attribute this lack of effect to a few different factors. High recorder effort in the British Isles means that even ephemeral annual hybrids are routinely found (Preston and Pearman 2015), and we accounted for fewer annual than perennial plant species in the British flora by modelling all pairwise intrageneric species pairs. We also disentangled hybrid formation from hybrid persistence by using observed hybrid occurrences. Lastly, it may be that the annual genus *Euphrasia* is having a marked effect, as it is one of the few solely annual genera with many hybrids recorded (Preston and Pearman 2015; Metherell and Rumsey 2018). Hybridisation is also constrained and shaped by the biogeography of parental species (Lowry et al. 2008). We reveal that a hybrid is more likely to be formed when parental species distributions are highly overlapping due to increased opportunity for crossing events; a factor not yet statistically modelled in any flora-wide study to date. Parental species overlap remains a crude estimate due to lack of resolution on fine scale co-occurrence, and does not take into account habitat change through space, or levels of habitat disturbance. Britain has a postglacial flora with high levels of disturbance which is known to change the landscape of hybridisation (Guo 2014; Abbott 1992) and more studies are needed in undisturbed habitats to understand hybridisation in a comparative context.

Methods

We extracted information on hybrid taxa, their parental progenitors, and ploidy from the Hybrid Flora of the British Isles (Stace, Preston, and Pearman 2015) and used the latest plant taxonomy according to the New Flora of the British Isles (Stace 2019). The Hybrid Flora excludes complex, apomictic groups like *Hieracium*, *Taraxacum* and *Rubus*. Further ploidy information was extracted from the BSBI Cytology database and the Kew C-value database using custom python scripts (see https://github.com/Euphrasiologist/web_mining). Ploidy levels were inferred using R scripts (https://github.com/Euphrasiologist/Floristic_DNA_Barcoding) and each was checked manually, using only chromosome counts based on native UK material. Species with multiple ploidy levels were excluded unless it was exactly known which cytotype contributed to the hybrid. We excluded: (a) hybrids known not to have formed in the British Isles (e.g. taxa introduced as hybrids), (b) triple hybrids, (c) dubious or doubtful hybrids, (d) crosses at below specific rank (subspecies, varieties) and (e) hybrids where at least one parent was a recently introduced non-native species (however archaeophytes,

introduced pre-1500, were included). We also removed the rare cases of intergeneric hybridisation (some Rosaceae, Poaceae and Orchidaceae) due to model scaling issues associated with looking at all possible species combinations across the flora. Downstream, hybrids were excluded if there was no barcode data associated with the parental species. Hybrid propensity was calculated by counting the total number of hybrid taxa that a particular parental taxa had participated in. We did not scale the hybrid propensity by genus size (sensu Whitney et al. (2010)) as the data was only used for visual interpretation.

We estimated phylogenetic relationships from the Barcode UK dataset (Jones et al. In Prep.), which includes a three locus DNA barcode for British native flowering plant species. Complex or apomictic groups, as omitted from the Hybrid Flora, were not sequenced (except for *Sorbus*, as there was well sampled DNA barcodes for the group). Due to the different diversities and alignment success of plastid and nuclear ribosomal DNA we used a single alignment of plastid sequences to infer relationships between all taxa, while ITS was aligned separately for each genus and thus only used to infer congeneric relationships. Plastid DNA was aligned for all taxa using ZZZ while ITS was aligned by genus, padded with Ns, and gapped using the program catfasta2phym (https://github.com/nylander/catfasta2phym). Phylogenetic inferences were made using IQ-TREE (Nguyen et al. 2015) in an analysis with three partitions allowing models of molecular evolution to differ between loci, and including a multifurcating constraint tree based on APGIV relationships generated with Phylomatic (Webb and Donoghue 2005). Tree support was estimated using 1000 ultrafast bootstraps (Hoang et al. 2018). Tree-based genetic distances were inferred using the R function cophenetic.phylo() from the package ape (Paradis, Claude, and Strimmer 2004) while separate pairwise distances for ITS and plastid DNA were calculated with the R function dist.alignment() from the seqinr package (Charif and Lobry 2007). The resulting values were the square root of pairwise distances. Tree manipulation took place in R, with the circular plot made with the R package circize (Gu et al. 2014); the phylogeny was coerced into a circular dendrogram for visualisation. Other plots were generated with the R package ggplot2 (Wickham 2016) and lattice (Sarkar 2008). All other data manipulation took place in R version 3.6.1 using base R, and packages data.table (Computer Program 2019) and dplyr (Wickham et al. 2020).

Our final dataset for the phylogenetic analysis contained a maximum of 466 hybrid combinations in 6103 unique congeneric pairwise combinations. In the analysis of genetic distance, the plastid genetic distance dataset contained 413 hybrid combinations, and the ITS genetic distance dataset 279. We used phylogenetic generalised linear mixed models implemented in the R package MCMCglmm as it allows for addition of a phylogeny with flexible variance structures for the random effects (Hadfield 2010). The response variable was a binary response of whether two congeneric species produced recorded hybrids or not, and was assumed to have residuals approximated by a probit distribution. We used parameter expanded priors for better mixing and fixed the residual variance at 1. The models were run for 1.3 million iterations with a thinning interval of 1000 and a burn-in of 300000. We used five fixed effect covariates to understand their contribution to explaining the variation in hybrid formation. Firstly, pairwise branch length between parental species calculated from the phylogeny (above) was added to understand the contribution of intrageneric relatedness. Pairwise overlap in geographical distribution (number of 10x10km²; generated from data at https://database.bsbi.org/) accounted for extent of overlap of parental species. Genus size for each species was calculated from species present in the phylogeny. Lastly, whether the parental species were of the same ploidy level or not, was added as a two level categorical factor; same ploidy level (homoploid) or not (heteroploid). We ran two models which differed in their fixed effect structure only, with the addition of the ploidy level and one without. This was because ploidy data was limited either due to missing counts or if ploidy was difficult to estimate; the number of species analysed decreased from 1098 to 684 upon addition of ploidy level. 148 out of 244 families contained missing ploidy data, with a few genera containing high amounts of missing data, e.g. *Sorbus*, *Alchemilla*, *Juncus*, and especially *Carex*. The phylogenetic (species level) best linear unbiased predictors (BLUP's; means of the posterior distribution) were extracted from the model excluding ploidy level and are equivalent to the per species point estimates of the probability of hybridisation.

The inverse relatedness matrix (unscaled phylogeny) and species were fitted as random effects in a multi-

membership model structure, as each hybrid event is the outcome of two parental species. The effect of phylogeny was added to the fixed effect predictions by calculating:

$$\sigma_u^2 + 2\sigma_s^2 + (\sigma_p^2)\overline{B} \quad (1)$$

Where u is the residual variance, s is the species variance and p is the phylogenetic variance. \overline{B} is defined as the average tree of all species pairs:

$$\overline{B} = +2\overline{d} + 2\overline{o} \quad (2)$$

Where \overline{d} is the average species phylogenetic variance and \overline{o} is the sum of the pairwise species phylogenetic covariances divided by the number of possible combinations of species multiplied by two. As we did not allow intergeneric hybridisation, we calculated \overline{B} at the species level constrained to genus (\overline{B}_S):

$$\overline{B}_S = \frac{2 \sum o_i n c_i + 2 \sum d_i n_i}{\sum n c_i} \quad (3)$$

i represents the i th genus. Variables o and d are defined as above but calculated for each genus, n_i is the number of individuals in a genus and $n c_i$ is the number of possible pairwise combinations within genus. Using this method, we were able to account for the size of genus in our phylogenetic variance estimates. We implement this algorithm in the R package VCVglimm (Brown 2019). Lastly, phylogenetic signal was calculated using:

$$\frac{\sigma_p^2 \overline{B}_S}{\sigma_u^2 + 2\sigma_s^2 + 2\sigma_p^2 \overline{B}_S} \quad (4)$$

All parameters are sampled from the posterior distribution of each coefficient, and distributions are summarised using modes and highest posterior density intervals at the 95% level. P-values were taken directly from model output for continuous covariates or categorical covariates with only two levels, otherwise (for life history effects) Wald Tests jointly tested all factor levels (Brown 2019). Genetic distance comparisons were made using Wilcoxon tests on the pairwise genetic distances, which does not assume that the distances approximate any distribution. All data and code for analyses are available at https://github.com/Euphrasiologist/Floristic_DNA_Barcoding.

Abbott, R. J. 1992. "PLANT Invasions, Interspecific Hybridization and the Evolution of New Plant Taxa." Journal Article. *Trends in Ecology & Evolution* 7 (12): 401–5. [https://doi.org/10.1016/0169-5347\(92\)90020-c](https://doi.org/10.1016/0169-5347(92)90020-c).

Abbott, R. J., A. C. Brennan, J. K. James, D. G. Forbes, M. J. Hegarty, and S. J. Hiscock. 2009. "Recent Hybrid Origin and Invasion of the British Isles by a Self-Incompatible Species, Oxford Ragwort (*Senecio Squalidus* L., Asteraceae)." Journal Article. *Biological Invasions* 11 (5): 1145–58. <https://doi.org/10.1007/s10530-008-9382-3>.

Abbott, R. J., and A. J. Lowe. 2004. "Origins, Establishment and Evolution of New Polyploid Species: *Senecio Cambrensis* and *S-Eboracensis* in the British Isles." Journal Article. *Biological Journal of the Linnean Society* 82 (4): 467–74. <https://doi.org/10.1111/j.1095-8312.2004.00333.x>.

Ainouche, M. L., P. M. Fortune, A. Salmon, C. Parisod, M. A. Grandbastien, K. Fukunaga, M. Ricou, and M. T. Misset. 2009. "Hybridization, Polyploidy and Invasion: Lessons from *Spartina* (Poaceae)." Journal Article. *Biological Invasions* 11 (5): 1159–73. <https://doi.org/10.1007/s10530-008-9383-2>.

- Bardy, K. E., P. Schonswetter, G. M. Schneeweiss, M. A. Fischer, and D. C. Albach. 2011. “Extensive Gene Flow Blurs Species Boundaries Among *Veronica Barrelieri*, *V. Orchidea* and *V. Spicata* (Plantaginaceae) in Southeastern Europe.” Journal Article. *Taxon* 60 (1): 108–21. <https://doi.org/10.1002/tax.601010>.
- Becker, M., N. Gruenheit, M. Steel, C. Voelckel, O. Deusch, P. B. Heenan, P. A. McLenachan, O. Kardailsky, J. W. Leigh, and P. J. Lockhart. 2013. “Hybridization May Facilitate in Situ Survival of Endemic Species Through Periods of Climate Change.” Journal Article. *Nature Climate Change* 3 (12): 1039–43. <https://doi.org/10.1038/nclimate2027>.
- Beddows, I., and L. E. Rose. 2018. “Factors Determining Hybridization Rate in Plants: A Case Study in Michigan.” Journal Article. *Perspectives in Plant Ecology Evolution and Systematics* 34: 51–60. <https://doi.org/10.1016/j.ppees.2018.07.003>.
- Bittencourt, N. S. 2019. “Reproductive Systems and Low Outbreeding Barriers Between *Jacaranda Cuspidifolia* and *J. Mimosifolia* (Jacarandaeae, Bignoniaceae).” Journal Article. *Nordic Journal of Botany* 37 (11): 12. <https://doi.org/10.1111/njb.02558>.
- Brown, J. H. 1984. “ON the Relationship Between Abundance and Distribution of Species.” Journal Article. *American Naturalist* 124 (2): 255–79. <https://doi.org/10.1086/284267>.
- Brown, Max. 2019. *VCVglimm: Manipulating Model Outputs from Mcmcglmm and Lme4 Objects, usually Involving Variance-Covariance Matrices*.
- Brys, R., A. Vanden Broeck, J. Mergeay, and H. Jacquemyn. 2014. “THE Contribution of Mating System Variation to Reproductive Isolation in Two Closely Related *Centaureum* Species (Gentianaceae) with a Generalized Flower Morphology.” Journal Article. *Evolution* 68 (5): 1281–93. <https://doi.org/10.1111/evo.12345>.
- Chapman, M. A., and R. J. Abbott. 2010. “Introgression of Fitness Genes Across a Ploidy Barrier.” Journal Article. *New Phytologist* 186 (1): 63–71. <https://doi.org/10.1111/j.1469-8137.2009.03091.x>.
- Chapman, M. A., and J. M. Burke. 2007. “Genetic Divergence and Hybrid Speciation.” Journal Article. *Evolution* 61 (7): 1773–80. <https://doi.org/10.1111/j.1558-5646.2007.00134.x>.
- Charif, D., and J.R. Lobry. 2007. “SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis.” Book Section. In *Structural Approaches to Sequence Evolution*, edited by U. Bastolla, M. Porto, H.E. Roman, and M. Vendruscolo, 207–32. Springer Verlag.
- Computer Program. 2019. <https://CRAN.R-project.org/package=data.table>.
- De Hert, K., H. Jacquemyn, S. Van Glabeke, I. Roldan-Ruiz, K. Vandepitte, L. Leus, and O. Honnay. 2012. “Reproductive Isolation and Hybridization in Sympatric Populations of Three *Dactylorhiza* Species (Orchidaceae) with Different Ploidy Levels.” Journal Article. *Annals of Botany* 109 (4): 709–20. <https://doi.org/10.1093/aob/mcr305>.
- Edmands, S. 2002. “Does Parental Divergence Predict Reproductive Compatibility?” Journal Article. *Trends in Ecology & Evolution* 17 (11): 520–27. [https://doi.org/10.1016/s0169-5347\(02\)02585-5](https://doi.org/10.1016/s0169-5347(02)02585-5).
- Elkington, T. T. 1984. “CYTOGENETIC Variation in the British Flora - Origins and Significance.” Journal Article. *New Phytologist* 98 (1): 101–18. <https://doi.org/10.1111/j.1469-8137.1984.tb06100.x>.
- Ellstrand, N. C., and K. A. Schierenbeck. 2000. “Hybridization as a Stimulus for the Evolution of Invasiveness in Plants?” Journal Article. *Proceedings of the National Academy of Sciences of the United States of America* 97 (13): 7043–50. <https://doi.org/10.1073/pnas.97.13.7043>.

- Ellstrand, N. C., R. Whitkus, and L. H. Rieseberg. 1996. "Distribution of Spontaneous Plant Hybrids." Journal Article. *Proceedings of the National Academy of Sciences of the United States of America* 93 (10): 5090–3. <https://doi.org/10.1073/pnas.93.10.5090>.
- Ennos, R. A., R. Whitlock, M. F. Fay, B. Jones, L. E. Neaves, R. Payne, I. Taylor, N. De Vere, and P. M. Hollingsworth. 2012. "Process-Based Species Action Plans: An Approach to Conserve Contemporary Evolutionary Processes That Sustain Diversity in Taxonomically Complex Groups." Journal Article. *Botanical Journal of the Linnean Society* 168 (2): 194–203. <https://doi.org/10.1111/j.1095-8339.2011.01206.x>.
- Evans, A. M. 1962. "SPECIES Hybridization in *Trifolium* .1. METHODS of Overcoming Species Incompatibility." Journal Article. *Euphytica* 11 (2): 164 –&. <https://doi.org/10.1007/bf00033789>.
- Felsenstein, J. 1976. "THEORETICAL Population-Genetics of Variable Selection and Migration." Journal Article. *Annual Review of Genetics* 10: 253–80. <https://doi.org/10.1146/annurev.ge.10.120176.001345>.
- Gomez-Rodriguez, C., A. Crampton-Platt, Mjtn Timmermans, A. Baselga, and A. P. Vogler. 2015. "Validating the Power of Mitochondrial Metagenomics for Community Ecology and Phylogenetics of Complex Assemblages." Journal Article. *Methods in Ecology and Evolution* 6 (8): 883–94. <https://doi.org/10.1111/2041-210x.12376>.
- Gu, Z. G., L. Gu, R. Eils, M. Schlesner, and B. Brors. 2014. "Circlize Implements and Enhances Circular Visualization in R." Journal Article. *Bioinformatics* 30 (19): 2811–2. <https://doi.org/10.1093/bioinformatics/btu393>.
- Guo, Q. F. 2014. "Plant Hybridization: The Role of Human Disturbance and Biological Invasion." Journal Article. *Diversity and Distributions* 20 (11): 1345–54. <https://doi.org/10.1111/ddi.12245>.
- Hadfield, J. D. 2010. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The Mcmcglmm R Package." Journal Article. *Journal of Statistical Software* 33 (2): 1–22. <Go to ISI>:WOS:000275203300001.
- Hegde, S. G., J. D. Nason, J. M. Clegg, and N. C. Ellstrand. 2006. "The Evolution of California's Wild Radish Has Resulted in the Extinction of Its Progenitors." Journal Article. *Evolution* 60 (6): 1187–97. <https://doi.org/10.1554/05-634.1>.
- Heinze, B. 2011. "Towards a Quantitative Description of Landscape, Demography and Flowering Phenology Effects on Realized Hybridization Potential." Journal Article. *Molecular Ecology* 20 (11): 2233–5. <https://doi.org/10.1111/j.1365-294X.2011.05098.x>.
- Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." Journal Article. *Molecular Biology and Evolution* 35 (2): 518–22. <https://doi.org/10.1093/molbev/msx281>.
- Husband, B.C., S.J. Baldwin, and J. Suda. 2013. "The Incidence of Polyploidy in Natural Plant Populations: Major Patterns and Evolutionary Processes." Book Section. In *Plant Genome Diversity*, edited by Greilhuber J, Dolezel J, and Wendel J. Vol. 2. Vienna: Springer.
- Johnson, S. D. 2018. "Natural Hybridization in the Orchid Flora of South Africa: Comparisons Among Genera and Floristic Regions." Journal Article. *South African Journal of Botany* 118: 290–98. <https://doi.org/10.1016/j.sajb.2018.01.011>.
- Lamont, B. B., T. He, N. J. Enright, S. L. Krauss, and B. P. Miller. 2003. "Anthropogenic Disturbance Promotes Hybridization Between *Banksia* Species by Altering Their Biology." Journal Article. *Journal of Evolutionary Biology* 16 (4): 551–57. <https://doi.org/10.1046/j.1420-9101.2003.00548.x>.
- Lexer, C., M. E. Welch, J. L. Murphy, and L. H. Rieseberg. 2003. "Natural Selection for Salt Tolerance Quantitative Trait Loci (Qtls) in Wild Sunflower Hybrids: Implications for the Origin of *Helianthus Paradoxus*,

- a Diploid Hybrid Species.” Journal Article. *Molecular Ecology* 12 (5): 1225–35. <https://doi.org/10.1046/j.1365-294X.2003.01803.x>.
- Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. “The Strength and Genetic Basis of Reproductive Isolating Barriers in Flowering Plants.” Journal Article. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363 (1506): 3009–21. <https://doi.org/10.1098/rstb.2008.0064>.
- Marques, I., J. Loureiro, D. Draper, M. Castro, and S. Castro. 2018. “How Much Do We Know About the Frequency of Hybridisation and Polyploidy in the Mediterranean Region?” Journal Article. *Plant Biology* 20: 21–37. <https://doi.org/10.1111/plb.12639>.
- Martin, M. D., and T. C. Mendelson. 2018. “Hybrid Sterility Increases with Genetic Distance in Snubnose Darters (Percidae: Etheostoma).” Journal Article. *Environmental Biology of Fishes* 101 (2): 215–21. <https://doi.org/10.1007/s10641-017-0672-4>.
- Metherell, C., and F. Rumsey. 2018. *Eyebrights (Euphrasia) of the Uk and Ireland*. Book. Botanical Society of Britain; Ireland.
- Mitchell, N., L. G. Campbell, J. R. Ahern, K. C. Paine, A. B. Giroldo, and K. D. Whitney. 2019. “Correlates of Hybridization in Plants.” Journal Article. *Evolution Letters*, 16. <https://doi.org/10.1002/evl3.146>.
- Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. “Patterns of Reproductive Isolation in Three Angiosperm Genera (Vol 58, Pg 1195, 2004).” Journal Article. *Evolution* 58 (9): 2110–0. <Go to ISI>://WOS:000224339500021.
- Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. “IQ-Tree: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” Journal Article. *Molecular Biology and Evolution* 32 (1): 268–74. <https://doi.org/10.1093/molbev/msu300>.
- Novak, S. J., D. E. Soltis, and P. S. Soltis. 1991. “OWNBEY Tragopogons - 40 Years Later.” Journal Article. *American Journal of Botany* 78 (11): 1586–1600. <https://doi.org/10.2307/2444984>.
- Paradis, E., J. Claude, and K. Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” Journal Article. *Bioinformatics* 20 (2): 289–90. <https://doi.org/10.1093/bioinformatics/btg412>.
- Preston, C. D., and D. A. Pearman. 2015. “Plant Hybrids in the Wild: Evidence from Biological Recording.” Journal Article. *Biological Journal of the Linnean Society* 115 (3): 555–72. <https://doi.org/10.1111/bij.12577>.
- Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. “Components of Reproductive Isolation Between the Monkeyflowers *Mimulus Lewisii* and *M-Cardinalis* (Phrymaceae).” Journal Article. *Evolution* 57 (7): 1520–34. <Go to ISI>://WOS:000184589300007.
- Rhymer, J. M., and D. Simberloff. 1996. “Extinction by Hybridization and Introgression.” Journal Article. *Annual Review of Ecology and Systematics* 27: 83–109. <https://doi.org/10.1146/annurev.ecolsys.27.1.83>.
- Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. “Transgressive Segregation, Adaptation and Speciation.” Journal Article. *Heredity* 83: 363–72. <https://doi.org/10.1038/sj.hdy.6886170>.
- Sarkar, Deepayan. 2008. *Lattice: Multivariate Data Visualization with R*. Book. New York: Springer.
- Stace, C.A. 1975. *Hybridization and the Flora of the British Isles*. Book. Academic Press.
- . 2019. *New Flora of the British Isles*. Book. 4th ed. C & M Floristics.
- Stace, C A., C D. Preston, and D Pearman. 2015. *Hybrid Flora of the British Isles*. Book. Bristol::Botanical Society of Britain & Ireland.
- Stelkens, R., and O. Seehausen. 2009. “GENETIC Distance Between Species Predicts Novel Trait Expression in Their Hybrids.” Journal Article. *Evolution* 63 (4): 884–97. <https://doi.org/10.1111/j.1558-5646.2008.00599.x>.

- Tate, J.A., D.E. Soltis, and P.S. Soltis. 2005. “Polyploidy in Plants.” Book Section. In *The Evolution of the Genome*, edited by R. Gregory, 371–426. Academic Press.
- Webb, C. O., and M. J. Donoghue. 2005. “Phyloomatic: Tree Assembly for Applied Phylogenetics.” Journal Article. *Molecular Ecology Notes* 5 (1): 181–83. <https://doi.org/10.1111/j.1471-8286.2004.00829.x>.
- Whitney, K. D., J. R. Ahern, L. G. Campbell, L. P. Albert, and M. S. King. 2010. “Patterns of Hybridization in Plants.” Journal Article. *Perspectives in Plant Ecology Evolution and Systematics* 12 (3): 175–82. <https://doi.org/10.1016/j.ppees.2010.02.002>.
- Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H., R. François, L. Henry, and K. Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Widmer, A., C. Lexer, and S. Cozzolino. 2009. “Evolution of Reproductive Isolation in Plants.” Journal Article. *Heredity* 102 (1): 31–38. <https://doi.org/10.1038/hdy.2008.69>.
- Young, M. R., H. C. Proctor, J. R. deWaard, and P. D. N. Hebert. n.d. “DNA Barcodes Expose Unexpected Diversity in Canadian Mites.” Journal Article. *Molecular Ecology*, 13. <https://doi.org/10.1111/mec.15292>.