**The genetic landscape of hybridisation in the UK flora**

Max R. Brown1,5,6, Peter M. Hollingsworth2, Laura Forrest2, Michelle Hart2, Laura Jones3, Col Ford3, Natasha de Vere3,4, Alex D. Twyford1,2,6

1 University of Edinburgh, Institute of Evolutionary Biology, Charlotte Auerbach Road, Edinburgh, EH9 3FL, UK

2 Royal Botanical Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, UK

3 Conservation and Research Department, National Botanic Garden of Wales, Llanarthne, Carmarthenshire, SA32 8HG, UK

4 Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, UK.

5 Current address: Wellcome Trust Genome Campus, Hinxton, CB10 1RQ

6 Author for correspondence (e-mail mb39@sanger.ac.uk); ORCID id - <https://orcid.org/0000-0003-2561-516X>

**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 et al., 1996), the ecological factors and species traits that promote or prevent hybridisation across species 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 use phylogenetic mixed effect models to combine estimates of hybridisation from extensive field observations (Stace et al., 2015), with DNA barcoding data and ecological attributes for over 1,000 British 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 impacts the likelihood of hybrid formation, with a congeneric hybrid pair more likely to hybridise when at least one of the species is perennial. 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 et al., 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 et al., 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 specific results may generalise to diverse natural plant communities and entire floras. 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 et al., 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.). 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 (Young et al., 2019). 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, 2019). 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 et al., 2015, Stace, 1975, Preston and Pearman, 2015). The hybrid flora also summarises their ecology, distribution, cytology and parentage (Stace et al., 2015). We use the Barcode UK dataset (Jones et al., 2021), a new DNA barcoding resource of a three locus DNA barcode of *rbcL*, *matK*, and ITS2 for the native and archaeophyte seed plants of the UK. We use this data to build a species-level phylogeny, and subsequently integrate this 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), 7.8% produce hybrids. From 244 genera containing multiple species, 96 contain hybrids, and the 480 recorded hybrids are disproportionately concentrated in five genera, with 45.8% of all 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 parent to 14+ hybrids), while others include widespread species that disproportionately contribute to the number of hybrids (e.g. *Rumex crispus* isinvolved in 12 hybrid combinations). We then explored whether the number of hybrids is a simple function of genus size as would be predicted given the greater opportunity of hybridisation in large genera with more possible pairwise congeneric combinations (Johnson, 2018), using phylogenetic mixed models. Our models show that the probability of 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; 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 0.54 (0.31-0.69 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, and rendered into a time tree using treePL (see Methods). 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. *Rumex*, 2. *Euphrasia*, 3. *Epilobium*, 4. *Potamogeton*, and 5. *Dactylorhiza*), 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 et al., 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). From the 15 annual genera where only annual parental combinations are reported, 69% of these hybrid reports were from a single genus, *Euphrasia*.

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 et al., 1996). We find a significant effect of life history of parental species on the probability of hybrid formation (χ2 = 8.57, df=2, P=0.0138; Supplementary Table 1), which shows that even a single parental species being perennial has an impact on the probability of hybrid formation.

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 (Ireland, 2015). 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 ±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 (Brown, 1984) may be secondary to intrinsic genetic factors in determining hybridisation at a broad-spatial scale, 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 et al., 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 et al., 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 et al., 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). 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 (Martin and Mendelson, 2018). Overall these results show parental genetic divergence is a good predictor of hybrid formation. However, low parental genetic divergence may either be a cause of hybrid formation, or a consequence of subsequent 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 and divergence time 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. ITS and plastid loci are shown, along with divergence time between taxa generated from our phylogenetic tree.

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 et al., 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 et al., 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 et al., 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 et al., 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 et al., 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 et al., 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 (BSBI, 2019) and the Kew C-value database (Leitch, 2019) using custom python scripts (see <https://github.com/Euphrasiologist/web_mining>). Ploidy levels were inferred using only chromosome counts based on native UK material, and each was checked manually. 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., 2021), 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*). 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 catfasta2phyml (<https://github.com/nylander/catfasta2phyml>). 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). We used the phylogenetic software TreePL (https://github.com/blackrim/treePL) to infer divergence times for our phylogeny with fossil calibration dates taken from (Ramirez-Barahona et al., 2020). Tree-based genetic distances were inferred using the R function cophenetic.phylo() from the package ape (Paradis et al., 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 circlize (Gu et al., 2014); the phylogeny was coerced into a circular dendrogram for visualisation. Other plots were generated with the R package ggplot2 (Wickham, 2009) and lattice (Sarkar, 2008). All other data manipulation took place in R version 3.6.1 using base R, and packages data.table (Dowle and Srinivasan, 2019) and dplyr (Wickham and Francois, 2016).

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:

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

Where is the average species phylogenetic variance and 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 at the species level constrained to genus ():

represents the th genus. Variables and are defined as above but calculated for each genus, is the number of individuals in a genus and 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 VCVglmm (Brown, 2019). Lastly, phylogenetic signal was calculated using:

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/XXX>.

**References: [not yet formatted]**

ABBOTT, R. J. 1992. PLANT INVASIONS, INTERSPECIFIC HYBRIDIZATION AND THE EVOLUTION OF NEW PLANT TAXA. *Trends in Ecology & Evolution,* 7**,** 401-405.

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**,** 1145-1158.

ABBOTT, R. J. & LOWE, A. J. 2004. Origins, establishment and evolution of new polyploid species: Senecio cambrensis and S-eboracensis in the British Isles. *Biological Journal of the Linnean Society,* 82**,** 467-474.

AINOUCHE, M. L., FORTUNE, P. M., SALMON, A., PARISOD, C., GRANDBASTIEN, M. A., FUKUNAGA, K., RICOU, M. & MISSET, M. T. 2009. Hybridization, polyploidy and invasion: lessons from Spartina (Poaceae). *Biological Invasions,* 11**,** 1159-1173.

BARDY, K. E., SCHONSWETTER, P., SCHNEEWEISS, G. M., FISCHER, M. A. & ALBACH, D. C. 2011. Extensive gene flow blurs species boundaries among Veronica barrelieri, V. orchidea and V. spicata (Plantaginaceae) in southeastern Europe. *Taxon,* 60**,** 108-121.

BECKER, M., GRUENHEIT, N., STEEL, M., VOELCKEL, C., DEUSCH, O., HEENAN, P. B., MCLENACHAN, P. A., KARDAILSKY, O., LEIGH, J. W. & LOCKHART, P. J. 2013. Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change,* 3**,** 1039-1043.

BEDDOWS, I. & ROSE, L. E. 2018. Factors determining hybridization rate in plants: A case study in Michigan. *Perspectives in Plant Ecology Evolution and Systematics,* 34**,** 51-60.

BITTENCOURT, N. S. 2019. Reproductive systems and low outbreeding barriers between Jacaranda cuspidifolia and J. mimosifolia (Jacarandeae, Bignoniaceae). *Nordic Journal of Botany,* 37**,** 12.

BROWN, J. H. 1984. ON THE RELATIONSHIP BETWEEN ABUNDANCE AND DISTRIBUTION OF SPECIES. *American Naturalist,* 124**,** 255-279.

BROWN, M. 2019. VCVglmm: Manipulating model outputs from MCMCglmm and lme4 objects, usually involving  variance-covariance matrices. 0.1.0 ed.

BRYS, R., VANDEN BROECK, A., MERGEAY, J. & JACQUEMYN, H. 2014. THE CONTRIBUTION OF MATING SYSTEM VARIATION TO REPRODUCTIVE ISOLATION IN TWO CLOSELY RELATED CENTAURIUM SPECIES ( GENTIANACEAE) WITH A GENERALIZED FLOWER MORPHOLOGY. *Evolution,* 68**,** 1281-1293.

BSBI. 2019. *BSBI Cytology database* [Online]. <http://rbg-web2.rbge.org.uk/BSBI/cytsearch.php>. [Accessed 2019].

CHAPMAN, M. A. & ABBOTT, R. J. 2010. Introgression of fitness genes across a ploidy barrier. *New Phytologist,* 186**,** 63-71.

CHAPMAN, M. A. & BURKE, J. M. 2007. Genetic divergence and hybrid speciation. *Evolution,* 61**,** 1773-1780.

CHARIF, D. & LOBRY, J. R. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *In:* BASTOLLA, U., PORTO, M., ROMAN, H. E. & VENDRUSCOLO, M. (eds.) *Structural approaches to sequence evolution.* Springr Verlag.

DE HERT, K., JACQUEMYN, H., VAN GLABEKE, S., ROLDAN-RUIZ, I., VANDEPITTE, K., LEUS, L. & HONNAY, O. 2012. Reproductive isolation and hybridization in sympatric populations of three Dactylorhiza species (Orchidaceae) with different ploidy levels. *Annals of Botany,* 109**,** 709-720.

DOWLE, M. & SRINIVASAN, A. 2019. data.table: Extension of `data.frame`. R package version 1.12.8 ed.

EDMANDS, S. 2002. Does parental divergence predict reproductive compatibility? *Trends in Ecology & Evolution,* 17**,** 520-527.

ELKINGTON, T. T. 1984. CYTOGENETIC VARIATION IN THE BRITISH FLORA - ORIGINS AND SIGNIFICANCE. *New Phytologist,* 98**,** 101-118.

ELLSTRAND, N. C., WHITKUS, R. & RIESEBERG, L. H. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 5090-5093.

ENNOS, R. A., WHITLOCK, R., FAY, M. F., JONES, B., NEAVES, L. E., PAYNE, R., TAYLOR, I., DE VERE, N. & HOLLINGSWORTH, P. M. 2012. Process-Based Species Action Plans: an approach to conserve contemporary evolutionary processes that sustain diversity in taxonomically complex groups. *Botanical Journal of the Linnean Society,* 168**,** 194-203.

EVANS, A. M. 1962. SPECIES HYBRIDIZATION IN TRIFOLIUM .1. METHODS OF OVERCOMING SPECIES INCOMPATIBILITY. *Euphytica,* 11**,** 164-&.

FELSENSTEIN, J. 1976. THEORETICAL POPULATION-GENETICS OF VARIABLE SELECTION AND MIGRATION. *Annual Review of Genetics,* 10**,** 253-280.

GOMEZ-RODRIGUEZ, C., CRAMPTON-PLATT, A., TIMMERMANS, M., BASELGA, A. & VOGLER, A. P. 2015. Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. *Methods in Ecology and Evolution,* 6**,** 883-894.

GU, Z. G., GU, L., EILS, R., SCHLESNER, M. & BRORS, B. 2014. circlize implements and enhances circular visualization in R. *Bioinformatics,* 30**,** 2811-2812.

GUO, Q. F. 2014. Plant hybridization: the role of human disturbance and biological invasion. *Diversity and Distributions,* 20**,** 1345-1354.

HADFIELD, J. D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software,* 33**,** 1-22.

HEGDE, S. G., NASON, J. D., CLEGG, J. M. & ELLSTRAND, N. C. 2006. The evolution of California's wild radish has resulted in the extinction of its progenitors. *Evolution,* 60**,** 1187-1197.

HEINZE, B. 2011. Towards a quantitative description of landscape, demography and flowering phenology effects on realized hybridization potential. *Molecular Ecology,* 20**,** 2233-2235.

HOANG, D. T., CHERNOMOR, O., VON HAESELER, A., MINH, B. Q. & VINH, L. S. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution,* 35**,** 518-522.

HUSBAND, B. C., BALDWIN, S. J. & SUDA, J. 2013  The Incidence of Polyploidy in Natural Plant Populations: Major Patterns and Evolutionary Processes. *In:* J, G., J, D. & J, W. (eds.) *Plant Genome Diversity.* Vienna: Springr.

IRELAND, B. S. O. B. A. 2015. *Taxon-hectad validation (NBN record cleaner ruleset)* [Online]. Available: <https://database.bsbi.org/object.php?entityid=2cd4p9h.3hb0st&class=ChecklistInstance> [Accessed 2020].

JOHNSON, S. D. 2018. Natural hybridization in the orchid flora of South Africa: Comparisons among genera and floristic regions. *South African Journal of Botany,* 118**,** 290-298.

JONES, L., TWYFORD, A. D., FORD, C. R., RICH, T. C. G., DAVIES, H., FORREST, L. L., HART, M. L., MCHAFFIE, H., BROWN, M. R., HOLLINGSWORTH, P. M. & DE VERE, N. 2021. Barcode UK: A complete DNA barcoding resource for the flowering plants and conifers of the United Kingdom. *Molecular Ecology Resources,* 21**,** 2050-2062.

LAMONT, B. B., HE, T., ENRIGHT, N. J., KRAUSS, S. L. & MILLER, B. P. 2003. Anthropogenic disturbance promotes hybridization between Banksia species by altering their biology. *Journal of Evolutionary Biology,* 16**,** 551-557.

LEITCH, I. J. J., E. PELLICER, J. HIDALGO, O. BENNETT, M.D. 2019. *Plant DNA C-values Database* [Online]. [Accessed 2020].

LEXER, C., WELCH, M. E., DURPHY, J. L. & RIESEBERG, L. H. 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. *Molecular Ecology,* 12**,** 1225-1235.

LOWRY, D. B., MODLISZEWSKI, J. L., WRIGHT, K. M., WU, C. A. & WILLIS, J. H. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 363**,** 3009-3021.

MARQUES, I., LOUREIRO, J., DRAPER, D., CASTRO, M. & CASTRO, S. 2018. How much do we know about the frequency of hybridisation and polyploidy in the Mediterranean region? *Plant Biology,* 20**,** 21-37.

MARTIN, M. D. & MENDELSON, T. C. 2018. Hybrid sterility increases with genetic distance in snubnose darters (Percidae: Etheostoma). *Environmental Biology of Fishes,* 101**,** 215-221.

METHERELL, C. & RUMSEY, F. 2018. *Eyebrights (Euphrasia) of the UK and Ireland*, Botanical Society of Britain and Ireland.

MITCHELL, N., CAMPBELL, L. G., AHERN, J. R., PAINE, K. C., GIROLDO, A. B. & WHITNEY, K. D. 2019. Correlates of hybridization in plants. *Evolution Letters***,** 16.

MORGAN, M. T. 2001. Consequences of life history for inbreeding depression and mating system evolution in plants. *Proceedings of the Royal Society B-Biological Sciences,* 268**,** 1817-1824.

MOYLE, L. C., OLSON, M. S. & TIFFIN, P. 2004. Patterns of reproductive isolation in three angiosperm genera (vol 58, pg 1195, 2004). *Evolution,* 58**,** 2110-2110.

NGUYEN, L. T., SCHMIDT, H. A., VON HAESELER, A. & MINH, B. Q. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution,* 32**,** 268-274.

NOVAK, S. J., SOLTIS, D. E. & SOLTIS, P. S. 1991. OWNBEY TRAGOPOGONS - 40 YEARS LATER. *American Journal of Botany,* 78**,** 1586-1600.

PARADIS, E., CLAUDE, J. & STRIMMER, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics,* 20**,** 289-290.

PRESTON, C. D. & PEARMAN, D. A. 2015. Plant hybrids in the wild: evidence from biological recording. *Biological Journal of the Linnean Society,* 115**,** 555-572.

RAMIREZ-BARAHONA, S., SAUQUET, H. & MAGALLON, S. 2020. The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution,* 4**,** 1232-+.

RAMSEY, J., BRADSHAW, H. D. & SCHEMSKE, D. W. 2003. Components of reproductive isolation between the monkeyflowers Mimulus lewisii and M-cardinalis (Phrymaceae). *Evolution,* 57**,** 1520-1534.

RHYMER, J. M. & SIMBERLOFF, D. 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics,* 27**,** 83-109.

RIESEBERG, L. H., ARCHER, M. A. & WAYNE, R. K. 1999. Transgressive segregation, adaptation and speciation. *Heredity,* 83**,** 363-372.

SARKAR, D. 2008. *Lattice: Multivariate Data Visualization with R,* New York, Springr.

STACE, C. A. 1975. *Hybridization and the Flora of the British Isles.*, Academic Press.

STACE, C. A. 2019. *New Flora of the British Isles*, C & M Floristics.

STACE, C. A., C D. PRESTON & PEARMAN, D. 2015. *Hybrid Flora of the British Isles,* Bristol, :Botanical Society of Britain & Ireland.

STELKENS, R. & SEEHAUSEN, O. 2009. GENETIC DISTANCE BETWEEN SPECIES PREDICTS NOVEL TRAIT EXPRESSION IN THEIR HYBRIDS. *Evolution,* 63**,** 884-897.

TATE, J. A., SOLTIS, D. E. & SOLTIS, P. S. 2005. Polyploidy in Plants. *In:* GREGORY, R. (ed.) *The Evolution of the Genome.* Academic Press.

WEBB, C. O. & DONOGHUE, M. J. 2005. Phylomatic: tree assembly for applied phylogenetics. *Molecular Ecology Notes,* 5**,** 181-183.

WHITNEY, K. D., AHERN, J. R., CAMPBELL, L. G., ALBERT, L. P. & KING, M. S. 2010. Patterns of hybridization in plants. *Perspectives in Plant Ecology Evolution and Systematics,* 12**,** 175-182.

WICKHAM, H. 2009. ggplot2: elegant graphics for data analysis. Springer New York.

WICKHAM, H. & FRANCOIS, R. 2016. dplyr: A Grammar of Data Manipulation

WIDMER, A., LEXER, C. & COZZOLINO, S. 2009. Evolution of reproductive isolation in plants. *Heredity,* 102**,** 31-38.

YOUNG, M. R., PROCTOR, H. C., DEWAARD, J. R. & HEBERT, P. D. N. DNA barcodes expose unexpected diversity in Canadian mites. *Molecular Ecology***,** 13.

YOUNG, M. R., PROCTOR, H. C., DEWAARD, J. R. & HEBERT, P. D. N. 2019. DNA barcodes expose unexpected diversity in Canadian mites. *Molecular Ecology***,** 13.

**Supplementary Information:**

**Source code:**

Source code for all the analyses and figures can be found at https://github.com/Euphrasiologist/Floristic\_DNA\_Barcoding .

**Model outputs:**

Table 1: Model 1: Probability of hybridisation with genetic distance, hectads shared and life history of parental species as fixed covariates. The posterior mean of the distribution of each coefficient is given, along with lower and upper 95% Credible Intervals. The p-value (pMCMC) is also reported and given in bold where significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Posterior mean | l-95% CI | u-95% CI | Effective sample size | pMCMC |
| (Intercept) | -1.78 | -3.71 | 0.32 | 1000 | 0.096 |
| Branch length between species pairs | -0.22 | -0.24 | -0.19 | 415 | **0.0010** |
| Hectads shared between species pairs | 0.0006 | 0.0004 | 0.0008 | 1000 | **0.0010** |
| Annual-perennial parent pair | -0.44 | -1.23 | 0.40 | 872 | 0.28 |
| Perennial-perennial parent pair | 0.61 | -0.41 | 1.51 | 1000 | 0.21 |
| Genus size | 0.0081 | -0.031 | 0.049 | 1000 | 0.69 |

To test the joint effect of life history (the levels annual-perennial parent pair and perennial-perennial parent pair), a Wald test was performed on the posterior mean distributions and the covariances between these two factor levels to determine whether the joint effect of life history was significantly different from zero. The result of this test is:

(χ2 = 8.57, df=2, P=0.0138)

Therefore, despite the individual tests shown in the summary in the table above being non significant, we see an overall significant effect of life history. This is because the factor levels are correlated in their effect.

**Phylogenetic signal:**

Table 2: Phylogenetic signal of probability of hybridisation and the species variance independent of phylogenetic effects. 95% Credible Intervals of the variances are also presented. See methods for calculation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variance Component** | **Posterior Mode** | **Lower Credible Interval** | **Upper Credible Interval** |
| Model 1 Phylogenetic Variance | 0.54 | 0.31 | 0.69 |
| Model 1 Species Variance | 0.43 | 0.25 | 0.60 |
| Model 2 Phylogenetic Variance | 0.58 | 0.30 | 0.72 |
| Model 2 Species Variance | 0.43 | 0.13 | 0.45 |

Table 3: Model 2: Probability of hybridisation with ploidy, genetic distance, hectads shared and life history of parental species as covariates. The posterior mean of the distribution of each coefficient is given, along with lower and upper 95% Credible Intervals. The p-value (pMCMC) is also reported and given in bold where significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Posterior mean | l-95% CI | u-95% CI | Effective sample size | pMCMC |
| (Intercept) | -0.51 | -2.52 | 1.64 | 1000 | 0.62 |
| Branch length between species pairs | -0.23 | -0.26 | -0.19 | 7134 | **0.0010** |
| Crosss ploidy effect | -0.76 | -1.05 | -0.47 | 1204 | **0.0010** |
| Hectads shared between species pairs | 0.0008 | 0.0006 | 0.0011 | 1000 | **0.0010** |
| Annual-perennial parent pair | -0.60 | -1.74 | 0.40 | 1000 | 0.28 |
| Perennial-perennial parent pair | 0.55 | -0.55 | 1.84 | 1000 | 0.39 |
| Genus size | -0.018 | -0.080 | 0.036 | 1000 | 0.57 |

**Figures:**

Figure 1: Trees with root nodes containing the highest and lowest posterior mean probability of hybridisation from Model 1 (BLUP’s of nodes in the phylogeny). A is the top tree (subset of Polygonaceae) whilst B is the tree with lowest probability of hybridisation (Fabaceae and Polygalaceae).



Figure 2: The joint probability of hybridisation between two parental species give both branch length between species (tree based divergence time between species measured in millions of years) and geographical overlap between parental species (measured as overlap in occupancy of 10x10km grid squares in the UK). The degree of shading in the scale bar and tiles represent the posterior probability of hybridisation from Model 1 given parameter values for each variable. Estimates are visualised at mean genus size, for annual-perennial parental combinations and accounting for phylogenetic relationships between species.



Figure 3: Predicted fit of probability of hybridisation given hectad sharing and ploidy difference of parental species from Model 2. Dashed lines indicate the 95% Credible Intervals, and the bold lines represent the posterior mode of the coefficients of congeneric pairs of species hybridising as a function of pairwise overlap in distribution, conditional on parental ploidy status. The effect is visualised at mean divergence time between between all pairs of species for annual-perennial parent combinations and accounting for phylogenetic effects. The bold red dashed line indicates mean pairwise overlap in distribution (10x10km2).



Figure 4: Predicted fit of probability of hybridisation given divergence time between parental species and ploidy difference of parental species from Model 2. Homoploid indicates parental species of the same ploidy level, and heteroploidy indicates parental species of different ploidy levels. Dashed lines indicate the 95% Credible Intervals, and the bold lines represent the posterior mode of the coefficients of congeneric pairs of species hybridising as a function of pairwise divergence time, conditional on parental ploidy status. The effect is visualised at mean hectad sharing for annual-perennial parent combinations and accounting for phylogenetic effects. The red dashed line indicates mean pairwise divergence time between all pairs of species.



Figure 5: Predicted fit of probability of hybridisation given divergence time between parental species from Model 1. Black dashed lines are the 95% Credible Intervals, bold line is the posterior mean of the coefficient for the probability of congeneric pairs of species hybridising as a function of branch length. This effect is visualised at mean hectad sharing, for annual-perennial parent combinations and accounting for phylogenetic effects. The bold red dashed line indicates mean genus level divergence time between pairs of species.

