Chapter 4 - Is there evidence of diploid-tetraploid hybridisation in a *Euphrasia* contact zone?

Max R. Brown Hannes Becher Sebastian Williams Alex D. Twyford

Abstract

Cross-ploidy hybridisation is expected to be rare due to the strong ploidal prezygotic barrier, however in some cases cross-ploidy hybridisation has led to the creation of new hybrid species. Hybridisation between species that differ in ploidy level however, has been reported across flowering plants. British eyebrights (*Euphrasia*) represent a good study system to investigate cross-ploidy hybridisation, as at least two putative hybrid species (*E. vigursii* and *E. rivularis*) have formed from progenitor species of different ploidy levels. In this study, we analysed a contact zone between the diploid species *E. rostkoviana* and the tetraploid species *E. arctica* in Wales. We sequenced the internal transcribed spacer region (ITS1), and used Genotyping by Sequencing (GBS) to look for evidence of cross-ploidy hybridisation and introgression. All sites in the ITS1 region were fixed between diploids and tetraploids, indicating a strong barrier to hybridisation. Further, analysis of the GBS data using PCA, STRUCTURE, and AMOVA across 270 SNPs, indicated clear separation between the ploidy levels. While the global F_{ST} between species was high at 0.44, the distribution across all SNPs was bimodal, indicating potential differential selection on loci between diploids and tetraploids. Only using demographic inference with $\delta a \delta I$ did we find evidence of limited gene flow – around one or fewer migrants per generation. Overall, our results are consistent with cross-ploidy hybridisation being rare or absent. Our work shows that the ploidy barrier to hybridisation in *Euphrasia* is strong using a variety of methods, and adds to the growing body of research on cross-ploidy hybridisation in plants.

Introduction:

Natural hybridisation is an important evolutionary phenomenon with wide ranging consequences, from extinction (Rhymer and Simberloff 1996), to hybrid speciation (Hegarty and Hiscock 2005). Most studies to date have investigated hybridisation between diploid species, while hybridisation between species that differ in their ploidy level (cross-ploidy hybridisation) has generally received less attention. While this may in part be due to technical issues with inferring homology relationships between diploids and polyploids, there are also clear biological reasons. Contrasting ploidy levels represent a known, highly effective barrier to hybridisation (Husband and Sabara 2004). The main barriers are abnormal endosperm ratios of maternal:paternal genomes at fertilisation which prevent hybrid seed formation (Johnston et al. 1980), and later hybrid sterility caused by irregularities in chromosome pairing at meiosis leading to aneuploid gametes (Tate, Soltis, and Soltis 2005). Both of these factors prevent hybridisation and introgression between species with contrasting ploidy levels. These barriers can be overcome through unreduced gamete production in the lower ploidy parent, or where a triploid (or other intermediate ploidy) F1 hybrid is formed, by either backcrossing to one of the parental species (Ramsey and Schemske 1998), or by whole genome duplication to restore fertility (Abbott and Lowe 2004). Cross ploidy hybridisation may be an important as a mechanism for maintaining genetic variation in polyploid species (although distinguishing this from recurrent polyploidisation can be difficult Shimizu-Inatsugi et al. (2009)), exchanging adaptive alleles between species (Chapman and Abbott 2010), and has been shown to generate new polyploid cytotypes or species (Abbott and Lowe 2004).

A number of natural cross-ploidy hybridisation examples exist in the literature and these cover a variety of phylogenetically distinct taxa. For example, there is cross ploidy hybridisation reported in *Dactylorhiza* (De Hert et al. 2012), *Mercurialis* (Buggs and Pannell 2007), and *Epidendrum* (Pinheiro et al. 2010), the latter two of which often form hybrid zones. Where hybridisation is particularly common, hybrid swarms can develop, as seen in co-occurring diploid and tetraploid species of *Cochlearia* (Fearn 1977). Cross ploidy hybridisation in the genera *Senecio* and *Mimulus* have led to the creation of three hybrid species endemic to Britain (Abbott and Lowe 2004; Vallejo-Marin 2012). Two of these hybrid species have resulted from whole genome duplication of initial triploid F1 hybrids (*Senecio cambrensis* and *Mimulus peregrinus*), whilst the other species was created through introgression to the tetraploid parent (*Senecio eboracensis*). Although far from exhaustive, these examples highlight that cross-ploidy hybridisation involves mostly diploid and tetraploid species and introgression, when it occurs, is usually in the direction of the higher ploidy parental species. Here, the higher likelihood of unreduced gametes produced by the lower ploidy parent may allow it to form fertile lineages with the higher ploidy parent (Stebbins 1971; Baduel et al. 2018).

Euphrasia (Orobanchaceae) is a large temperate genus of hemiparasitic plants, with around 263 species worldwide (Nickrent pers. comm.). In Britain and Ireland, there are 21 species of Euphrasia, which and are considered a taxonomically complex group characterised by recent postglacial divergence (Wang et al. 2018), plastic phenotypes (Brown et al. 2020), and the widespread occurrence of natural hybridisation (Metherell and Rumsey 2018). The genus in Britain and Ireland consists of five diploid and sixteen tetraploid species. This ploidy difference is associated with contrasting mating systems where the tetraploids are mixed maters, or highly selfing, while the diploids outcross more extensively (French et al. 2005). The tetraploid species are allotetraploids, with two subgenomes that are around 5% divergent. One subgenome of the tetraploid species is only 0.2% divergent from the diploid species genome, making it likely that the British diploid species are one of the parents of British tetraploid species (Becher et al. 2020). Out of 72 hybrid Euphrasia combinations reported in the British flora, 13 are reported to be diploid-tetraploid hybrids based on morphology (Stace, Preston, and Pearman 2015). Unusually, these cross ploidy hybrids are purported to be diploids derived from triploid F1s that backcross to the diploid parent (Yeo 1956). This contrasts with most prior predictions as to the directionality of cross ploidy hybridisation, where the hybrids are tetraploids. In the proposed scenario, genetic material from the tetraploid species is expected to introgress into the diploid species through backcrossing to the diploid parent. The diploid-like subgenome of the tetraploid species is homologous to extant diploid species and therefore expected to successfully pair and recombine with diploid chromosomes.

Here, we present a genetic analysis of a contact zone between diploid *E. rostkoviana* and tetraploid *E. arctica*. Both of these species have large corollas suggestive of higher rates outcrossing (Metherell and Rumsey 2018), and hybrids have been identified widely across Britain (Stace, Preston, and Pearman 2015), however putative F1 triploid hybrids are yet to be found in the wild. We look for any evidence of hybridisation and introgression between ploidy levels. First, we used Sanger sequencing of a locus that shows diagnostic differences between diploids and tetraploids to see if there is evidence of hybridisation. We then used Genotyping by Sequencing (GBS) to look for evidence of introgression. We also used demographic modelling to investigate the most likely scenario of historical gene flow to explain the observed genetic structure. The results enable us to discuss the processes governing reproductive isolation in a diploid-tetraploid contact zone. We predict that as reproductive barriers in *Euphrasia* are low, and as cross ploidy hybridisation is being increasingly found in other plant groups using genomic data, that there will be cross ploidy hybrids and evidence for introgression.

Materials and methods

Population sampling and DNA extraction

Population sampling took place in July 2017, at the managed hay meadow at Cae Trawscoed in the National Botanical Garden Wales (lat/long: 51.8447/-4.14531) where there is a mixed population of diploid *Euphrasia rostkoviana* and tetraploid *Euphrasia arctica*. These species can be easily separated by morphology based on the presence or absence of long-stalked flexuous glandular hairs. Putative hybrids show a low density of long glandular hairs and are intermediate for a range of other traits. Both species were abundant and distributed amongst each other, with *E. rostkoviana* present amongst taller vegetation, and *E. arctica* more dominant in shorter cropped vegetation. A total of 95 individuals were sampled from the mixed population, 45 being identified based on morphology as diploid and 50 being tetraploid. Plants were sampled evenly along transect of approximately 12m where plants were highly intermixed. Specimens were transferred into silica bags to dry until DNA extraction. DNA was extracted using the DNeasy Plant Mini kit (Qiagen), following manufacturer's protocols.

Sanger sequencing and sequence analysis

ITS sequencing was used as there is a known diagnostic difference between diploid and tetraploid species, while plastid sequencing used the most variable locus that is widely used in population genetic studies of *Euphrasia*. The sample size for the ITS sequencing was 70 individuals with eight removed due to low sequence quality, while for plastid sequencing eight individuals were used. PCRs were performed in 25 μL reactions; DNA amplification protocols and conditions for PCRs are given in Supplementary Tables 1 and 2. To check the quality of PCR products, the DNA was visualised on a 1% agarose gel. PCR products were then cleaned with Exo-SAP (Affometrix) using standard protocols and submitted to Edinburgh Genomics for sequencing reactions using BigDye Terminator Cycle Sequencing chemistry and Sanger Sequencing on an ABI 3730. ITS1 PCR products were sequenced in the forwards direction only (with the ITS4 primer), whilst rpL32-trnLUAG was sequenced in both directions.

ITS1 and *trnL* spacer chromatograms were aligned and edited in Geneious (version 9.0.5). Low quality bases were trimmed at the beginning and ends of sequences; six ITS1 sequenced were excluded due to poor sequence quality. The 558bp trnL spacer alignment of eight sequences showed no variable sites and was therefore not analysed further. The final ITS1 alignment included 62 individuals and was 658 bp in length with 58 variable sites in total. 26 diploids and 36 tetraploids were present in the alignment. An outgroup species *E. transmorrisonensis* (diploid; from Taiwan), was added from NCBI (GenBank accession number: AY165615) to polarise the ITS1 phylogeny for visualisation but was not used in downstream analyses. We constructed a Maximum Likelihood phylogeny using IQ-TREE (version 1.5.5; Nguyen et al. (2015)) using ModelFinder to find the most suitable substitution model (using the TESTNEWMERGE model flag) with 1000 ultra-fast bootstraps (Hoang et al. 2018). The resulting newick file was visualised in ggtree (version 2.1.2; Yu et al. (2017)). To further characterise sequence variation, this alignment was read into R version 3.6.1 (R Core Team 2019) using the function read.dna() from the package ape (version 5.4; Paradis and Schliep (2019)) and converted to a genind object with adegenet (version 2.1.2; Jombart (2008)). We tested how much genetic variation was partitioned between ploidy levels using Analysis of Molecular Variance (AMOVA) implemented in the poppr (version 2.8.4; Kamvar, Tabima, and Grunwald (2014)) function poppr.amova().

GBS & SNP discovery

GBS library preparation was performed following the protocol in (Elshire et al. 2011). We used the enzyme ApeKI to fragment the genome and create cut sizes for adapter ligation. Samples were then pooled and cleaned before PCR amplification and sequencing. A single well was used as a water control. The pooled library was submitted to the University of Oregon for 100bp single end sequencing on the Illumina HiSeq 4000, generating 17,397,350 reads. We then used the TASSEL 5 pipeline version 2 to discover SNPs using default settings, except the k-mer length was increased to 75 (Glaubitz et al. (2014); see https://github.com/Euphrasiologist/GBS_V2_Tassel_5). Master sequence tags (n = 1,001,272) were aligned to a *Euphrasia arctica* reference genome (Becher et al. 2020) using BWA with default settings (version 0.7.17; Li and Durbin (2009)). All scaffolds in the reference genome greater than 1000bp were merged into a single scaffold to reduce computational time. The VCF file was filtered for variants with >50% missing data and individuals with >75% missing data using vcftools (version 0.1.16; Danecek et al. (2011)).

We focused our analyses of genetic structure, hybridisation and introgression on regions that are homologous between diploids and tetraploids, and not in the other more divergent tetraploid subgenome where gene exchange with diploids is unlikely. As such, variants were excluded if they were not located on the 'conserved' set of scaffolds of presumed disomic inheritance that were identified in genome-wide sequence comparisons of diploid and tetraploid Euphrasia by Becher et al. (2020). The conserved set comprises 46 Mbp (n = 3454) of sequence and is likely to represent (a part of) the conserved subgenome. A minor allele frequency filter was applied to remove invariant sites using vcftools. One variant was kept per scaffold to ensure variants were not tightly linked, using PLINK (version 1.9; Purcell et al. (2007)). This left 92 individuals (42 diploids and 50 tetraploids) and 270 variable sites for analyses.

Identifying hybridisation between ploidy levels

To investigate hybridisation between the two species of *Euphrasia* present in the contact zone, we first conducted a Principal Component Analysis (PCA) of our genomic GBS dataset. First, the VCF was loaded into R using the package vcfR (version 1.10; Knaus and Grunwald (2017)) and the PCA was carried out using the adegenet (Jombart 2008) function dudi.pca() where missing values were substituted by the mean allele frequencies. The PCA was visualised using ggplot2 (version 3.2.1; Wickham (2016)). Second, we performed a Bayesian admixture analysis in the program STRUCTURE (version 2.3.4; Pritchard, Stephens, and Donnelly (2000)) using the same GBS dataset. The VCF was converted to a STRUCTURE file format using PGDSpider (version 2.1.1.5; Lischer and Excoffier (2012)). We set the K-value to be 2 as there were two divergent species of differing ploidies, and the run was set with a burn-in of 100,000 for 1,000,000 iterations on the 'admixture' option. In addition, 90% probability intervals were stored using the ANCESTPINT option. Samples where probability intervals overlapped zero or one were considered non-hybrids. The Q-matrix, and probability intervals, were extracted from the output and plotted using a custom R script (https://github.com/Euphrasiologist/StructuRe). Third, we explicitly attempted to identify hybrid individuals using the program NEWHYBRIDS (version 1.1; Anderson and Thompson (2002)), which classifies individuals into one of six potential categories (parent A, parent B, F1, F2, backcross (BC)1 to parent A, BC1 to parent B) based on their SNP genotypes. The model was run with a burn in of 100,000 iterations and a run length of 100,000 sweeps.

Quantification of genetic variability within and between ploidy levels

We computed several population genetic statistics on the GBS dataset to understand population structure in the contact zone, and used AMOVA (as above) to detect regions of the genome that may have introgressed.

Weir and Cockerham's estimator of F_{ST} was calculated for each SNP in the GBS data using vcftools (Danecek et al. 2011) and visualised in ggplot2 (Wickham 2016). The average F_{ST} across all SNPs was reported as the global F_{ST}. An AMOVA was run for both the ITS1 and GBS datasets using the function poppr.amova() from the R package poppr (Kamvar, Tabima, and Grunwald 2014), which was used to understand the partitioning of genetic variability both within and between ploidy levels. P-values were then derived from the output using the randtest() function from ade4 (version 1.7-15; Bougeard and Dray (2018)), randomly permuting sample matrices 9999 times.

Demographic inference with $\delta a \delta I$

We carried out demographic model fitting using the package $\delta a \delta I$ to see what the best fitting model of hybridisation was for the disomic GBS data (Gutenkunst et al. 2009). Missing data was handled by scaling down the size of the joint site-frequency spectra to 24 (haploid) genomes per species. We implemented four models, each involving one ancestral population splitting into two sub-populations corresponding to diploids and tetraploids, which could differ in size. Model parameters corresponding to the diploid and tetraploid sub-populations are denoted with the subscripts D and T. The models differ by the amount of gene flow allowed between the sub-populations: (1) constant gene flow with five parameters (two population sizes NeD and NeT, two migration rates MDT and MTD, and the time of the population split T0), (2) historic gene flow only with six parameters (two population sizes, two migration rates, the time when gene flow ceased T1, and the time difference between T1 and the time of the split denoted T0), (3) secondary contact as (2) but with gene flow in T0 and not in T1, and (4) no gene flow with three parameters (two population sizes and T0, see Figure 3A for schematics). We fixed F at 0.75 and 0.81 for *E. arctica* and *E. rostkoviana* according to the empirical estimates. To assess the uncertainty of the model fitting, we used 99 individual sub-samplings of our data set. Because we found the model fitting results to depend strongly on the initial conditions, we ran 99 replicates with randomly perturbed starting values for each model and down-sampled data set resulting in 39,204 optimisations. From each set of 99 replicates, we selected the one with the best log likelihood. In order to compare these nested models with different numbers of parameters, we computed the Akaike Information Criterion (AIC) of each fit, and we plotted the results with matplotlib (Hunter 2007).

Results

ITS patterns of genetic structure

The 658bp alignment of the ITS1 sequence data included 26 diploid individuals and 36 tetraploid individuals. The chromatograms revealed no evidence of sequence additivity or double peaks, which might have indicated hybrid individuals or retained duplicate copies in the polyploids. The maximum likelihood phylogenetic tree showed two distinct clades of *Euphrasia* that were highly supported (Figure 1A). All 58 SNPs in the alignment were fixed between the ploidy groups. Accordingly, an analysis of molecular variance (AMOVA) showed that 99.5% of the ITS1 variation in the samples was explained by ploidy ($p < 0.001$; Table 1A). Limited sampling of four individuals from each species for the trnL spacer showed no differences between the species.

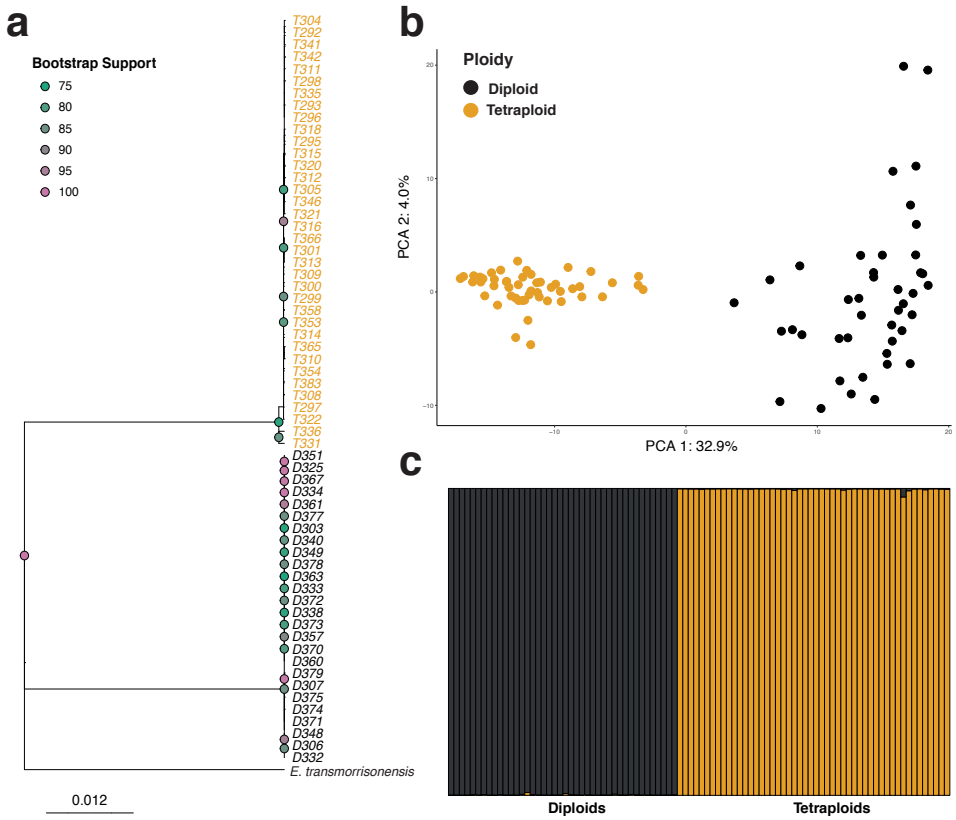

Figure 1: ITS1 marker and GBS analysis suggest strong barriers to gene flow between diploid (black) and tetraploid (orange) species of Euphrasia. A) Phylogenetic tree of the ITS1 marker generated from IQ-TREE; *E. transmorrisonensis* is used as an outgroup species. Bootstrap support is visualised on a colour scaled point at each node and only support values above 75 are shown. B) A Principal Component Analysis of 92 *Euphrasia* individuals across 270 loci located on different scaffolds of the reference *E. arctica genome*. The first two Principal Components are plotted against each other. C) Admixture analysis performed in STRUCTURE on the same 92 individuals, where K = 2.

Patterns of genetic structure from GBS data

Analysis of SNP data generated using GBS corroborated the findings from the ITS1 dataset. In total, 270 SNPs were analysed representing putatively disomically inherited scaffolds in the Euphrasia arctica reference genome shared between diploids and tetraploids (see Methods). Principal Component analysis (PCA) showed that the first principal component explained 32.9% of the genomic variation and clearly separated individuals by ploidy, with two separate clusters and no evidence of intermediate genotypes (Figure 1B). The same pattern was also present in the STRUCTURE analysis with a K value of 2, which reported Q values all above 0.98 (Figure 1C) that assigned individuals to clusters consistent with their morphological identification. The probability interval of each Q value overlapped either zero or one, indicating no hybrid genotypes. The NEWHYBRIDS analysis assigned each individual as 100% either parental species, with no evidence of F1 hybrids or backcrossed individuals. An AMOVA on the GBS data showed 78.4% of genomic variation was explained by ploidy ($p < 0.001$; Table 1B). The remainder of the genomic variation (21.6%) was due to

differences within ploidy level.

Table 1: Hierarchical analysis of molecular variance (AMOVA) for both ITS1 and GBS data sets. The total variation is partitioned between and within ploidy level of individuals in the analysis. Degrees of freedom (df), the variance of each of the observations (Sum sq), and percentage of variation explained by each level of variation (% Var) are reported. The significance of the components of variance are reported as p-values.

Data set	Variation	DF	Sum Sq	% Var	p-Monte Carlo
A) ITS1	Between ploidy	1	454.9	99.5	0.0001
	Within ploidy	60	4.9	0.5	
	Total	61	459.8	100	
B) GBS	Between ploidy	1	56	78.4	0.0001
	Within ploidy	90	30.3	21.6	
	Total	91	86.3	100	

Detecting hybridisation across the genome

We calculated population genetic parameters from GBS data to further investigate potential hybridisation between the two species differing in ploidy level. The global F_{ST} between species was high at 0.44, indicating that diploid *E. rostkoviana* and tetraploid *E. arctica* were highly differentiated across the genome. The distribution of F_{ST} however showed a bimodal distribution with high count of SNPs that were either mostly shared or private (Figure 3). Altogether, the AMOVA, PCA, STRUCTURE, and NEWHYBRIDS results showed a clear signal of strong differentiation between the two species.

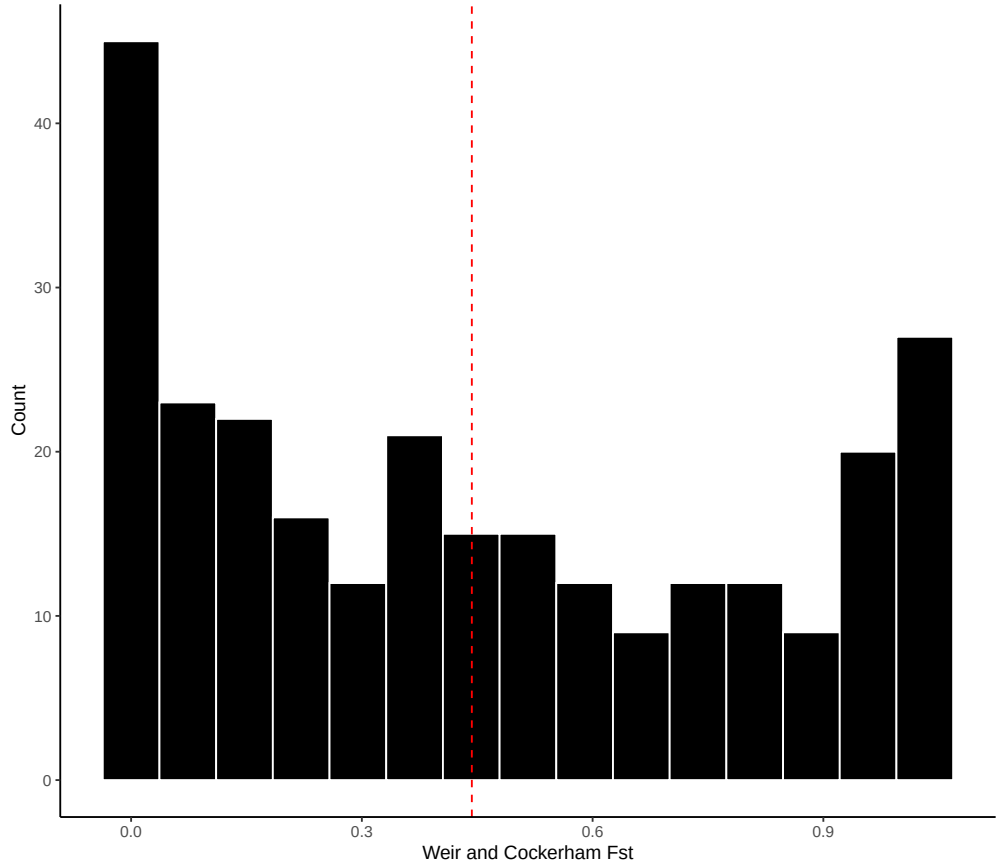

Figure 2: The distribution of F_{ST} for each of the 270 SNPs in the GBS dataset across all 92 individuals between the two ploidy levels. The red dashed vertical line indicates the mean global F_{ST} across all SNPs (0.44).

Model selection suggests low levels of gene flow

Demographic model fitting with $\delta a \delta I$ resulted in the highest overall support for models with constant gene flow and with secondary contact (Figure 3B). These models consistently scored low AIC values. The difference in AIC between the best model (constant gene flow, median AIC 425.77) and the runner-up (secondary contact, median AIC 425.89) was not significant (t-test t=-1.38, p=0.17). The other two models had significantly higher median AICs than the best model (historic gene flow: 462.83, t=-7.29, p<0.001; no gene flow: 452.05, t=-52.31, p<0.001). However, for some realisations of the data re-sampling, the model with no gene flow scored the best AIC coinciding with generally low estimates of the age of the population divergence time (T0 tends to be lower without gene flow than in alternative models, see Supplementary Figure 1).

The levels of gene flow fitted to our models tend to be low with M, the number of migrants per generation, of the order of 1 or less. The mean values over all replicates for the constant gene flow model were 0.3 (diploid->tetraploid, sd=0.09) and 0.4 (tetraploid->diploid, sd=0.13). For the model with secondary contact the migration rates were fitted to be 1.3 (diploid->tetraploid, sd=1.02) and 0.3 (tetraploid->diploid, sd=0.09). In all models, the effective population size tends to be slightly higher in the diploids than in the tetraploids

(Supplementary Figure 1, left-hand panels). While our data suggest the presence of gene flow, the similar AIC values we obtained under different models show how different demographic scenarios may produce similar genetic patterns.

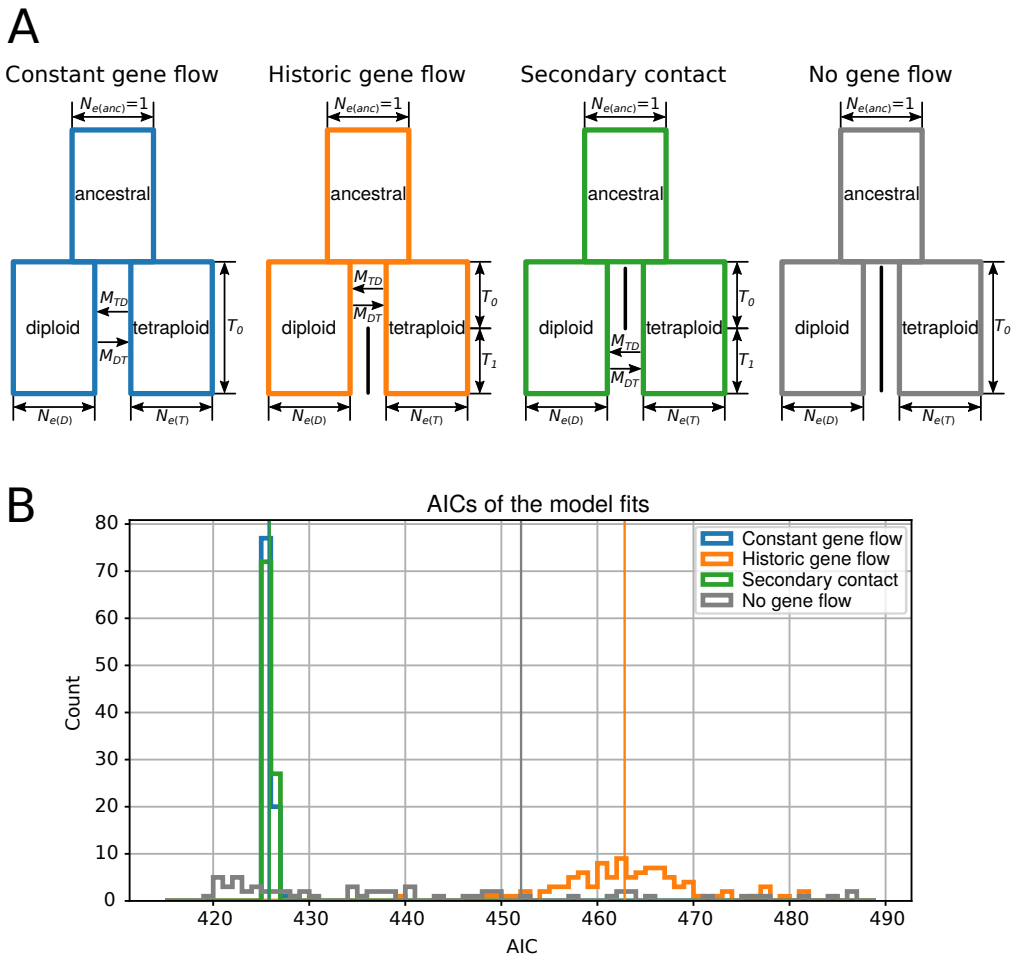

Figure 3: A) Schematics of the demographic models fitted indicating their parameters. Ne, the effective population size, is scaled to be 1 in the ancestral population. N_{eD} and N_{eT} indicate the ratios of the present-day effective population sizes of the diploid and tetraploid populations respectively, relative to the ancestral one. T is the number of generations to coalescence, which is subdivided in two epochs in the models with ancestral gene flow and secondary contact. M is the number of migrants per generation with subscripts indicating the direction of gene flow. B) The distributions of AIC values for model fits to 99 sub-sampled data sets. The models with constant gene flow and secondary contact are not distinguishable. Vertical lines indicate the distribution medians.

Discussion:

In this study, we investigated hybridisation between diploid *E. rostkoviana* and tetraploid *E. arctica* in a contact zone in south Wales. Neither ITS1, which is known to have diagnostic differences between species, nor the GBS dataset, provide evidence for recent hybridisation or introgression between these two species

of *Euphrasia*. This contrasts with the extensive hybridisation known to occur between *Euphrasia* species of the same ploidy level (Metherell and Rumsey 2018; Stace, Preston, and Pearman 2015). For example, using GBS data, Zlonis and Gross (2018) showed that there is extensive gene flow from native to invasive tetraploid species of *Euphrasia* in North America. Although we expected to find cross ploidy hybrids, to date there has only been a single wild triploid hybrid *Euphrasia* individual that has been found (Yeo 1956), and previous attempts to synthesise cross ploidy hybrids artificially have been unsuccessful (Yeo 1968). This result is consistent with ploidy representing a strong reproductive barrier between species (Husband and Sabara 2004). While we are confident that recent hybridisation is rare (or absent), based on the small sample of individuals we cannot exclude the possibility that early generation hybrids may have been overlooked if they are rare - our data are consistent with anything up to 4% of plants being first generation hybrids.

Although conventional population genetic analyses show little evidence of recent gene flow between the two Euphrasia species in this study, demographic modelling using $\delta a \delta I$, and inspection of the distribution of F_{ST} values across all SNPs indicates that there may be limited ongoing gene flow. $\delta a \delta I$ gives low estimates for bidirectional gene flow, which is known to occur in other cross ploidy hybrid systems (Bleeker, 2003). In particular, the higher estimate for tetraploid to diploid gene flow in the constant gene flow model may give support to the hypothesis that Yeo outlined more than 60 years ago (Yeo 1956), and corroborates a recent study on British Euphrasia which suggested possible diploid-tetraploid gene flow (Becher et al., 2020). Although the signature of gene flow is low this may be because it is limited to specific regions of the genome. Here, selection may be operating differentially in the diploid and tetraploid species, which may explain the bimodal F_{ST} distribution seen in our dataset (Whitlock, 2008).

The results from this study are consistent with some cross-ploidy hybrid systems where hybridisation is very rare, or where hybrids are strongly selected against. For example, in diploid *Centaurea pseudophrygia* and tetraploid *C. jacea*, only targeted sampling was able to reveal cross ploidy hybrids which were otherwise not found by random sampling in 12 different contact zones (Koutecky et al. 2011). In a contact zone of diploid *Senecio madagascariensis* and tetraploid *S. pinnatifolius* no hybrids have been detected in the field, however genetic analysis of the seeds revealed hybrid genotypes (Prentis et al. 2007). It is possible in the *Euphrasia* contact zone that hybrid seed is being formed but either does not germinate, or hybrid seedlings do not survive to maturity. Our results stand in strong contrast to those found in *Dactylorhiza*, where triploid F1 cross ploidy hybrids, and backcrossed individuals are found frequently (De Hert et al. 2012; Balao et al. 2017). The mechanisms underlying these differences in the frequency of cross ploidy hybrids in divergent species are yet to be fully established, and are likely an area for fruitful research.

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.

Anderson, E. C., and E. A. Thompson. 2002. "A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data." Journal Article. *Genetics* 160 (3): 1217–29. <Go to ISI>://WOS:000174790600033.

Baduel, P., S. Bray, M. Vallejo-Marin, F. Kolar, and L. Yant. 2018. "The "Polyploid Hop": Shifting Challenges and Opportunities over the Evolutionary Lifespan of Genome Duplications." Journal Article. *Frontiers in Ecology and Evolution* 6: 19. https://doi.org/10.3389/fevo.2018.00117.

Balao, F., M. Tannhauser, M. T. Lorenzo, M. Hedren, and O. Paun. 2017. "Genetic Differentiation and Admixture Between Sibling Allopolyploids in the Dactylorhiza Majalis Complex (Vol 116, Pg 351, 2016)." Journal Article. *Heredity* 118 (2): 210–10. https://doi.org/10.1038/hdy.2016.119.

Becher, H., M. Brown, G. Powell, C. Metherell, N.J. Riddiford, and A.D. Twyford. 2020. "Maintenance of Species Differences in Closely Related Tetraploid Parasitic Euphrasia (Orobanchaceae) on an Isolated Island." Journal Article. *Biorxiv*.

Bougeard, S., and S. Dray. 2018. “Supervised Multiblock Analysis in R with the Ade4 Package.” Journal Article. *Journal of Statistical Software* 86 (1): 1–17. https://doi.org/10.18637/jss.v086.i01.

Brown, M.R., N. Frachon, E.L.Y. Wong, C. Metherell, and A.D. Twyford. 2020. “Life History Evolution, Species Differences, and Phenotypic Plasticity in Hemiparasitic Eyebrights (Euphrasia).” Journal Article. *American Journal of Botany* 107 (3): 1–10.

Buggs, R. J. A., and J. R. Pannell. 2007. “Ecological Differentiation and Diploid Superiority Across a Moving Ploidy Contact Zone.” Journal Article. *Evolution* 61 (1): 125–40. https://doi.org/10.1111/j.1558-5646.2007.00010.x.

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.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. “The Variant Call Format and Vcftools.” Journal Article. *Bioinformatics* 27 (15): 2156–8. https://doi.org/10.1093/bioinformatics/btr330.

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.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. “A Robust, Simple Genotyping-by-Sequencing (Gbs) Approach for High Diversity Species.” Journal Article. *Plos One* 6 (5). https://doi.org/10.1371/journal.pone.0019379.

Fearn, G. M. 1977. “MORPHOLOGICAL and Cytological Investigation of Cochlearia Populations on Gower Peninsula, Glamorgan.” Journal Article. *New Phytologist* 79 (2): 455–58. https://doi.org/10.1111/j.1469-8137.1977.tb02226.x.

French, G. C., R. A. Ennos, A. J. Silverside, and P. M. Hollingsworth. 2005. “The Relationship Between Flower Size, Inbreeding Coefficient and Inferred Selfing Rate in British Euphrasia Species.” Journal Article. *Heredity* 94 (1): 44–51. https://doi.org/10.1038/sj.hdy.6800553.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire, Q. Sun, and E. S. Buckler. 2014. “TASSEL-Gbs: A High Capacity Genotyping by Sequencing Analysis Pipeline.” Journal Article. *Plos One* 9 (2): 11. https://doi.org/10.1371/journal.pone.0090346.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. “Inferring the Joint Demographic History of Multiple Populations from Multidimensional Snp Frequency Data.” Journal Article. *Plos Genetics* 5 (10): 11. https://doi.org/10.1371/journal.pgen.1000695.

Hegarty, M. J., and S. J. Hiscock. 2005. “Hybrid Speciation in Plants: New Insights from Molecular Studies.” Journal Article. *New Phytologist* 165 (2): 411–23. https://doi.org/10.1111/j.1469-8137.2004.01253.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.

Hunter, J. D. 2007. “Matplotlib: A 2D Graphics Environment.” Journal Article. *Computing in Science & Engineering* 9 (3): 90–95. https://doi.org/10.1109/mcse.2007.55.

Husband, B. C., and H. A. Sabara. 2004. “Reproductive Isolation Between Autotetraploids and Their Diploid Progenitors in Fireweed, Chamerion Angustifolium (Onagraceae).” Journal Article. *New Phytologist* 161 (3): 703–13. https://doi.org/10.1046/j.1469-8137.2003.00998.x.

Johnston, S. A., T. P. M. Dennijs, S. J. Peloquin, and R. E. Hanneman. 1980. "THE Significance of Genic Balance to Endosperm Development in Interspecific Crosses." Journal Article. *Theoretical and Applied Genetics* 57 (1): 5–9. https://doi.org/10.1007/bf00276002.

Jombart, T. 2008. "Adegenet: A R Package for the Multivariate Analysis of Genetic Markers." Journal Article. *Bioinformatics* 24 (11): 1403–5. https://doi.org/10.1093/bioinformatics/btn129.

Kamvar, Z. N., J. F. Tabima, and N. J. Grunwald. 2014. "Poppr: An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction." Journal Article. *Peerj* 2: 14. https://doi.org/10.7717/peerj.281.

Knaus, B. J., and N. J. Grunwald. 2017. "VCFR: A Package to Manipulate and Visualize Variant Call Format Data in R." Journal Article. *Molecular Ecology Resources* 17 (1): 44–53. https://doi.org/10.1111/1755-0998.12549.

Koutecky, P., T. Badurova, M. Stech, J. Kosnar, and J. Karasek. 2011. "Hybridization Between Diploid Centaurea Pseudophrygia and Tetraploid c. Jacea (Asteraceae): The Role of Mixed Pollination, Unreduced Gametes, and Mentor Effects." Journal Article. *Biological Journal of the Linnean Society* 104 (1): 93–106. https://doi.org/10.1111/j.10958312.2011.01707.x.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." Journal Article. *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Lischer, H. E. L., and L. Excoffier. 2012. "PGDSpider: An Automated Data Conversion Tool for Connecting Population Genetics and Genomics Programs." Journal Article. *Bioinformatics* 28 (2): 298–99. https://doi.org/10.1093/bioinformatics/btr642.

Metherell, C., and F. Rumsey. 2018. *Eyebrights (Euphrasia) of the Uk and Ireland*. Book. Botanical Society of Britain; Ireland.

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.

Paradis, E., and K. Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." Journal Article. *Bioinformatics* 35 (3): 526–28. https://doi.org/10.1093/bioinformatics/bty633.

Pinheiro, F., F. de Barros, C. Palma-Silva, D. Meyer, M. F. Fay, R. M. Suzuki, C. Lexer, and S. Cozzolino. 2010. "Hybridization and Introgression Across Different Ploidy Levels in the Neotropical Orchids Epidendrum Fulgens and E-Puniceoluteum (Orchidaceae)." Journal Article. *Molecular Ecology* 19 (18): 3981–94. https://doi.org/10.1111/j.1365-294X.2010.04780.x.

Prentis, P. J., E. M. White, I. J. Radford, A. J. Lowe, and A. R. Clarke. 2007. "Can Hybridization Cause Local Extinction: A Case for Demographic Swamping of the Australian Native Senecio Pinnatifolius by the Invasive Senecio Madagascariensis?" Journal Article. *New Phytologist* 176 (4): 902–12. https://doi.org/10.1111/j.1469-8137.2007.02217.x.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." Journal Article. *Genetics* 155 (2): 945–59. <Go to ISI>://WOS:000087475100039.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." Journal Article. *American Journal of Human Genetics* 81 (3): 559–75. https://doi.org/10.1086/519795.

Ramsey, J., and D. W. Schemske. 1998. "Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants." Journal Article. *Annual Review of Ecology and Systematics* 29: 467–501. https://doi.org/10.1146/annurev.ecolsys.29.1.467.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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.

Shimizu-Inatsugi, R., J. Lihova, H. Iwanaga, H. Kudoh, K. Marhold, O. Savolainen, K. Watanabe, V. V. Yakubov, and K. K. Shimizu. 2009. "The Allopolyploid Arabidopsis Kamchatica Originated from Multiple Individuals of Arabidopsis Lyrata and Arabidopsis Halleri." Journal Article. *Molecular Ecology* 18 (19): 4024–48. https://doi.org/10.1111/j.1365-294X.2009.04329.x.

Stace, C A., C D. Preston, and D Pearman. 2015. *Hybrid Flora of the British Isles*. Book. Bristol::Botanical Society of Britain & Ireland.

Stebbins, GL. 1971. *Chromosomal Evolution in Higher Plants*. Book. London, UK: Edward Arnold.

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.

Vallejo-Marin, M. 2012. "Mimulus Peregrinus (Phrymaceae): A New British Allopolyploid Species." Journal Article. *Phytokeys* 14: 1–14. https://doi.org/10.3897/phytokeys.14.3305.

Wang, X. M., G. Gussarova, M. Ruhsam, N. de Vere, C. Metherell, P. M. Hollingsworth, and A. D. Twyford. 2018. "DNA Barcoding a Taxonomically Complex Hemiparasitic Genus Reveals Deep Divergence Between Ploidy Levels but Lack of Species-Level Resolution." Journal Article. *Aob Plants* 10 (3): 13. https://doi.org/10.1093/aobpla/ply026.

Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Yeo, P F. 1956. "Hybridisation Between Diploid and Tetraploid Species of Euphrasia." Journal Article. *Watsonia* 3: 253–69.

Yeo, PF. 1968. "The Evolutionary Significance of the Speciation of Euphrasia in Europe ." Journal Article. *Evolution* 22 (4): 736–47.

Yu, G. C., D. K. Smith, H. C. Zhu, Y. Guan, and T. T. Y. Lam. 2017. "GGTREE: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data." Journal Article. *Methods in Ecology and Evolution* 8 (1): 28–36. https://doi.org/10.1111/2041-210x.12628.

Zlonis, K. J., and B. L. Gross. 2018. "Genetic Structure, Diversity, and Hybridization in Populations of the Rare Arctic Relict Euphrasia Hudsoniana (Orobanchaceae) and Its Invasive Congener Euphrasia Stricta." Journal Article. *Conservation Genetics* 19 (1): 43–55. https://doi.org/10.1007/s10592-017-0995-x.