THE UNIVERSITY OF READING
School of Biological Sciences

Co-occurring Mediterranean orchids: insights from morphometrics, genetic diversity and introgression in two species: *Ophrys fusca* and *Ophrys dyris*

A thesis submitted by
JOANA ISABEL ALBERTO ABREU
for the degree of
Doctor of Philosophy

SUPERVISOR: Professor Julie A. Hawkins

CO-SUPERVISORS: Dr. Michael F. Fay, Dr. Helena C. Cotrim

29 October 2021

ABSTRACT

Ophrys is amongst the best known orchid genera and is an established system for the study of pollinator mediated floral evolution. Two species, *Ophrys fusca* s.l. and *Ophrys dyris* (= *O. omegaifera* subsp. *dyris*), belonging to *Ophrys* section Pseudophrys are the focus of this study. An integrative study of morphological and genetic diversity of *O. fusca* and *O. dyris* was made, based on data collected from six populations in Central Portugal in five years. The populations were characterized a priori as "mixed" or pure". The data included morpho-anatomical traits (272 plants) and phenological records and pollination success assessments (887 flowers from 260 plants). Additionally, sampling for DNA (165 plants) and leaf and pollinarium collection for cytological study (67 plants) was carried out.

Multivariate analysis of the whole morphological dataset recovered two groups. When including only the plants common both to morphological and genetic datasets and a reduced number of diagnostic characters, three groups were found. Traits that better discriminate morphological groups are qualitative ones, namely the pilosity of the different lobes of the labellum. Univariate analysis of morpho-anatomical traits showed that vigour characters and depth of labellum indentation were the most variable, including between populations. Labellum indentation also showed greatest differences between the three types of populations (*O. fusca*/mixed/*O. dyris*). The study of flowering behaviour revealed incomplete floral isolation between the two taxa. Discrepancy between morphotypes in the average time for the flowering period in different years might threaten the synchrony with the supposed specialist pollinator, required for an effective pollination by sexual deception. This type of disruption being confirmed, it could represent an advantage for the *fusca* plants regarding the availability of pollinators. Percentages of pollination success obtained range from 2% to 20%. No significant differences in pollination success were found between plants confirmed as parental species or hybrids based on genetic data. Flow cytometry methods revealed constancy of nuclear DNA content (1C = 11.19 pg) in all the specimens analysed, including species and putative hybrids. Constancy of cytotypes was also confirmed, based on chromosome counts from the roots of two specimens, one of each species (2n = 4x = 72, 74). 72 for *fusca*, 74 for *dyris*. Following microsatellite data analysis, all the populations considered included individuals with mixed genotypes, representing a total of 44.8% of the plants. Genetic hybrids were found in all but one population, refuting the initial assumption of "pure" (one taxon) versus "mixed" populations. When comparing genetic and morphological datasets, despite a positive correlation, strong discrepancies have been found between the composition of the morphological groups and the genetic clusters. Implications are considered, in terms of predicting the persistence of putative hybrids and the conservation of evolutionary processes.

TABLE OF CONTENTS

ABSTRACT 1

TABLE OF FIGURES 4

TABLE OF TABLES 7

ACKNOWLEDGEMENTS 9

1. INTRODUCTION 10

1.1. HYBRIDISATION AND RELATED CONCEPTS 10
1.2. REPRODUCTIVE BARRIERS 12
1.3. OUTCOMES OF HYBRIDISATION 13
1.4. STUDYING HYBRIDISATION 15
1.4.1 Morphological studies 15
1.4.2. Cytogenetic studies 16
1.4.3. Genetic studies 17
1.4.4. Case studies of plant hybrids 18
1.5. CONSERVATION 21
1.6. THE STUDY GROUP AND STUDY SITES 22
1.6.1. Ophrys dyris Maire and Ophrys fusca Link 24
1.7. THESIS OUTLINE AND MAIN AIMS 28

2. MORPHOLOGY, PHENOLOGY AND REPRODUCTIVE SUCCESS 30

2.1 INTRODUCTION 30
2.1.1 Studying hybridisation 30
2.1.2. Orchids: spotlighting floral morphology 30
2.1.3. Phenological barriers to hybridisation 31
2.1.5. Pollination success as a proxy of reproductive success 32
2.1.6. Aims and scope of the chapter 33
2.2. MATERIALS AND METHODS 34
2.2.1. Plant sampling 34
2.2.2. Sampling strategy 36
2.2.3. Data Analysis 42
2.3. RESULTS AND DISCUSSION 45
2.3.1 Morphometry 45
2.3.2. Phenology 61
2.3.2. Pollination success 68
2.4. CONCLUSIONS 70
2.4.1. Morphology 70
2.4.2 Phenology 71
2.4.3 Reproductive success 72

3. CYTOGENETICS 73

3.1. INTRODUCTION 73
3.2. METHODS 76
3.2.1. Flow cytometry 76
3.2.2. Chromosome counts 77
3.3. RESULTS AND DISCUSSION 78
3.3.1. The use of alternative tissues to estimate GS and infer ploidies in Ophrys 78
3.3.2. Tetraploid cytotypes in Ophrys fusca and O. dyris across central Portugal 79
3.3.2. The origin of Ophrys dyris 81

4. GENETIC DIVERSITY AMONG THE STUDIED PLANTS – MICROSATELLITES 82

4.1 INTRODUCTION 82
4.1.1 Aims and scope of the chapter 83
4.2 METHODS 84

4.2.1 Plant Material *84*
4.2.2. DNA isolation and molecular analysis *85*
4.2.3. Data Analysis *88*
4.3 RESULTS AND DISCUSSION 95
4.3.1. Summary diversity statistics *95*
4.3.2. Allelic patterns, Inbreeding, Selfing rate and Linkage Disequilibrium *96*
4.3.3. Population differentiation *97*
4.3.4. Gene flow and hybridization *101*
4.3.5. Multivariate statistics *102*
4.3.6 Mantel Test *103*
4.3.7. Bringing together seed set and the genetic data analysis *104*
4.4. CONCLUSIONS 104

5. GLOBAL DISCUSSION AND CONCLUSION 106

5.1. The understanding of pollinator-mediated floral evolution *107*
5.2. Are these separate entities? Is hybridization occurring between them? *108*
5.3. Considerations for Taxonomy *109*
5.4. Absence of cytological isolation *110*
5.5. Is Ophrys dyris a hybrid? *111*
5.6. Evolutionary Processes and Conservation *111*
5.7. Future work *113*
5.8. Reflections on the work done *114*
Lab work and data exploration *115*

REFERENCE LIST 117

TABLE OF FIGURES

FIGURE 1. DISTRIBUTION OF THE STUDIED TAXA IN EUROPE, ADAPTED FROM PEDERSEN & FAURHOLDT (2007). A. DISTRIBUTION OF *OPHRYS FUSCA*. B. DISTRIBUTION OF *OPHRYS DYRIS*. 24
FIGURE 2. FLORAL STRUCTURE OF *OPHRYS FUSCA* (A) AND *O. DYRIS* (B) 26
FIGURE 3. GEOGRAPHIC REGION OF THE POPULATIONS STUDIED, EACH NAMED BY THE ACRONYM USED: AV – ARRUDA DOS VINHOS; CF – CASALINHO DO FACHO; ME – MENDIGA; PI – PINHEIRINHOS; SA – SERRA DE SANTO ANTÓNIO; MJ – MONTEJUNTO. 27
FIGURE 4. GEOGRAPHIC LOCATION OF POPULATIONS STUDIED, NAMED BY THE ACRONYM USED. EXPECTED PRESENCE OF THE STUDIED TAXA, ACCORDING TO PREVIOUS FIELD OBSERVATIONS (F – *OPHRYS FUSCA*; D – *O. DYRIS*; I – INTERMEDIATE MORPHOTYPES): SA – D; ME – D+I+F; MJ – D; AV – F; CF – F; PI – F+D+I 34
FIGURE 5 SOME OF THE SCREENS OF THE DIGITAL DATABASE USED TO REGISTER INFORMATION IN THE FIELD. A. FIRST INPUT SCREEN FROM THE MENU FLOWERS. B. MAIN MENU, INCLUDING THE SUBMENUS SITES, PLANTS, SEASONS, FLOWERS, OBSERVATIONS, GPS DEVICE AND BUGS. 39
FIGURE 6. FLORAL STRUCTURE OF *OPHRYS FUSCA* AND *O. DYRIS*, ILLUSTRATING SOME OF THE MACROMORPHOLOGICAL CHARACTERS USED FOR MORPHOLOGICAL ANALYSIS. 5 – MAXIMUM LENGTH OF LATERAL SEPALS; 6 – MAXIMUM WIDTH OF LATERAL SEPALS; 9 – DEPTH OF INDENTATION; 10 - CONCAVITY OF THE MID LOBE; 11 – MAXIMUM LABELLUM LENGTH; 12 – MAXIMUM LABELLUM LENGTH WIDTH; 13, 14 – PILOSITY OF THE CENTRAL LOBE MEASURED 2MM INSIDE THE BOTTOM MARGIN (13) IN A VERTICAL LINE OR (14) AT FOURTY-FIVE DEGREES TO A VERTICAL LINE; 15, 16 - PILOSITY OF THE LATERAL LOBE MEASURED 1MM INSIDE THE BOTTOM MARGIN (15) IN A VERTICAL LINE OR (16) AT FOURTY-FIVE DEGREES TO A VERTICAL LINE; 18 – SPECULUM MAXIMUM LENGTH; 19 – SPECULUM MAXIMUM WIDTH; 20-W-SHAPE DELIMITATION; 21, 22 – LENGTH AND WIDTH OF THE GYMNOSTEMIUM; 23 – LENGTH OF THE STIGMATIC SURFACE; 24 – WIDTH OF THE STIGMATIC SURFACE. 41
FIGURE 7. PRINCIPAL COORDINATES PLOTS OF THE TWO FIRST AXIS, FOLLOWING PRINCIPAL COORDINATES ANALYSIS IN MSVP SOFTWARE USING GOWER GENERAL SIMILARITY COEFFICIENT. EIGEN VALUES OF 4.722 (AXIS 1) AND 1.989 (AXIS 2); CUMULATIVE PERCENTAGES OF 26% AND 38% FOR AXIS 1 AND 2. PCOA PERFORMED IN 62 PLANTS AND 26 CHARACTERS. .45
FIGURE 8. DISCRIMINANT FUNCTION ANALYSIS PERFORMED IN XLSTAT, RUN ON THE TWO GROUPS IDENTIFIED AFTER THE PCOA. 62 PLANTS, 26 TRAITS, NUMBERED BY DESCENDING ORDER OF ACCOUNTABILITY FOR MORPHOLOGICAL GROUPS DISCRIMINATION: 1 – PILOSITY OF THE MID LOBE, 45 DEGREES; 2 – PILOSITY OF THE MID LOBE, VERTICAL LINE; 3 - PILOSITY OF THE LATERAL LOBES, 45 DEGREES; 4 - PILOSITY OF THE LATERAL LOBES, VERTICAL LINE; 5 - SPECULUM W-SHAPE DELIMITATION; 6 - DOWNCURVING OF THE MID LOBE; 7 – DEPTH OF INDENTATION; 8 - CURVATURE OF LABELLUM; 9 - EXTENT OF PILOSITY IN THE BASE OF LABELLUM; 10 - PRESENCE OF BASAL GROOVE; 11 – LENGTH OF STIGMATIC SURFACE; 12 - SPECULUM BRIGHTNESS; 13 - LIP WIDTH; 14 – LIP LENGTH; 15 – STEM LENGTH; 16 – SPECULUM MARBLED; 17 - WIDTH OF GYMNOSTEMIUM; 18 – NUMBER OF FLOWERS; 19 - INFLORESCENCE LENGTH; 20 - SPECULUM MAXIMUM LENGTH; 21 - LENGTH OF LATERAL SEPALS; 22 - MAXIMUM WIDTH OF LATERAL SEPALS; 23 – WIDTH OF STIGMATIC CAVITY; 24 - SPECULUM MAXIMUM WIDTH; 25 - LENGTH OF GYMNOSTEMIUM; 26 - PRESENCE OF TRICHOMES IN THE STIGMATIC CAVITY. 46
FIGURE 9. (A). RESULTS OF THE PRINCIPAL COMPONENTS ANALYSIS OF 82 PLANTS CLASSIFIED BY MORPHOTYPE (1- DYRIS; 2 – INTERMEDIATE; 3 . FUSCA), FOR THE DATASET OF 26 CHARACTERS. CENTRAL (MEAN PER DIMENSION) POINTS SHOWN. (B) SCREE PLOT WITH THE PERCENTAGE OF EXPLAINED VARIANCE BY EACH OF THE DIMENSIONS, THE TWO FIRST DIMENSIONS ACCOUNTING FOR APPROXIMATELY 52 % OF THE VARIATION. (C) AND (D). HISTOGRAMS OF PERCENTAGE CONTRIBUTIONS FOR THE FIRST (C) AND SECOND DIMENSION (D) OF THE PCA. THE VARIABLES NUMBERING FOLLOWS TABLE 4.....47
FIGURE 10. PRINCIPAL COORDINATES ANALYSIS USING GOWER GENERAL SIMILARITY COEFFICIENT FOR THE SIMILARITY MATRIX, CONSIDERING THE PLANTS OF THE GENETIC DATASET (49 PLANTS) AND DIAGNOSTIC CHARACTERS (15), PERFORMED IN MSVP SOFTWARE. EIGEN VALUES OF 6.346 (AXIS 1) AND 2.646 (AXIS 2); CUMULATIVE PERCENTAGES OF 36.4% AND 51.5%, AXIS 1 AND 2, RESPECTIVELY. 51
FIGURE 11. DISCRIMINANT FUNCTION ANALYSIS PERFORMED IN XLSTAT, RUN ON THE THREE GROUPS IDENTIFIED IN THE PCOA WITH PLANTS FROM BOTH THE MORPHOLOGICAL AND THE GENETIC DATASET (49 PLANTS, 15 CHARACTERS). A - PILOSITY OF THE MID LOBE 45°; B - PILOSITY OF THE STIGMATIC CAVITY; C - DOWNCURVING OF THE MID LOBE; D - CURVATURE OF THE LABELLUM; E - SPECULUM MARBLED; F- INFLORESCENCE LENGTH; G - PILOSITY OF THE LATERAL LOBE, VERTICAL LINE; H - PILOSITY OF THE LATERAL LOBE, 45 °; I - PILOSITY OF THE MID LOBE, VERTICAL LINE; J - DEPTH OF INDENTATION; K - SPECULUM W-SHAPE DELIMITATION; L - NUMBER OF FLOWERS; M - STEM LENGTH; N - LIP WIDTH; O - LIP LENGTH. TRAITS A, B AND C ARE THOSE WHO ACCOUNT MOST FOR MORPHOLOGICAL GROUPS DISCRIMINATION.....52
FIGURE 12. COMPARISON OF THE POPULATIONS STUDIED REGARDING THE QUANTITATIVE MORPHOLOGICAL TRAITS ASSESSED: 1 – STEM LENGTH; 2 – INFLORESCENCE HEIGHT; 3 – NUMBER OF FLOWERS; 4 – LATERAL SEPALS (WIDTH/LENGTH); 5– DEPTH OF INDENTATION (MM); 6 – LIP WIDTH/LENGTH; 7 – SPECULUM LENGTH/WIDTH; 8 - GYMNOSTEMIUM LENGTH (MM); 9 – GYMNOSTEMIUM WIDTH (MM); 10 - STIGMATIC SURFACE LENGTH/WIDTH. CENTRAL "X" ARE MEAN VALUES. THE BOXES ARE TERMINATED BY THE TWENY-FIFTH PERCENTILE BELOW AND SEVENTY-FIFTH PERCENTILE ABOVE. FOR POPULATION CODES AND TAXA PRESENT IN EACH POPULATION, PLEASE SEE TABLE 1. 54

FIGURE 13. COMPARISON OF THE THREE DIFFERENT MORPHOTYPES ATTRIBUTED TO THE PLANTS STUDIED, *DYRIS* (1), HYBRID (2) AND *FUSCA* (3) REGARDING THE QUANTITATIVE MORPHOLOGICAL TRAITS ASSESSED: 1 – STEM LENGTH; 2 – INFLORESCENCE LENGTH; 3 – NUMBER OF FLOWERS; 4 – LATERAL SEPALS (WIDTH/LENGTH); 5– DEPTH OF INDENTATION (MM); 6 – LIP WIDTH/LENGTH; 7 – SPECULUM LENGTH/WIDTH; 8 - GYMNOSTEMIUM LENGTH (MM); 9 – GYMNOSTEMIUM WIDTH (MM); 10 - STIGMATIC SURFACE LENGTH/WIDTH. CENTRAL "X" ARE MEAN VALUES. THE BOXES ARE TERMINATED BY THE TWENTY-FIFTH PERCENTILE BELOW AND SEVENTY-FIFTH PERCENTILE ABOVE. 55
FIGURE 14. EXPRESSION OF THE QUALITATIVE CHARACTERS ASSESSED PART 1, PER POPULATION (LEFT SIDE) AND MORPHOTYPE (RIGHT SIDE), PART 1: - LONGITUDINAL CURVATURE OF LABELLUM; CONCAVITY OF THE MID LOBE; SPECULUM MARBLED; W-SHAPE DELIMITATION; FOR POPULATION CODES AND TAXA PRESENT IN EACH POPULATION, PLEASE SEE TABLE 4 57
FIGURE 15. EXPRESSION OF THE QUALITATIVE CHARACTERS ASSESSED, PART 2, PER POPULATION (LEFT SIDE) AND MORPHOTYPE (RIGHT SIDE), PART 2: PILOSITY OF THE MID LOBE, VERTICAL LINE; PILOSITY OF THE MID LOBE, 45º; PILOSITY OF THE LATERAL LOBES, VERTICAL LINE; PILOSITY OF THE LATERAL LOBES, 45º; EXTENT OF THE PILOSITY AT THE BASE OF THE LABELLUM; FOR POPULATION CODES AND TAXA PRESENT IN EACH POPULATION, PLEASE SEE TABLE 4. 58
FIGURE 16. TIME (JULIAN DAY) DISTRIBUTION FOR THE OBSERVATIONS OF ALL FLOWERING STAGES COMBINED – RECENTLY FLOWERING (RFL), FULLY OPEN (FOP), FLOWER SENESCENT (FLS) - FOR THE THREE MORPHOTYPES, IN THE POPULATIONS SAMPLED FOR PHENOLOGY IN 2011: MJ, PI, ME AND CF. OBSERVATIONS OF FLOWERS AT THE STAGE OF BUD (FLB) OR CAPSULE (CPS) HAVE BEEN EXCLUDED. 61
FIGURE 17. PROPORTIONS OF PLANTS WITH FULLY OPEN FLOWERS (FOP) PER WEEK, IN 2011 – CF (*O. FUSCA*), ME (MIXED), MJ (*O. FUSCA*) AND PI (MIXED). 62
FIGURE 18. JULIAN WEEKS WITH OVERLAPPING OF DIFFERENT MORPHOTYPES FOR FULLY OPEN FLOWERS (FOP PHENOPHASE), IN EACH POPULATION. EACH COLOUR REPRESENTS A COMBINATION OF MORPHOTYPES, AS DESCRIBED IN THE LEGEND OF THE GRAPH: M12 – *DYRIS* + INTERMEDIATE; M123 – *DYRIS* + INTERMEDIATE + *FUSCA*; M13 – *DYRIS* + *FUSCA*; M23 – INTERMEDIATE + *FUSCA*. 64
FIGURE 19. NUMBER OF FLOWERS AT ALL THE FLOWERING PHENOPHASES - RECENTLY FLOWERING (RFL), FULLY OPEN (FOP) AND FLOWER SENESCENT (FLS) – FOR THE THREE MORPHOTYPES, IN EACH POPULATION (VERTICAL AXIS), PER JULIAN WEEK (HORIZONTAL AXIS). OVERLAPPING OF COLOURS IN THE SAME ROW MEANS SIMULTANEOUS FLOWERING OF DIFFERENT MORPHOTYPES IN THE SAME POPULATION. MORPHOTYPES: 1 – *DYRIS*; 2 – INTERMEDIATE; 3 – *FUSCA*. 65
FIGURE 20. MEAN PEAK FLOWERING DATES BY YEAR OF SAMPLING (ERROR BARS OMITTED), PLOTTED BY MORHOTYPE (A) - 1 (*DYRIS*), 2 (INTERMEDIATE) AND 3 (*FUSCA*) AND BY POPULATION (B) – *ARRUDA DOS VINHOS*, AV; *CASALINHO FACHO*, CF; *MENDIGA*, ME; *MONTEJUNTO*, MJ; *PINHEIRINHOS*, PI, *SERRA DE SANTO ANTÓNIO*, SA. YEARS CONSIDERED: 2011, 2012, 2013, 2014 AND 2015. 67
FIGURE 21. AVERAGE VALUES FOR REPRODUCTIVE SUCCESS IN EACH OF FOUR POPULATIONS, MEASURED AS THE NUMBER OF INFLATED CAPSULES PER NUMBER OF FLOWERS SURVEYED IN EACH PLANT, PER YEAR (2011-2012). POPULATIONS CONSIDERED: ME, MJ, PI AND SA. 68
FIGURE 22. AVERAGE PERCENTAGE OF CAPSULES PER MORPHOTYPE (1-*DYRIS*; 2-INTERMEDIATE; 3-*FUSCA*), MEASURED AS THE NUMBER OF INFLATED CAPSULES PER NUMBER OF FLOWERS SURVEYED IN EACH PLANT. IN EACH MORPHOTYPE, COLUMNS CORRESPOND TO DIFFERENT YEARS, BETWEEN 2011 AND 2015. ERROR BARS (STANDARD ERROR) ARE DISPLAYED IN EACH COLUMN. 69
FIGURE 23. *OPHRYS FUSCA* LINK SUBSP. *FUSCA* (A) AND *OPHRYS DYRIS* MAIRE = *O. OMEGAIFERA* H.FLEISCHM. SUBSP. *DYRIS* (MAIRE) DEL PRETE (B). 74
FIGURE 24. DISTRIBUTION MAP OF POPULATIONS SAMPLED IN CENTRAL PORTUGAL. 1. ARRUDA DOS VINHOS (AV); 2. CASAL FACHO (MARQUES ET AL.); 3. MONTEJUNTO (MJ); 4. SERRA DE STO ANTÓNIO (SA); 5. PINHEIRINHOS (PI); 6. MENDIGA (ME). ..75
FIGURE 25. REPRESENTATIVE FLUORESCENCE HISTOGRAMS IN *OPHRYS* (* = 2C AND 4C PEAKS OF THE INTERNAL STANDARD USED (*PISUM SATIVUM* 'CTIRAD'). (A) *OPHRYS DYRIS* LEAF ANALYSIS (2C ~ 21.99 PG). NOTE THAT AN ADDITIONAL HISTOGRAM RUN ON LOGARITHMIC SCALE IS SHOWN FOR ILLUSTRATING THE IMPACT OF ENDOREPLICATION (2C, 4C AND 8C PEAKS DISPLAYED). (B) *OPHRYS DYRIS* POLLINARIUM ANALYSIS (1C ~ 11.17 PG). 79
FIGURE 26. METAPHASE CELLS FROM ROOT TIPS OF *OPHRYS*, OBSERVED ON A ZEISS AXIOPLAN IMAGING MICROSCOPE. (A) *OPHRYS FUSCA* (CF02) 2N = 72, (B) *OPHRYS DYRIS* (PI15) 2N = 74. SCALE BAR = 10 MM. 80
FIGURE 27. MAP OF PORTUGAL WITH THE APPROXIMATE LOCATION OF THE POPULATIONS SAMPLED. 85
FIGURE 28. ALLELE PATTERNS ACROSS POPULATIONS. DATA COMPUTED IN GENALEX FROM THE BINARY DATA MATRIX, 163 SAMPLES, 13 LOCI. 97
FIGURE 29. GRAPHIC PLOTTED BY STRUCTURE HARVESTER SHOWING THE VARIATION OF DELTAK WHEN TESTING SEVERAL VALUES FOR K (1-7) IN STRUCTURE SOFTWARE. BELOW IS THE EVANNO TABLE PROVIDED BY THE SAME APPLICATION (ADAPTED). 99
FIGURE 30. INDIVIDUAL ADMIXTURE PROPORTIONS (Q) GIVEN BY STRUCTURE SOFTWARE FOR EACH INDIVIDUAL OF THE SIX POPULATIONS (K=2; 13 LOCI). DATA PLOTTED BY ORIGINAL ORDER (A) AND BY Q (B), FOR 163 INDIVIDUALS FROM ALL THE POPULATIONS SAMPLED. NUMBERS 1 TO 6 CORRESPOND TO FIELD POPULATIONS: 1. AV; 2.CF; 3. ME; 4. MJ; 5.PI; 6.SA 100

FIGURE 31. ADMIXTURE PROPORTIONS PER POPULATION, AS GIVEN BY STRUCTURE SOFTWARE, WITH K=2. THRESHOLD VALUES FOR MIXED GENOTYPES OF 0.1 – 0.9. THRESHOLD VALUES FOR HYBRIDS OF 0.35-0.65. CL. 1: CLUSTER 1; CL. 2: CLUSTER 2; CL. 3: CLUSTER 3; MIX. GEN.: MIXED GENOTYPES BETWEEN CLUSTERS 1 AND 2. 102
FIGURE 32. PCA PERFORMED ON BRUVO DISTANCE OBTAINED IN POLYSAT, ON THE 163 INDIVIDUALS FROM THE GENOTYPES DATA MATRIX (C). EIGEN VALUES AND COORDINATES X, Y OBTAINED IN POYSAT. DOTS PLOTTED IN EXCEL. THE FIRST AXIS EXPLAINS 12.0% OF THE TOTAL VARIATION IN THE DATASET, THE SECOND 10 %. COLLORED SYMBOLS ARE ACCORDING TO SAMPLING POPULATIONS: AV AND CF ARE POPULATIONS OF *O. FUSCA*; MJ AND SA ARE POPULATIONS OF *O. DYRIS;* ME AND PI ARE MIXED POPULATIONS.....103
FIGURE 33. MANTEL TEST PERFORMED ON THE DISTANCE MATRICES OF THE MORPHOLOGICAL AND GENETIC DATASETS. A. SCATTERPLOT OF THE CASES OF BOTH MATRICES; B – DISTRIBUTION OF R (AB), ESTIMATED FROM 10000 PERMUTATIONS.103
FIGURE 34. PERCENTAGE OF CAPSULES PER GENETIC GROUP (AVERAGE VALUES), DYRIS (1), FUSCA (2) AND HYBRID (H), MEASURED AS THE NUMBER OF INFLATED CAPSULES PER NUMBER OF FLOWERS IN EACH PLANT. THE LEFT-HAND SIDE BARS, ORANGE, REPRESENT "FILLED" CAPSULES, THE RIGHTHAND SIDE, IN GREY, THE TOTAL CAPSULES. VERTICAL LINES IN EACH BAR CORRESPOND TO STANDARD DEVIATION. 104

TABLE OF TABLES

TABLE 1. LOCATION, TAXONOMIC COMPOSITION AND NUMBER OF PLANTS SAMPLED PER POPULATION. PM – NUMBER OF PLANTS SAMPLED FOR MORPHO-ANATOMICAL DATA (CHAPTER 2); PF – NUMBER OF PLANTS ASSESSED REGARDING PHENOLOGICAL STAGES AND POLLINATION SUCCESS, TOTAL NUMBER OF FLOWERS, "FL", BETWEEN () (CHAPTER 2); PP – NUMBER OF PLANTS SAMPLED FOR PLOIDY ANALYSIS (CHAPTER 3); PS – NUMBER OF PLANTS SAMPLED FOR DNA (CHAPTER 4). 28

TABLE 2. DESIGNATION CODE, LOCATION, TAXONOMIC COMPOSITION AND NUMBER OF PLANTS SAMPLED PER POPULATION. PM – NUMBER OF PLANTS SAMPLED FOR THE MORPHO-ANATOMICAL DATASET; NUMBER OF PLANTS (PF) AND TOTAL NUMBER OF FLOWERS (FP) ASSESSED REGARDING FLOWERING STAGES AND POLLINATION SUCCESS. 36

TABLE 3. PLANT LIFESTAGES AND FLOWER DEVELOPMENTAL STAGES CONSIDERED. 38

TABLE 4. MORPHO-ANATOMICAL FLORAL CHARACTERS INCLUDED IN THE ANALYSIS. DIFFERENT CHARACTER TYPES (T) WERE CONSIDERED: BINARY (B), CONTINUOUS (C) AND QUALITATIVE (Q). UNITS OF MEASUREMENT AND ITS DEGREE OF ACCURACY (CONTINUOUS CHARACTERS) OR CHARACTER STATES (QUALITATIVE AND BINARY CHARACTERS) ARE PROVIDED. BELOW THE STATES OF QUALITATIVE CHARACTERS, THE CODES USED IN THE QUANTITATIVE APPROACH/ANALYSIS ARE DISPLAYED. FLOWER POSITION INDICATES THE POSITION OF THE FLOWER FOR WHICH MEASUREMENTS WERE CONSIDERED, IN THE CASE OF PLANTS WITH MORE THAN ONE FLOWER ASSESSED AND/OR MEASURED IN SEVERAL YEARS. 40

TABLE 5. PERCENTAGES OF VARIANCE ACCOUNTED BY EACH OF THE FIRST TEN COORDINATES AFTER PRINCIPAL COORDINATES ANALYSIS PERFORMED IN M;SVP SOFTWARE. 46

TABLE 6. CORRELATION TABLE FOR PEARSON'S TWO-TAILED TEST. VALUES FOR THE FOUR VARIABLES CONSIDERED. N = 226 48

TABLE 7. VIF VALUES FOR COLLINEARITY BETWEEN VARIABLES AND INTERACTIONS 49

TABLE 8. RESULTS OF THE FINAL GLMM MODEL FOR THE PREDICTION OF MORPHOTYPES, SELECTED BY STEPWISE MODEL REDUCTION. LIP WIDTH HAS REVEALED TO BE THE MOST SIGNIFICANT PREDICTOR. COLLINEARITY VALUES ADDED: VIF (PER PREDICTOR) AND KAPPA FOR THE GLOBAL MODEL 49

TABLE 9. QUANTITATIVE MORPHO-ANATOMICAL TRAITS ASSESSED, PER INDIVIDUAL PLANT (STEM LENGTH, INFLORESCENCE LENGTH AND LENGTH OF THE BASAL BRACT) OR PER FLOWER (ALL THE TRAITS BELOW). VALUES AVERAGED OVER MORPHOTYPES, PER INDIVIDUAL PLANT FOR THE FIRST THREE TRAITS OR PER ALL THE FLOWERS FOR THE REMAINING, STANDARD DEVIATION AND THE RANGE IN PARENTHESIS 56

TABLE 10. SUMMARY TABLE OF-QUALITATIVE VARIABLES ASSESSED AND THE STATE OF CHARACTER BEST REPRESENTED IN EACH MORPHOTYPE, OR THE TWO MOST REPRESENTED, IN CASE THE DIFFERENCE BETWEEN THEIR PROPORTIONS IN THE SAMPLE DOESN'T REACH 10%. BETWEEN BRACKETS ARE THE CORRESPONDING PROPORTIONS IN THE SAMPLE. N = NUMBER OF PLANTS ASSESSED. 60

TABLE 11. RESULTS FOR THE GLM TESTING FOR THE EFFECT OF THE NUMBER OF FULLY OPEN FLOWERS (FOP) IN THE DIFFERENT POPULATIONS. ALL MORPHOTYPES CONSIDERED. THE NUMBER OF OPEN FLOWERS IN THE POPULATIONS OF MJ (P=0.015) AND ME (P=0.029) DIFFER SIGNIFICANTLY FROM THE OTHERS. 63

TABLE 12. RESULTS OF GLM ANALYSIS TESTING FOR DIFFERENCES IN THE NUMBER OF FLOWERS AT FOP FROM THE THREE MORPHOTYPES, IN THE POPULATIONS WITH TWO OR THREE MORPHOTYPES SAMPLED IN 2011. SIGNIFICANT VALUES IN BOLD. INTERACTIONS WERE EXCLUDED FROM THE MODEL DUE TO VARIANCE DEFICIT. 65

TABLE 13. RESULTS FOR THE GLM TESTING FOR THE EFFECT OF YEAR AND POPULATION AS PREDICTORS ON THE NUMBER OF FLOWERS AT THE FOP STAGE, FOR ALL POPULATIONS AND ALL YEARS. 68

TABLE 14. NUCLEAR DNA CONTENTS FROM ALL THE PLANTS ASSESSED, FROM POLLINARIA (67). 77

TABLE 15. SUMMARY OF THE NUCLEAR DNA CONTENTS FROM FLOW CYTOMETRY USING SILICA-DRIED POLLINARIA AND FRESH LEAVES OF *O. FUSCA* AND *O. DYRIS*. CHROMOSOME NUMBERS FOR THE CORRESPONDING SAMPLES ARE PROVIDED. 78

TABLE 16. LOCATION, TAXONOMIC COMPOSITION AND PLANT NUMBERS OF EACH POPULATION SAMPLED. PT – NUMBER OF PLANTS TAGGED; PS – NUMBER OF PLANTS SAMPLED FOR DNA; PP – NUMBER OF PLANTS SAMPLED FOR PLOIDY ANALYSIS 84

TABLE 17. PRIMER PAIRS SUCCESSFULLY AMPLIFYING AND RUN FOR 166 INDIVIDUALS. LOCI MARKED * CORRESPOND TO THOSE ANALYSED IN THE CURRENT REPORT. ANN. TEMP. – ANNEALING TEMPERATURE USED (TEMPERATURES WHERE EFFICIENCY OF PCR AMPLIFICATION WAS MAXIMAL. 86

TABLE 18. CHARACTERIZATION OF THE EIGHT MICROSATELLITE MARKERS PREVIOUSLY DEVELOPED FROM *O. FUSCA* BY COTRIM ET AL. (2009). ADAPTED FROM THE SAME AUTHORS. 87

TABLE 19. GENETIC DIVERSITY SUMMARY STATISTICS PER POPULATION (# NUMBER OF INDIVIDUALS SAMPLED), COMPUTED IN SPAGEDI SOFTWARE WITH DATA FROM 13 LOCI: NUMBER OF ALLELES WITH NON-ZERO (N_A); NUMBER OF EFFECTIVE ALLELES (N_{AE}); ALLELIC RICHNESS CORRECTED FOR SAMPLE SIZE (K=4 GENE COPIES) (A_R); NEI'S GENE DIVERSITY CORRECTED FOR SAMPLE SIZE (H_E); OBSERVED HETEROZYGOSITY (H_O); INDIVIDUAL INBREEDING COEFFICIENT (F_I). 95

TABLE 20. GENETIC DIVERSITY SUMMARY STATISTICS BY PRIMER. MEAN VALUES ACROSS ALL LOCI GIVEN IN TABLE 1, AS OVERALL VALUES. 96

TABLE 21. GENETIC DIFFERENTIATION INDICES CALCULATED IN SPAGEDI 163 INDIVIDUALS, FROM THE GENOTYPES MATRIX (C). MEAN F_{ST}, R_{ST} – MEAN POPULATION DIFFERENTIATION; F_{IS}, R_{IS} – INBREEDING COEFFICIENTS FOR F AND R STATISTICS; PAIRWISE RHO –PAIRWISED POPULATION DIFFERENTIATION (RONFORT ET AL., 1998), USING THE ANOVA FRAMEWORK. 98
TABLE 22. AMOVA TABLE COMPUTED IN GENALEX. 99
TABLE 23. NUMBERS (#) AND PERCENTAGE (%) OF INDIVIDUALS STRONGLY ASSIGNED (Q>0.95) TO EACH CLUSTER. K = 2 100

ACKNOWLEDGEMENTS

After such a long journey I would like to thank to my supervisors, Julie, Mike and Helena, for their precious help and tremendous patience. Julie overall, for sorting out so many issues derived from the particular circumstances in which this project was carried out. Without that, the current submission would not have been possible.

Also, to all the colleagues from whom I learned and shared experience with: all the people in Kew, all the colleagues at Lisbon University and the Natural History Museum and the colleagues in Reading and Kew. A special thanks to Victor Rodriguez-Garcia and his help in the lab throughout the first period at Reading University, Robyn Cohan and Jaume Pellicer for supervision and assistance at the Jodrell Laboratories.

Professor Richard Bateman for the insightful comments in the beginning of the study and suggestions when defining the set of morphological characters to sample.

Peter Rooney, for help and comments with the statistical analysis of morphological and phenological dataset.

All those who occasionally shared field days with me, thank you very much.

My husband Rui, for help in the field, with the software, in developing a programme and database that allowed me to leave loads of field sheets and machinery at home and carry almost only... a tablet to the field! Also for data verification, assistance with material to tag and measure plants, detect plants markings, assuming the care of the children when mum is working... and sorting out so many details.

Henrique and Maria Inês, my children, those who most diverted me from the project but also my permanent inspiration.

1. INTRODUCTION

1.1. Hybridisation and related concepts

Hybridisation is defined as reproduction between members of genetically distinct populations, sensu Barton & Hewitt (Barton and Hewitt). It was already considered by Linnaeus (Linnaeus) and by Darwin (Darwin). Nevertheless, despite botanists having long accepted hybridisation as a process (Anderson and Stebbins, 1954, Arnold, 1992 Abbott, Albach et al. 2013, Abbott et al., 2013), the assumptions of the biological species concept and its focus on reproductive isolation caused a delay in the wider recognition of hybridisation as an important evolutionary mechanism (Coyne and Orr, 2004, Marques et al., 2017b). The phenomenon is now widely recognized as an important evolutionary force that can cause biodiversity loss or lead to the emergence of new biodiversity (Quilodran et al., 2020). It is involved in adaptation to novel environmental conditions (Lewontin and Birch, 1966, L Morjan and H Rieseberg, 2004, Baskett and Gomulkiewicz, 2011, Arnold and Kunte, 2017) and speciation (Anderson, 1948, Rieseberg, 1997, Abbott et al., 2013, Mallet, 2007), hence it is also relevant to conservation programmes (Ennos et al., 2012).

Hybridisation is common on a per-species, but very rare on an individual basis: interspecific hybridisation has been indicated for 25% of plant species and 10% of animal species (Mallet, 2007). A recent analysis of hybridisation in the Mediterranean region focusing mainly on the Iberian Peninsula (Marques et al., 2017b), reported 1032 accepted hybrid taxa for the latter region, representing 13% of the species in the total flora. Regarding the British Isles, 720 hybrids were documented in 1975 (Stace, 1975) as all known hybrids from Flora of Britain and Ireland at that time. An updated edition brought together the existing taxonomy and information on geographic distribution, accounting for 909 hybrids (Stace et al., 2015), and interest in hybridisation has expanded since the frequency of hybridisation is predicted to increase under global change (Vallejo-Marin and Hiscock, 2016).

Introgression is frequently associated with hybridisation. Sensu Anderson (Anderson, 1949, Anderson and Stebbins, 1954), it refers to the incorporation of alleles from one entity (species) into the gene pool of a different entity, usually via backcrossing to one or both parents after hybridisation (Anderson, 1948, Arnold, 1992). It is overall a relative term, as genetic material is introgressed with respect to the genetic material of the second entity. Part of the genetic material remains thus constant, only a portion of the gene pool of each of the hybridising taxa being transferred (Harrison and Larson, 2014).

The study of hybrids and the recognition of species are necessarily intertwined (Soltis and Soltis, 2009). Species are the most important units of study in classical biology, the species rank being the fundamental taxonomic rank. Most often, when hybridisation events are described, the genetically distinct populations involved are different species. Nonetheless, it is now known that barriers between morphologically, ecologically or genetically distinct organisms are semipermeable (Harrison and Larson, 2014, Behie and Groves, 2015, vonHoldt et al., 2018, Quilodran et al., 2020), the boundaries between species being widely discussed. From a contemporary perspective, it is precisely the knowledge of the extent of hybridisation and introgression that will help to define such boundaries (Harrison and Larson, 2014). In a review published in 1997, Mayden (1997) recognized 27 different species concepts. They diverge in scope, but three main basic principles can be synthesised, according to Bateman (2010):

- Similarity of appearance – morphology. This is the most commonly used concept. Individuals of one species should have a combination of reliable characters that consistently separates them from all others.
- Reproductive isolation. This is the main criterion of the biological species concept. As long as individuals naturally interbreed without any significant reduction in their fitness, they are considered to belong to the same species. Demonstration of interbreeding can be ecological (crossing evidence detected in the field) or genetic (determining if, and on what basis, individuals exchange genes – if gene flow is occurring)
- Monophyly. This approach identifies natural groups, both self-defining and the product of evolutionary change, including all the descendants of a single presumed ancestor.

Conflicting perspectives on species concepts can have a high impact on the number of entities considered. For example, in *Ophrys* L., the orchid genus that is the focus of this study, the number of species ranges from > 250 following the narrowest concepts (Delforge, 2002), to 19 species and 65 subspecies (Pederson and Faurholdt, 2007). In *Ophrys*, such disparity is also a consequence of the huge variability of floral morphology, variation that might be partitioned into discrete entities when geographically circumscribed and supposedly with a specific pollinator (Bateman et al., 2010), or grouped into morphologically broader species when gene flow is thought to occur and a more inclusive concept of species is used. Despite preferring the perspective of Pederson and Faurholdt (2007), one tend to agree with Coyne (2004) and Mallet (1995, 2007), who advocated a genetic version of Darwin's definition, the genotypic cluster method definition in which species are seen as distinguishable groups of genotypes that remain distinct in the face of potential or actual hybridisation and gene flow.

1.2. Reproductive barriers

Despite its widely reported occurrence, hybridisation is not inevitable whenever sympatric species come into contact, because natural pre- and post-zygotic barriers restrain it, limiting geneflow between species (Mayr, 1940, Keller et al., 2020, Coyne and Orr, 2004). Pre-zygotic barriers include those acting before pollen deposition, such as differences in habitat, phenology or pollinators, as well as those acting after pollen has been transferred. Post-pollination pre-zygotic barriers include interactions between pollen and pistil that may restrict pollen tube growth. Post-zygotic barriers include hybrid inviability or sterility, that can be due to genetic and chromosomal incompatibilities, and reduced hybrid fitness (Pickup et al., 2019).

The strongest barriers that are known to limit hybridisation are pre-mating, acting before pollen deposition. Phenotypical traits – morphology, behaviour or ecological traits - can establish preliminary boundaries which play an important role in isolating species (Harrison and Larson, 2014). Amongst the most effective is isolation by behaviour (Keller et al., 2020), orchids being an illustrative example, with the specific pollination syndrome being the key for species isolation within the group.

Pre- and post-zygotic barriers may also act to limit backcrossing between new hybrids and parental species. There are several cases described in literature where isolating mechanisms have been modulated to favour hybrids. Marques, Jurgens et al. (2016) reported one such case of recruitment of a novel pollinator by hybrids of *Narcissus* species – the hybrids are pollinated by an ant, whereas parents are pollinated by non-ant hymenopterans of larger size. In another study of reproductive isolation between a hybrid and its parental taxa, Lowe and Abbott (2004) concluded that both pre- and post-zygotic barriers contribute to the reproductive isolation between the hybrid *Senecio eboracensis* R.J.Abott & A.J.Lowe (2n=40) and its parents, *S. vulgaris* L. (2n=40) and *S. squalidus* L. (2n=20). Strong post-zygotic barriers in this case include the fact that the hybrid is triploid and highly sterile because of meiotic chromosome mispairing, whereas pre-zygotic isolation is mainly due to self-fertilisation being the predominant reproduction mechanism of both parent species.

Recent literature has focussed on mechanisms behind the evolution of reproductive barriers, integrating data from both ecological and genetic studies to highlight that the effectiveness of reproductive barriers in preventing hybridisation or determining the fate of hybrids is strongly affected by selective forces related to the mating systems of the hybrid species. In 2016, following a workshop on mechanisms of plant speciation, Lafon-Placette et al. (2016) referred to some of the most recent research and considered its effects on plant speciation paradigms, specifically on

reproductive barriers, making a synthesis of the genetic elements underlying reproductive barriers in the cases of *Dianthus*, *Petunia*, *Phlox*, *Solanum*, *Mimulus*, *Arabidopsis*, *Aegilops* and *Capsella*. In 2019, Pickup et al. (2019) detailed how different mating systems can alter the balance between gene flow and selection against hybrids. As summarised by the authors, the breeding system of the hybrids and parental species can influence the resulting hybrid zone mode (uni or bimodal, with or without asymmetries) and the level and direction of gene flow in different ways, depending on whether they are self-incompatible, self-compatible (ranging from predominant selfing to predominant outcrossing) or dioecious. Reproductive isolation is thus known to be driven by sexual selection (Lafon-Placette et al., 2016) - selection involving intrasexual competition and female choice for males, this being of paramount importance in plant speciation processes. A high sexual specificity will accelerate the evolution of traits that influence mating success and subsequently the evolution of traits responsible for pre- and post-mating hybridisation barriers between species (Lafon-Placette et al., 2016).

1.3. Outcomes of hybridisation

Natural hybridisation, which is easier to detect after previous allopatric populations have come into contact, frequently results in a hybrid zone (Harrison and Larson, 2014). Here, populations of individuals that are distinguishable on the basis of one or more heritable characters overlap spatially and temporally and cross to form viable and at least partially fertile offspring (Arnold, 1992). In these regions, parental types, F_1 hybrids, and multiple generation hybrids and backcrosses are present in varying proportions (Harrison and Larson, 2014). The presence of such variety of genotypes that result from many generations of recombination provide unique opportunities for genetic studies (Rieseberg, 1997, Rieseberg et al., 1999, Harrison and Larson, 2014). The study of natural hybridisation has thus been used to address several evolutionary questions.

The role hybridisation might have in generating new biodiversity and promoting speciation or, on the other hand, causing biodiversity loss, varies widely between different hybridising taxa and at different stages of divergence (Barton, 2001, Abbott et al., 2013, Osborne et al., 2016). Even the fitness of descendent hybrids varies widely (Burke and Arnold, 2001). Evolutionary outcomes of natural hybridisation (Abbott et al., 2013) are hence the subject of numerous studies. Following natural hybridisation, introgression may lead to (i) merging of the hybridising forms, either if hybrids exhibit reduced fitness and the growth rate is not enough for the replacement of parental taxa (demographic swamping) (Wolf et al., 2001) or when hybrids are fertile with no fitness reduction but genetic homogenization of the parental taxa takes place, also named genetic swamping (Nieto Feliner et al.,

Wolf et al., 2001) reinforcement of reproductive barriers through assortative (conspecific) mating, selection favouring extreme phenotypes (Hamilton and Miller, 2016, Burke and Arnold, 2001, Howard, 1986)) or, (iii) adaptive introgression, whereby fit, introgressed genotypes are produced, that expand into a novel habitat (Arnold, 1992, Rieseberg, 1997, Abbott et al., 2013). Recent research has demonstrated that these scenarios are not exclusive. The first accounts for the loss of biodiversity, being considered an example of "hybrid inferiority" and the erosion of variability; while the second and third stand for hybrid success, the third being a classic way to speciation. In the literature, case studies of hybridisation fitting all three types of outcomes are described. For example, (i) a case where hybridising forms have merged with parental taxa after a few generations has recently been described in *Cakile* (Li et al., 2020). Though the result of artificial crosses, after several generations of backcrossing, hybrids were barely distinguishable from the original parental plants, displaying the same phenotypes and breeding systems (Li et al., 2020). Reinforcement after speciation (Nieto Feliner et al., Butlin, 1987) depends on the evolution of barriers to gene flow due to natural selection against the production of new hybrids (Nieto Feliner et al., 2017). This mating discrimination prevents further introgression and may result in gradual loss of biodiversity. Mallet (2007) suggested that reinforcement should not be considered as hybrid speciation, because no third cluster of genotypes stabilises or becomes distinct in contact with either parent, meaning that a third species does not form. In the case of adaptive introgression (iii), when fit introgressed genotypes expand to a novel habitat, an opportunity for adaptation and evolution of new biodiversity is created, facilitating speciation (Lewontin and Birch, 1966, Arnold, 1992, Briggs and Walters, 2016). It is believed that, in such cases, exogenous selection plays a central role in the establishment of fit hybrids (Burke and Arnold, 2001). Positive selection may fix adaptive alleles whereas negative selection can remove detrimental ones (Schumer et al., 2016).

More classical views of the theme stated that for the success of new hybrid species, the evolution and maintenance of reproductive barriers are particularly crucial (Rieseberg, 1997). Taking this view, it is fundamental that the new species, whatever its particular habitat, stays isolated from parentals. From this perspective, the documentation of presence and the extent of reproductive isolation between the entities being studied is extremely important in hybridisation studies, adding to the fact that it can inform taxonomy and influence conservation strategies (Ennos et al., 2005).

1.4. Studying hybridisation

1.4.1 Morphological studies

When Anderson and Stebbins published their work in the late forties (Anderson, 1948, Anderson, 1949) and early fifties (Anderson and Stebbins, 1954), a new conceptual framework was being proposed that included natural hybridisation as responsible for new species and adaptations. The authors emphasized the origin and transfer of adaptations through natural hybridisation. Despite natural barriers that restrain hybridisation (section 1.2.), gene introgression amongst closely related species in sympatric populations can lead to complex patterns of morphological variation. Where fertile or partially fertile F1 hybrids are generated, a wide range of morphological, genetic and ecological variation can be released in backcross, F2, and later generation progeny (Lowe and Abbott, 2015). Hybrids may display intermediate trait values, combine traits from both parents and/or have extreme trait values. First-generation hybrids usually show a mosaic of parental and intermediate characters (Rieseberg et al., 1993). Morphological identification of hybrids can also be more difficult as ancestral polymorphism or mutations at a few genes related to phenotypical traits can make non-hybrid specimens resemble hybrids (Mallet, 2005).

In 1949, Anderson pioneered a method to display variation in hybridising populations, naming it the Hybrid Index (Anderson, 1949). The Hybrid Index method produces a scale of variation, with plants of pure parental taxa obtaining the highest and lowest scores, and plants of intermediate morphology having intermediate scores. It can thus be seen as a means of describing the degree of separation of plants of different morphology (Briggs and Walters, 2016).

Populations in hybrid zones which contain a variety of hybrid variants have been labelled “hybrid swarms” (Anderson, 1949). In fact, due to fertility constraints in hybrid swarms, different genomic combinations can affect normal expression of characters and give rise to new recombinants and introgressants (Yakimowski and Rieseberg, 2014, Anderson and Stebbins, 1954). Trait coherence can thus be broken, with a transgressive expression in hybrid derivates – mean values higher or lower than those of either parent – creating novel morphological variation, reproductive isolation and adaptation to new habitats (Whitney et al., 2010, Arnold et al., 2012, Yakimowski and Rieseberg, 2014).

Different authors agree that the occurrence of hybrid swarms can nowadays be related to the disturbance of habitat caused by human activities, as this generates new or graded ecological niches that allow the persistence of stabilised introgressants and/or homoploid hybrid species (Lowe and Abbott, 2015).

1.4.2. Cytogenetic studies

Methodological advances have placed genetic and cytogenetic methods at the centre of most studies focusing on hybridisation, as this phenomenon usually induces fast genomic changes. Cytotype characterisation of naturally hybridising species can be an important component of hybridisation studies (Petit et al., 1999). Such characterisation can reveal of more relevance in studies using co-dominant molecular markers such as the current work, in the sense that it contributes to its robust interpretation (Pellicer et al., 2012). In a hybridisation scenario, genetic incompatibilities attributed to different ploidies of the parental taxa may create reproductive barriers that will reduce the level of introgression (Abreu et al., 2017). Both cytogenetic and genomic studies have become particularly important in highlighting the occurrence of homoploid hybrid speciation phenomena, bringing strong evidence for these processes (Gross and Rieseberg, 2005, Taylor and Larson, 2019), of which there is an increasing number of reported examples.

1.4.2.1. Genome size

The size of genomes is measured as the C-value. Variations in C-value are not always correlated with the complexity of organisms or changes of ploidy, an observation that has intrigued some authors for some time, and has been termed as the 'C-value enigma' (Gregory, 2001). It is now known that repetitive elements of the genome – the non-protein encoding fraction – is correlated with genome size (Chooi, 1971) and accounts for half or more of the genome of species with large genomes (Bennetzen, 2000). Results from a study of the evolution of nuclear DNA content in homoploid hybrid species of *Helianthus* L. has confirmed that hybrid-derived species have 50% more nuclear DNA than the parental species. Nevertheless, first-generation hybrids differed in DNA content according to the maternal parent and it is not hybridization itself that leads to increased nuclear DNA content in this case, the evolutionary forces responsible therefore remaining mysterious (Baack et al., 2005).

Cytogenetic studies have also shown that increases in the size of genomes can be favoured or selected against as a result of local environments. There is a correlation between genome size and latitude, the direction of this relation depending on the species and environmental traits considered (Knight et al., 2005). The cited authors have proposed the "large genome constraint hypothesis", according to which species with large genomes, i.e. the highest C-values, tend to have restricted ecological distributions, being progressively excluded from harsh environments above a midlatitude (Knight et al., 2005). Plants with large genomes seem to be at disadvantages in the extremes of both temperature and precipitation ranges (Knight et al., 2005, Knight and Ackerly, 2002). Within the intermediate range, plants with larger genomes can, however, be favoured in the extremes of ecological ranges. This may

happen in places where cell expansion is a more efficient way of growing than cell division, such as in the cold weather (Knight and Ackerly, 2002).

The study of genome sizes of plants has also offered opportunity to detect intraspecific variation, sometimes with important consequences in the phenotypic variation of conspecifics (Bennett and Leitch, 2005). Reported cases include duplications, aneuploidy, or B chromosomes (Bennett and Leitch, 2005). Other examples of genuine intraspecific variation include differences in telomeric heterochromatic segments generated by losses or gains visible under the light microscope (Gustafson et al., 1983), for example. Other example was reported by Greilhuber (Greilhuber) for subspecies of *Scilla bithynica* Boiss. subsp. *bithynica* with many large C-bands (1C = 29.20 pg) and subsp. *radkae* (Davidov) Speta with few small C-bands (1C = 22.90 pg).

1.4.3. Genetic studies

Microsatellites [also referred to as simple sequence repeats (ssrs)], the genetic markers used in this study, are co-dominant molecular markers commonly used to test for evidence of hybridisation and gene flow. Consisting of tandemly repeating units of DNA of 1 or 2 to 6 bp in length, they are widely distributed throughout the nuclear genomes of eukaryotes (Bhargava and Fuentes, 2010). With a good performance in individual genotyping, gene flow studies and population differentiation (Lowe et al., 2007), they also have the advantage of allowing assessment of inbreeding levels (Lowe et al., 2007), and usually perform better than SNPs in short temporal or spatial scale studies, used when both good resolution and cross-species range are required, in taxa that are slowly evolving or highly clonal or to study genome evolution (Putman and Carbone, 2014).

Recent studies using new sequencing methods, allowing the access to longer genomic windows, permit detailed assessment of the role of gene flow and its evolutionary consequences during species diversification (Osborne et al., 2016). New molecular tools have reinforced the conclusions that species boundaries are semipermeable (Harrison and Larson, 2014). In addition, genome sequencing methods and new study systems have been unravelling the genetic architecture of phenotypical traits involved in hybridisation such as flower colour (Hopkins and Rausher, 2011, Hopkins and Rausher, 2012) or flower scent (reviewed in Sheehan et al. (2012), allowing to relate them with the efficiency of pre-zygotic barriers to hybridisation (Hopkins, 2013).

1.4.4. Case studies of plant hybrids

The literature on plant hybridisation includes a vast number of studies of closely related species or species complexes, where cases of hybridisation have been documented in natural populations, having become models for understanding hybridisation. These model systems include natural populations of irises (Arnold, 1992), sunflowers (Rieseberg et al., 2003, Rieseberg et al., 1999), ragworts (Lowe and Abbott, 2015, Lowe and Abbott, 2004, Brennan et al., 2012, Osborne et al., 2016, Nevado et al., 2020) and daffodils (Marques et al., 2010, Marques et al., 2016, Marques et al., 2012, Marques et al., 2007).

1. Iris hybrids

The ecology and evolutionary genetics of some species of the genus *Iris* L. have been studied for many decades. Already referred to by Anderson as a "typical case" of introgression (Anderson, 1949), each of the species of this complex is found in different regions of the United States, but all their distributions overlap in southern Louisiana, and thus this complex is called "the Louisiana irises" (Arnold, 1994). Asymmetric introgressive hybridisation has been extensively documented in natural populations of the genus (Arnold et al., 2010), in particular the introgression of genes from *Iris fulva* Ker Gawl. into both *Iris brevicaulis* Raf. and *Iris hexagona* Walter. Authors have also detected that determined pre-zygotic and post-zygotic reproductive barriers seem to contribute to it (Arnold et al., 2010). In this model system it was concluded that introgression provided some hybrid genotypes with adaptations to extreme environments (Martin et al., 2006), namely flooded conditions of more than 1 m of water for several months, thus having an adaptive effect on the recipient. Introgression positively affected flooding tolerance, a key-trait for the survival of these irises. Multiple genomic regions were detected to be involved (Martin et al., 2006)

2. Sunflower hybrids

The North American genus *Helianthus* L. is a well-documented case of homoploid hybrid speciation (Rieseberg et al.). Studies involving wild sunflower plants provided evidence that hybridisation facilitated ecological divergence and adaptation to novel habitats (Rieseberg et al., 2003), contributing to the adaptation to extreme habitats.

In 1999, Rieseberg and co-authors combined random amplified polymorphic DNAs with a linkage map of those markers to study introgression across a hybrid zone, having identified chromosomal segments with reduced introgression across three replicate, supposedly independent hybrid zones (Rieseberg et al., 1999), and the results seemed to indicate that reduced introgression was a consequence of selection and, in fact, several chromosomal segments were associated with hybrid pollen sterility, an important barrier to gene flow in this genus. Pioneering the analysis of differential introgression in the

context of a genetic map, the authors laid the groundwork for further genetic studies. Buerkle and Rieseberg (Buerkle and Rieseberg) observed a remarkable consistency of introgression patterns across different hybrid zones, suggesting that habitat isolation is another of the barriers to gene exchange. In another genome-wide study of introgression, Scascitelli et al. (2010) analysed the variation of 88 genetically mapped microsatellite loci. Studying introgression between *Helianthus annuus* L. (common sunflower), *H. debilis* Nutt. and their putative hybrid, the authors concluded that this phenomenon is genome-widespread and asymmetric, not restricted to some genomic regions, with higher migration rates from the hybrid into the two parental taxa than vice-versa.

3. Ragwort hybrids

The genus *Senecio* L. (Osborne et al.), Asteraceae, represents another case that has been extensively studied, with geographically identified hybrid regions where hybrid swarms occur. Studies have documented a large number of hybrid speciation events (Lowe and Abbott, 2004, Brennan et al., 2012, Chapman and Abbott, 2010, Lowe and Abbott, 2015, Osborne et al., 2016, Nevado et al., 2020), this being common throughout this genus.

In the hybrid swarms under study in this species complex, several stable tetraploid hybrid derivates co-exist, e.g. those from *Senecio vulgaris* L. and *S. squalidus* L., the origin of which was investigated in 2015 (Lowe and Abbott, 2015). Considered was the possibility of a polytopic vs. a single origin followed by dispersal. The authors highlighted that the first would demonstrate the adaptive character of hybridisation, allowing allele-sharing between species and reproductively isolated recombinants under suitable conditions, while a single-origin scenario would underpin the particular nature of hybridisation, placing the emphasis on the study of long-distance dispersal and its consequences in generating and maintaining biodiversity (Lowe and Abbott, 2015). A polytopic origin was referred as the most likely origin for one of the derivate hybrids, *S. vulgaris* var. *hibernicus* Syme, while for the other taxon, *S. eboracensis* Abbott and Lowe, the strong possiblity of a local hybridisation event was suggested.

The genomic basis of the reproductive isolation and of the morphological differences between other two closely related species within the genus, *Senecio aethnensis* Jan ex DC .and *S. chrysanthemifolius* Poir., was revealed in a genomic approach of linkage mapping from SNPs (Chapman et al., 2016). In the same year, Osborne et al. (2016) fully resolved the phylogeny of eight *Senecio* species, demonstrating cases of introgressive hybridisation between multiple pairs of taxa across the species tree. Using a whole-genome gene-space dataset these species were thus confirmed as another study system for diversification with gene flow. Another chapter in the evolutionary history of *Senecio* has been written very recently, when Nevado et al. (2020) found that the homoploid hybrid species *S.*

squalidus had its origin in an event of hybridisation at the end of the 17th century, while its two parental species were in cultivation in two British gardens, shedding light into a very recent process of homoploid hybrid speciation. This new species spread rapidly, colonising the novel habitat in a process that has been very well documented. Speciation seems to be the outcome of genetic incompatibilities between hybrid and parental taxa and of transgressive phenotypes arisen by hybridisation that allowed extension for novel habitats (Nevado et al., 2020).

4. Poplars

Poplars (*Populus* spp.) are a recent case study of hybridisation, specifically between the species *Populus alba* L. and *Populus tremula* L. (Rajora and Dancik), ecologically divergent forest trees. The study of introgression across the hybrid zone of these two species has already contributed with relevant insights into the processes underlying hybridisation. Works from Lexer (2007, 2010) using a whole genome approach, have revealed that, despite reproductive isolation between these species being stronger than previously assumed, introgression of neutral or advantageous alleles still takes place. The same authors also found unexpected gaps between recombinant hybrids and parental taxa (Lexer et al., 2010). The reason for these widespread genotypic patterns was assortative mating and strong postzygotic barriers, rather than recent population history (Lexer et al., 2010). Both differential and asymmetric introgression was detected by Lindtke at al. (2012), when studying the genomics of reproductive isolation by genotyping 77 mapped microsatellites in individuals from three replicate hybrid zones. Notwithstanding a very strong differentiation between the genomes of the parental taxa, many loci in recombinant hybrids showed "greatly increased between species heterozygosity", contradicting expectations based on the assumption of selection against intermediate genotypes (Lindtke et al., 2012). The authors pointed out that epistatic interactions within genomes are likely to contribute significantly to the maintenance of reproductive isolation between these divergent species, parental taxa maintaining their integrity despite gene flow (Lindtke et al., 2012). Results from Stolting *et al.* (2013), based on SNPs assayed by restriction site associated DNA (RAD) sequencing from two allopatric populations, pointed to introgression increasing in an uncontrolled way. The authors found evidences of allele sharing in several genomic blocks, hypothesizing that this was caused by recurrent gene flow, instead of shared ancestral polymorphism. A complex history of admixture is suggested, as the genomes have been affected by hybridization and introgression "since thousands if not tens of thousands of years". Assortative mating and post-zygotic isolation are confirmed as barriers to gene exchange. Another study within this complex showed that high levels of hybridisation and substantial hybrid fitness may allow for the maintenance of species integrity (Lindtke et al., 2014). By assessing the incidence of intraspecific and hybrid matings within a mosaic hybrid zone of these poplar species, the authors revealed that seedlings included a full range of hybrids, in agreement with weak

reproductive barriers in the early stages of the life cycle of *Populus*. The maintenance of species boundaries in later stages was therefore attributed to post-zygotic selection. Many hybrid seedlings have thus to be removed between the seedling stage and maturity, so that a large amount of selection takes place between the seedling stage and maturity.

5. Daffodils

Narcissus L. (daffodils) is another genus in which species are known to readily hybridise in nature, multiple hybridisation events having been identified (Marques et al., 2017a). Molecular data have revealed the complexities of the evolution of these plants, one of the most popular ornamental bulbs, tackling issues of evolution and systematics already acknowledged by many previous authors.

In 2010, a study combining DNA sequencing, cytological studies, crossing experiments and niche modelling revealed introgression processes within the genus (Marques et al., 2010). In this study, aiming to unravel the origin of a controversial hybrid and explain the occurrence of orphan hybrid populations, the authors highlighted the weakness of reproductive barriers to the fertilization by foreign pollen and point out extirpation via demographic competition as the most likely explanation for the origin of hybrid disjunct populations. More recently, the hybrid nature of several putative hybrids, including allopolyploids, has been confirmed (Marques et al., 2017b) and morphological and cytogenetic results also support that previous detected phylogenetic inconsistencies can be attributed to hybridisation (Marques et al., 2017b). Such results underpin the important role that reticulate evolution has had within the genus.

1.5. Conservation

Most conservation programs, based on prioritised species lists to identify priority habitats for conservation, are not appropriate for actively evolving groups and complex populations that show atypical levels of morphological diversity (Ennos et al., 2005). Conservationists today recognise that conservation of evolutionary processes that generate biodiversity can be as important or even more so than conservation of taxonomic entities (Ennos et al., 2005, Ennos et al., 2012). When species have naturally coexisted for many generations, hybridisation should not be seen as a threat (Rieseberg et al., 2007), and regions where those species occur in sympatry can be important areas in which to preserve evolutionary processes (Cozzolino et al., 2006). On the other hand, as hybridisation can also be a force acting towards extinction (Outcomes of hybridisation, section 1.3) (Wolf et al., 2001),

extinction risk should be evaluated, in the case of vulnerable species. Conservation and management plans should thus consider whether hybridisation is occurring and the type of outcomes being generated.

1.6. The study group and study sites

This thesis is a study of a pair of species, *Ophrys dyris* Maire and *Ophrys fusca* Link. The genus *Ophrys* belongs to the family Orchidaceae Adans (1763). Orchidaceae are a morphologically diverse and widespread family of monocots, placed within the order Asparagales, an order comprising plants with inferior ovaries, simultaneous microsporogenesis and the presence of sepal nectaries. After Asteraceae, Orchidaceae are the second-largest family of flowering plants, and one of the most recent families undergoing a major evolutionary and dynamic radiation (Bateman et al., 2003, Chase et al., 2003). Estimates of the number of species vary from 17500 to 35000 (Swarts and Dixon, 2009, Mabberley, 2017), with the Plant List including 27801 accepted species names (2013). Orchids are collectively distributed over most of the ecosystems of the world, excluding the driest deserts and Antarctica, being particularly numerous in wet tropics (Fay and Chase, 2009). Orchidaceae are divided into five subfamilies, sensu Pridgeon (Pridgeon et al., 2001): i) subfamily Apostasioideae, which include two genera, considered to be the most primitive orchids (e.g. *Apostasia* Blume); ii) subfamily Cypripedioideae, composed of five genera (e.g. *Cypripedium* L.); iii) subfamily Orchidoideae, containing the orchids with a single, erect, basitonic fertile anther – monandrous (e.g. *Ophrys*); iv) subfamily Vanilloideae (e.g. *Vanilla* Plum. ex Mill.); and subfamily Epidendroideae, the largest subfamily, embracing a higher number of genus (e.g. *Dendrobium* Sw., *Phalaenopsis* Blume, *Cymbidium* Sw.).

First described by Carl von Linnaeus (1753), *Ophrys* belongs to subtribe Orchidinae, tribe Orchideae, subfamily Orchidoideae. Based on morphological characters (Bernardos et al., 2005) and on molecular methods (Pridgeon et al., 1997, Soliva et al., 2001, Bernardos et al., 2005), it is regarded as a monophyletic group. The strong support for monophyly of the genus is in contrast to its species relationships, which are poorly resolved due to high morphological divergence, potential for rapid speciation or paraphyly events (Schlüter et al., 2011). There is also a lack of agreement about the number of species in the genus, with the number of species recognised between 19 (Pederson and Faurholdt, 2007) and > 250 (e.g. Delforge 2002). A lack of clear species boundaries has been attributed to interspecific hybridisation and introgression (Devey et al., 2008). *Ophrys* occurs mostly around the

Mediterranean Basin, its distribution also including parts of North Africa (Morocco, Algeria, Libya and Tunisia), Cyprus, the Middle East, the Caucasus and Anatolia and regions of the Near East (Caspian Sea and Persian Gulf) (Soliva et al., 2001, Pederson and Faurholdt, 2007).

All species of *Ophrys* are perennial herbs with a rootstock consisting of two, occasionally three, tubers. In each autumn a new tuber begins to develop, its growth being slow during the winter months, size increasing more rapidly in spring. When fully developed the new tuber replaces the previous, old tuber. Stems are glabrous with ovoid or elliptical stolons and foliaceous bracts, leaves usually held in a basal rosette. Flowers are spurless and non-rewarding, each one including a downward-pointing labellum and a hairless marking designated "speculum", that resembles the body and/or wings of an insect. Species of the *fusca* complex are referred as essentially pollinated by insects belonging to *Andrena* genus (Delforge, 2002, Stökl et al., 2005), despite *Colletes cunicularius* (Colletidae bees) being also mentioned as a pollinator (Delforge, 2002). In a few *Ophrys* species, other bee species besides the main pollinator can be occasionally engaged, these males being sporadically attracted by different sympatric *Ophrys* species and hybrid production being possible (Paulus and Gack, 1990). Different pollinators have been regionally reported across *O. fusca* range: *Andrena flavipes* in southern Spain (Delforge, 2002, Paulus and Gack, 1990), *A. nigroaenea* in Crete, Italy and France *A. cinereophila* in Cyprus and Crete or *Colletes cunicularius* in Portugal (Delforge, 2002) and southern Spain (Paulus and Gack, 1990). *Anthophora atroalba* is the reported pollinator for O*. dyris* in southern Spain. (Paulus and Gack, 1990, Delforge, 2002).

Ophrys is emerging as a model system for the study of pollinator-mediated floral evolution (Breitkopf et al., 2014,(Paulus, 2019 #262, Paulus and Gack, 1990(Stökl, 2005 #381(Vereecken, 2014 #202))). Its mode of pollination, sexual deception, has attracted a great deal of research and popular interest. Sexual deception describes the plants mimicking their pollinator's female mating signal and subsequent pollination by mating attempts – "pseudo-copulation". This mechanism has been studied in detail for *Ophrys sphegodes*, having been confirmed that *Ophrys* flowers emit a floral bouquet that strongly resembles that of virgin female insect (Schiestl and Ayasse, 2002, Schiestl et al., 1997, Schiestl et al., 1999, Schiestl et al., 2000). Male bees become familiar to the floral bouquet with time and having no gain for continuing visiting orchid flowers, pollinators are only attracted for a short period after emergence. Pollination by the above-mentioned insects rely on male bees emerging before females and before flowering, in parallel with the orchid flowering before female bee hatching. Time discrepancy between both male and female emergence and male and orchid flowering, avoids competition between the flower and the female insect for the young males. As the plants offer no reward to the insects, the reproductive success is pollinator- and pollination-limited, rather than

resource-limited, and usually low (Tremblay et al., 2005, Vandewoestijne et al., 2009). The fact that expert pollinators learn to avoid non-rewarding flowers (Dukas and Real, 1993) limits even more the occurrence of pollination events.

1.6.1. *Ophrys dyris* Maire and *Ophrys fusca* Link

Both the narrower (Delforge, 2002) and the broader (Pederson and Faurholdt, 2007) taxonomic concepts of *Ophrys* - consider the studied entities, *O. fusca* Link s.l. and *O. omegaifera* H. Fleischm. subsp. *dyris* (Maire) Del Prete [names following the taxonomy of Pederson and Faurholdt (2007)], to represent different species complexes. For a matter of simplicity, we have chosen to adopt the designations of *Ophrys fusca* and *Ophrys dyris* throughout the study.

Ophrys dyris and *O. fusca* are one pair of *Ophrys* species found in sympatry that offer opportunities for the study of hybridisation in the context of a specialised orchid-pollinator system. These two species are included in the groups *omegaifera* and *fusca* of section *Pseudophrys* Godfery (Orchidaceae), and are closely related species (Bernardos et al., 2005, Devey et al., 2008, Cotrim et al., 2016). They are found across the Iberian Peninsula and northern Africa, *O. fusca* being quite common and widespread, with a geographical range that reaches western Asia (**Figure 1**A), and *O. dyris* having a narrow and more localised distribution in this region (**Figure 1**B). Flowers of *O. fusca* are extremely variable in morphology, leading to the segregation of the species into more than ten different species (e.g. Delforge 2002), though other authors have taken a more conservative approach, recognising these two species in the broad sense, with the morphological variants being treated at lower taxonomic levels (Pederson and Faurholdt). Both species are listed in protected habitats, as components and bioindicators of habitat 6210 of Directive 92/43/CEE, under which habitats are prioritised when any listed species is numerous.

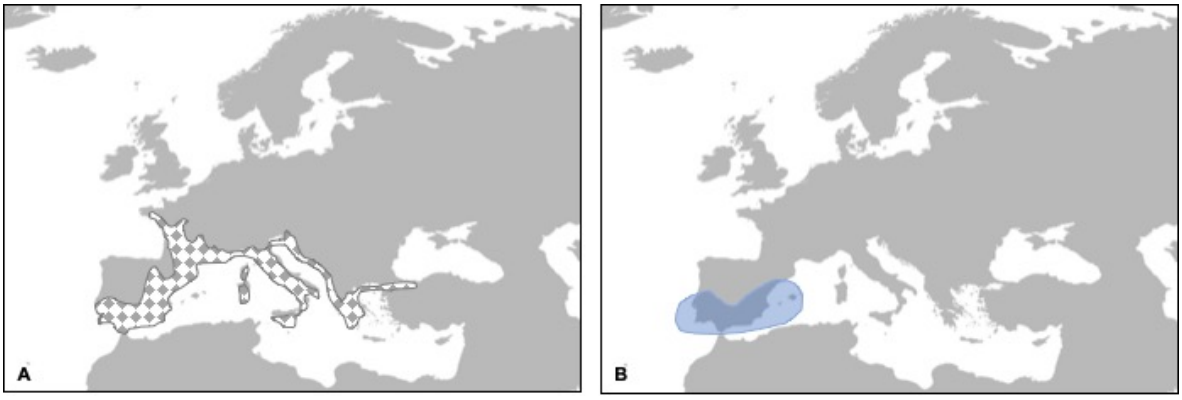


Figure 1. Distribution of the studied taxa in Europe, adapted from Pedersen & Faurholdt (2007). A. Distribution of *Ophrys fusca*. B. Distribution of *Ophrys dyris.*

Ophrys dyris and *Ophrys fusca* are found across the Iberian Peninsula and northern Africa, *O. dyris* having a narrower distribution in this region, circumscribed to western Mediterranean - occurrences reported in the southwestern part of the Iberian Peninsula, Morocco and Mallorca and *O. fusca* being more common and widespread, with a geographical range that reaches western Asia (**Figure 1**A). They can be found in sympatry and offer an opportunity for the study of hybridisation in the context of a specialised orchid-pollinator system. These species have similar morphological characters, and careful inspection is needed for field identification (**Figure 2**), particularly as intermediate individuals are known. Populations of individuals with intermediate morphology may be found, characterised either as *O. vasconica* in the western part of the range or *O. sitiaca* in the eastern part (Pederson and Faurholdt, 2007). However, these authors consider the intermediate populations a partially stabilised hybrid complex, using the name *O.* x*brigittae* H.Baumann to refer to the hybrids between *O. dyris* and *O. fusca* in the eastern part of the range. Aside from these hybrids, several "paleohybrid" species considered to result from introgression between *O. dyris* and *O. fusca* have been described for Portugal, based on morphological characters. These are *O. algarvensis* D.Tyteca, Benito & M.Walravens, *O. vasconica* (O.Danesch & E.Danesch) P.Delforge and *O. lenae* M.R.Lowe & D.Tyteca (Lowe and Tyteca, 2012). "Paleohybrids", unlike *O.* x*brigittae,* are found in populations where neither parent is present, this suggesting that hybridisation occurred in the past. Specimens classified as *O. algarvensis* were not segregated from *O. dyris* in Cotrim et al. (2016) using nuclear microsatellite Bayesian analysis; the same authors also reported that *O. dyris* is not genetically isolated from *O. fusca* (Cotrim et al., 2016). The putative hybrids above mentioned are considered to form part of a "partly stabilized hybrid complex" between the *omegaifera* and *fusca* groups by Pederson and Faurholdt (2007). The possibility of *O. dyris* itself being of hybrid origin has been suggested (Devey et al., 2008, Cotrim et al., 2009). Devey *et al.* (2008) postulated that *O. dyris* could be an intersectional hybrid between sections *Pseudophrys* and *Ophrys*, based on ITS and AFLP data. However, as only one sample was used, a general conclusion about the origin of the species could not be drawn at that time.

Figure 2. Floral structure of *Ophrys fusca* (A) and *O. dyris* (B)

In Portugal, *O. fusca* and *O. dyris* (Figure 2.) may occur in sympatry, overlapping in some limestone regions. The distribution of putative hybrids and introgressed individuals amongst these populations is intriguing. Pollinator specificity reported to the genus would not make one expect it. There are some populations where both taxa occur in sympatry and are clearly distinguished, but other populations that appear to comprise parental taxa and numerous putative interspecific hybrids, displaying a range of intermediate phenotypes. One can also find populations where only one morphological type can be found. Initial assumptions of taxonomic composition of each population were based on work by Tyteca (1997), previous work done by Cotrim (2007) and other sources (personal field observations; M. Porto and H. Cotrim *pers. comm.*). From the regions of the country where the two taxa occur, it is possible to select areas to include "pure" populations - where only one studied species is thought to occur - and populations where both species live in sympatry, the limestone massif in Central Portugal and Arrábida mountain. This region (Figure **3**) offers an opportunity to study hybridization and isolation barriers and to assess if cytotype variation accounts for isolation.

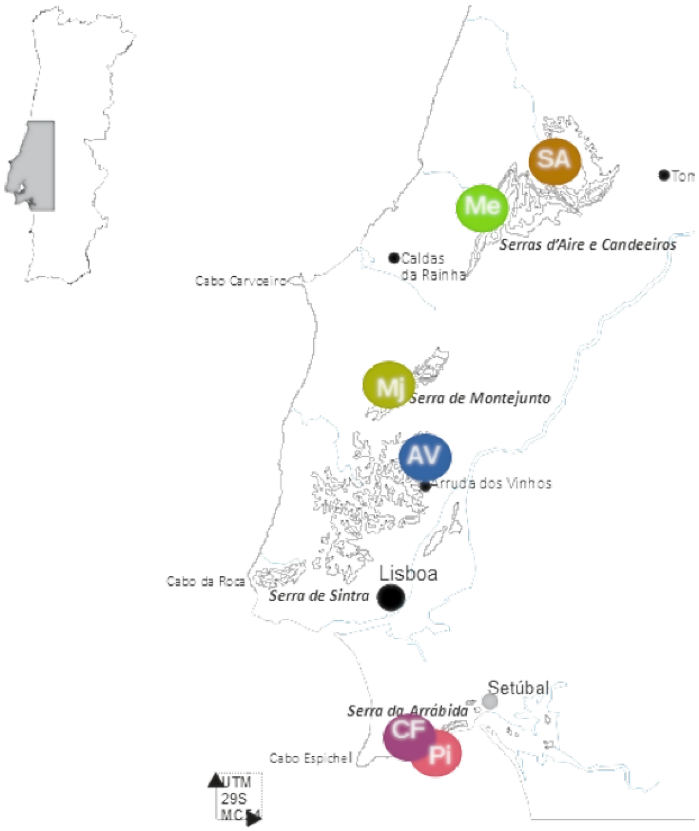


Figure 3. Geographic region of the populations studied, each named by the acronym used: AV – Arruda dos Vinhos; CF – Casalinho do Facho; Me – Mendiga; Pi – Pinheirinhos; SA – Serra de Santo António; Mj – Montejunto.

Populations of Arruda dos Vinhos (AV) and Casalinho do Facho (CF) were considered as "pure" populations for *Ophrys fusca* and Serra de Santo António (SA) plus Montejunto (Mj) populations as "pure" populations for *O. dyris*. Mendiga (Me) and Pinheirinhos (Pi) were considered as "mixed" populations of both taxa (Table 1). Pure populations were thus considered as a proxy for the species to which their individuals belong to.

At the outset of this project, populations were located and circumscribed, and a representative number of plants (usually between 50 and 60), to be included in the study were tagged by setting a physical marker in the ground, and georeferenced. Each plant was thus tagged individually and uniquely identified. These plants were considered for different datasets, and included in multiple studies, here presented in different chapters. Table 1 summarizes putative taxonomic affinity and plant numbers for each dataset, per each population.

Table 1. Location, taxonomic composition and number of plants sampled per population. Pm – number of plants sampled for morpho-anatomical data (Chapter 2); Pf – number of plants assessed regarding phenological stages and pollination success, total number of flowers, "fl", between () (Chapter 2); Pp – number of plants sampled for ploidy analysis (Chapter 3); Ps – number of plants sampled for DNA (Chapter 4).

POP. CODE	NAME, COUNTY	TAXA REPRESENTED	PM	PF (FL)	PP	PS
AV	Arranhó, Arruda dos Vinhos	*O. fusca*	38	12 (69)	35	35
CF	Casalinho do Facho, Sesimbra	*O. fusca*	60	60 (244)	5	30
Me	Mendiga, Porto de Mós	*O. fusca* + *O. dyris*	68	68 (212)	5	35
Mj	Serra de Montejunto, Alenquer	*O. dyris*	65	65 (216)	9	26
Pi	Pinheirinhos, Sesimbra	*O. fusca* + *O. dyris*	39	38 (95)	3	26
SA	Serra Sto António, Alcanena	*O. dyris*	23	17 (51)	10	13

1.7. Thesis outline and main aims

The current thesis is structured in five different chapters, including three data chapters following the same format (Introduction, Materials and Methods, Results and Discussion). Chapter 3 consists of an article that has already been published (Abreu et al., 2017), which format has been adapted to be consistent with the rest of the thesis. The chapters presenting new data or analyses in this thesis are as described below.

In **Chapter Two** the following questions are addressed:

- Are there clearly separated morphological groups within the dataset? Which morphological traits distinguish them?
- Does flowering time significantly diverge between these two entities?
- What levels of pollination are occurring in these populations? Is pollination associated with specific morphological traits or genetic groups?

Chapter Three consists of an article published in *New Journal of Botany*. The goal of this part of the work is to know whether there are different cytotypes represented amongst the plants and how does genome size vary in the populations considered. Flow cytometry and cytogenetic methods to assess genome size and cytotype diversity are described.

Chapter Four presents an analysis of genetic diversity parameters with the aim of understanding genetic structure and differentiation within and among populations, as well as assessing gene flow and detect hybrids (indirect methods) and introgression. In the current work we have used Bayesian clustering methods to study patterns of hybridization and detect individuals with admixed genotypes due to introgression between clusters, as suggested by Burgarella *et al.* (2009b), Field *et al.* (2011) and Vaha & Primmer (2006). Comparisons between datasets are also included. The main question to address is thus if these taxonomic entities hybridise and to what extent is gene flow happening.

Chapter Five comprises a comprehensive overview of the results of the project, as well as a few reflections on data collection, giving perspectives for future work.

Addressing questions about morphology, phenology and genetics allows a synthetic view of the processes occurring for these species at these sites, in the same way that multiple approaches have been valuable for the hybrid populations that have become model studies, e.g. Arnold (1994); Rieseberg (1999) or Buerkle and Rieseberg (2001). The different approaches fit together as pieces of a jigsaw, so that one completes another. As well as providing a case study of the generality of the processes of hybridisation, this study will be of particular interest to the study of reproductive isolation in *Ophrys* (Stebbins and Ferlan, 1956, Ayasse et al., 2000, Soliva and Widmer, 2003, Bateman et al., 2003, Bradshaw et al., 2010, Breitkopf et al., 2014, Paulus, 2019), among others). Finally, it is important to understand the dynamics and ecological processes driving the relations between these two species in nature.

2. MORPHOLOGY, PHENOLOGY AND REPRODUCTIVE SUCCESS

2.1 Introduction

2.1.1 Studying hybridisation

Different genomic combinations present in hybrid zones can affect normal expression of characters and give rise to new recombinants and introgressants, breaking trait coherence (Anderson and Stebbins, 1954, Yakimowski and Rieseberg, 2014). Hybrid derivates can thus have a transgressive expression – mean values higher or lower than those of either parent – creating novel morphological variation and triggering reproductive isolation and adaptation to new habitats (Arnold et al., 2012, Whitney et al., 2010, Yakimowski and Rieseberg, 2014). In the study of hybrid zones or closely related species suspected of introgression, any assessment of whether interspecific hybridisation is involved depends on characterising the variation expected in the supposed parental species (Briggs and Walters, 2016). Aside from the merging of the hybridising forms and the confusing taxonomic patterns that can result from hybridisation and gene flow in hybrid zones, there are several factors that can contribute to divergence between parental species when these occur in sympatry. Reinforcement, ecological character displacement or local adaptation are evolutionary processes that might explain divergence of species in these cases, by influencing trait evolution in sympatry (Hopkins, 2013). Divergence decreases mating or hybridisation between divergent species and might drive speciation. Two of traits that are assumed to correlate positively with hybridisation are flowering time and floral morphology (Hopkins, 2013). Their divergence can decrease mating opportunities between sympatric taxa, while their overlapping positively influences hybridisation rate (Silvertown et al., 2005). In addition, local adaptation to abiotic or biotic variation can also increase divergence between species (Hopkins, 2013).

2.1.2. Orchids: spotlighting floral morphology

The orchids are characterised by rapid production of an incredibly diverse array of floral traits (Chase et al., 2003). These floral traits are generally considered as shaped by pollinator-mediated selection (Xu et al., 2012). Floral traits include traits such as morphology, colour, phenology, and scent, since these are all involved in plant-pollinator interactions and account for the specificity between orchids and pollinators (Jersáková et al., 2007), thus being involved in the rapid diversification of these plants

(Streinzer et al., 2009, Xu et al., 2012). Ecological isolation between different species is the result of coevolution with pollinators and is characterized by the maintenance of fine morphological distances among putative species (Scopece et al., 2007).

Ophrys flowers are amongst the most charismatic and intriguing floral morphologies, since they are spectacular structures from the point of view co-evolution with pollinators. They display a global insect-like appearance, achieved by a combination of different features but most importantly a highly modified labellum which contains several different particular traits. It is on that morphologically complex petal, the labellum, and on scent traits that rest the most important characteristics concerning pollinator attraction (Bateman and Rudall, 2006, Bateman et al., 2018c, Vereecken and Francisco, 2014). Olfactory signals to the insect are secreted as pseudo-pheromones (Ayasse et al., 2011, Sedeek et al., 2014, Sedeek et al., 2016, Vereecken and Francisco, 2014, Bateman et al., 2018c), and morphological traits that give the insect both visual and tactile indications. The centre of the upper lip surface holds a structure of varying size and shape that has a smooth appearance but is in fact formed of minute trichomes with flattened bases and longitudinal cuticular striations, the mirror or speculum (Pederson and Faurholdt, 2007). The visual signs are known to be related to the speculum and the complex markings that surround it (Streinzer et al., 2009, Vignolini et al., 2012a) and the reflectivity of the different areas of the labellum (Ascensão et al., 2005, Bradshaw et al., 2010, Francisco and Ascensão, 2013, Vereecken and Francisco, 2014), while the tactile cues come from the shape and orientation of labellum trichomes (Francisco and Ascensão, 2013, Vereecken and Francisco, 2014, Bateman et al., 2018a).

2.1.3. Phenological barriers to hybridisation

The most common isolation mechanism between sympatric species is phenological, with divergence in flowering time permitting speciation (Hopkins, 2013). Flowering time divergences act as a pre-zygotic reproductive barrier between plant species, naturally limiting gene flow (Jacquemyn et al., 2011). In many hybrid zones, the occurrence of floral time divergence is very important for direct consequences of hybridization and reinforcement (Petit et al., 1999, Jacquemyn et al., 2011). Many phenological studies aim to understand the impact of climate change and other disturbing factors on phenological processes (Sparks and Carey, 1995, Hutchings et al., 2018, Robbirt et al., 2011, Robbirt et al., 2014). The impact of climate change in the disruption of phenological events has already been assessed for *O. sphegodes* by Robbirt *et al.* (2011, 2014) and Hutchings *et al.* (2018). Global change is thought to have a larger effect on pre-zygotic barriers than post-zygotic barriers (Lafon-Placette et al., 2016). There could be more hybridisation events under climate change when phenological change

occurs at different rates in species with phenological barriers to hybridisation (Vallejo-Marin and Hiscock, 2016).

Species with highly specialized pollination mechanisms are likely to be especially vulnerable to climate change due to risks of loss of synchrony with their pollinators (Hutchings et al., 2018). As regard to the species being studied, considering a changing climate scenario, reproductive success might be differently impacted in the two species, if they have different pollinators. *Ophrys* is generally considered to show pollinator-mediated reproductive isolation (Bateman and Rudall, 2006), and it can be inferred that such disruptive effect might cause that they occasionally will be visited by different pollinators.

2.1.5. Pollination success as a proxy of reproductive success

Reproductive success (RS) in orchids, particularly in sexual-deceptive such as *Ophrys*, tends to be extremely low (Neiland and Wilcock, 1998, Ayasse et al., 2000, Vandewoestijne et al., 2009) – recent average values for seed set (measured as the proportion of inflated capsules) in *O. fuciflora* were 4.26 $\pm$ 0.71%, in *O. insectifera* 13.79 $\pm$ 1.31% and 8.92 $\pm$2.09 in *O. sphegodes* (Vandewoestijne et al., 2009). Although from a different genus, percentages of 14-16% were registered for natural fruit set in *Anacamptis papilionacea* (Pellegrino et al., 2010).

Some morphological traits have been reported to influence reproductive success (Vandewoestijne et al., 2009). In *Ophrys*, plant traits such as plant height, inflorescence size, flower position, nearest neighbour distance, patch geography and population density affect RS (Vandewoestijne et al., 2009).

In this study, pollination success was considered as a proxy of both female (1) and male (2) reproductive success. Seed set has been checked as the number of filled capsules per plant (1) and the removal of pollinia as a way of assessing the visit from the pollinator (2). In nectarless species such as in *Ophrys* species, pollinia removal and receipt occurs mostly in the same visit, being almost coincident in time (Jersakova and Kindlmann, 2004). Deceptive orchids are known to have lower reproductive success than the rewarding ones (Neiland and Wilcock, 1998, Jersakova and Kindlmann, 2004, Kindlmann and Jersáková, 2006), and to flower earlier (Kindlmann and Jersáková, 2006). In a study comparing rewarding and non-rewarding species, Jersáková and Kindlmann (2004) found approximately half as many pollinator visits were made to rewardless flowers.

Some morphological traits have been reported to influence positively reproductive success in *Ophrys* (Vandewoestijne et al., 2009) such as plant height, inflorescence size or flower position. Other traits

that also have a positive influence in RS are the nearest neighbour distance, patch geography and population distance (Vandewoestijne et al., 2009).

2.1.6. Aims and scope of the chapter

The main goal of the present study is to make a morphological characterisation of the plants of *Ophrys dyris* Maire and *O. fusca* Link that are found in selected populations in central Portugal, and to document the phenology and reproductive success for these plants in different contexts. As described in Chapter 1, these plants have similar morphologies and may occur in sympatry. For the purpose of morphological characterisation, mixed populations (populations with all individuals clearly assigned to one of the taxa, and with both taxa present), intermediate populations (populations mostly with plants presenting intermediate morphologies) and pure populations (populations where only one morphological type occurs) were sampled.

In order to have a morphological portrait of these two species of orchids in central Portugal, floral characters were considered most important for characterization. They are the most useful source of morphological information, since vegetative traits are highly invariant throughout *Ophrys* genus (Pers. Obs., Stern et al., 2014, Stern, 2008). Furthermore, high intraspecific variation of some floral traits exists, as is common within *Ophrys*, and has already been reported by other authors (Ascensão et al., 2005, Paulus, 2006). Finally, as indicated above, floral traits contribute to pollinator specificity in *Ophrys,* so any study of hybridisation would be usefully focussed on these traits.

The two target species have been observed to flower in approximately in the same period. However, there has been no formal assessment of the different phenological stages, so documentation of the flowering period in each population according to morphology is carried out here. In this study, pollination success will be considered as a proxy of both female (1) and male (2) reproductive success. Seed set has been checked as the number of filled capsules per plant (1) and the removal of pollinia as a way of assessing the visit from the pollinator (2).

Main aims of this chapter are:

(1) Analyse morphological variation within and between populations and morphotypes;

(2) Understand if there are clearly separated morphological groups

(3) Which morphological traits allow us to better differentiate them.

(4) Date the flowering period of the target plants and check for differences between morphotypes

(5) Investigate whether there are phenological barriers or opportunity for geneflow between morphotypes.

(6) Record pollination success of target plants, in terms of capsule filling (female reproductive success), as a proxy for reproductive success at the level of populations and morphotypes

(6) By considering the populations of the target plants, and whether they are mixed or comprise predominantly one morphotype, to ask whether:

(i) Population-level flowering behaviour in mixed populations differs from population-level flowering in predominantly "pure" populations
(ii) Capsule filling is different between populations and morphotypes

(7) Where the field schedule and data permit, to compare flowering and reproductive success between years

2.2. Materials and Methods

2.2.1. Plant sampling

Natural populations of *Ophrys fusca* and *O. dyris* in Central Portugal were sampled and monitored (Figure 4) between March and June, from 2011 to 2015. Initial choice of sampling sites was based on the purpose of comparing populations where different situations of the two taxa could be studied: populations with individuals clearly assigned to both taxa (A), populations with a range of intermediate morphologies (B) and populations where only one morphological type occurs (C). Data were collected from a total of six populations, locations being shown in Figure 4. *Arruda dos*

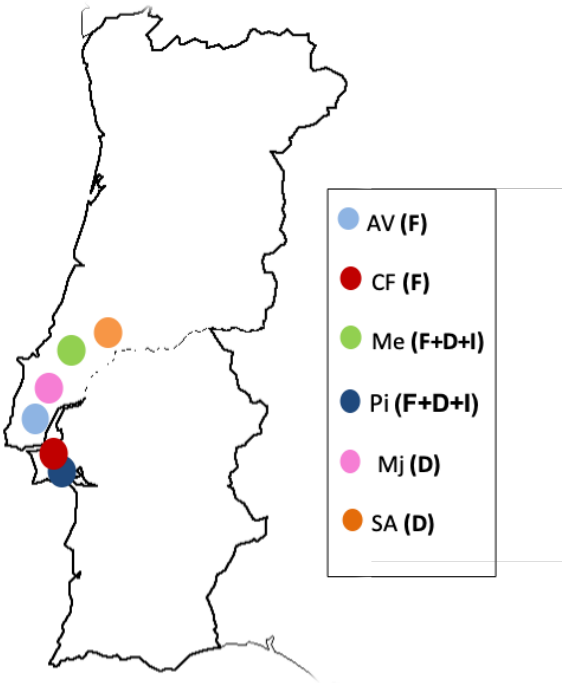


Figure 4. Geographic location of populations studied, named by the acronym used. Expected presence of the studied taxa, according to previous field observations (F – *Ophrys fusca*; D – *O. dyris*; I – Intermediate morphotypes): SA – D; Me – D+I+F; Mj – D; AV – F; CF – F; Pi – F+D+I

Vinhos (coded AV throughout the study) and *Casalinho do Facho* (coded CF), in Arruda dos Vinhos and Sesimbra municipality, were initially assumed as populations exclusively of *O. fusca*. *Montejunto* (Mj) and *Serra de Santo António* (SA), in Cadaval and Alcanena municipalities, considered as populations were only *O. dyris* occurs. In both *Pinheirinhos* (Pi) and *Mendiga* (Me) we can find individuals assigned to each of the taxa, and also a wide range of intermediate morphotypes. Table 2 summarizes the sampling in each population. A total of 342 plants were tagged across the six populations. Morpho-anatomical traits were assessed in 272 individual plants, phenological stage and capsule filling in 260. For physical identification of each plant in the terrain, nails covered by coloured anchors where used, numbered with alcohol marking pen, and bearing a small red ribbon on the top. Location of plants was georeferenced and associated to a GPS waypoint. A minimum of 20 was set in the beginning, for the number of individual plants per population to be measured and followed throughout the study (please see Table 2 for the number of plants assessed). Morphometric data have been collected from 2011 onwards. In the two latest field seasons (2014 and 2015), plants were sampled mainly for phenological data, ploidy assessment and chromosome counts; morphometric characters were only assessed for plants which had not been previously measured for a particular character or to complete the sampling.

The number of plants assessed in each population ranges between 23 (SA) and 68 (Me) (Table 2), varying according to population density and geographic area initially circumscribed. In each population, sampled plants represented the morphological variation in the population; in populations where only one morphotype was thought to exist, the first criterion for determining sampling was the density of plants, also considering their clonal ability; in those where more than one morphotype is present, plants were sampled in equal proportions covering the range of morphological variation displayed, regardless of the taxonomic affiliation to one of the studied taxa.

The occurrence of several different intermediate morphotypes made taxonomic classification of many of these plants difficult. This is why I decided to adopt a first “blind analysis” regarding taxonomic classification of each plant, considering populations as geographical units, and plants regardless of the taxon. In a first approach, morphological variation within populations and between different populations was assessed (analysis by population). In a second approach we assigned plants to morphotypes: *dyris* and *fusca* for the plants resembling the respective species and “intermediates” for the plants with intermediate morphotypes between them.

Table 2. Designation code, location, taxonomic composition and number of plants sampled per population. Pm – number of plants sampled for the morpho-anatomical dataset; number of plants (Pf) and total number of flowers (Fp) assessed regarding flowering stages and pollination success.

POP. CODE	NAME, COUNTY	TAXA REPRESENTED	PM	PF	FP
AV	Arranhó, Arruda dos Vinhos	*O. fusca*	38	12	69
CF	Casalinho do Facho, Sesimbra	*O. fusca*	60	60	244
Me	Mendiga, Porto de Mós	*O. fusca* + *O. dyris*	68	68	212
Mj	Serra de Montejunto, Alenquer	*O. dyris*	65	65	216
Pi	Pinheirinhos, Sesimbra	*O. fusca* + *O. dyris*	39	38	95
SA	Serra Sto António, Alcanena	*O. dyris*	23	17	51

2.2.2. Sampling strategy

In each flowering season (March-beginning of June), field measurements were performed in reproducing individuals previously tagged and geolocated. In-situ measurements were adopted with the purpose of minimising disturbance to the plants and to avoid potential errors occurring when measuring excised structures, such as the shrinkage of desiccation-prone structures (Bateman and Rudall, 2006). Such a strategy also allows relevant observations of the habitat and ecological context during sampling (Bateman and Rudall, 2011). Phenological stages were assessed during the same field visits. In the sense that the proportion of reproducing plants per population varies significantly throughout the years, and also considering the individual variation in the number of flowers, each plant developmental stage (Table 3) was also taken into account in the phenological record for each season.

Regarding morpho-anatomical measurements, 5 (2011) or 32 (between 2012 and 2015) traits were assessed, for 272 individual plants (Table 4). From the dataset of 32 traits adopted in the season of 2012, six colour traits were later excluded from further data sampling and analysis. Collection of these data proved to be too time consuming in the field, increasing the difficulty in monitoring all the populations during the flowering period and adding to a great observer-error associated. The exclusion of colour traits left a set of 26 morpho-anatomical characters. The initial set of five characters included stem length, number of flowers, flower position, lip width and lip length. Some environmental variables were also assessed for each plant (altitude, number of neighbour plants in 2 m2, presence of other species of Ophrys, distance to the nearest plant with the same morphotype,

distance to the nearest tagged plant), although these were not considered for subsequent analysis due to inconsistencies in the measurement procedures adopted in the different years.

The characters list which included the first 5 initially suggested by Cotrim (pers. comm.) was further extended, mainly based on that of Devey (2007), adapted to the identification of these particular taxa, also including some suggestions by Professor Richard Bateman (pers. comm.). The final selection also took into account the usefulness of the characters in distinguishing the studied entities, the stability of each character state, and the feasibility of repeated observations. The set of morpho-anatomical traits comprises 15 quantitative characters, three presence/absence and eight qualitative characters. The first group includes 13 continuous variables, measured using a digital vernier-caliper (+- 0.01 mm), and two discrete variables, number of flowers (3) and flower position (4), this referring to the sequential order of each flower in the inflorescence, from the base (expressed as "1") to the top (with a maximum of "11"). From the 15 quantitative traits, three refer to the plant (stem length, 1; inflorescence length, 2; number of flowers, 3) and 12 to each flower. Out of these 12 characters, five are concerned to the labellum - lip length and width (11, 12), speculum length and width (18, 19), extent of indentation at the base of the labellum (9); two (measurements of length and width) to each one of the following structures: lateral sepals (5, 6), gymnostemium (21, 22) and stigmatic surface (23, 24); one treating flower position (4). Presence/ absence traits designate the presence of marbled appearance of the speculum (17), presence of a basal groove (8) and the presence of trichomes surrounding the stigmatic cavity at the base of the labellum (25). Character number 8, "basal groove", refers to the longitudinal furrow in the base of the lip, a trait used for traditional morphological delimitation of these species, identifying *O. fusca*. "Depth of indentation" (9) means the extent (mm) of that furrow, measured as the hypotenuse of the triangle formed by its distal and lateral extremities. For the qualitative characters, character states were empirically estimated within the pre-defined range of variation. They reproduce prominence, extent, position or orientation of a given structure in the flower. The set of qualitative characters is composed by eight traits that refer to the labellum and the stigmatic cavity: the type of delimitation of the "W-shape" that surrounds the speculum (not visible/ visible not sharp/ sharp) – character number 20; the relative extent of the trichomes that cover the central and lateral lobes of the labellum, either observed in a vertical line or at forty five degrees to the vertical (very short/ short/ long) - characters 13 to 16; the extent of the pilosity in the base of the labellum (none/ very short/ short/ long) – character 26; the degree of curvature of the labellum viewed transversely from base – 7 and the concavity of the mid lobe - 10. The curvature of the labellum of these plants can be flat or set at 90° to the stigmatic cavity, or it can have a knee-like bend at the base, with side lobes abruptly downcurved. In what concerns the curvature of the mid lobe (10),

it ranges from slightly concave and shaped upwards, as can be rarely found in *O. fusca*, to having the central and side lobes strongly recurved inwards.

With the aim of recording the flowering period of these plants, a total of 887 flowers from 260 different plants of the same sampling set were also assessed for phenological stages, data being collected between 2011 and 2015. Overall, this dataset included 12 plants (69 flowers) from AV population, 60 plants (244 flowers) from CF, 60 plants (212 flowers) from Me, 65 plants (216 flowers) from Mj, 38 plants (95 flowers) from Pi and 17 plants (51 flowers) from SA (Table 2). Plants were monitored regarding lifestages - "non sprout", "sprout not reproducing" or "reproducing", the number of flowers being registered in this case. To each flower was produced a different record, coded by its position in the inflorescence (numbered from bottom to top) and its phenological stage was recorded according to the following code: 1 - "flower bud" (FLB); 2 - "recently flowering" (RFL); 3 - "fully open" (FOP); 4 - "senescent flower" (FLS) (**Table 3**). FLB corresponds to a bud stage, ranging from early buds until one day before the anthesis, when a small opening can already be spotted in the bud; RLS classifies flowers at the beginning of anthesis, between the previous stage and the time when the labellum achieves its final position; FOP accounts for fully open flowers, after the labellum had already reached the final alignment; FLS refers to flowers from the moment when desiccation of labellum structures starts to be visible onwards.

Table 3. Plant lifestages and flower developmental stages considered.

LIFEST		PHEN. STAGE
Non sprout		
Sprout not reproducing		
Sprout reproducing	Flower bud (FLB)	
	Recently flowering (RFL)	
	Fully open (FOP)	
	Flower senescent/ capsule (FLS/CPS)	

To estimate reproductive success, capsule filling and pollinaria removal were considered. As a part of the field inspection of each plant, along with the recording of the phenological stage, capsule filling and pollinaria removal were checked in each flower. Individual reproductive success was assessed as the number of inflated capsules per inflorescence.

Data collection in the field started with field sheets (paper copies) designed in preparation of 2011 field season, data being digitised afterwards. In 2012, from the original tables, a database was built using MEMENTO 1.9.3.2 software for Android, with which data were directly registered in each field visit, by the use of a tablet device in the field. Each category of data was assigned to a different menu: "Sites" (population location information)/ "Plants"/ "Flowers"/ "Observations" (for the phenological recordings), plus a different menu for the field year and the characters that vary yearly - "Season" (Figure 5). For each tagged specimen, at least one photo was taken of the whole plant and three photographs of each of the flowers, one facing upwards towards the labellum and one from each of side. Images were captured using a digital camera (NikonD90 + AF Nikkor 28-80 mm; f/ 3.3-5.6G), and the tablet device camera for the whole plant photograph from 2012 onwards, as the database created included image capture and editing.

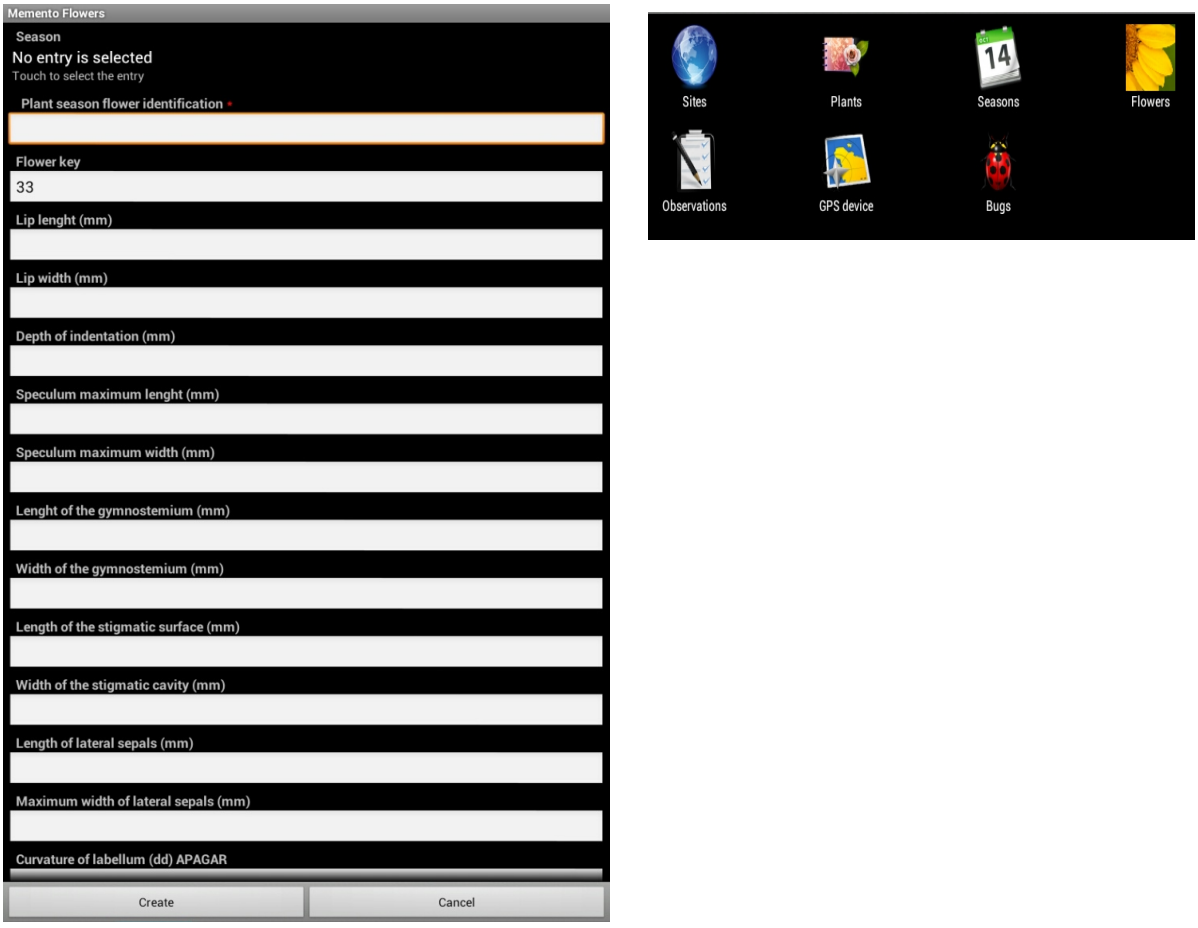


Figure 5 Some of the screens of the digital database used to register information in the field. A. First input screen from the menu Flowers. B. Main menu, including the submenus Sites, Plants, Seasons, Flowers, Observations, GPS device and Bugs.

Table 4. Morpho-anatomical floral characters included in the analysis. Different character types (T) were considered: binary (B), continuous (C) and qualitative (Q). Units of measurement and its degree of accuracy (continuous characters) or character states (qualitative and binary characters) are provided. Below the states of qualitative characters, the codes used in the quantitative approach/analysis are displayed. Flower position indicates the position of the flower for which measurements were considered, in the case of plants with more than one flower assessed and/or measured in several years.

Type	Nr	Character	Units of measurement/ Range of character states
		Stem and Inflorescence	
C	1	Stem height*	0.01 mm
C	2	Inflorescence length*	0.01 mm
Q	3	Number of flowers*	1-11
Q	4	Flower position	1-11
		Lateral sepals and petals	
C	5	Maximum length of lateral sepals	0.01 mm
C	6	Maximum width of lateral sepals	0.01 mm
		Labellum	
Q	7	Degree of longitudinal curvature of labellum viewed transversely from base	Slightly concave-flat-moderately recurved- highly rec.-strongly rec. (1-2-3-4-5)
B	8	Presence of basal groove	Yes/No 1/0
C	9	Depth of indentation	0.01 mm
Q	10	Concavity of the mid lobe	Flat-gently convex-mod convex-strongly conv-lat. lobes strongly rec. inwards 1-2-3-4-5
C	11	Maximum labellum length	0.01 mm
C	12	Maximum labellum width	0.01 mm
Q	13	Pilosity of the central lobe 2mm in a vertical line inside the bottom margin	Very short-short-long 1-2-3
Q	14	Pilosity of the central lobe 2mm inside the margin, measured at 45º to the vertical	Very short-short-long 1-2-3
		Lateral labellum lobes	
Q	15	Pilosity of the lateral lobes 1 mm inside the bottom margin measured in a vertical line	Very short-short-long 1-2-3
Q	16	Pilosity of the lateral lobes 1 mm inside the bottom margin measured at 45º to the vertical	Very short-short-long 1-2-3
		Speculum	
B	17	Presence of marbled appearance	Yes/No 1/0

Type	Nr	Character	Units of measurement/ Range of character states
C	18	Maximum length	0.01 mm
C	19	Maximum width	0.01 mm
Q	20	W-shape delimitation	Not visible-not sharp-sharp 1-2-3
		Gymnostemnium and stigmatic cavity	
C	21	Length of the gymnostemium	0.01 mm
C	22	Width of the gymnostemium	0.01 mm
C	23	Length of the stigmatic surface	0.01 mm
C	24	Width of the stigmatic cavity	0.01 mm
B	25	Presence of trichomes surrounding the stigmatic cavity, base of the labellum	Yes/No 1/0
Q	26	Extent of the pilosity in the base of the labellum	None-very short-short-long 0-1-2-3-

* In cases where the same plant was assessed in more than one field season, mean values were used.

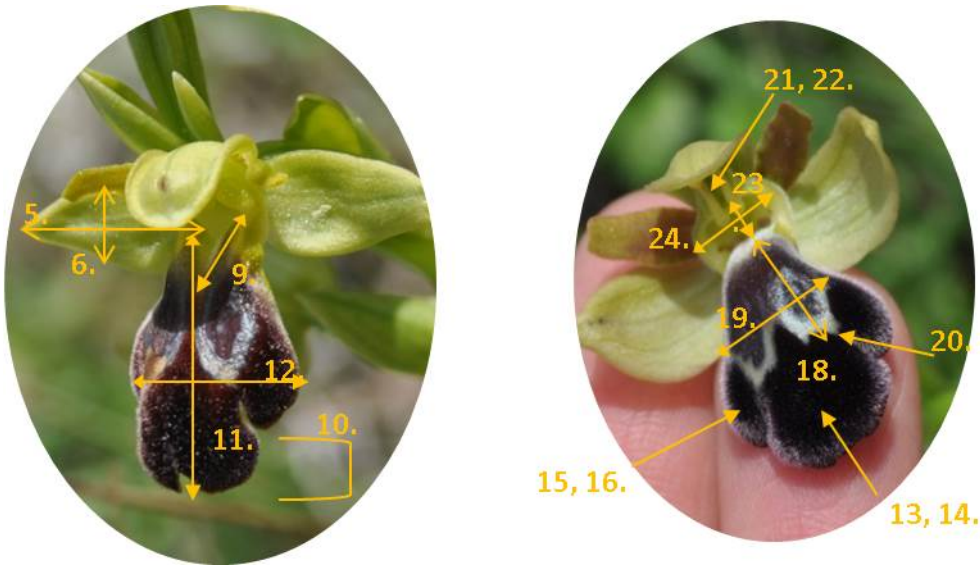


Figure 6. Floral structure of *Ophrys fusca* and *O. dyris*, illustrating some of the macromorphological characters used for morphological analysis. 5 – Maximum length of lateral sepals; 6 – Maximum width of lateral sepals; 9 – Depth of indentation; 10 - Concavity of the mid lobe; 11 – Maximum labellum length; 12 – Maximum labellum length width; 13, 14 – Pilosity of the central lobe measured 2mm inside the bottom margin (13) in a vertical line or (14) at fourty-five degrees to a vertical line; 15, 16 - Pilosity of the lateral lobe measured 1mm inside the bottom margin (15) in a vertical line or (16) at fourty-five degrees to a vertical line; 18 – Speculum maximum length; 19 – Speculum maximum width; 20-W-shape delimitation; 21, 22 – Length and width of the gymnostemium; 23 – Length of the stigmatic surface; 24 – Width of the stigmatic surface.

2.2.3. Data Analysis

2.2.3.1. Morphological data

Data extracted from the digital database were summarized in an Excel v. 14.0 spreadsheet. For characters measured in more than one field season for the same plant (e.g. stem height, number of flowers), mean values were used. 26 characters were left (Table 4), after the exclusion of the colour characters and "length of the basal bract", due to inconsistencies in the measurements and missing data for a large number of samples. For the numerical analysis, presence/absence characters (8, 17, 25) were coded as binary (1/0) and discrete qualitative characters (7, 10, 13-16, 20, 26) transformed into a numbered scale, from one to three or one to five.

Data analysis was initially performed independently of taxonomic affiliation of each plant; the analysis was population-based, considering plants of all morphotypes equally. To check for the presence of different groups within the complete morphological dataset, a PCoA was performed in Multivariate Statistical Package (MSVP) version 3.22 (Kovach, 2013), using Gower similarity coefficient (Gower, 1971) as the algorithm for the similarity matrix. This approach can cope with a variety of different data types, including dichotomous and qualitative characters without any reprogramming. The matrix obtained was subsequently used to calculate principal coordinates, effective for simultaneously analysing heterogeneous sets of morphological characters, with the advantage of accommodating missing values. First and second axes were plotted. (**Figure 7**) Data consisted of the complete set of 26 characters used for 62 plants. To understand which traits best discriminate between the groups found, stepwise Discriminant Analysis was run in XLSTAT add-in for Excel, version 2018.2 (Addinsoft, 2018) after each PCoA, samples being grouped according to PCoA ordination. At a second stage, all the plants included in the study were classified by morphotype. For simplicity, despite the wide spectrum of morphological variability, all the morphological variations found were assigned to three morphotypes: ***dyris*** (1) for plants with a morphology closer to *Ophrys dyris,* ***fusca*** (3) for those attributed to *O. fusca*, and **intermediate** (2) for those presenting an intermediate or mixed morphology between the preceding groups. For taxon assignment segregation, the description of characters of Pederson (2007) was used. Re-classification of plants was based on careful inspection of photographs and field notes relating to each sampled specimen.

The same type of exploratory analysis was then performed by morphotype for 82 plants with information for the 26-character set, missing values being replaced with average values. A simple PCA was run on all characters using function *dudi.pca* in the R package ade4 (Bougeard and Dray, 2018) (please see **Figure 9**, pg. 47). A general linear mixed model (GLMM) approach with all the characters was attempted, but the model failed to resolve because there were too many factors and too few

observations. Also, the number of missing values was very high, influencing any possible outcomes. A GLMM was hence run for a subset of the data, in R environment, considering the four variables measured from the most plants: lip length, lip width, stem length and number of flowers. 226 plants were included in the analysis. Normality for each variable was checked by plotting its distribution against the normal curve. Resulting histograms showed small deviations from normality, namely for the variables stem length and number of flowers, but this is not a reason for concern as it could be compensated by GLMM. Therefore, a Gaussian distribution was used. Correlation was initially tested using two-tailed Pearson's r (significant at the 0.01 level), high correlation (0.0305) being found between stem length and number of flowers (**Table 6**), thus indicating the likelihood of collinearity. To confirm this, collinearity was checked using r functions kappa adapted by Becker (Becker, 2011) and VIF (variance inflation factor) (Belsley, 1991), standard measures of collinearity in regression type models. Stepwise model reduction was performed both in packages cAIC4 (Saefken et al., 2014) and lmerTest (Kuznetsova et al., 2017) in R, with similar results for collinearity reduction (Kappa values of 3.32 for the first, 3.33 with the second), but the second turning out a lower value for AIC (standard measure for model fit, representing the relative amount of information lost by the model), resulting in a better model. Interactions were excluded by backward elimination, keeping two interactions between fixed factors - Stem Length:Lip Length + Lip Length:Lip Width - and the only random factor (population) retained in all the tested models (**Table 8**). Variables were standardised prior to the analysis.

As some plants appear in different datasets of the study (morphological and genetic, in this case), the same procedure, to check for the presence of groups and the relative contribution of the different values, was adopted with plants, common to the morphological and genetic datasets: a second PCoA was performed with this reduced universe of plants sharing morphological and genetic data (49). Fifteen morpho-anatomical traits, those revealed to be more informative after the first discriminant analysis, were now considered. The same statistical package (Kovach, 2013) was used, with Gower coefficient, to calculate the new matrix (**Figure 8**). To understand which variables (morphological traits) better explain the latest ordination, a second discriminant analysis was then run for this subset of data (49 samples and 15 characters), classified following the new groups from this PCoA. For the univariate analysis of morpho-anatomical characters, either per population and per morphotype, XLSTAT was used. Box-and-whisker plots were built in Excel version 14.0 software.

2.2.3.2. Phenological data

Concerning the phenological observations, one started by considering only the registers of fully open flowers (FOP), as field visits during early (FLB, RFL) and latest stages (FLS) of flowers were not regular

enough for documentation of the different phenological stages in these populations to be feasible. However, despite acknowledging this sampling bias, observations for phenological stages of early (RFL) and late flowering (FLS) were later considered to check for the possibilities of hybridization. To allow data collected in different years to be compared, civil dates for each phenological record were converted to the Julian date in the corresponding year (Wolf *et al.* (2017). Julian weeks – periods of seven days starting from day one in each year - were also used. Whenever the same plant was recorded at its flowering peak on two different occasions, this meaning two different dates being recorded, intermediate values were used. The time range of the flowering period of each morphotype (Figure 16) was plotted in XLSTAT add-in for Excel, version 2018.2 (Addinsoft, 2018). Acknowledging that it is at the FOP flowering stage that the plant has the right signals for insect attraction, I firstly plotted only flowers at this phenological phase. However, in order to include all the opportunities for hybridization, such as the possibilities of rare pollination events taking place at earlier or later stages, I subsequently plotte all the flowering phases, namely RFL, FOP and FLS. To identify opportunities for geneflow between morphotypes, I have thus assessed whether plants of different morphotypes were flowering at the same time – all flowering stages - within the same population (Figure 19) and plotted the number of flowers from different morphotypes simultaneously fully open in each population (Figure 13). Charts were produced in R, using colour palettes from RColourBrewer (Harrower and Brewer, 2003) and R package viridis (Garnier et al., 2018).

To understand if different morphotypes and populations respond differently throughout the years, the average Julian date for FOP flowers for each morphotype and population between 2011 and 2015 (Figure 20) was analysed in XLSTAT add-in for Excel, version 2018.2 (Addinsoft, 2018). Statistical analysis of flowering time used GLM approaches in R (base functions) (Table 10, 11, 12). To assess whether the inter-annual variation is causing differences in the number of plants with fully open flowers, a GLM testing for the effect of Year and Population was run (Table 8). All morphotypes were considered.

2.2.3.3. Reproductive success

Concerning the study of reproductive success, capsule data was coded as presence (filled capsule)/absence (no capsule). Individual reproductive success was calculated as the number of filled capsules over the total number of flowers per inflorescence, in the different field seasons considered, in XLSTAT, version 2018.2 (Addinsoft, 2018). Values for population and morphotype reproductive success were averaged from all the flowering plants assessed.

2.3. Results and Discussion

2.3.1 Morphometry

2.3.1.1 Multivariate analysis

The plot of the two first principal coordinates (Figure 7) shows that the distribution of individuals is not homogeneous: phenotypically, there are groups within the dataset, the two first coordinates together explaining 37 % of the total variance. The first coordinate (PCo1) accounts for most of this variance, 26%, globally separating one of the populations considered as pure of *O. dyris,* Mj, on the left side, from both of *O. fusca*, AV and CF, on the right side. SA, the other population initially assumed as pure for *O. dyris*, has its individuals in between both groups, disagreeing with what would be expectable. The distribution of Me individuals in the plot seems to corroborate field observations, of a wide range of different morphotypes, spanning a large proportion of the morphological range. The same is not observed for the other population with seeming mixed morphotypes according to field observations, Pi, where plants appear mostly closer to *O. fusc*a populations, only two plants appear closer to the individuals of *O. dyris*. Despite the inequality between the number of plants in each side of the ordination, most likely due to skewed sampling regarding the proportion of morphotypes, this may be the result of some geographic segregation between morphotypes in this particular population That doesn't happen in Me, where it would not be possible to separate areas corresponding to different morphotypes, also in a larger number than in Pi. Most within-populations variation for *O. fusca*, particularly in AV and CF, seems to be explained by the second coordinate, PCo2, with its individuals spanning the range of this axis. This is mainly influenced by vigour traits - inflorescence length, number of flowers or stem length – which increase upwards, and not so much by floral characters.

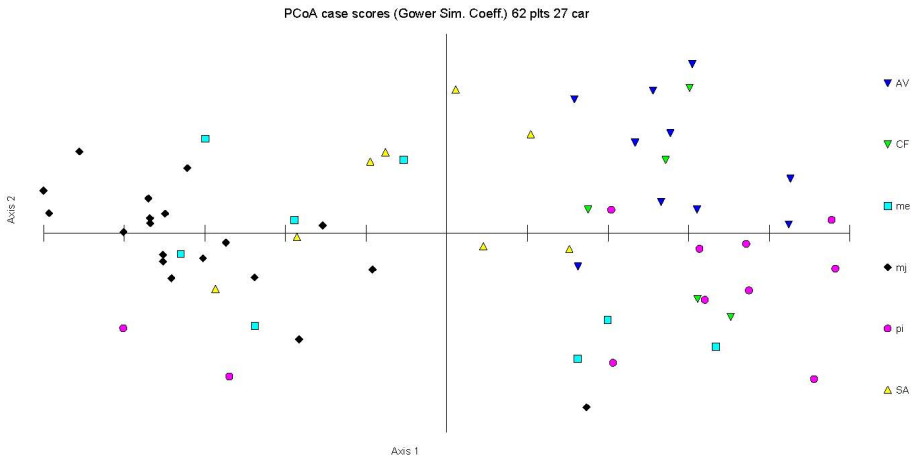


Figure 7. Principal Coordinates plots of the two first axis, following principal coordinates analysis in MSVP software using Gower General Similarity coefficient. Eigen values of 4.722 (axis 1) and 1.989 (axis 2); cumulative percentages of 26% and 38% for axis 1 and 2. PCoA performed in 62 plants and 26 characters.

Table 5. Percentages of variance accounted by each of the first ten coordinates after Principal Coordinates Analysis performed in M;SVP software.

Coordinate	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
% of variance accounted for	26.24	11.37	8.84	6.89	4.28	4.07	2.76	2.44	2.29	2.11
Cumulative % of variance	26.24	37.60	46.44	53.33	57.61	61.68	64.44	66.88	69.17	71.28

Discriminant function analysis computed on the same samples of the first PCoA (Figure 8), identified qualitative characters as the variables that better discriminate between these. The traits assessing pilosity of the lobes of the labellum – the mid lobe (characters numbered 1 and 2 in the analysis) and the lateral lobes (3, 4) as well as delimitation of the W-shape surrounding the speculum (5), a trait used in the traditional segregation of both taxa, are those that best contribute to the segregation of morphological groups, in descending relevance order.

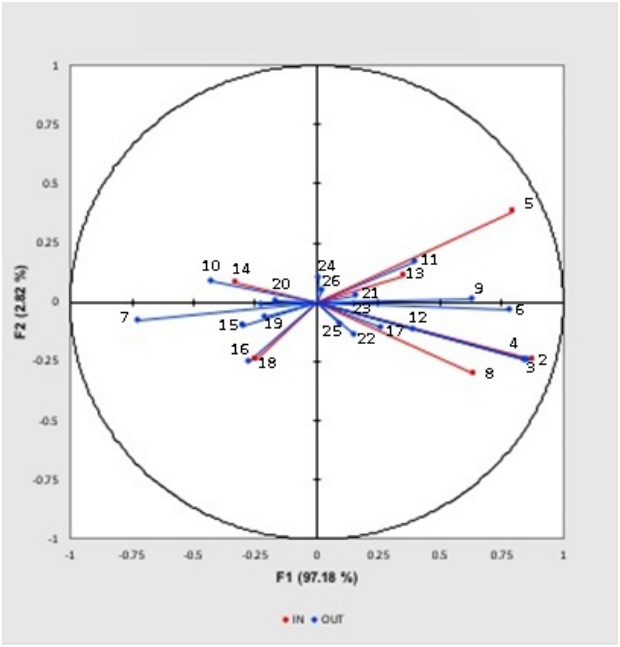


Figure 8. Discriminant function analysis performed in XLSTAT, run on the two groups identified after the PCoA. 62 plants, 26 traits, numbered by descending order of accountability for morphological groups discrimination: 1 – Pilosity of the mid lobe, 45 degrees; 2 – Pilosity of the mid lobe, vertical line; 3 - Pilosity of the lateral lobes, 45 degrees; 4 - Pilosity of the lateral lobes, vertical line; 5 - Speculum w-shape delimitation; 6 - Downcurving of the mid lobe; 7 – Depth of indentation; 8 - Curvature of labellum; 9 - Extent of pilosity in the base of labellum; 10 - Presence of basal groove; 11 – Length of stigmatic surface; 12 - Speculum brightness; 13 - Lip width; 14 – Lip length; 15 – Stem length; 16 – Speculum marbled; 17 - Width of gymnostemium; 18 – Number of flowers; 19 - Inflorescence length; 20 - Speculum maximum length; 21 - Length of lateral sepals; 22 - Maximum width of lateral sepals; 23 – Width of stigmatic cavity; 24 - Speculum maximum width; 25 - Length of gymnostemium; 26 - Presence of trichomes in the stigmatic cavity.

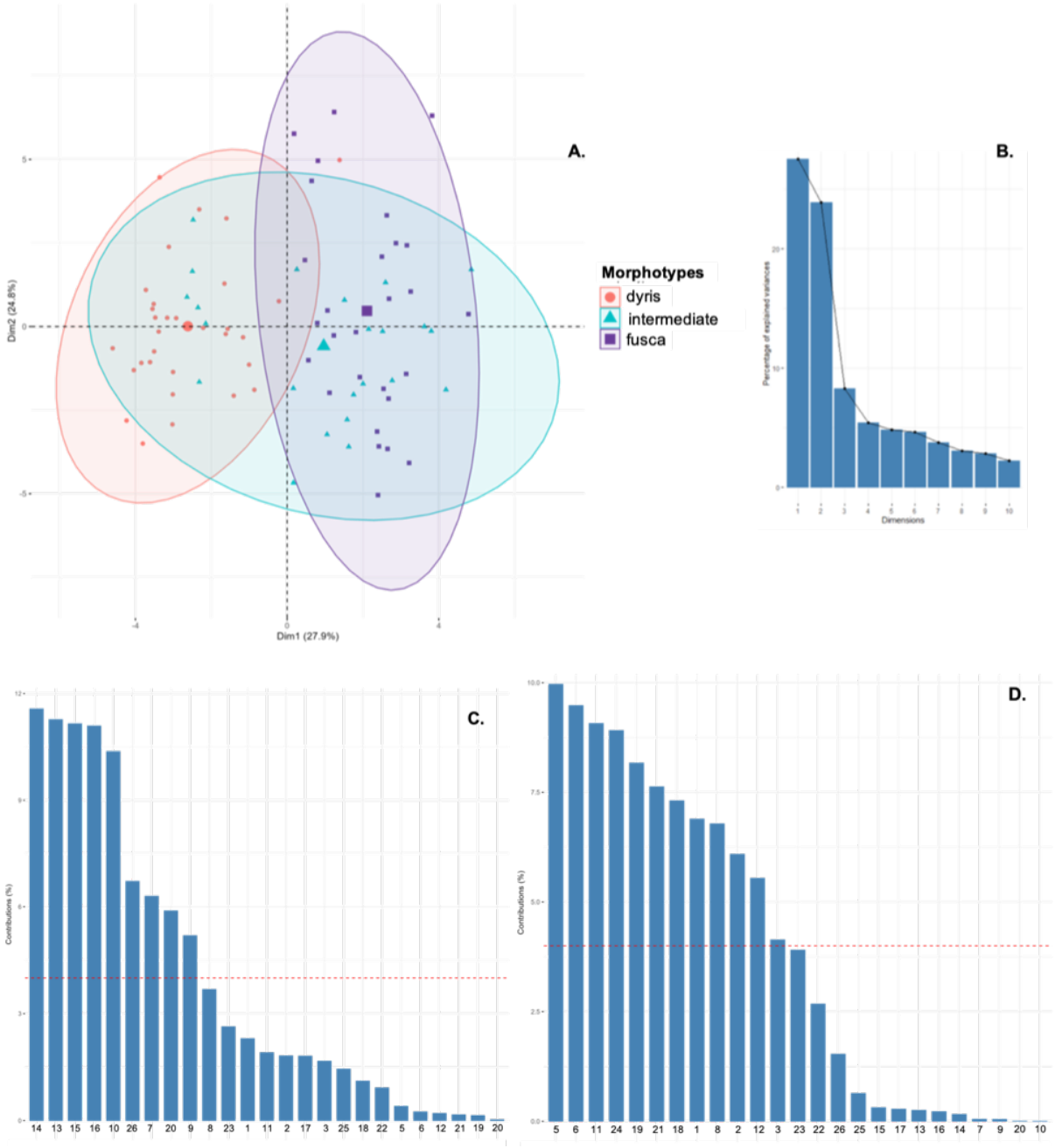


Figure 9. (A). Results of the principal components analysis of 82 plants classified by morphotype (1- dyris; 2 – intermediate; 3 - fusca), for the dataset of 26 characters. Central (mean per dimension) points shown. (B) Scree plot with the percentage of explained variance by each of the dimensions, the two first dimensions accounting for approximately 52 % of the variation. (C) and (D). Histograms of percentage contributions for the first (C) and second dimension (D) of the PCA. The Variables numbering follows Table 4.

In the ordination of plants by morphotype, different clusters can be recognised. Plants of morphotypes dyris and fusca) are well differentiated in Figure 9A, the first displayed in the left side of the plot and the second on the right side. Plants of the intermediate morphotype span in the middle area, throughout a large proportion of the morphological range. There are more plants of the intermediate morphotype overlapping the fusca cluster than those overlapping dyris cluster. Also, the intermediate

cluster spans a larger proportion of the second axis than the dyris cluster. The two first dimensions are clearly very influential, explaining approximately 52% of the variation (**Figure 9**A).

2.3.1.2 GLM approach

A mixed generalised linear model was tried with the full morphological dataset (226 plants and 26 variables), but the model failed to resolve due to the reduced number of observations for the high number of factors. The high number of missing values would also influence the results. Results of the subsequent analysis, with the four variables measured for more plants – lip length, lip width, stem length and count of flowers - are described hereafter.

Table 6. Correlation table for Pearson's two-tailed test. Values for the four variables considered. N = 226

		Morphotype	Stem length	Lip Length	Lip Width	Count Flowers
Morphotype	Pearson correlat. Sign. (2-tailed)*	1.00				
Stem length	Pearson correlat. Sign. (2-tailed)*	0.29 0.0000	1.00			
Lip Length	Pearson correlat. Sign. (2-tailed)*	0.21 0.0014	0.59 0.0000	1.00		
Lip Width	Pearson correlat. Sign. (2-tailed)*	-0.19 0.0053	0.31 0.0000	0.65 0.0000	1.00	
Count Flowers	Pearson correlat. Sign. (2-tailed)*	0.14 0.0305	0.70 0.0000	0.41 0.0000	0.23 0.0004	1.00

*Significant at the 0.01 level (2-tailed)

Pearson's correlation test revealed high correlation between the variables stem length and count of flowers (Table 6), indicating the likelihood of collinearity in the GLMM. The outcome of the collinearity check was a Kappa value for collinearity of the full model of 11.92, which signifies moderate collinearity (Kappa<10 is reasonable collinearity;<30 means moderate collinearity; 30 and above means there should be reason for concern). Variance inflation factor (VIF) values also revealed high collinearity (**Table 7**), as results above 2 represent concern and above 5 means action should be taken (Becker, 2011).

Table 7. VIF values for collinearity between variables and interactions

Predictor variables and interactions	VIF
StemLength	2.362870
LipLength	2.685479
LipWidth	2.982316
CountFlowers	2.551353
StemLength :LipLength	4.191288
StemLength : LipWidth	5.664500
LipLength : LipWidth	2.823802
StemLength : CountFlowers	3.318652
LipLength : CountFlowers	4.546272
LipWidth : CountFlowers	4.929746
StemLength : LipLength : LipWidth	5.333223
StemLength : LipLength : CountFlower	5.366976
StemLength : LipWidth : CountFlowers	5.756850
LipLength : LipWidth : CountFlowers	5.903325
StemLength : LipLength : LipWidth : CountFlowers	3.972894

After stepwise model reduction, two interactions between fixed factors were retained (stem length:lip length, lip length:lip width) and one of the predictors, count of flowers, was left behind. final GLMM model in Table 8. Results from the model revealed a highly significance (<0.001) of lip width in segregating morphotypes, also confirming there are some interactions between the fixed factors that revealed significant.

Table 8. Results of the final GLMM model for the prediction of morphotypes, selected by stepwise model reduction. Lip width has revealed to be the most significant predictor. Collinearity values added: VIF (per predictor) and kappa for the global model

	Morphotype					
Predictors	*Estimates*	*std. error*	*CI*	*P*	*VIF*	*Kappa (model)*
(Intercept)	1.94	0.26	1.43 – 2.46	**<0.001**	1.199519	3.33
StemLength	-0.06	0.05	-0.15 – 0.03	0.164	2.161371	
LipLength	0.09	0.06	-0.04 – 0.21	0.178	1.891113	
LipWidth	-0.21	0.05	-0.3 – -0.11	**<0.001**	1.150789	
StemLength*LipLength	0.11	0.03	0.04 – 0.18	**0.001**	1.237678	
LipLength*LipWidth	0.08	0.03	0.02-0.15	**0.006**	1.199519	
Random Effects						
σ^2	0.25					
$\sigma_{00\ Population}$	0.40					
ICC	0.61					
N Population	6					
Observations	226					
Marginal R^2 / Conditional R^2	0.097 / 0.652					
AIC	386.751					

2.3.1.3. Morphometric PCoA with plants included in the genetic dataset

Analysis of the morphological data from the subset of plants shared with the genetic dataset, revealed more tightly circumscribed groups than when including the whole dataset. Figure 12 displays the distribution of the 49 plants both included in morphological and genetic analysis considering fifteen diagnostic characters. The three different groups in which the plants are gathered can be clearly identified, the analysis being more robust than when considering the whole dataset (**Figure 7**). The two first coordinates together explain 51.5% of the total variance, the first coordinate (PCo1) accounting for most of this variance, 36.4%, and PCo2 for the remaining 15.1%. On the extreme left of the plot one can find most of the plants from the *O. dyris* populations, all the Mj plants considered and most of the plants from SA, except two. The right-hand side cluster includes all of the CF plants considered, together with the great majority of those from AV, representing an *O. fusca* cluster. Despite having more plants in the *dyris* cluster, Me samples are scattered throughout the three different groups, corroborating the initial assumption of a mixed population. Plants from Pi, also a mixed population, are the core of the third group, in the lower right-hand portion of the plot. Two individuals from AV population, accounting for *O. fusca*, are also placed in this group, designated hereafter as the mixed cluster. This placement of plants from AV in different clusters goes against our initial assumption of this being considered a “pure” population for that taxon.

shows the subsequent discriminant function analysis performed on the same, the 49 plants that shared the morphological and genetic dataset. Although considering three different groups, qualitative characters are the variables with more discriminating influence. Once again, the traits assessing pilosity of the lobes of the labellum are amongst those that better separate groups, but when we reduce the dataset to the plants that were sampled for DNA with information from only the diagnostic characters, other variables come into play, such as the pilosity of the stigmatic cavity (B),the downcurving of the mid lobe (C), the curvature of the labellum (D) or the marbling of the speculum (E), second to fifth by decreasing order of influence.

Figure 10 Principal Coordinates Analysis using Gower General Similarity coefficient for the similarity matrix, considering the plants of the genetic dataset (49 plants) and diagnostic characters (15), performed in MSVP software. Eigen values of 6.346 (axis 1) and 2.646 (axis 2); cumulative percentages of 36.4% and 51.5%, axis 1 and 2, respectively.

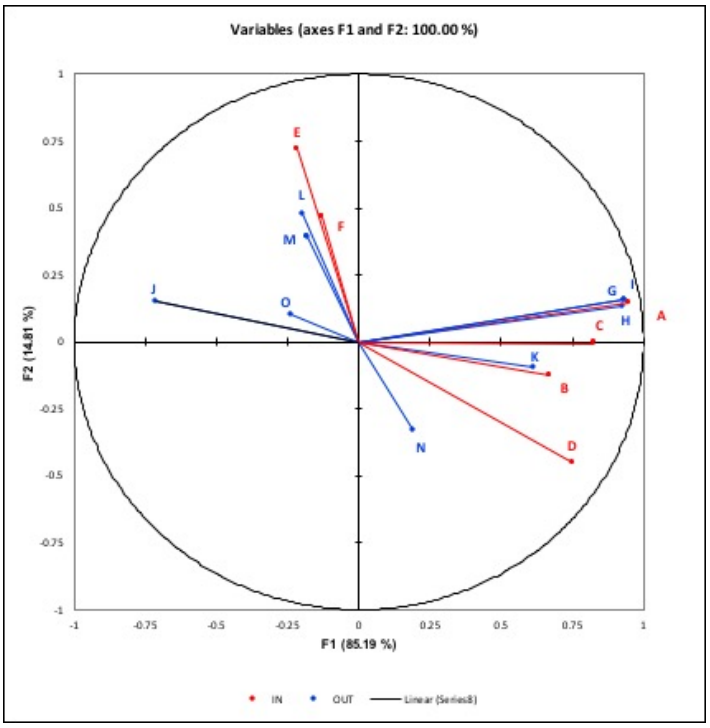


Figure 11 Discriminant function analysis performed in XLSTAT, run on the three groups identified in the PCoA with plants from both the morphological and the genetic dataset (49 plants, 15 characters). A - Pilosity of the mid lobe 45□; B - Pilosity of the stigmatic cavity; C - Downcurving of the mid lobe; D - Curvature of the labellum; E - Speculum marbled; F- Inflorescence length; G - Pilosity of the lateral lobe, vertical line; H - Pilosity of the lateral lobe, 45 □; I - Pilosity of the mid lobe, vertical line; J - Depth of indentation; K - Speculum W-shape delimitation; L - Number of flowers; M - Stem length; N - Lip width; O - Lip length. Traits A, B and C are those who account most for morphological groups discrimination.

2.3.1.4 Univariate analysis of morphological characters

Univariate distribution of morphological characters is presented from Figurre 12 to Figure 15. All the characters are displayed both by population and by morphotype. Quantitative trait expression is displayed by population in Figure 12 , by morphotype in Figure 13, characters listed in Table 4. Qualitative and binary characters distribution are presented in both Figure 14 and Figure 15. When data are presented by population, these are displayed in the following sequence: the two populations for *Ophrys fusca* (AV and CF) on the left, populations for *O. dyris* on the rigth, and mixed populations (Me and Pi) in the middle. As for morphotypes, morphotype 1 corresponds to plants attributed to *O. dyris*, morphotype 3 to plants attributed to *O. fusca*, and 2 to individuals of intermediate morphotypes.

The large stature of the plants of *Arruda dos Vinhos* (AV) is revealed in the results of the vigour characters 1 (stem length), 2 (inflorescence length) and 3 (number of flowers). From plots number 1 to 3 (), AV (first box in each plot) is the population which presents the most vigorous plants, with higher values for inflorescence length and higher number of flowers. In these three characters, absolute lower values appear in a supposedly mixed population, Pi. Traits number 1 and 2 show a similar pattern between populations, apart from the fact that inflorescence length in CF are clearly lower than in AV, and the values for Me span along a wider range. Going against what would be expected, distribution of values for inflorescence length (2) and number of flowers (3) has different patterns. Accounting for this is the lower flower density per plant in Me and SA, where flowers are sparser in the inflorescence. Results by morphotype (Figure 13) are of higher absolute values for stem (1) and inflorescence length (2) in the *fusca* morphotype, but globally higher number of flowers (3) for plants with hybrid morphotype. Both morphotypes may have 11 flowers (Table 9), but a high number of flowers can be found in more hybrid plants. *Dyris* is the morphotype with a shorter range of values for these three characters. Regarding the size of lateral sepals (4), four of the populations seem to have a similar pattern, with close median values, differently from the two populations considered for *O. dyris*, which display a large amount of variability - plants from *Montejunto* (Mj) and *Serra de Santo António* (SA). Nevertheless, absolute values measured for both these populations fall within the range given by Pederson and Faurholdt (2007). When referring to absolute values, the lowest values for both length and width are found at *Pinheirinhos* (Pi), displayed as outliers in the whisker plot. When plants are classified by morphotype (Figure 13), intermediate plants are those which display a shorter range for proportions of lateral sepals. For the two other morphotypes most of this trait expression falls on similar values between them.

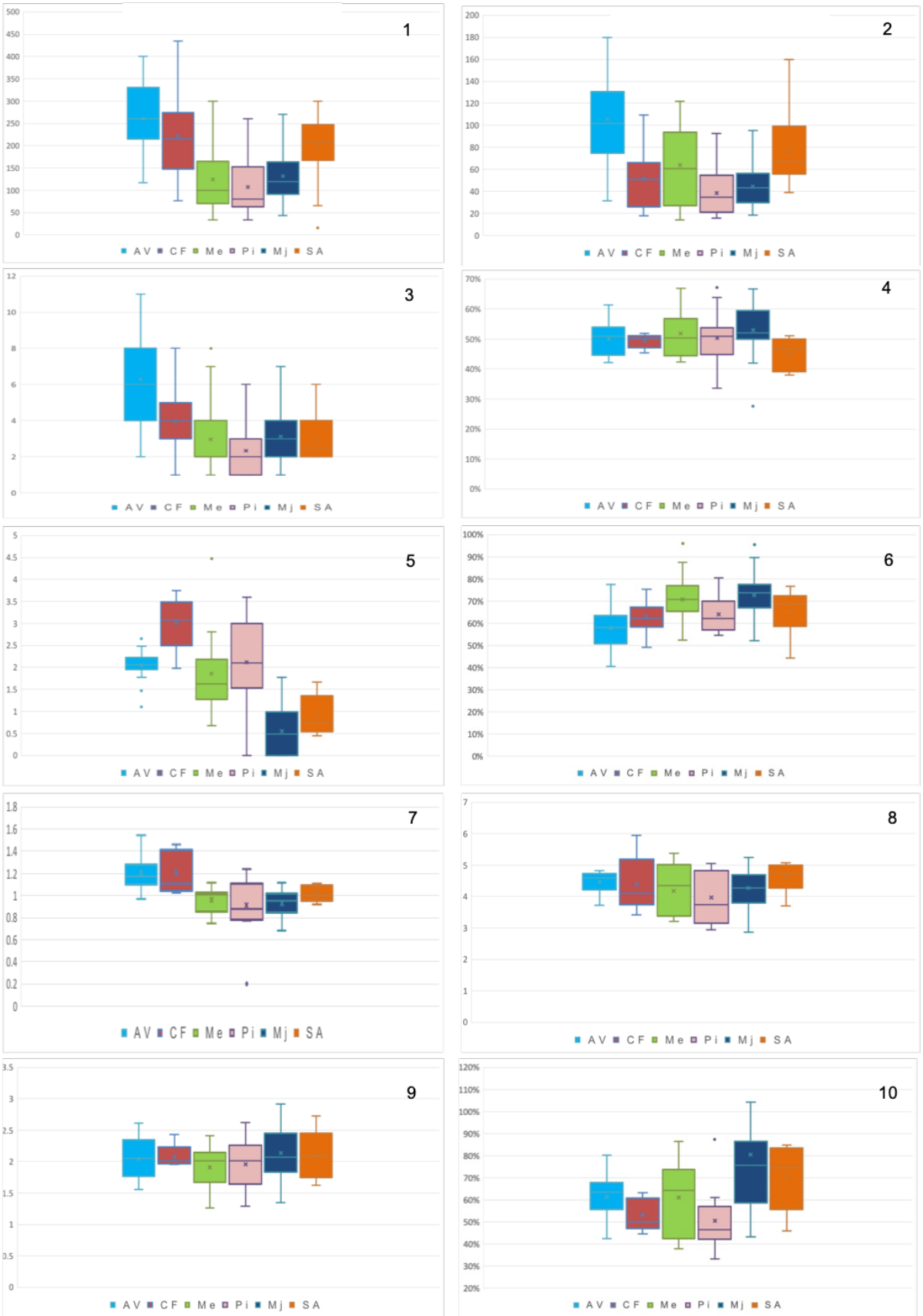


Figure 12. Comparison of the populations studied regarding the quantitative morphological traits assessed: 1 – Stem length; 2 – Inflorescence height; 3 – Number of flowers; 4 – Lateral sepals (width/length); 5– Depth of indentation (mm); 6 – Lip width/length; 7 – Speculum length/width; 8 - Gymnostemium length (mm); 9 – Gymnostemium width (mm); 10 - Stigmatic surface length/width. Central "x" are mean values. The boxes are terminated by the tweny-fifth percentile below and seventy-fifth percentile above. For population codes and taxa present in each population, please see Table 1.

The following character, depth indentation (5) - the main diagnostic character in the traditional taxonomy - displays mixed populations in the middle of the plot, with intermediate values; Mj and SA populations with values close to zero, and CF, accounting for *O. fusca* with higher values in the whole dataset. AV values have a narrow distribution range, similar to the median values of the mixed population of Pi. The same trend is observed by morphotype (Figure).

Regarding lip dimensions, values distribution in the different populations does not seem to follow any recognisable pattern (Figure 12, plot 9). Median values of plants in mixed populations are similar to those of *O. dyris* populations, and the values in the two populations accounting for *O. fusca* are more different from each other than from any of the other two groups. On the contrary, after plants having been classified by morphotype (Figure), the three ranges of values can be distinguished: intermediate plants have most of its expression for this trait at an intermediate range, *fusca* plants presenting lower values and *dyris* plants higher proportions between width and length of the lip. On what concerns speculum dimensions (width/length) (Figure 12, plot 7), *O. fusca* populations have clearly lower values than the remaining, with boxes from mixed and *O. dyris* populations displaying very similar proportions. The pattern for *fusca* is confirmed when plants are displayed by morphotype (Figure 13, plot 9), ranges of the two other morphotypes being now more clearly separated, intermediate-morphotype plants also exhibiting intermediate values. As for gymnostemium dimensions, all the populations express these traits (8 and 9) in a similar way, not being possible to distinguish ranges of values of plants from *O. fusca*, *O. dyris* and mixed populations using this character. Such assortment does not globally change by morphotype, except from the fact that morphotype *fusca* displays the lower values for gymnostemium. A different relative position between populations is found when looking at stigmatic surface measurements plot (Table 4, trait 10), the higher values having been recorded in the populations of Mj and SA (*O. dyris*). The lowest values for the proportions of the stigmatic cavity occurs at Pi. This trend is confirmed when measurements are plotted by morphotype: *dyris* morphotype express the higher values, the lowest measurements for this character belonging to the intermediate morphotype. One of the reasons to explain the short distribution range of AV values regarding some of the characters, namely 5, 8 and 10, might be due to the occurrence of presumable clonality in this population, leading to a reduced variability in the expression of some morphological traits.

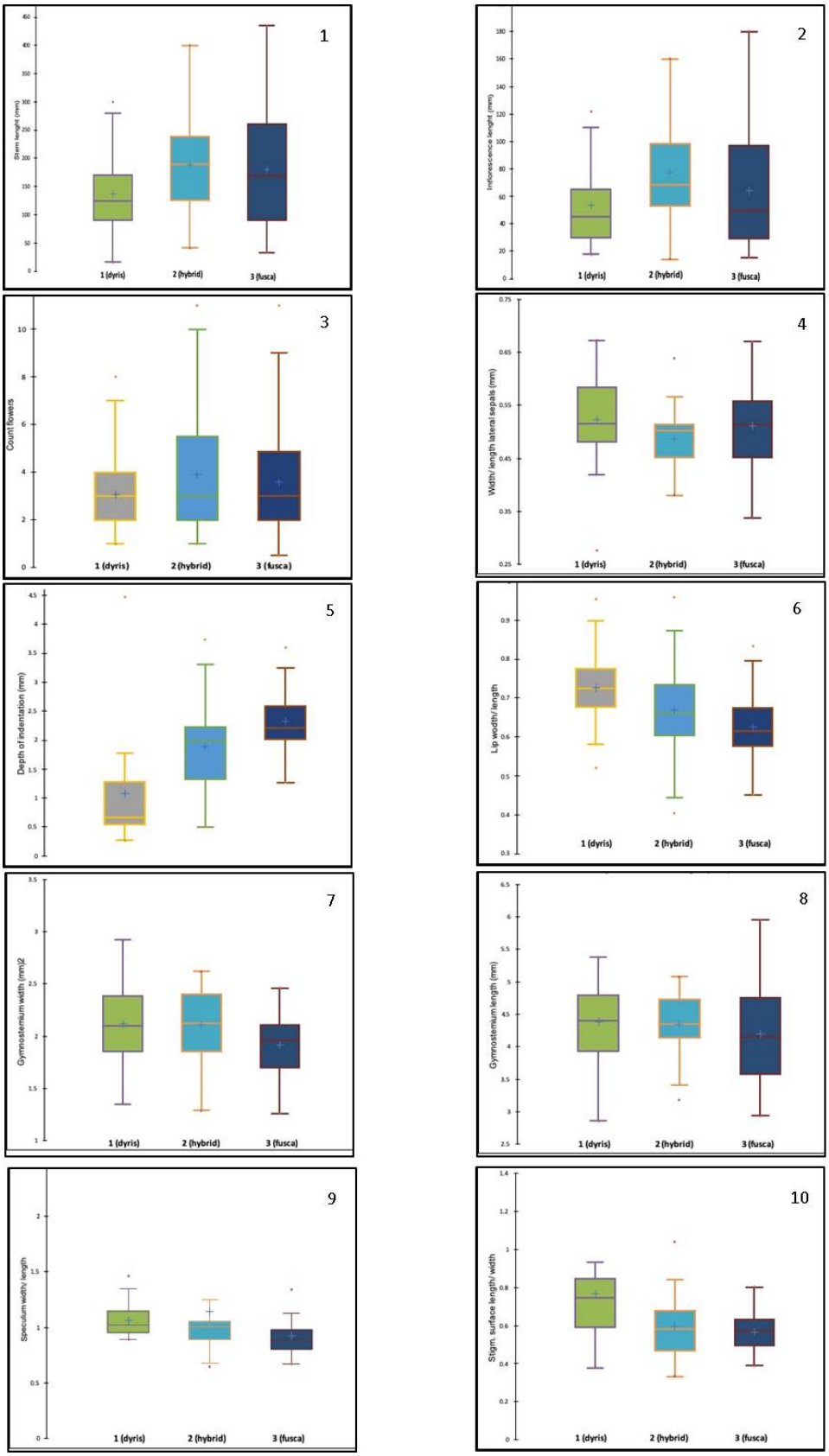


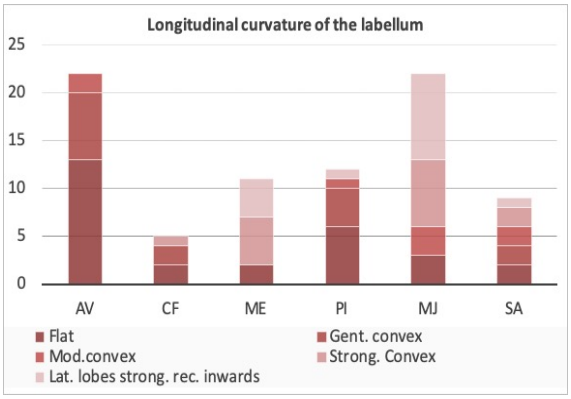
Figure 13. Comparison of the three different morphotypes attributed to the plants studied, *dyris* (1), hybrid (2) and *fusca* (3) regarding the quantitative morphological traits assessed: 1 – Stem length; 2 – Inflorescence length; 3 – Number of flowers; 4 – Lateral sepals (width/length); 5– Depth of indentation (mm); 6 – Lip width/length; 7 – Speculum length/width; 8 - Gymnostemium length (mm); 9 – Gymnostemium width (mm); 10 - Stigmatic surface length/width. Central "x" are mean values. The boxes are terminated by the twenty-fifth percentile below and seventy-fifth percentile above.

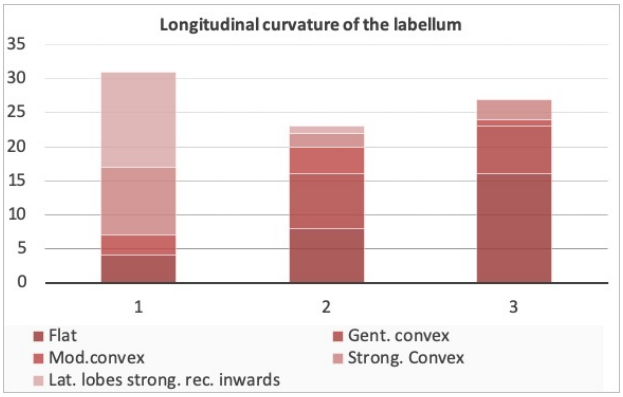
.

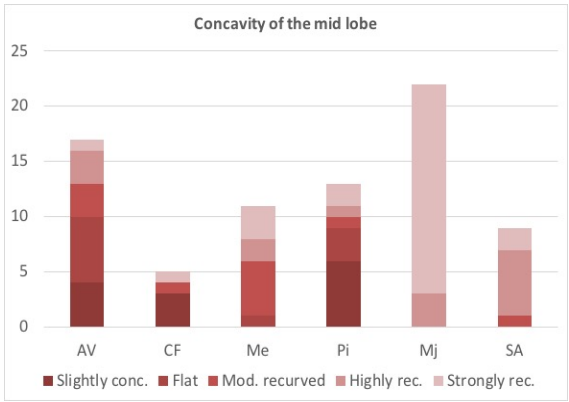
Characters 4, 6, 8 and 9 revealed not useful in differentiating populations, the last three also showing a large variation within each population. This assumption is not valid when it refers to morphotypes, with lip proportions (6) clearly segregating the three groups and gymnostemium length (9) with smaller values for *fusca* than for the morphotypes of *dyris* and intermediate. Depth of indentation (5) is the single quantitative character who seems to separate the three different groups of populations as well as morphotypes. In the analysis per population, this is the only trait in which the plots representing the supposedly mixed populations display intermediate values between the ones representing *Ophrys fusca* and *O. dyris*. With regards to morphotypes, lip proportions also account for this differentiation. Apart from these segregating traits and the vegetative vigour characters (stem length) – in the expression of which these populations have slightly lower values – the range of values for mixed populations box plots appear either close to *O. fusca* ones (traits number 4 and 10), to *O. dyris* (traits number 6, 7) distribution ranges, or even not differentiating from any of the groups (traits number 8 and 9).

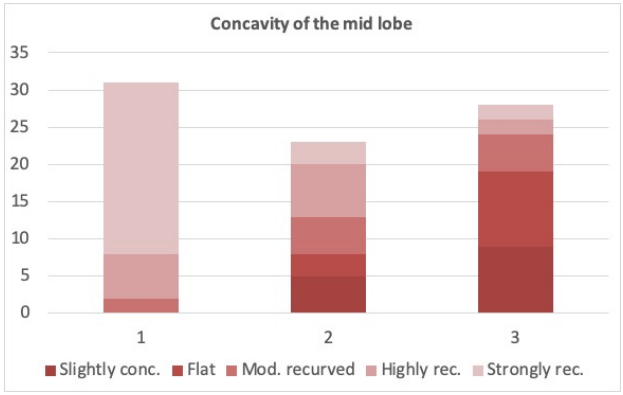
Table 9. Quantitative morpho-anatomical traits assessed, per individual plant (stem length, inflorescence length and length of the basal bract) or per flower (all the traits below). Values averaged over morphotypes, per individual plant for the first three traits or per all the flowers for the remaining, standard deviation and the range in parenthesis

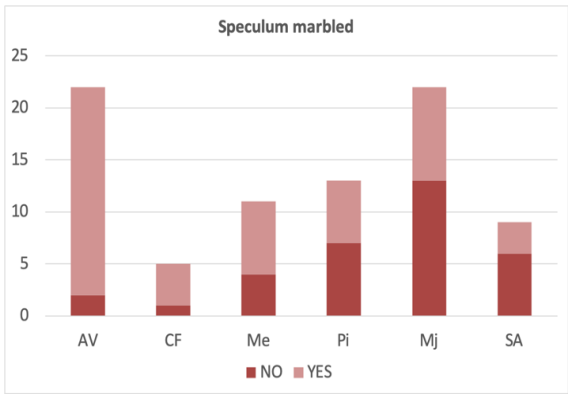
	Morphotype 1 (dyris) (n=99)	**Morphotype 2 (hybrid)** (n=71)	**Morphotype 3 (fusca)** (n=118)
Stem length (mm)	136.57 ± 61.43 (16-300)	188.07 ± 83.35 (41.2-400)	179.57 ± 97.55 (33-435)
Inflorescence length (mm)	53.08 ± 28.12 (17.62-121.85)	76.51 ± 36.89 (14.14-160)	64.62 ± 44.38 (15.5-180)
Length basal bract (mm)	53.54 ± 24.67 (14.77-145.91)	74.51 ± 28.04 (12.34-132.96)	60.07 ± 34.26 (10.61-147.79)
Nb of flowers	3.05 ± 1.47 (1-8)	3.91 ± 2.30 (1-11)	3.60 ± 2.29 (0.5-11)
Lip length (mm)	14.52 ± 1.53 (9.47-18.9)	15.01 ± 2.46 (8.9-19.3)	15.68 ± 2.94 (9.5-20.6)
Lip width (mm)	10.52 ± 1.37 (7.46-14.6)	9.94 ± 1.79 (4.26-14.5)	9.77 ± 2.04 (6-14.4)
Depth of indentation (mm)	0.83 ± 0.92 (0-4.47)	1.73 ± 0.92 (0-3.74)	2.33 ± 0.55 (1.27-3.6)
Speculum max length (mm)	7.44 ± 1.06 (5.09-9.46)	8.24 ± 2.11 (1.72-11.49)	8.53 ± 1.79 (4.47-11.93)
Speculum max width (mm)	7.80 ± 0.84 (5.9-9.74)	8.14 ± 0.96 (5.93-9.74)	7.63 ± 1.13 (5.9-9.83)
Length gymnostemium (mm)	4.38 ± 0.58 (2.86-5.38)	4.34 ± 0.52 (3.18-5.08)	4.20 ± 0.73 (2.94-5.95)
Width gymnostemium (mm)	2.12 ± 0.40 (1.35-2.92)	2.11 ± 0.35 (1.29-2.62)	1.92 ± 0.31 (1.26-2.46)
Length stigmatic surface (mm)	2.99 ± 1.14 (1.17-7.37)	2.39 ± 0.69 (1.25-3.76)	2.18 ± 0.53 (1.16-3.01)
Width stigmatic cavity (mm)	3.95 ± 0.72 (2.41-5.18)	4.02 ± 0.60 (2.57-5.38)	3.88 ± 0.72 (1.99-4.92)
Length lateral sepals (mm)	12.7 ± 2.16 (8.15-17.35)	13.09 ± 2.00 (9.29-17.46)	11.57 ± 2.60 (4.57-14.99)
Max width lateral sepals (mm)	6.54 ± 1.08 (4.17-9.01)	6.33 ± 1.02 (4.55-8.36)	5.91 ± 1.41 (1.54-8.08)

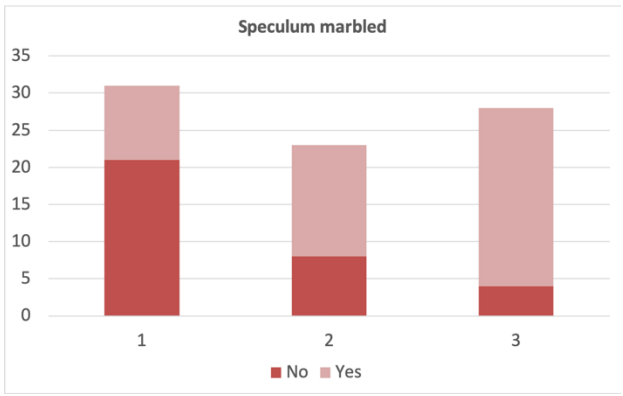


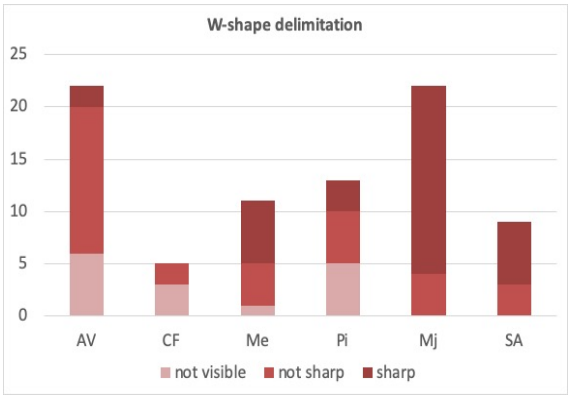












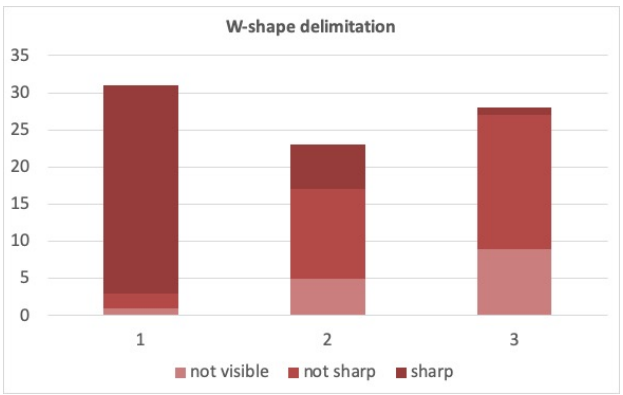


Figure 14. Expression of the qualitative characters assessed part 1, per population (left side) and morphotype (right side), part 1: - Longitudinal curvature of labellum; Concavity of the mid lobe; Speculum marbled; W-shape delimitation; For population codes and taxa present in each population, please see

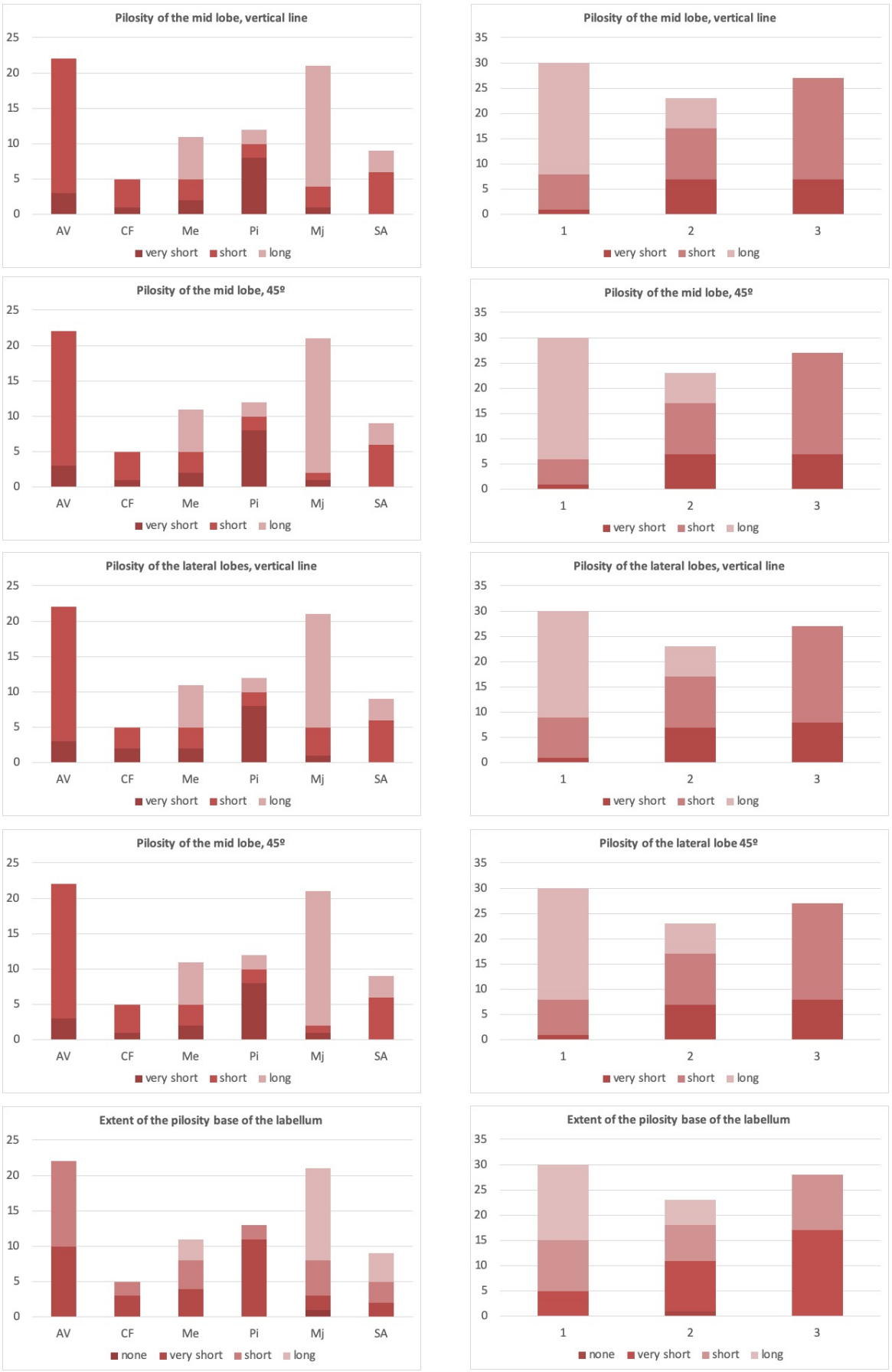


Figure 15. Expression of the qualitative characters assessed, part 2, per population (left side) and morphotype (right side), part 2: Pilosity of the mid lobe, vertical line; Pilosity of the mid lobe, 45º; Pilosity of the lateral lobes, vertical line; Pilosity of the lateral lobes, 45º; Extent of the pilosity at the base of the labellum; For population codes and taxa present in each population, please see

Regarding longitudinal curvature of labellum and concavity of mid lobe (Figure 14, part 1), characters more remarkably exhibited in *O. dyris* plants, the two populations in the left side of these plots, theoretically accounting for *O. fusca*, do not or hardly include individuals displaying the upper levels of each scale, i.e., lateral lobes strongly recurved inwards and the mid lobe strongly recurved. Mj, accounting for *O. dyris*, is the population where we can find most plants with both these features. Mixed population Me has either plants with flat (few), strongly curved or strongly recurved inwards labella, whereas Pi includes flat (the majority), gently convex, moderately convex and strongly recurved inwards labellum plants. Surprisingly, SA (considered for *O. dyris*) includes plants with all the states for curvature of the labellum, in very similar proportions. Classification of the sampled plants per morphotypes results in a similar pattern for this trait, with the singularity that amongst the group of plants of intermediate morphotype (2), the states typically associated to *fusca* are expressed in a higher proportion of plants: thirty-five per cent of the plants have a flat or gently convex labellum (Table 10). As for concavity of the mid lobe, after Mj, the second higher proportion of plants with strongly recurved mid lobe appears in Me, although SA exceeds its proportion in plants with highly recurved lobes, level 4 of the 5 levels defined concavity scale. Accordingly, the extremes of this trait expression are found in higher proportions in morphotypes 1 (*dyris*) and 3 (*fusca*), more expressively in the *dyris* morphotype (the strongly recurved phenotype).

Regarding W-shape delimitation, a trait also traditionally used to distinguish *O. dyris*, plants with sharply delimitated W-shapes are largely found in Mj and not represented at all in CF (accounting for *O. fusca*). Mixed populations (Me and Pi) have the three states represented, although Me has the same proportion of sharply delimitated plants than SA. The same trend highlights from the bar charts displaying morphotypes measurements, with this character being more strongly expressed in the *dyris* than in the *fusca* morphotype, i.e. a higher proportion of the parental phenotype.

Records regarding the marbling of the speculum also display a predictable pattern in the different populations. Such “marbling” is usually associated to morphotypes closer to *O. fusca*, and this is confirmed in the graphical representation of the values for this morphotype (3). Plants with marbled speculum occur mostly in AV and CF populations, while most assessed plants in Mj and SA are not marbled, as also displayed for the *dyris* morphotype (1). In Me the number of marbled individuals overpasses the non-marbled, while in Pi the proportions are very similar, with a slightly higher proportion of non-marbled plants, this relative proportion being repeated within the range of the intermediate morphotype (2).

The results for the characters assessing the pilosity of the lobes of the labellum (Figure 15) show remarkable consistency amongst the different measurement angles and lobes: higher amount of

plants with long hairs in Mj, followed by Me and then by SA; absence of long hairs recorded in AV and CF populations, where most plants have the labellum lobes with short hairs. In Pi the majority of plants also has the labellum with very short hairs, despite a few plants with simply "short" or even long hairs (two plants each). The same consistency amongst the different variables describing pilosity of the labellum is observed by morphotypes. Further described discriminant analysis has revealed pilosities of the labellum related traits as the characters that most influence the separation between morphotypes. Presence of long hairs is also a feature commonly associated to *Ophrys dyris* plants, which seems to be supported by both univariate and discriminant analysis. As for the pilosity at the base of the labellum surrounding the stigmatic cavity, long trichomes appear in three of the populations: in a larger number of plants in Mj, followed by SA and by Me. The two supposedly mixed populations have much different proportions of plants with long and short hairs. In Pi there is a large majority of plants with very short hairs and a small proportion with short hairs, long hairs not having been registered. At Me the amount of plants with very short, short or long hairs is similar, the latest being slightly less represented than the former two types. When looking at the results by morphotype, one of the extremes of the expression range for this character is only represented in the intermediate morphotype (2), even if in a very small proportion (~1%). Also not expected was the fact that a non-negligible proportion of plants (~5%) of the *dyris* morphotype (1) display a "very short hairs" phenotype, which is more characteristic of *fusca* (3).

Table 10. Summary table of-qualitative variables assessed and the state of character best represented in each morphotype, or the two most represented, in case the difference between their proportions in the sample doesn't reach 10%. Between brackets are the corresponding proportions in the sample. n = number of plants assessed.

	Type	**Morpho 1 (dyris)** (n=31 plts)	**Morpho 2 (hybrid)** (n=23 plts)	**Morpho 3 (fusca)** (n=23 plts)
Longit. curvature of labellum	Qualitative	Lat. lobes strong. recurved inw., 5 (45%)	Flat-gently convex, 1/2 (35%)	Flat, 1 (61%)
Concavity of the mid lobe	Qualitative	Strongly recurved, 5 (74%)	Moderat-highly recurved, 4-3/1 (30-22%)	Flat, 2 (39%)
Speculum marbled	Binary	No, 0	Yes, 1	Yes, 1
W-shape delimitation	Qualitative	Sharp,3	Not sharp,2	Not sharp,2
Pilosity mid lobe vert. line	Qualitative	Long, 3 (73.3%)	Short, 2 (43.5%)	Short, 2 (82.6%)
Pilosity mid lobe 45º	Qualitative	Long, 3 (80%)	Short, 2 (43.5%)	Short, 2 (82.6%)
Pilosity lat. lobes vert. line	Qualitative	Long, 3 (70%)	Short, 2 (43.5%)	Short, 2 (78.3%)
Pilosity lat. lobes 45 degrees	Qualitative	Long, 3 (73.3%)	Short, 2 (43.5%)	Short, 2 (78.3%)
Extent of pilosity at the base of labellum	Qualitative	Long, 3 (50%)	None-very short-short, 1-2-3 (44-30-22%)	None-very short, 1-2 (52-48%)

2.3.2. Phenology

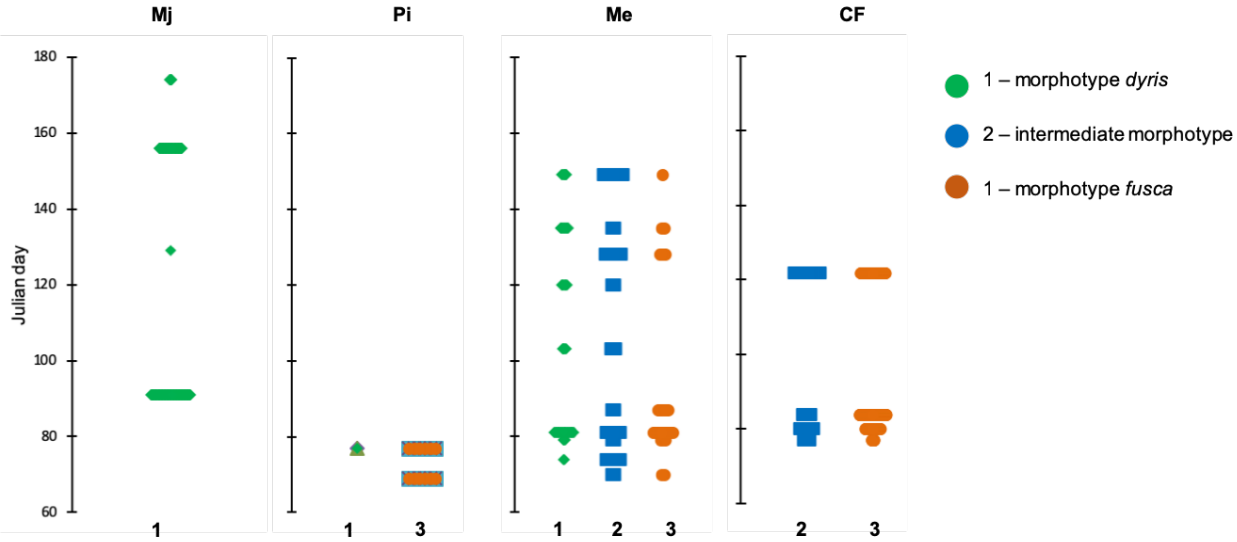


Figure 16. Time (Julian day) distribution for the observations of all flowering stages combined – recently flowering (RFL), fully open (FOP), flower senescent (FLS) - for the three morphotypes, in the populations sampled for phenology in 2011: Mj, Pi, Me and CF. Observations of flowers at the stage of bud (FLB) or capsule (CPS) have been excluded.

Amongst the three morphotypes in all the populations monitored in 2011 (Figure 16), the extended flowering period - from recently open flowers to senescence - started as early as in the 69th Julian day, in the second week of March, for plants with the morphotype *fusca* in the population of Pi. The last senescent flower was registered that year in Mj population, of *dyris* morphotype, in the day 174, late June. In this field season only four populations had been considered for phenological assessment, excluding AV and SA. For morphotype *dyris* the extended flowering period spans between Julian day 74 (third week of March) and 174 (late June), while in morphotype *fusca* between day 69 (second week of March) and 149 (last week of May). Flowers presenting an intermediate morphotype were observed in a time range similar to that of *fusca* plants, the earlier flowering date being day 70 in 2011 (second week of March) and the latest day the 149th. From the years when flowers inspection occurred, 2011 was chosen to look at flowering behaviour per population, due to the more thoroughgoing sampling. In CF population the majority of flowers were fully open in the second-third week of March, while in Mj the fourth week of April (JD 97 to 103) is when more flowers have that phenological stage (FOP) (Figure 17). Another assumption supported by the data displayed in Figure 16 is that intermediate morphotype can also be found in populations previously considered as "pure", such as CF.

It is likely that the interspaced flowering time displayed in the three morphotypes has been influenced by some skewness in the temporal distribution of the sampling visits. Regarding *fusca* morphotype, a

previous assumption of a later flowering period for this taxon, based in the initial recognition visits and previous work done by other colleagues in the region, made us carry out most field visits for data collection in AV population later in the calendar. Nevertheless, it seems reasonable to consider the possibility of a second flowering peak, as CF also display FOP flowers in the fifteenth week (Figure 13), three weeks after the flowering period of most flowers. Not surprisingly, the flowering peak in one of mix populations spans throughout a larger period than in populations assumed as "pure" (Figure 17 and Figure 13): in Me population, FOP flowers were recorded in six different weeks, in significant proportions. FOP flowers of *O. dyris* were also registered in Mj in the same number of weeks, but with much different proportions between weeks, with a clear tendency to flower in the end of March (13th week). When comparing the flowering periods in populations assumed as "pure" for each species, *O. dyris* flowering period seems to be longer than that of *O. fusca*: in Mj FOP flowers were registered throughout six weeks, while in CF the period between anthesis and senescence did not last more than five weeks. Also interesting is the fact that first registers of flowering plants have been done in mixed populations (Pi and Me), with no registers of flowering plants in the remaining populations at that time.

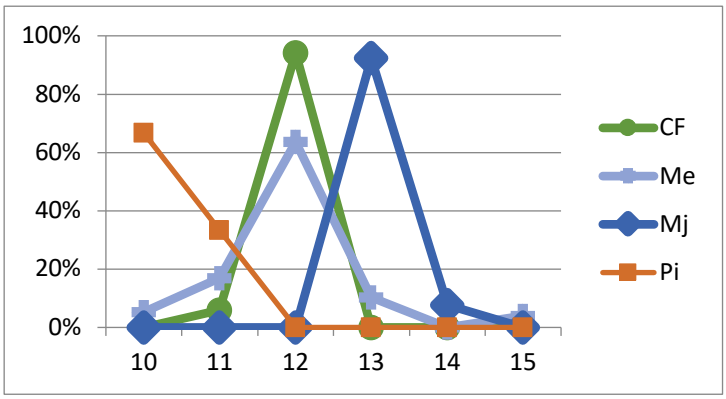


Figure 17. Proportions of plants with fully open flowers (FOP) per week, in 2011 – CF (*O. fusca*), Me (mixed), Mj (*O. fusca*) and Pi (mixed).

Table 11. Results for the GLM testing for the effect of the number of fully open flowers (FOP) in the different populations. All morphotypes considered. The number of open flowers in the populations of Mj (p=0.015) and Me (p=0.029) differ significantly from the others.

	Julian day		
Predictors	*Estimates*	*CI*	*P*
(Intercept)	95.69	73.77 – 117.61	**<0.001**
Population [Me]	17.78	-7.06 – 42.61	0.155
Population [Mj]	41.81	8.57 – 75.04	**0.015**
Population [Pi]	-17.36	-67.97 – 33.25	0.490
FOP	-0.38	-1.39 – 0.62	0.444
Population [Me] * FOP	-2.48	-4.69 – -0.28	**0.029**
Population [Mj] * FOP	-0.21	-1.45 – 1.02	0.729
Population [Pi] * FOP	-3.62	-38.96 – 31.73	0.836
Observations	41		
R^2 Nagelkerke	1.000		

GLM results for the number of fully open flowers (FOP) in the different populations (Table 11) reveals significant differences (0.015) in the Julian day for the population of Mj. The total number of FOP flowers (belonging to the three morphotypes) also influence meaningfully (0.029) the Julian Day in Me population. GLM searching for differences between populations in the number of flowers at all the flowering stages combined (RFL, FOP and FLS) in 2011, failed to resolve.

Figure 13 displays the number of flowers at the flowering peak (FOP) when there were plants from two or more morphotypes flowering at the same time, allowing to identify in which time periods there could have been opportunity for gene flow between morphotypes. Mixed population of Me has overlapping of morphotypes for the longer time range, thus presenting this opportunity for a longer period. All the combinations of morphotypes can be found to flower simultaneously in this population, being the most frequent the combination of three morphotypes - dyris (M1), intermediate (M2) and fusca (M3) – fully open in the same Julian week (50% of the fully open flowers). CF is the population in which there are more flowers from different morphotypes open in the same week, although for a short period, and only for morphotypes *fusca* and intermediate. From all the populations, CF is where the number of FOP flowers when in overlapping is significantly higher. According to these results, overlapping of the peak of flowering between morphotypes *dyris* and *fusca* in the same population is very rare. It happens only in Me and Pi.

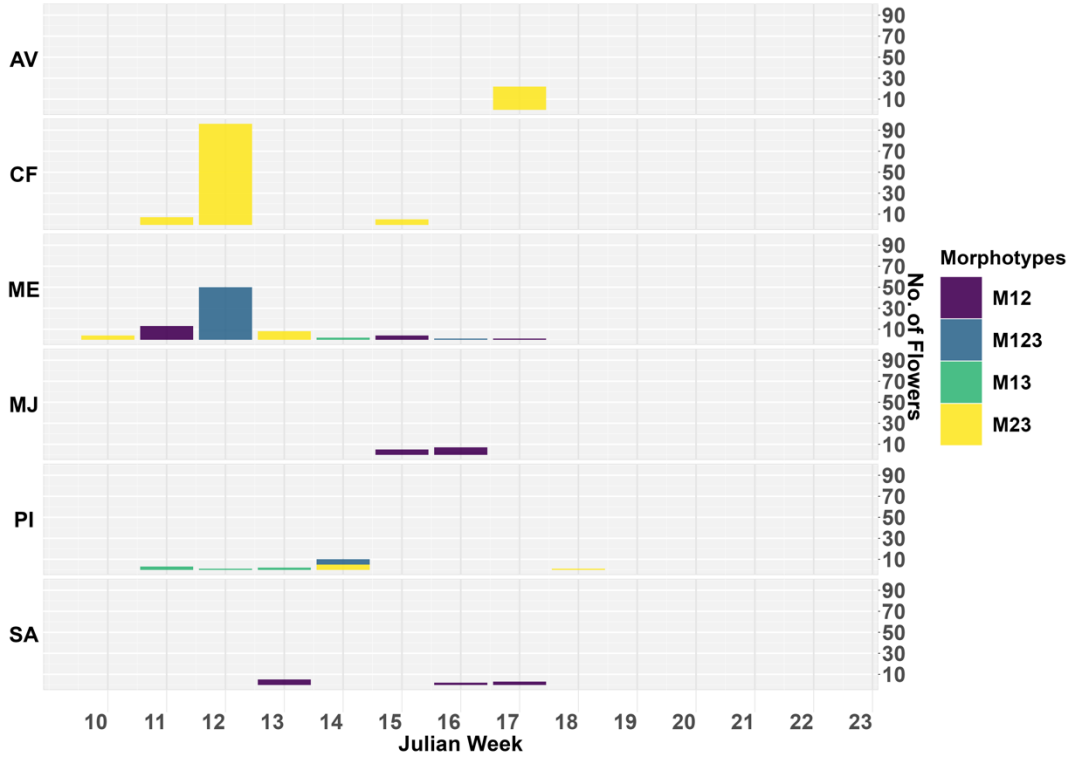


Figure 18. Julian weeks with overlapping of different morphotypes for fully open flowers (FOP phenophase), in each population. Each colour represents a combination of morphotypes, as described in the legend of the graph: M12 – *dyris* + intermediate; M123 – *dyris* + intermediate + *fusca*; M13 – *dyris* + *fusca*; M23 – intermediate + *fusca.*

Figure 19 displays morphotypes separately and gathers all the flowering phenophases, from recently open flowers (RFL) to senescent flowers (FLS), thus allowing to include any early or late pollination event that could cause gene flow. From the observation of the ridges one can recognize that percentages of overlapping between intermediate and *fusca* morphotypes are much higher than between intermediate and *dyris* morphotype. Reduced overlapping between morphotypes *dyris* (1, purple colour) and *fusca* (3, yellow colour) is detected, although in a higher proportion than when considering only the flowering peak, FOP phenophase (Figure 13). This was registered in Me population in the 12th, 18th and 20th Julian week. In all these three periods of overlapping, open flowers of the intermediate morphotype (2, green colour) were also registered. The first period, 12th week, was when a higher number of flowers from all morphotypes was registered at the same time (from both Figure 13 and Figure 19), approximately two weeks after the flowering peak had been observed for the first time in that population. Overlapping between *dyris* and *fusca* morphotypes was also observed in Pi population, at a much smaller proportion than in Me. Between the 11 and 13th Julian week for only these two morphotypes, and at the 23th week for the three morphotypes simultaneously. SA is the population where less overlapping between morphotypes occurred, i.e. less flowers from different morphotypes open in the same Julian week.

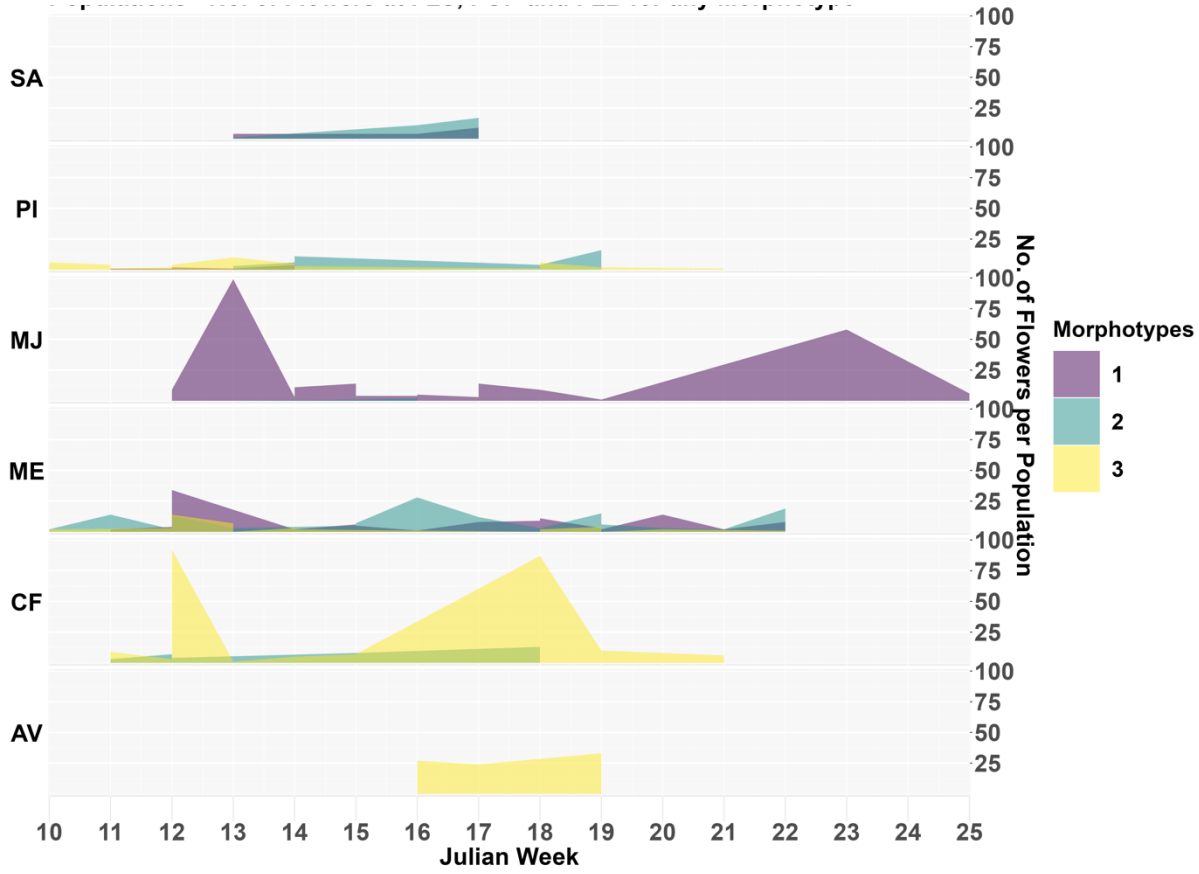


Figure 19. Number of flowers at all the flowering phenophases - recently flowering (RFL), fully open (FOP) and flower senescent (FLS) – for the three morphotypes, in each population (vertical axis), per Julian week (horizontal axis). Overlapping of colours in the same row means simultaneous flowering of different morphotypes in the same population. Morphotypes: 1 – *dyris*; 2 – intermediate; 3 – *fusca*.

Table 12. Results of GLM analysis testing for differences in the number of flowers at FOP from the three morphotypes, in the populations with two or three morphotypes sampled in 2011. Significant values in bold. Interactions were excluded from the model due to variance deficit.

Predictors	*Estimates*	*CI*	*P*	*std. error*
(Intercept)	105.93	74.08 – 137.77	<0.001	16.25
FOP1	-0.72	-1.34 – -0.11	**0.038**	0.31
FOP2	-3.85	-7.53 – -0.17	0.061	1.88
FOP3	-0.16	-1.29 – 0.97	0.786	0.58
Population [Me]	9.92	-24.57 – 44.41	0.582	17.60
Population [Mj]	45.69	5.04 – 86.33	**0.046**	20.74
Population [Pi]	-32.41	-75.90 – 11.09	0.168	22.19
Observations	20			
R^2 Nagelkerke	1.000			
AIC:	200.745			

From the GLM results, when two or three morphotypes flower at the same time, *dyris* is the only morphotype in which significant differences are found regarding the number of flowers at the flowering peak (FOP1). Accordingly, significant values were found for Mj population, where plants of *dyris* morphotype are dominant.

When plotting average values of Julian day per year of sampling one can see that the three morphotypes respond slightly differently to the different conditions between years (Figure 20A), this difference increasing in the two last years, 2014 and 2015. In the three first years of sampling, from 2011 to 2013, intermediate (M2) and *fusca* (M3) morphotypes don't seem to diverge in the averaged flowering peak.

Plotting the average time of the flowering peak per population (Figure 20B), differences are more meaningful within years. Years 2012, 2013 and 2015 were found to differ significantly from the others regarding the number of FOP flowers with P values of 0.006, 0.014 and 0.025. Populations Me (0.021) and Pi (0.040) are meaningfully distinguishable from the other populations. Unevenness of the sampling strategy for some populations throughout the years resulted in the absence of a full record in some years or populations, not allowing complete lines for all the populations in Figure 20.

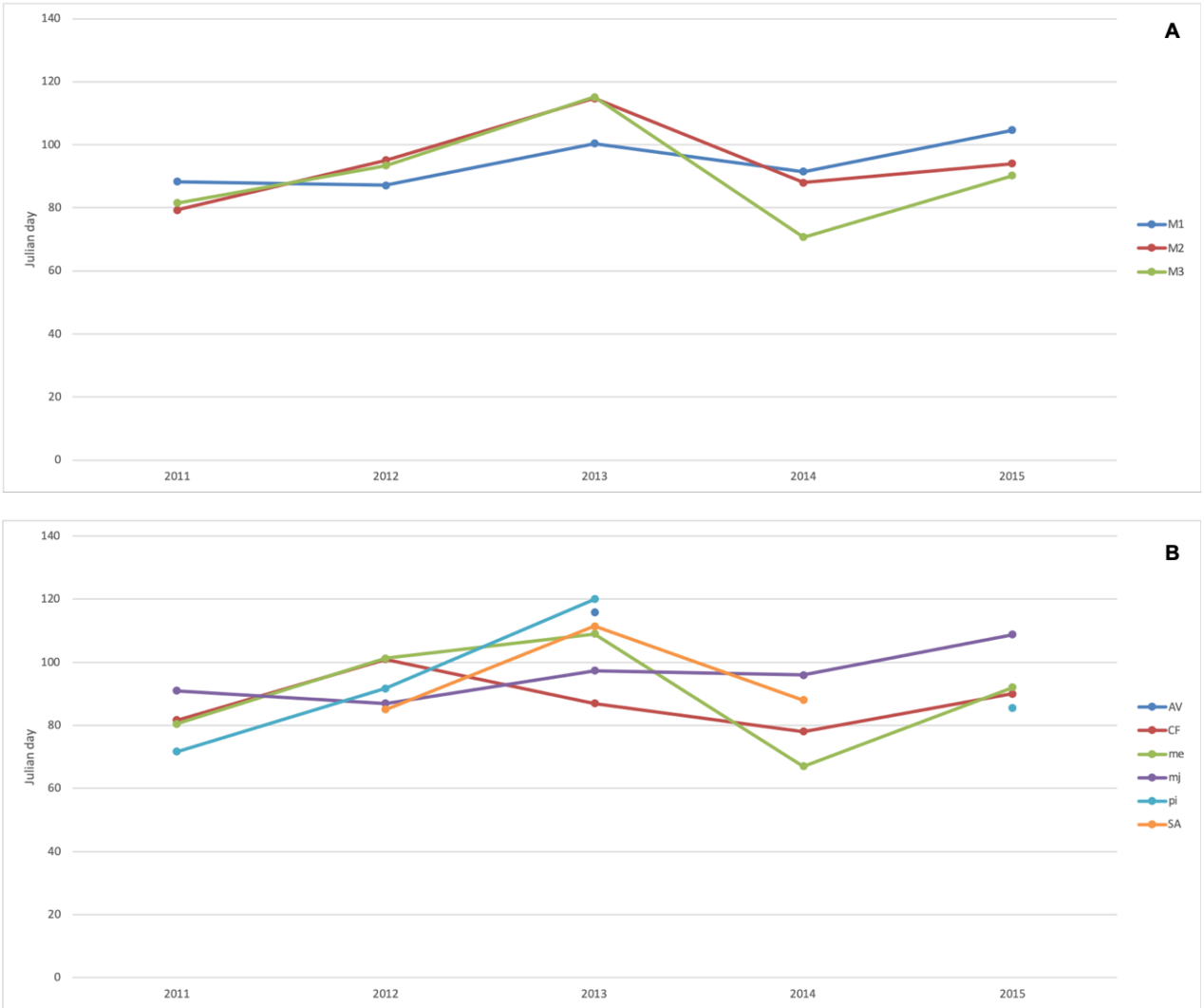


Figure 20. Mean peak flowering dates by year of sampling (error bars omitted), plotted by morhotype (A) - 1 (*dyris*), 2 (intermediate) and 3 (*fusca*) and by population (B) – *Arruda dos Vinhos*, AV; *Casalinho Facho*, CF; *Mendiga*, me; *Montejunto*, mj; *Pinheirinhos*, pi, *Serra de Santo António*, SA. Years considered: 2011, 2012, 2013, 2014 and 2015**.**

From the results of the GLM (Table 13) testing for the effect of Year and Population, the number of open flowers in the populations of Mj (p=0.015) and Me (p=0.029) differs significantly from the remaining, Mj with higher significance than Me. Regarding years, 2012 was the year that most influences FOP flowers distribution.

Table 13. Results for the GLM testing for the effect of Year and Population as predictors on the number of flowers at the FOP stage, for all populations and all years.

	FOP			
Predictors	*Estimates*	*CI*	*P*	*std. error*
(Intercept)	15.29	6.28 – 24.30	0.001	4.60
Population [CF]	-5.46	-14.98 – 4.06	0.263	4.86
Population [Me]	-10.62	-19.50 – -1.74	**0.021**	4.53
Population [Mj]	-5.77	-14.53 – 2.98	0.199	4.47
Population [Pi]	-9.29	-18.06 – -0.51	**0.040**	4.48
Population [SA]	-7.13	-17.27 – 3.01	0.171	5.17
Date.YearO [2012]	-5.55	-17.27 –	**0.006**	2.00
Date.YearO [2013]	-6.29	-17.27 – 3.01	**0.014**	2.53
Date.YearO [2014]	-5.72	-12.95 – 1.51	0.124	3.69
Date.YearO [2015]	-6.54	-12.19 – -0.88	**0.025**	2.88
Observations	129			
R^2 Nagelkerke	1.000			
AIC:	934.423			

2.3.2. Pollination success

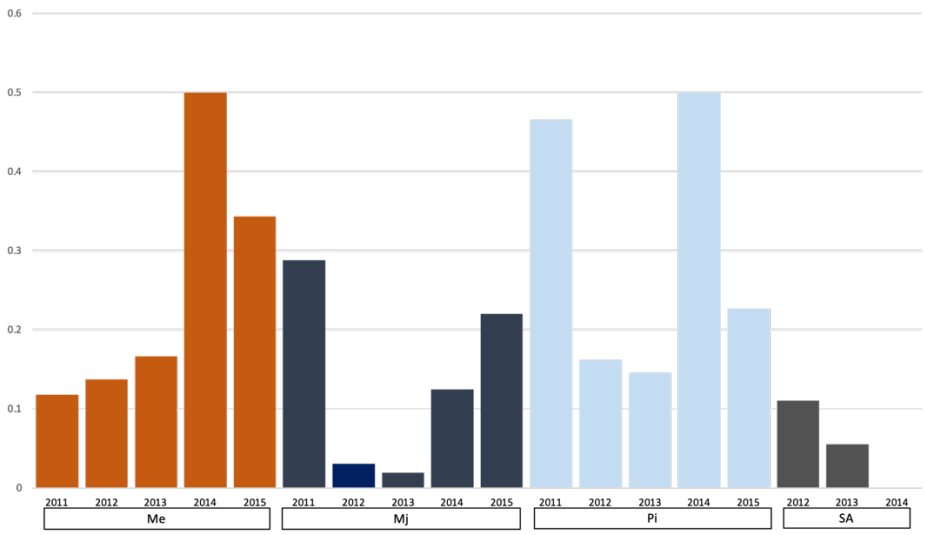


Figure 21. Average values for reproductive success in each of four populations, measured as the number of inflated capsules per number of flowers surveyed in each plant, per year (2011-2012). Populations considered: Me, Mj, Pi and SA.

Figure 21 shows the percentage of capsules filled averaged over total number of flowers assessed in each population, in each field season. Populations of CF and AV have been excluded from this analysis by the fact that comprehensive sampling regarding capsule filing has been done in no more than two years (CF) or one year (AV). Mj and SA, considered as populations where only the *dyris* morphotype occurs, display lower percentages for seed set than populations considered as of mixed morphotypes. We believe that the higher percentages in these populations of different taxonomic affinities, Pi and Me, are more likely related to population parameters such as density, population size or patch visibility, that have already been acknowledge to influence pollination success (Tremblay et al., 2005, Vandewoestijne et al., 2009), than to taxonomic affinity or other individual traits.

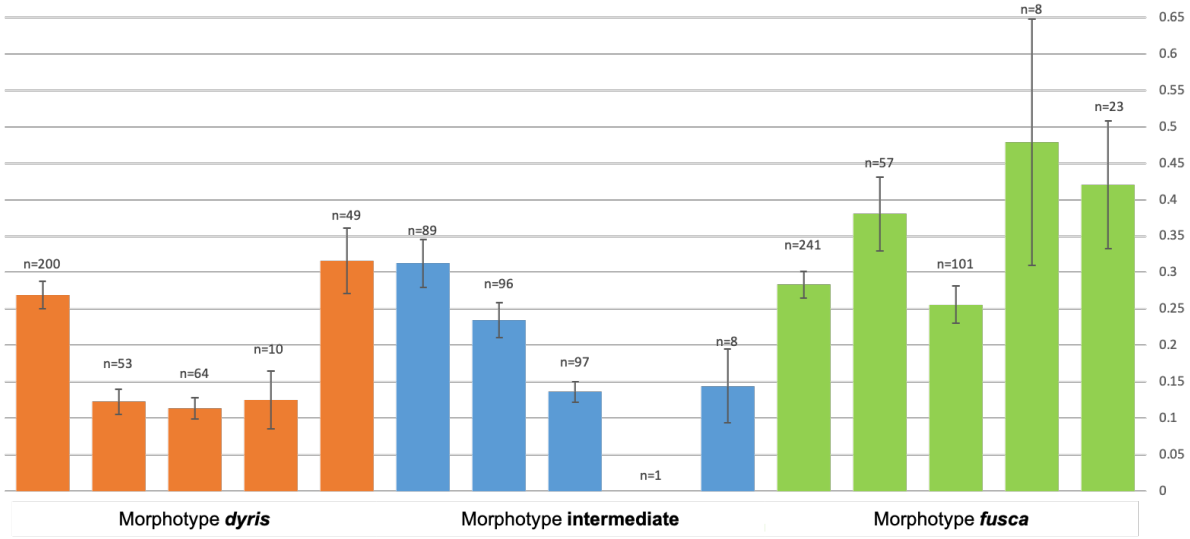


Figure 22. Average percentage of capsules per morphotype (1-*dyris*; 2-intermediate; 3-*fusca*), measured as the number of inflated capsules per number of flowers surveyed in each plant. In each morphotype, columns correspond to different years, between 2011 and 2015. Error bars (standard error) are displayed in each column.

When the same plants are assigned to the different morphotypes, regardless of the population where they have been sampled, percentages of capsule filling obtained are as displayed in Figure 22. The *fusca* morphotype (1) has the higher value amongst the three morphotypes in the years 2012, 2013, 2014 and 2015, while *dyris* (1) is the morphotype that has more capsules filled in 2011. The intermediate morphotype (2) has either the lowest (2011, 2015) or intermediate (2012, 2013, 2014) percentages for the proportion of capsules between the three.

The current results globally show high values for seed set percentages, when compared to previous studies. Values range between 3% for the *dyris* morphotype in 2013 up to 48% for the *fusca* morphotype in 2014, with average values of 11.4% (*dyris* morphotype), 23.4% (*fusca* morphotype) and 7.7% (hybrid morphotype) over all the years sampled. Despite from different species, these values

are not far from the range of previous published average values for other species of *Ophrys: O. fuciflora* - 4.26 ± 0.71%, *O. insectifera* - 13.79 ± 1.31%, or *O. sphegodes* - 8.92 ±2.09 (Vandewoestijne et al., 2009). A rate of 14-16% was recorded by Pellegrino (2010) for another deceptive orchid, *Anacamptis papilionacea* (L.) R.M. Bateman, Pridgeon & M.W.Chase.

2.4. Conclusions

2.4.1. Morphology

The analysis of the morphological dataset revealed that, despite the occurrence of several intermediate phenotypes, there are still morphological groups that preserve some isolation in terms of the expression of morpho-anatomical traits. When considering the complete morphological dataset and all the characters, two groups of plants were recovered. These two groups are referred as the *fusca* group and the *dyris* group. In the exploratory analysis including 49 plants common to both the morphological and genetic datasets and 15 diagnosing characters, three groups were identified.

Variables that best discriminate between morphological groups are qualitative, related to the tactile and visual clues given to the pollinator by the labellum structure. Most notable is the pilosity of the mid lobe of the labellum, whether measured along a diagonal angle (highest contribution to discrimination) or vertical line (second highest), followed by the pilosity of the lateral lobes of the labellum. Also relevant for the segregation of morphological groups is the delimitation of the W-shape, used for the separation of both taxa in traditional taxonomy. These markings surround the reflective speculum, which is considered to mimic the female of the pollinating insect species (Paulus, 2006, Bradshaw et al., 2010, Vignolini et al., 2012b) and thought to provide the male with visual cues (Streinzer et al., 2010, Vignolini et al., 2012b). In the analysis including only the plants represented in both the morphological and genetic dataset, once again, the qualitative traits that contribute most to the separation of groups are those which also play a key role in pollinator attraction due to the visual and tactile cues they give to the insect. Differences in such characters may account for pre-zygotic isolation barriers between the different morphological groups. On the grounds of the highlighted relevance of such traits for segregation between these groups, one can say there are still pre-zygotic isolation barriers acting between these taxa in the studied region. Surprisingly, the trait most used to separate the two taxa in traditional taxonomy, the longitudinal furrow, is not amongst those.

From the univariate analysis we conclude that vigour characters differ most strongly between populations (, plots 1 to 3). Vigour is likely explained by local adaptation processes. Arruda dos Vinhos (AV) - a population considered for the *fusca* morphotype – is the one which presents the higher values

for vigour characters, and Pinheirinhos (Pi) the lowest. The quantitative characters showing greatest differences between groups of populations (*O. fusca*/ mixed/ *O. dyris*) and morphotypes are depth of indentation (5), the ratio of the speculum dimensions (7) and the ratio between length and width of the stigmatic surface (10). Of the qualitative characters, pilosity of both the mid and lateral lobes of the labellum are the characters which vary the most between morphotypes *fusca*, mixed and *dyris* populations.

Some characters recently considered as relevant for phylogenetic separation between different species of *Ophrys* and interacting with pollinator species (Streinzer et al., 2010, Bateman et al., 2018b), have not been considered in the present study. Lateral sepal colour has recently been suggested to have more influence on the pollinator than speculum size or shape and it was found to be the most homoplasic character in a tree of fourteen analysed taxa of *Ophrys* (Bateman et al., 2018c). Analysis of this trait would probably be useful in distinguishing morphotypes within our dataset. Nonetheless, due to insufficient data, observer errors, time constraints in the field and difficulties with scoring colours, its registration had been abandoned for data analysis.

Following the morphological analysis, it seems sensible to question our initial assumption of "pure" populations of each taxon. Even considering only the characters that best segregate morphotypes such as the pilosity of the labellum lobes or the delimitation of the speculum, there are overlapping character states found in supposed pure *fusca* and *dyris* populations. For example, a small number of plants with a sharply delimited speculum, more characteristic of *O. dyris*, can be found in AV population. As for pilosity as diagnostic character, in the SA population (initially considered as *O. dyris* population) most of the plants have short, and not long hairs as one would expect, on both mid and lateral lobes of the labellum, regardless of the measurement angle. It also curious that most SA individuals have intermediate placement in the multivariate analyses. In addition to withdrawing the designation of pure and mixed populations, one should question the initial assumption regarding the taxonomic affinity of two of the populations: of SA as a population of *O. dyris* and AV as one of *O. fusca*.

2.4.2 Phenology

One conclusion is that there is no complete isolation between floral times of the two morphotypes. Overlapping flowering times occur in a small quantity of flowers, for short periods, and then only in populations previously considered as mixed, Me and Pi. This is despite intermediate morphotypes

being found in every population, meaning that no population should be considered as "pure". The population with least flowering overlap between morphotypes was found to be SA.

There is a marked tendency for plants in *O. fusca* populations to flower earlier, mostly in March and for *O. dyris* populations to start flowering from the 13th week onwards. However, a few plants in Mj population were fully open in the 12th week (2012), and what could be a second flowering peak in CF seems to occur in the 15th week, when there are still *O. dyris* flowers open in Mj population. This is consistent with the published flowering times for these species, from March to early April in *O. fusca* and starting about the same time and extending to mid-April in *O. dyris* (Lowe and Tyteca, 2012).

Flowering period is longer for the *fusca* morphotype than for the *dyris* morphotype. It is my belief that plants that flower for a longer period have an advantage over those with a shorter flowering span in terms of pollinator attraction. As long as there are open flowers in the population, the insect will have alternative flowers and will visit the population frequently. Consequently, another local trait that might influence pollination success is population density, not used for analysis in the current study. Vandewoestijne (2009) concluded that it negatively affects the reproductive success, which decreases with an increase in population density (Tremblay et al., 2005, Vandewoestijne et al., 2009).

2.4.3 Reproductive success

We believe that the results obtained for pollination success are more reasonably explained by population or certain individual vigour traits than by any factor related to taxonomic affinity. Hence, for a robust interpretation of data, other traits should have been considered for data analysis, such as population density, and the occurrence of other species of *Ophrys* in each population, identified as rewarding or non-rewarding. This should be of great relevance, as there are cases when nonrewarding species may be dependent on other species in the community to provide rewards (Wilcock and Neiland, 2002). To better understand the pollination success ecologically in these populations, it would have been significant if we had had more information about the pollinator of the plants. Does the same insect may occasionally visit plants attributed to the different taxa? Is there the possibility that the plants are visited by other organisms that may occasionally serve as pollinators? These are relevant questions that should be considered in a further study.

3. CYTOGENETICS

The content of this chapter has been published in the following research article:

Abreu, J. I. A., Hawkins, J. A., Cotrim, H. C., Fay, M. F., Hidalgo, O., & Pellicer, J. (2017). *Ophrys fusca* and *Ophrys dyris* (Orchidaceae) - constancy of tetraploidy amongst populations in Central Portugal. *New Journal of Botany*. doi: 10.1080/20423489.2017.1408185

The author has carried out field work and sampling, pre-treatment of the roots for analysis, flow cytometry measurements and the chromosome counts. The two last steps had the close supervision of Dr. Jaume Pellicer. Pre-tests with pollinaria stored under different conditions were not performed by the author but by Dr. Jaume Pellicer and Dr. Oriane Hidalgo, on material collected and posted by the author. Main topics on specific procedures related to the laboratory work were also first defined by Dr. Jaume Pellicer. All the remaining co-authors read and insightfully commented on the manuscript, particularly Dr. Julie Hawkins.

3.1. Introduction

Ophrys is a distinctive genus supported by morphological and molecular characters, but there is a lack of agreement about the number of species in the genus that has been attributed to interspecific hybridisation and introgression (Stebbins and Ferlan, 1956, Devey et al., 2008) . Better documentation of the extent of reproductive isolation between species could inform taxonomy and conservation strategies and contribute to our understanding of pollinator-mediated floral evolution in these charismatic, sexually deceptive orchids. Cytotype characterisation of hybridising species at the population level can be an important component of studies of hybridisation and introgression (interspecific gene exchange, through repeated hybridisation and backcrossing, following Anderson (1949), contributing to the robust interpretation of the co-dominant molecular marker data used to measure gene flow (Pellicer et al., 2012). Furthermore, in a hybridisation scenario, isolation barriers due to genetic incompatibilities attributed to the ploidy of the parental taxa may account for low incidence of introgressed individuals. Since homoploid hybrid species may show only weak post-zygotic isolation, ploidy influences the frequency of introgressed individuals. Discriminating between polyploidy and other variables influencing reproductive isolation of hybrids (fertile or partially fertile F1 individuals resulting from the interbreeding between the two species), such as pollinator behaviour, could be important in these plants.

Ophrys dyris Maire and *Ophrys fusca* Link are one pair of species found in sympatry and offering opportunity for the study of hybridisation in the context of a specialised orchid-pollinator system.

These two species are respectively included in the groups *omegaifera* and *fusca* of section *Pseudophrys* Godfery (Orchidaceae) and are closely related (Bernardos et al., 2005, Cotrim et al., 2016, Devey et al., 2008). Despite the different approaches and reported relationships pointed out below, we refer to these taxonomic entities as separate species, as our preliminary morphological analysis seem to maintain the segregation between specimens presenting trait expression of the diagnostic characters in the mean values of its description range.

These species are found across the Iberian Peninsula and northern Africa, *O. dyris* having a narrow and more localised distribution in this region, and *O. fusca* being more common and widespread, with a geographical range that reaches western Asia. Flowers of the latter species are extremely variable in morphology, leading to the segregation of the species into more than ten different species (Delforge, 2002), though other authors have taken a more conservative approach, recognising these two species in the broad sense, with the morphological variants being treated at lower taxonomic levels (Pederson and Faurholdt, 2007). *Ophrys dyris* is less variable in floral morphology and less abundant throughout its distribution range. Both species are listed in protected habitats, as components and bioindicators of habitat 6210 of Directive 92/43/CEE, under which habitats are prioritised when any listed species are numerous.

Figure 23. *Ophrys fusca* Link subsp. *fusca* (A) and *Ophrys dyris* Maire = *O. omegaifera* H.Fleischm. subsp. *dyris* (Maire) Del Prete (B).

Ophrys dyris and *O. fusca* have similar morphological characters, and careful inspection is needed for field identification (Figure 23), particularly as intermediate individuals are known, some of which have been characterised as hybrids using co-dominant molecular marker data (Cotrim et al., 2016). In 1981, the name *O. ×brigittae* H.Baumann was used to refer to the hybrids between *Ophrys dyris* and *Ophrys fusca.* Aside from *O. ×brigittae*, several other species considered to result from introgression between

O. dyris and *O. fusca* have been described for Portugal. These are *O. algarvensis* D.Tyteca, Benito & M.Walravens, *O. vasconica* (O.Danesch & E.Danesch) P.Delforge and *O. lenae* M.R.Lowe & D.Tyteca (Lowe and Tyteca, 2012). They were referred to as "paleohybrids" by Lowe and Tyteca (2012), as neither parent is present in the populations, such absence suggesting that hybridization occurred in the past. As far as we know, there are no molecular studies focused on the origin of these putative hybrids which form part of a "partly stabilized hybrid complex" between the *omegaifera* and *fusca* groups (Pederson and Faurholdt, 2007). *Ophrys dyris* itself might be of hybrid origin. Devey *et al.* (2008) postulated that the plant which represented *O. dyris* in their study was an intersectional hybrid between sections *Pseudophrys* and *Ophrys*, and based on ITS and AFLP data it seems likely that the specimen in cause was an F1 hybrid. However, as they included only one sample, more data are needed to interpret on the potential hybrid origin of the species.

In Portugal, *O. fusca* and *O. dyris* overlap in at least three limestone regions: in the Aire and Candeeiros Mountains, in Arrábida, and in the Algarve. The distribution of putative hybrids and introgressed individuals varies (Abreu *et al.*, in prep.), but it is not known whether post-zygotic isolation due to incompatible parental ploidy could account for these differences. O*phrys fusca* is known to exist at diploid (2*n* = 2*x* = 36) and tetraploid (2*n* = 4*x* = 72) levels (Greilhuber and Ehrendorfer, 1975), whereas tetraploid and pentaploid (2*n* = 4*x* = 90) cytotypes of *O. dyris* have been described. Greilhuber and Ehrendorfer (1975) first described tetraploids for Mallorcan specimens of the *O. fusca* aggregate, but also reported diploid Italian plants. In 2005, D'Emerico *et al.* reported that Italian *O. fusca* is diploid (D'Emerico et al., 2005). In 2007 and 2010, chromosome counts of 2*n* = 4*x* =72, 76 and 2*n* = 4*x* = 72 were reported for *O. fusca* from Arrábida region, Portugal, by García-Barriuso *et al.*(2010). Cotrim *et al.* (2009, 2016) inferred that specimens from the western Iberian Peninsula with four alleles per individual for some microsatellite loci were tetraploid. Considering *O. dyris*, following an early report of tetraploid plants (Kullenberg, 1961), Bernardos et al. (2003), later cited by Aedo and Herrero (2005) and Amich et al. (2007), reported counts of 2*n* = 4*x* =72 and 2*n* = 5*x* = 90. A confirmation of 2*n* = 4*x* =72 was reported by García-Barriuso *et al.*(2010) for specimens from Montejunto region, Portugal.

Figure 24. Distribution map of populations sampled in central Portugal. 1. Arruda dos Vinhos (AV); 2. Casal Facho (Marques et al.); 3. Montejunto (Mj); 4. Serra de Sto António (SA); 5. Pinheirinhos (Pi); 6. Mendiga (Me).

This study is part of a wider project describing the genetic and morphometric diversity of *O. fusca* and *O. dyris* in central Portugal (Figure 24) and documenting hybridisation and introgression between these species. Here we focus on describing the ploidy of these plants, characterising the accessions to be included in genetic and morphometric surveys with a view to integrative analysis in the future. Our new characterisations will permit robust interpretation of microsatellite data in preparation. Presently, knowledge of ploidy levels will clarify whether variants described in the literature are found in the Portuguese populations and may shed light on the hypothesis of hybrid origin for the species *O. dyris*, and on the possible dynamics of the *Ophrys fusca-O. dyris* populations known in Portugal.

3.2. Methods

3.2.1. Flow cytometry

As indicated above, nuclear DNA contents were estimated from a subset (67) of the plants that were measured for morphology and collected for DNA analysis in the scope of the broader study. Samples came from the six populations being monitored (Figure 24): *O. fusca* - 35 from *Arruda dos Vinhos* (north of Lisbon, AV) and 5 from *Casal Facho* (Arrábida, CF); *O. dyris* - 9 from *Montejunto* (Montejunto mountain, Mj) and 10 from *Serra de Sto António* (Santo António Mountain, SA); samples from populations where both species are present - 3 from *Pinheirinhos* (Arrábida, Pi) and 5 from *Mendiga* (Aire and Candeeiros Mountains, Me). Analyses were carried out by flow cytometry using pollinaria instead of leaf tissue, in order to avoid potential misinterpretation of the results given the differential release of nuclei in orchids (Garcia et al., 2014). To test the performance of pollinaria under different storage conditions, we analysed (i) fresh, (Nieto Feliner et al.) fixed in 3:1 ethanol:acetic acid and (iii) silica-dried pollinaria after collection in the field. Although fresh samples provided good quality results, after several days of storage the quality of the results decayed significantly, so we used silica-dried pollinaria instead, which provided better quality flow histograms and comparable relative fluorescence. Measurements were performed on the pollinaria of 67 plants (Table 14) that had been previously screened with microsatellites. Genome size was also assessed using fresh leaves. For the cytotype screening, one to three pollinaria were co-chopped with the selected internal standard [*Pisum sativum* 'Ctirad' (2C = 9.09 pg) or *Petroselinum crispum* 'Curled Moss' (2C = 4.45 pg)] in a Petri dish containing 1ml of Ebihara buffer (Ebihara et al., 2005) following the one-step procedure described in Dolezel et al. (2007). The nuclei suspension was filtered through a 30 µm nylon mesh and stained with 100 µl propidium iodide (1mg/ml). Samples were kept on ice for 15 min and analysed using a Partec Cyflow SL3 flow cytometer fitted with a 100 mW green solid state laser (Cobolt Samba). For

each run, 3,000 particles were analysed. Measurements from leaves followed the same procedures, with three replicates run for each of the accessions and analysing 5,000 particles per run.

3.2.2. Chromosome counts

Table 14. Nuclear DNA contents from all the plants assessed, from pollinaria (67).

Plant	Fluorescence Ratio (Sample/Standard)	Relative nuclear DNA content	DNA Ploidy	Internal standard
AV02	2.429	10.9305	4x	*Petroselinum*
AV03	2.518	11.331	4x	*Petroselinum*
AV04	FAIL			
AV05	2.398	10.791	4x	*Petroselinum*
AV06	2.586	11.637	4x	*Petroselinum*
AV07	2.488	11.196	4x	*Petroselinum*
AV08	2.514	11.313	4x	*Petroselinum*
AV09	2.598	11.691	4x	*Petroselinum*
AV10	2.411	10.8495	4x	*Petroselinum*
AV11	2.554	11.493	4x	*Petroselinum*
AV12	2.521	11.3445	4x	*Petroselinum*
AV13	2.594	11.673	4x	*Petroselinum*
AV14	2.649	11.9205	4x	*Petroselinum*
AV15	1.249	11.35341	4x	*Pisum*
AV16	2.561	11.5245	4x	*Petroselinum*
AV17	2.524	11.358	4x	*Petroselinum*
AV18	2.264	10.188	4x	*Petroselinum*
AV19	2.604	11.718	4x	*Petroselinum*
AV20	1.224	11.12616	4x	*Pisum*
AV22	2.531	11.3895	4x	*Petroselinum*
AV24	2.549	11.4705	4x	*Petroselinum*
AV25	2.578	11.601	4x	*Petroselinum*
AV26	2.567	11.5515	4x	*Petroselinum*
AV27	2.63	11.835	4x	*Petroselinum*
AV28	2.549	11.4705	4x	*Petroselinum*
AV29	2.547	11.4615	4x	*Petroselinum*
AV30	2.53	11.385	4x	*Petroselinum*
AV31	2.6	11.7	4x	*Petroselinum*
AV32	2.452	11.034	4x	*Petroselinum*
AV33	2.61	11.745	4x	*Petroselinum*
AV34	2.66	11.97	4x	*Petroselinum*
AV35	2.637	11.8665	4x	*Petroselinum*
AV36	2.587	11.6415	4x	*Petroselinum*
AV37	2.591	11.6595	4x	*Petroselinum*
AV38	2.629	11.8305	4x	*Petroselinum*
CF02	2.565	11.5425	4x	*Petroselinum*
CF37	2.574	11.583	4x	*Petroselinum*
CF42	1.244	11.30796	4x	*Pisum*
CF43	2.567	11.5515	4x	*Petroselinum*
CF58	2.541	11.4345	4x	*Petroselinum*
Me68	2.649	11.9205	4x	*Petroselinum*
Me69	2.617	11.7765	4x	*Petroselinum*
Me70	1.245	11.31705	4x	*Pisum*
Me71	2.658	11.961	4x	*Petroselinum*
Me72	2.625	11.8125	4x	*Petroselinum*
Mj02	FAIL			
Mj10	2.541	11.4345	4x	*Petroselinum*
Mj19	2.632	11.844	4x	*Petroselinum*
Mj28	2.591	11.6595	4x	*Petroselinum*
Mj31	1.194	10.85346	4x	*Pisum*
Mj54	FAIL			
Mj56	2.521	11.3445	4x	*Petroselinum*
Mj60	2.554	11.493	4x	*Petroselinum*
Mj61	1.167	10.60803	4x	*Pisum*
SA04	2.542	11.439	4x	*Petroselinum*
SA05	2.564	11.538	4x	*Petroselinum*
SA06	1.254	11.39886	4x	*Pisum*
SA07	1.249	11.35341	4x	*Pisum*
SA08	2.484	11.178	4x	*Petroselinum*
SA12	1.236	11.23524	4x	*Pisum*
SA13	1.285	11.68065	4x	*Pisum*
SA21	2.622	11.799	4x	*Petroselinum*
SA22	2.578	11.601	4x	*Petroselinum*
SA23	2.653	11.9385	4x	*Petroselinum*
Pi15	1.231	11.18979	4x	*Pisum*
Pi80	1.289	11.71701	4x	*Pisum*
Pi87	1.239	11.26251	4x	*Pisum*

Knowing that all the plants have approximate values for genome size, we chose reference samples from all the populations to assess ploidy level and chromosome numbers. To count the number of chromosomes, root tips of those plants, with genome sizes previously assessed (seven plants from all the populations sampled: two from Pi, one from AV, one from CF, one from Me, one from Mj and one from SA) and collected from the field were used. To promote active root growth, plants had been re-potted one week before pre-treatment with colchicine. After roots were harvested, c. 10-15 mm root tips were cut and placed into a tube of cold distilled water. Samples were then placed into a 0.05% colchicine (w/v) solution and placed at 21ºC for three hours, from where they were transferred to a freshly prepared fixative of 3:1 ethanol:acetic acid. After 48 hours at 4ºC, they were moved to Eppendorf tubes in 70% ethanol. Roots were then washed in distilled water for 5 to 10 minutes using a shaker at room temperature, and transferred to 1M hydrochloric acid (HCl) at 60º C to be hydrolysed for 5 to 6 minutes. For the following staining phase, we used Schiff's reagent for 20 minutes, and subsequently used acetic orcein for a further 20 minutes.

The root tips were excised with a razor blade under a stereo microscope and mounted on a microscope slide in one drop of 2 % aceto-orcein to be squashed. Plates were then observed on a Zeiss Axioplan Imaging microscope and the metaphase plates photographed with a digital camera (SPOT RT; Diagnostic Inc.). Images were edited with the software ProgRes Capture Pro v2.9.1 (Jenoptik Optical Systems GmbH, Jena, Germany).

3.3. Results and discussion

3.3.1. The use of alternative tissues to estimate GS and infer ploidies in *Ophrys*

In recent times, flow cytometry has become the preferred method to estimate nuclear DNA contents in plants, not only because it is fast, reproducible and enables analysis of thousands of particles, but also because it only requires small amounts of tissue (Dolezel et al., 2007). Nonetheless, this is not always straightforward, and several taxonomic groups pose challenges due to either the presence of cytosolic compounds or unbalanced numbers of nuclei in G_1-phase of mitosis, which may lead to misinterpretation of the resulting flow histograms (Travnicek et al., 2015). Orchids are sometimes challenging due to frequent rounds of endoreplication (including partial endoreplication), which might be tissue-specific, hence efforts have focused in searching for alternatives in order to overcome this problem (e.g. Pellicer & Leitch, 2014; Trávníček *et al.*, 2015). In our study, although *Ophrys* does not seem to represent a major challenge in this respect, we investigated the use of pollinaria for estimating nuclear DNA contents and allocating DNA-ploidies instead of leaves. We found that the results obtained not only were highly similar (Table 15), but the number of

particles released were always higher when using pollinaria [e.g. out of 5,000 particles (including debris): leaf (337 nuclei), pollinarium (780 nuclei)], thus making our inferences more robust. In addition, since fresh and silica-dried pollinaria resulted in similar relative fluorescences [e.g. ratio standard-sample 1.235(dried)-1.261(fresh)], we also overcome efficiently one of the limitations of flow cytometry, i.e. the need for fresh leaves for high-quality estimations.

Table 15. Summary of the nuclear DNA contents from flow cytometry using silica-dried pollinaria and fresh leaves of *O. fusca* and *O. dyris*. Chromosome numbers for the corresponding samples are provided.

	Pollinaria (1C value)	Leaves (2C value)	Chromosome number (2n)
CF02 (*O. fusca*)	11.125 pg	21.716 pg	72
Pi15 (*O. dyris*)	11.247 pg	22.120 pg	74

As reported in Trávníček *et al.* (2015), pollinaria did not show multiple rounds of endopolyploidy, with most samples only displaying G_1 peaks (= 1C-value peak, since these are haploid cells). By contrast, our tests on leaves showed higher levels of endopolyploidy with 2C, 4C and 8C peaks, with 4C peaks (G_2) nuclei more than three times more abundant than 2C peaks (G_1) (i.e. 300/2C *vs*. 1045/4C (Figure 25).

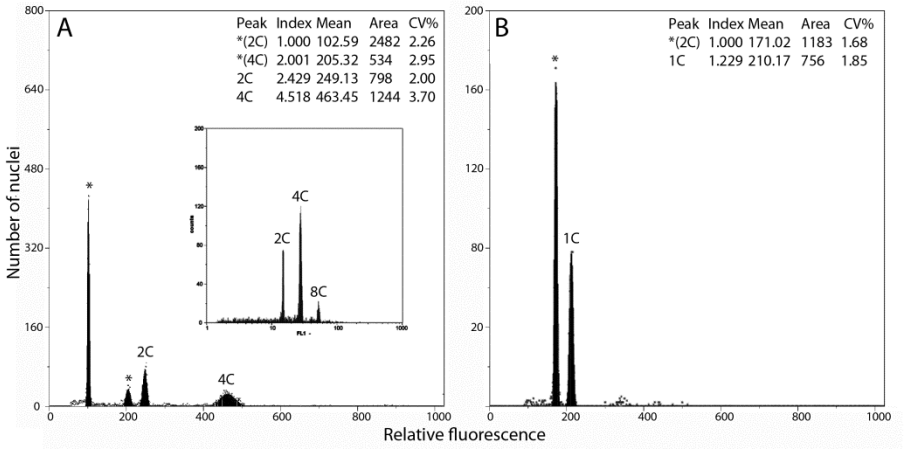


Figure 25. Representative fluorescence histograms in *Ophrys* (* = 2C and 4C peaks of the internal standard used (*Pisum sativum* 'Ctirad'). (A) *Ophrys dyris* leaf analysis (2C ~ 21.99 pg). Note that an additional histogram run on logarithmic scale is shown for illustrating the impact of endoreplication (2C, 4C and 8C peaks displayed). (B) *Ophrys dyris* pollinarium analysis (1C ~ 11.17 pg).

3.3.2. Tetraploid cytotypes in *Ophrys fusca* and *O. dyris* across central Portugal

Chromosome counts for the two plants from different populations (CF and Pi) and different species, *O. fusca* (CF02) and *O. dyris* (Pi15), revealed the same ploidy levels, with chromosome numbers of 2*n* = 4*x* = 72, 74, respectively (Figure 26). Since genome size from pollinaria and from leaves revealed approximately the same nuclear DNA content for all plants (average 1C-value = 11.19 pg), including typical *O. fusca* and *O. dyris*, specimens with intermediate morphotypes and putative hybrids following molecular analysis (data not shown), we inferred a tetraploid cytotype across central Portugal for the species *O. fusca* and *O. dyris* (Table 14) and their putative hybrids (Table 14). The difference between the chromosome numbers (2*n* = 72, 74) could be due to aneuploidy events (as identified previously in *Ophrys* by Greilhuber (1975), although chromosome breakage during preparation cannot be discounted.

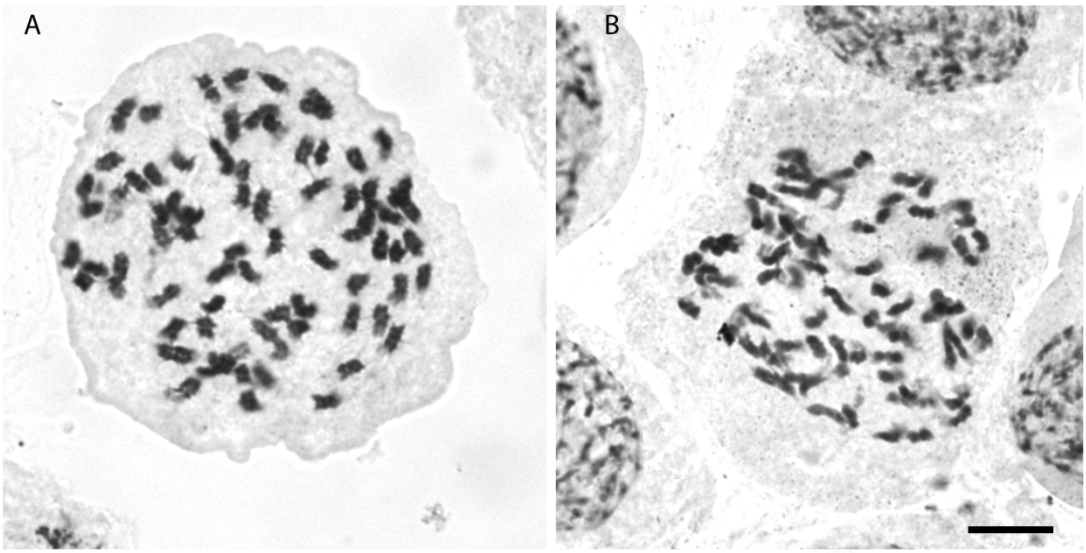


Figure 26. Metaphase cells from root tips of *Ophrys*, observed on a Zeiss Axioplan Imaging microscope. (A) *Ophrys fusca* (CF02) 2n = 72, (B) *Ophrys dyris* (Pi15) 2n = 74. Scale bar = 10 µm.

Within species of section *Pseudophrys*, García-Barriuso *et al.* García-Barriuso et al. (2010) observed a restricted geographic distribution of tetraploids, and interpreted their findings as suggestive of a young polyploid complex *sensu* Stebbins (1971). Our results provide support the view that polyploidy may have played an important role in shaping the evolutionary diversification of section *Pseudophrys* in the Iberian Peninsula (Amich et al., 2007, García-Barriuso et al., 2010), and that this area is a hotspot of polyploidisation in the section (García-Barriuso et al., 2010). It is notable that the remarkable cytotype consistency amongst the plants analysed here contrasts with the diversity of phenotypes and different genetic groups found (Abreu *et al.*, in prep.). It is also relevant to any discussion of possible fate of hybrids between *O. fusca* and *O. dyris*. Hybridisation in the case of species of the same ploidy could have a range of outcomes, including the establishment of a hybrid swarm, the transfer of traits through introgressive hybridisation, and the origin of new homoploid hybrid species (Yakimowski and Rieseberg, 2014), as already referred as a likely evolutionary mode in *Ophrys* by Paulus (Paulus and Gack, 1990). Since Stebbins and Ferlan (1956) reported solitary or few fully fertile F1 hybrid individuals in (another pair of) sympatric *Ophrys* populations but no hybrid swarms, the view has been that the breakdown of species-specific pollinator relationships in *Ophrys* hybrids would mediate against the establishment of complex hybrid populations. The fact that both species studied here have the same ploidy levels suggests that cytological isolation is not acting as a post-zygotic barrier, at least in this case, and a homoploid hybrid speciation process might be in the beginning. Future dissection of the hybridising populations to identify the frequency of F1, F2 and other hybrid progeny can be interpreted in this light.

3.3.2. The origin of *Ophrys dyris*

In 2008, when studying phylogenetic relations between several species of *Ophrys,* Devey et al. (2008) suggested that either the particular accession used in their analysis was of hybrid origin or the species itself could be of hybrid origin. If the latter hypothesis holds true, the likely parents would be from sections *Pseudophrys* and *Ophrys* (*fuciflora* aggregate), based on the phylogenetic results obtained by the authors. Despite being impossible to confirm the identification of such specimen (Bateman, *pers. comm.*) and the uncertainty of this conclusion, as assumed by the authors, our current molecular dataset under analysis supports the idea of a hybrid origin for *O. dyris*. To clarify this idea, plastid genes were analysed in the scope of this study (data not publ.), which seem to indicate *O. lutea* (section *Ophrys*) as one of the parents. In addition, Cotrim et al. (2016), supported by plastid haplotypes, reported introgressed individuals and hybridising populations between *O. fusca*-*O. lutea*, bearing out the tight relationship between these species previously pointed out by Soliva et al. (2001).

The different hypotheses illustrate different scenarios of breakage of reproductive barriers: the hybrid would be either the outcome of an intersectional cross or a plant with both parents from the same section. As the strength of reproductive isolation mechanisms (mainly prezygotic in sexual deceptive orchids) does not seem to be related with genetic distance or species divergence (Scopece et al., 2007), both scenarios would be equally plausible. In the light of previous counts and the results presented here, in either case the outcoming hybrid would be tetraploid with a diploid and a tetraploid parent. Such crosses are known to result from fusion of an unreduced 2*n* gamete from the diploid parent and a normally reduced 2*n* gamete from the tetraploid parent (Carroll and Borrill, 1965, Petit et al., 1999). Unreduced gametes are the result of abnormal meiotic division and, despite ease of production in controlled experiments, their frequency in natural populations is still largely unknown. The viability of the resulting hybrid would have been influenced by paternal:maternal ratios, the direction of the cross and by endosperm development (Burton and Husband, 2000, Sabara et al., 2013). Maintenance is likely to be grounded on competitive ability, higher fecundity, selfing and habitat segregation between cytotypes, as documented by Rodriguez (1996) and cited by Petit et al. (1999). As theorised by the latter author, the establishment of a species such as *O. dyris* might reflect the maintenance of viable populations for successive generations, or the rapid colonisation of new areas. Pre-zygotic isolating mechanisms, such as pollinator-mediated isolation, are of major importance for the establishment of such species (Petit 1999). The constancy of cytotypes, recurrent hybridisation events and gene flow between *O. dyris* and *O. fusca* suggest that isolation - at least from one of the putative parents – is not complete. To achieve a more comprehensive understanding of the evolution of this group including clarifying the origin of *O. dyris*, further genetic analyses are being conducted.

4. GENETIC DIVERSITY AMONG THE STUDIED PLANTS – MICROSATELLITES

4.1 Introduction

To fully understand patterns of introgression and hybridization in natural populations, knowledge of genetic variation and geneflow is needed in combination with morphological data. Genetic variation, as described by Lowe et al. (2007) is "the raw material upon which selection will act", which is "continually being created by mutation and at the same time eroded by selection and drift" (Lowe et al., 2007). Gene flow, on the other hand, it is what allows the maintenance of genetic connectivity between populations (Lowe et al., 2007). It can have significant effects on morphology and ecology of taxa and influence both their coexistence at contact zones and possible adaptation to novel environments (Stebbins, 1971, Chapman and Abbott, 2010). Therefore, information on the frequency of gene flow, the genetic structuring of populations and on the degree of isolation between the genetic entities, helps to understand how species complexes will be maintained (Kloda et al., 2008, Symonds et al., 2010) and how species can respond to selective pressures. Such information is obtained from the use of molecular markers.

Microsatellites were the molecular markers chosen for the genetic study to the plants and populations analysed in this study. Consisting of tandemly repeating units of DNA of 1 or 2 to 6 bp in length, microsatellites are widely distributed throughout the nuclear genomes of eukaryotes (Bhargava and Fuentes, 2010). They have proved to perform well in individual genotyping, gene flow studies and population differentiation (Lowe et al., 2007), with the advantage of allowing assessment of inbreeding levels (Weising et al., 1995, Lowe et al., 2007). Such markers were also chosen due to the possibility of detecting hybrids, as suggested by Vaha and Primmer (2006). Microsatellites can even be a reliable alternative to AFLPs in the individual genotyping of polyploids, if one can either resolve allele copy number ambiguity or analyse them as effective dominant markers (Pfeiffer et al., 2011, Dufresne et al., 2014). Although the uncertainty related to allele frequency estimation is also a disadvantage when compared to SNPs (Putman and Carbone, 2014), microsatellites usually perform better than such markers in short temporal- or spatial-scale studies, used when both good resolution and cross-species range are required, in taxa that are slowly evolving or highly clonal or to study genome evolution (Putman and Carbone, 2014). Despite difficulties related to the scoring process, determination of allele dosage in polyploids or null alleles, which represent disadvantages when compared to SNPs, microsatellites have been used as powerful markers in population genetics (Putman and Carbone, 2014), allowing the analysis of population structure and address phylogeographic questions in polyploids, with the possibility of including diploids and polyploids in the

same analysis (Dufresne et al., 2014). A drawback of using microsatellites in *Ophrys* is that high mutation rate, assortative mating in natural populations and the occurrence of hybridization events (Soliva et al., 2001) mean that loci do not to assort independently, preventing us from assuming strict Hardy-Weinberg equilibrium when measuring genetic variation and choosing the statistics to use.

As a consequence of being highly mutable markers, inferences based on such data should consider their mechanism of mutation. The theoretical mutation model that generally best describes microsatellites evolution is the stepwise mutation model (SMM) (Slatkin, 1995, Lowe et al., 2007, Putman and Carbone, 2014), according to which each mutation event acts on a single repeat unit, giving rise to selectively neutral alleles in a stepwise manner (Kimura and Ota, 1975). Regarding allele copy number uncertainty, recent analytical approaches considering allele dosage ambiguity such as the R packages POLYSAT (Clark and Jasieniuk, 2011) or POPPR (Kamvar et al., 2014), by allowing different ploidies in the same dataset and calculating distance matrices not based in allele frequencies, may help overcome some of the difficulties mentioned and make use of the full potential of such markers in what regards genetic distance-based statistics.

4.1.1 Aims and scope of the chapter

In this chapter one attempts to assess **gene flow (indirect methods) and detect hybrids and introgression** using microsatellites. Results from these markers will provide information to help in deciding which of three scenarios best represents the genetic evolution of the entities studied, *O. fusca* and *O. dyris*:

(A) The plants studied are clustered in **two main groups**, corresponding to the two species, maintaining their genetic integrity. Admixed genotypes correspond to sporadic hybridization events.

Three different genetic groups arise from the dataset, gene flow happening to such an extent that it has resulted in a third genetic group, in one of the following scenarios:

(B) The species exchange genetic material, gene flow happening equally in both ways, the third genetic group with ancestry equally shared between both species.

(C) One of the species is acting as a "genetic donor", whereas the other is gaining new genetic material, in a one-way introgression process. This species will tend loose its genetic integrity in the future, while the first tends to keep it. The third cluster is expected to include a higher proportion of genotypes from the "donor" than from the "receptor-species".

By making comparisons between datasets and analyses for subset of data, the following questions will be addressed: (i) Do genetic clusters coincide with morphological groups? (Nieto Feliner et al., 2017) Do levels of pollination success differ in different genetic clusters?

Microsatellite data were collected from 20 loci screened at Jodrell Laboratories in 2013 plus eight loci developed by (Cotrim et al., 2009) and tested at the Botanical Garden and Natural History Museum in Lisbon.

4.2 Methods

4.2.1 Plant Material

In 2011, 2012 and 2013, natural populations from the target species in the centre of Portugal were surveyed during the flowering season, regardless of the plant species (Figure 27). Populations sampled include sites where only one of the taxonomic entities is represented (*Montejunto* - Mj and *Serra de Santo António* - SA, for *Ophrys dyris*; *Arruda dos Vinhos* – AV and *Casalinho do Facho* - CF for *O. fusca*) and sites where the two taxa seem to occur (*Mendiga* – Me; *Pinheirinhos* – Pi). Populations are here designated by the local name and a code number based on this for each sample (Table 16). Aside with the measurement of morphological and phenological data from 272 tagged plants, unpollinated flowers from 168 individuals were collected for DNA extraction (dry weight ranging from 0.15 to 1.5 g per plant, corresponding to one to two flowers per individual). Since 2012 plants were sampled for both DNA and flow cytometry measurements, pollinia being detached from the flowers (see Flow cytometry in section III). Plants for DNA were randomly sampled amongst the tagged plants across each population area. Fresh material for DNA extraction was stowed in Falcon tubes with a few drops of water and taken to the lab within the following 24h, where they were stored at -80ºC.

Table 16. Location, taxonomic composition and plant numbers of each population sampled. Pt – number of plants tagged; Ps – number of plants sampled for DNA; Pp – number of plants sampled for ploidy analysis

Pop. Code	Name and county	Taxonomic composition	Pt	Ps	Pp
AV	Arranhó, Arruda dos Vinhos	*O. fusca*	38	35	35
CF	Casalinho do Facho, Sesimbra	*O. fusca*	44	30	5
Me	Mendiga, Porto de Mós	*O. fusca* + *O. dyris*	63	35	5
Mj	Serra de Montejunto, Alenquer	*O. dyris*	65	26	9
Pi	Pinheirinhos, Sesimbra	*O. fusca* + *O. dyris*	39	26	3
SA	Serra Sto António, Alcanena	*O. dyris*	23	13	10
			272	165	67

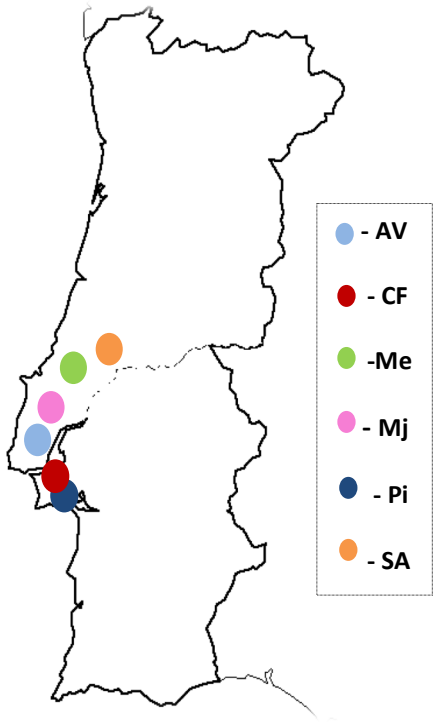


Figure 27. Map of Portugal with the approximate location of the populations sampled.

4.2.2. DNA isolation and molecular analysis

DNA was isolated from flowers preserved at -80C the Laboratory of Genetics of the National Museum of Natural History in Lisbon, using an adapted a 2x CTAB method (Doyle, 1987) as modified by Weising et al. (1995). Changes to the CTAB method include an increase in the amount of β-mercaptoethanol to 100 µl per sample, a decrease in the time of incubation at 65ºC to 15 minutes and always centrifuging at 4ºC. The amount of washing solution (ethanol and ammonium acetate) used was also adjusted according to the dimensions of the pellet for each sample.

The quality and approximate amount of total DNA obtained was initially assessed through agarose gel electrophoresis. According to the percentage of RNA visually estimated and the volume of DNA, samples were treated with RNAse at 100ul/µl and then the final amount of DNA quantified using a UV-Vis spectrophotometer (Nanodrop 2000, ThermoScientific). In total, 167 samples of total DNA were obtained, and 165 were analysed with microsatellites.

An outsourced microsatellite library was produced by Genoscreen, from a set of DNA samples chosen as representatives of the genetic variability found across all the target populations (seven samples equally pooled making 100µl of DNA at 10 ng/µl). Forty-eight primers have been developed, and 36 have been validated as polymorphic on agarose gels (**Table 17**). The PCR and amplification conditions were tested for those 36 primers on seven individuals chosen as reference-samples for the initial tests

– AV20, CF42, Me62, Me70, Mj61, Pi82, SA12 - plus a negative control. After the tests phase, 20 successfully amplifying primers were run for the whole set of samples (167), using standard polymerase chain reactions (PCRs).

Table 17. Primer pairs successfully amplifying and run for 166 individuals. Loci marked * correspond to those analysed in the current report. Ann. Temp. – annealing temperature used (temperatures where efficiency of PCR amplification was maximal.

Locus	Ann T (ºC)	Forward primer (5'-3')	Reverse primer (5'-3')	PCR prod. size (app.)	Rpt Motif
01*	**51**	GAGGAGGATGATCCAAGTTGT	CACCTTTTTCTCATCGATTTTG	140	AG
02*	**54**	GCATACTCGATATGGGGTCG	TCCCATGCAGTTAGGAGTGA	138	TTC
03	**55**	ATACCGGCAATCAAACCGAG	CAAGAATGACGCCTGCTAGA	202	GAG
05*	**55**	TCTCCAATTTTGAGGACTTGC	CACATACGACTTCAGTGTCTCG	141	GAA
06	**55**	TAGTTGCAAGAACCAAGGCA	CAACGTGTCCAAAATTATAACGA	166	CA
10*	**53**	TCAATTTATCCCTTGCGGTG	TCACCCTTATTTTCTTTCCCTTC	172	AGA
11*	**55**	AGATTTAGCATCCGTCAGCG	ACCTGATGGCCACATTTGAT	118	GGA
15	**55**	TAACAGAGGCGAAGAAGGGA	CCTGCAGATCTCACCACCTT	100	GAA
16	**55**	CTGTTGAGTGATGCAGAGCA	TCATCTCTGTGTCCCACCTT	245	AG
20	**55**	GGAGGGGGTTAGGTTGTTAGA	TCCATTCCTCTTGAGCTTCC	125	GAG
21*	**55**	ATCCCCAAAGGGCTGGTAG	CGATCACCTCCTCGTCTTGT	147	GT
24*	**55**	CAACATTATCAGGCCATCCA	CAGTCCTTGACACTGAGACCA	271	TC
28*	**53**	TCTAGGCCATGAGGACTTGG	CCCCACAATGATGTGATGAA	127	CAA
30*	**55**	TTGAGAGCGCGAGGTTAAAT	TAACCTTCCGAGACTGGTGC	202	CT
31	**53**	CGAACTTTGAAGGTTGATTCAG	CCACCCAATAAAAGCAGACAA	171	TG
35*	**54**	TGCATTGCAGCTACCAACAT	TATGCGGAGGAGTCTTTTGC	162	ACC
37	**51**	AAATTGGATTTAAATTGATCGGC	TTGATTAACCCATAAAATCCCG	240	GA
40*	**55**	AAATGCCGCTCCATTCTTCT	ACCTGGTGCACAAGAGGGTA	308	CT
41*	**55**	TCAAAGGACAATCTCCAAAGC	GCAACAGCCTATCCAAGCTC	143	GGA
45*	**55**	TGCGCCCTTATTATCCTCAA	TCCAGCGAGGAAGGAAGATA	139	TCT

For each microsatellite, forward primers have been modified by the addition of a 19bp M13 labelled tail (5'-AGGAAACAGCTATGACCAT-3') at the 5' end, labelled either with 6-JOE or FAM fluorescent dyes, following the method of Schuelke (2000). All genotyping reactions were performed in 10µl, hence using a three-primer protocol, containing approximately 10/20 ng of template DNA.

For the PCR cocktail one used MasterMix kit, 0.4µM of the forward primer, 0.8µM of the reverse primer, 1.6 of the M13-labelled tail (6-FAM or JOE) and 1µg/µl of BSA. Optimal annealing temperatures for each primer were previously tested based on its melting temperature, with standard

PCRs. Reactions followed the method previously described by Lavor (2000) for the standard cycling programme, slightly adapted to the one following described, with annealing temperatures ranging between 51 and 55°C (Table 17): initial denaturation at 94°C for 3 min followed by 40 cycles of 94°C for 20s, 55°C for 40s and 72° for 20s, before 10 cycles of 94°C for 1 min, 53°C for 1 min and 72°C for 1 min to incorporate the M13 tail, and a final extension of 30 min at 72°C. For a first check of amplification success, 2% agarose gel electrophoresis was performed for each genotyping reaction.
PCR products were resolved on a 3730 DNA Analyser (Applied Biosystems) and were sized with Genescan 500 ROX size standard, using Genemapper version 4.1 software (Applied Biosystems), at Jodrell Laboratory. Primers testing, polymerase chain reactions and locus screening of this microsatellite library were performed at Jodrell Laboratory, Royal Botanic Gardens Kew, from August to October 2013.

Eight microsatellite loci previously identified for the species *Ophrys fusca* (Cotrim et al., 2009) have also been tested and optimized for a small group of representative samples (Table 18). Amplifications were performed in 25µl reactions with 25 to 50 ng of template DNA, with 0.5 units/µl Taq polymerase, 200mM of each primer, 0.2mM of each primer, 0.2 DNTPs, 1x Taq buffer and 1.5mM $MgCl_2$ and 1 mg/µl BSA. Thermocycling conditions followed Cotrim et al. (2009). After the initial tests, only a few samples have been screened: 43 samples for OFCA101 and OFGA73, 15 samples for OFCA36 and OFGA83, 11 samples for OFGA8, eight samples for OFCA65 and OFGA32 and four samples to OFCA80. The remaining individuals are still to be screened. This part of the work was conducted at the Laboratory of Genetics of the Botanical Garden and Natural History and Science Museum (Lisbon).

Table 18. Characterization of the eight microsatellite markers previously developed from *O. fusca* by Cotrim et al. (2009). Adapted from the same authors.

Locus	Ann T (ºC)	Forward primer (5'-3')	Reverse primer (5'-3')	Repeat motif	Size of cloned allele
OFCA36	57	TGAGGGGAAATGGAAGGAC	GCAGCCTACTACCTTGATGA	(GA)12	254
OFCA65	57	TGGAGTGCCCCATATTTTGT	AGCACTCATGTTGTCACTTCAGA	(GATA)11 (GA)19	212
OFCA80	56	ATATCTTTACCCTTGGACCTC	GGGTATTAGGTTGAGGATTTAG	(GA)19	255
OFCA101	57	AGGGTATTAGGTTGAGGATTT	CTCGCCTTGCCCTTGCTA	(CT)19	184
OFGA8	57	ATACAAACACGAGACAAGC	AGTTAGGAGTGGGCGGTC	(CT)17	242
OFGA32	57	CCTTGGACTTATGTGTTGA	AAGAGGTCCAAAAGAACGG	(CT)12 CA(CT)7 (GA)8 AA(GA)25	197
OFGA73	57	GCAACCTGCCGAGACTGA	AGGAAGTGAAATGGTGGCA	GG(GA)8	223
OFGA83	57	TCTCCAACTCAAATACATCCA	GTAGGTGAATGTGGAAAGAT	(GA)18	196

4.2.3. Data Analysis

4.2.3.1 Scoring process

Microsatellite sample profiles were scored with GENEMAPPER software version 3.7. (Applied Biosystems). Genotype table automatically generated by the programme, when using the binning mode, was reviewed manually.

The scoring of the microsatellite data was repeated several times, particularly before having confirmed the ploidy level. Interpreting profiles proved to be difficult due to the following: i) profiles seemed to indicate different ploidies within the dataset, most samples appearing to be polyploids; ii) complexity of profiles; iii) having to decide which alleles occur in more than one copy in samples where the number of displayed alleles was fewer than the possible maximum number as displayed in other loci. The scarcity of software to analyse polyploid data was another drawback in the analysis process.

For the first analysis (five primers), considering different published assessments for the ploidy of *O. fusca* and the tetraploidy of *O. dyris* (Bernardos et al., 2003, Devey et al., 2008), any particular ploidy level was assumed, as previous attempts to count chromosomes had been unfruitful. After knowing the plants studied were all tetraploid (chapter III.), we could assume four sets of chromosomes when doing the scoring, simplifying further data analysis.

The absence of information regarding the inheritance mode of such plants was another aspect to consider, as this can be particularly relevant in population genetics when dealing with polyploids in the analysis (Wirth et al., 2009). In disomic inheritance as in allopolyploids, bivalent formation at meiosis causes fixed heterozygosity (Soltis and Soltis, 2000), whereas under polysomic inheritance – autopolyploids - multivalent formation in meiosis can result in random assortment of homologous chromosomes or in random assortment of sister chromatids (Bever and Felber, 1992).

Although we have no evidence of the inheritance mode on the studied individuals, analysis were conducted assuming polysomic inheritance as in autopolyploids. Despite biological differences Meirmans and Van Tienderen (2013) showed that assuming strict autopolyploidy may be valid even if the inheritance is partly disomic. According to these authors, to avoid the bias associated with strict disomy in estimating summary statistics such as population diversity or the amount of population divergence, a low rate of allele exchange between the composing subgenomes should be considered.

Considering all these variables, microsatellites scoring was performed from the beginning in three different ways, creating three different matrices, considering presence/absence of alleles (A) or allele sizes (B, C):

A) Binary data matrix
B) Matrix with allele sizes, assuming allele dosage ambiguity;
C) Matrix with allele sizes, inferring allele number of copies;

A) When scoring microsatellite as binary data, alleles at each locus are coded as present or absent (1/0), being treated as independent dominant loci (Andreakis et al., 2009). Although this removed the potential to treat microsatellites as co-dominant markers, as far as we know this is the only way of using statistical software conceived for diploid data such as GENALEX 6.501, with polyploids (Peakall and Smouse, 2012, Peakall and Smouse, 2006).

B) This scoring approach assumes both ploidy uncertainty and allele copy number ambiguity. It is similar to the scoring of haplotypes, each allele being recoded only once in each sample, independently of peak height and area. Subsequent population genetic parameters and distance-based statistical analysis were possible due to the incorporation of distance measures such as the one from Bruvo - adding virtual alleles and simulating different genotypes (Bruvo et al., 2004) and Lynch (1990) in the R packages POLYSAT (Clark and Jasieniuk, 2011) and POPPR (Kamvar et al., 2014), which calculated distance matrices to be exported to other softwares. Allele frequencies were also estimated in POLYSAT (Clark and Jasieniuk, 2011).

C) In order to consider the variability enclosed in co-dominant markers, allelic configurations have been devised when using this scoring approach. The highest number of peaks initially scored per individual was five (three samples), followed by four in most samples, suggesting tetraploidy. After ploidy had been confirmed, the scoring of such individuals who presented more than four peaks has been re-checked. Being lower polyploids, tetraploids still give us the possibility of determining allelic configurations and provide reliable estimates of heterozygosity (Wirth et al., 2009, Pfeiffer et al., 2011).
Despite some uncertainty when estimating allele dosage in polyploids, considering a matrix of estimated genotypes instead of a presence-absence binary matrix, in this way it is possible to proceed with the subsequent analysis with allele frequencies, rather than

with binary data. For this scoring method I considered both peak ratios and areas provided, although in a manual and a non-quantitative approach, based on the "microsatellite DNA allele counting-peak ratios" (MAC-PR) method, developed by Esselink et al. (2004) for polyploids, regardless of the inheritance mode (Esselink et al., 2004). Whenever proper assignment of alleles was not possible for samples with more than two alleles per locus (ratios of sizes and areas between alleles far from an integer value, or too variable), doubtful alleles were coded as null alleles. Reactions showing the higher number of peaks with similar height and area within the set of loci screened for each sample were considered as reference for the estimation of the minimum number of allele copies for that individual.

4.2.3.2. Genetic distance-based exploratory analysis

The starting point of the exploratory analysis of the genetic dataset was a pairwise matrix, calculated using a distance metric. Genetic distance-based analysis uses a distance or similarity measure to assess how far apart two individuals or populations are in a n-dimensional marker hyperspace (Lowe et al., 2007). The most used Euclidian distance is not applicable to ambiguous genotypes (B) and mixed ploidies. One have chosen the Bruvo distance measure (Bruvo et al., 2004) to compute the pairwise distance matrix used in subsequent analysis. This measure was specifically developed for microsatellites in polyploids and can be used for mixed ploidy levels. It allows dosage uncertainty, by computing the average over all possible allelic constitutions (Bruvo et al., 2004, Dufresne et al., 2014).

Bruvo distance matrix was calculated in POLYSAT (Clark and Jasieniuk, 2011), on one of the initial base matrices, assuming ambiguity in number of allele copies (B). A principal component analysis (PCA) was performed on Bruvo distance matrix from POLYSAT, to detect clusters of individuals, prior to defining the number of statistical populations. Axis coordinates and eigen values were therefore obtained from this software, graphics being plotted in MICROSOFT EXCEL 2010.

4.2.3.3. Genetic Diversity Summary statistics

Genetic diversity metrics computed for each population and locus include: number of alleles, $\mathbf{N_A}$; allelic richness, $\mathbf{A_R}$; observed heterozygosity, $\mathbf{H_o}$; Nei's gene diversity corrected for sample size, $\mathbf{H_E}$; number of effective alleles, $\mathbf{N_{AE}}$; individual inbreeding coefficient, $\mathbf{F_{IS}}$, calculated as kinship coefficients between gene copies within individuals. All the mentioned parameters were calculated for the dataset with inferred allele sizes (C) using SPAGeDi 1.5. (Hardy and Vekemans, 2002). Deviation from Hardy-Weinberg equilibrium (HW equilibrium) in each population were assessed with the F_{IS} individual

coefficient, which measures the reduction in heterozygosity of an individual: a positive value indicates a deficiency of heterozygotes compared to HW assumptions, whereas a negative value means excess of heterozygotes (Wright, 1949, Waples, 2015). Linkage of markers - linkage disequilibrium - was tested using the metric implemented in SPAGEDI, Identity Desequilibrium (g2g) (Hardy, 2015), per locus pair and population. Because of tetraploidy of plants, the assumption of multisomic inheritance in the statistical analysis, and the uncertainty on the segregation mode, deviations from HW proportions were not assessed directly, through exact tests, which is the common practice when working with codominant markers in diploid datasets. The same assumptions applied to LD-testing.

4.2.3.4. Genetic differentiation and structure

To understand how genetic variation is partitioned among populations (genetic differentiation), different approaches have been used:

1) As an alternative to F statistics, the *Rho* Statistic, ρ (Ronfort et al., 1998), was computed in SPAGEDI software for the data matrix with the inferred genotypes. This metric is frequently used when deviations from HW equilibrium occur, assuming the stepwise mutation model (SMM) for mutations. In recent simulation studies Meirmans and Van Tienderen (2013), *Rho*-st was found to be the only measure of population differentiation that was independent of the ploidy level, rate of double reduction, selfing rate and unbiased by the inheritance mode. Other authors reinforced this choice as the appropriate metric to assess population differentiation in polyploids with unknown segregation or mode of inheritance (Meirmans and Van Tienderen, 2013, Dufresne et al., 2014, Meirmans et al., 2018). R_{ST} (Slatkin, 1995), another estimator for genetic differentiation for microsatellite loci, was also calculated in the same software.

2) Analysis of molecular variance – AMOVA (Excoffier et al., 1992, Lowe et al., 2007). GENALEX was used for AMOVA on the binary data (A)
   Significance of variance components was obtained through 1000 permutations. With AMOVA we tested differentiation between the six populations. Phi-statistics - Φ (Cockerham, 1973, Cockerham, 1969) correlation statistics - analogous to F_{ST}, were generated. Other hierarchies will be further tested, with genotypes nested within taxa and taxa nested within populations.

3) Bayesian clustering methods were used to infer population structure, using STRUCTURE software v2.3.4. (Pritchard et al., 2000). The basis of such method is to assign individuals to

one or more clusters so that deviations from HW equilibrium are minimized (Pritchard et al., 2000). This software does not require prior information on the sampling locations neither on groups of individuals and uses Bayesian approach in order to find the best number of groups (K). As the dataset includes individuals from closely related species that may occur in sympatry, calculations were carried out under the admixture model of ancestry, assuming correlated allele frequencies.

Clustering methods were performed on data matrix where allele number of copies has been inferred (C), using 163 samples. Although individuals have been collected from geographically isolated populations, in the initial set of runs no prior assumptions were made on population membership or number of clusters (K), STRUCTURE being used to devise genetic clusters present in the global dataset. A set of independent runs was conducted for different values of K, between 1 and 7 (the number of sampling sites plus one). In each simulation 20 iterations where performed, with a burn-in length of 100 000 and a run length of 100 000 MCMC repetitions. The maximum ploidy level chosen was 4x. Individuals that had been recorded as having five alleles in previous analysis were carefully re-analysed, in the light of the confirmed tetraploidy.

The choice of K was carried out both following the criterion of the highest value of ln P(D) and α, as it is given by STRUCTURE, and the method of Earl and vonHoldt (2012), as implemented by Earl and vonHoldt in STRUCTURE HARVESTER software, version 6.94 (2012). Using the first criterion, once real *K* is reached, Ln P(D) at larger *K*s stabilizes or continues increasing slightly, and the variance between runs increases (Evanno et al., 2005). Ln(P) usually settles down once the MCMC converges, with a suggested range of 0.2 (Pritchard et al., 2000). From this plateau we have chosen the lowest value of ln-likelihood. STRUCTURE HARVESTER, a web-based software, uses STRUCTURE result files and employs an algorithm to determine if the Evanno method can be used (Earl and vonHoldt, 2012). Conditions being met, four plots are produced: mean likelihood and variance per K and three other graphics with the Evanno method results – mean difference between successive likelihood values L(K), second order rate of change of L(K) between successive K values and delta K, ΔK (mean of the absolute values of L''(K) (Evanno et al., 2005). The choice of the number of clusters is based on the last parameter – the K value which precedes the highest value of ΔK.

As recommended by Pritchard et al. (2010), the number of genetic clusters chosen using such statistics should be consistent with a biological interpretation of the assignments. After having chosen the value of K, this was used in the following runs performed, to obtain estimated allelic frequencies and membership proportions (q) of each individual in each inferred cluster.

Bayesian clustering methods implemented in STRUCTURE were also used to study patterns of hybridization and detect individuals with admixed genotypes due to introgression between clusters, as suggested by Vaha and Primmer (2006), Burgarella et al. (2009a) or Field et al. (2011), although authors refer this method as being reliable only for many loci, particularly when differentiation is low (Vaha and Primmer, 2006, Putman and Carbone, 2014). If considering genetic admixture model and K=2, each individual is assigned a proportion of genome (Q) inherited from each K cluster. Following this method, an admixed individual would have a Q value intermediate (to some extent) between the two clusters. Therefore, first-generation hybrids are expected to have a Q value of 0.5 (Vaha and Primmer, 2006).

The threshold Q values of 0.10 and 0.90 were chosen to separate introgressed from purebred individuals, 0.90 being the lower limit for a robust assignment of individuals to one of the clusters, and 0.1 the higher value for the other cluster. These values were proven to show the best performance in STRUCTURE software, although for a higher number of primers (Vaha and Primmer, 2006, Canestrelli et al., 2014). Following this criterion, any individual with Q between 0 and < 0.10, or > 0.9 and 1 is classified as purebred, and individuals with *Q*-value between 0.1 and 0.9 are classified as introgressed (Vaha and Primmer, 2006). From the introgressed individuals, hybrids were considered to be those which Q values ranged between 0.35-0.65, although this does not allow us to distinguish between F1 hybrids and backcrosses. For such introgression assessment and hybrids detection with STRUCTURE, a subset of reference samples was chosen to be considered as parental individuals. These were picked from the "purebred" individuals, which had been strongly assigned to each cluster in the previous runs (Q > 0.95) and phenotypical identification of which has been confirmed. Reference samples have been identified with the function POPFLAG and the option PFROMPOPFLAGONLY has been used, as described by Pritchard (2010). Twenty-five samples have been chosen from each cluster. Pie charts with the results were plotted in Microsoft Excel 2010 (Figure 31).

Although the number of clusters found (K=2) corroborates the initial assumption of the presence of the two taxa and intermediate individuals, this procedure could only be used for interspecific hybrid detection when confirming if the composition of each cluster matches the taxonomical groups, after individuals being assigned to each taxon. Admixture proportions of each individual only allow robust conclusions concerning hybridization between the two described taxonomic entities, *O. fusca* and *O. dyris*, to be drawn after such taxonomic assignment.

4.2.3.5. Comparing morphological and genetic datasets

To study the relation between genetic and morphological datasets, a Mantel test (Mantel, 1967, Mantel and Valand, 1970) was performed between distance matrices obtained for both datasets; a distance matrix (squared Euclidean distances) was computed in MSVP software for the morphological dataset, so that the comparison with the Bruvo distance matrix computed in Polysat for the genetic dataset would be possible. The test was performed with the data from 49 samples common to both datasets. Genetic data analysis and morphological analysis for the same individuals were also compared through plotting Q values against the morphological score (from Discriminant Analysis function). The test performed on the distance matrices used Pearson correlation, with r being estimated from 10000 permutations.

4.2.3.6. Comparing seed set with genetic data

In order to relate pollination success data to the genetic clustering of plants, plants from the first dataset have been re-classified according to the genetic groups in which they had been placed. Classification of plants followed the hybridization assessment, according to the groupings identified using software Structure software (Pritchard et al., 2000). Genetic groups defined by Q values obtained from such analysis - *hybrids*/*cluster 1*/*cluster 2* – were attributed to the plants shared between both datasets: 29 plants from cluster 1 (putatively *O. dyris*), six plants from cluster 2 (putatively *O. fusca*) and five hybrid plants. Percentages of pollination success per genetic group were then computed in XLSTAT, version 2018.2 (Addinsoft, 2018). Only data for flowers with both male (pollinia removal) and female (inflated capsule) pollination success were considered for this analysis. A Kruskal-Wallis test was performed in XLSTAT, for differences between the three groups distributions.

4.3 Results and Discussion

4.3.1. Summary diversity statistics

Table 19. Genetic diversity summary statistics per population (# number of individuals sampled), computed in Spagedi software with data from 13 loci: number of alleles with non-zero (N_A); number of effective alleles (N_{AE}); allelic richness corrected for sample size (k=4 gene copies) (A_R); Nei's gene diversity corrected for sample size (H_E); observed heterozygosity (H_O); individual inbreeding coefficient (F_I).

Population	#	N_A	N_{AE}	A_R	H_E	H_O	F_I
AV	(34)	8.46	4.26	2.71	0.7186	0.665	0.090
CF	(30)	8.92	4.08	2.67	0.6996	0.659	0.062
Me	(35)	10.31	4.90	2.76	0.7277	0.627	0.150
Mj	(26)	9.38	4.32	2.72	0.7122	0.622	0.146
Pi	(25)	7.54	3.90	2.62	0.6902	0.666	0.047
SA	(13)	6.69	3.93	2.60	0.6852	0.654	0.055
Overall	(163)	16.85	4.91	2.81	0.7387	0.648	0.134

In total, 17 alleles (N_A) were found across all the sampled populations and loci, in a total of 163 individuals included in the analysis. Mendiga (*Me*) was the population with the higher number of alleles, 10.31, followed by Montejunto (*Mj*), with a value of 9.38. The same population (*Me*) also had the highest number of effective alleles N_{AE}, 4.90, and allelic richness, A_R,2.76. The second highest value of N_{AE} was also in *Mj*, 4.32 and the lowest value was in Pinheirinhos (*Pi*). Despite also being low for *Pi*, the lowest figure for A_R was found in Serra Sto António (*SA).* When thinking of the taxonomic diversity found in these two populations, these relative values are not surprising. *Me* is, from all sites that have been sampled, the one where plants from both taxa interspaced with others of different mixed morphotypes are clearly observed, whereas *SA* is a population almost exclusively consisting of *O. dyris*, according to our expectations from field work. *Me* had the lowest H_o value among all populations, despite also having the highest value for H_E, 0.7277. The highest value for H_o was for *Pi.*

Table 20. Genetic diversity summary statistics by primer. Mean values across all loci given in Table 1, as overall values.

Locus	PCR prod. size (app.)	Rpt Motif	N_A	N_{AE}	A_R	H_E	H_O	F_I
01	140	AG	6	3.48	2.62	0.7123	0.802	-0.123
02	138	TTC	16	4.43	2.87	0.7742	0.497	0.356
05	141	GAA	15	4.94	2.97	0.7974	0.837	-0.042
10	172	AGA	24	7.36	3.28	0.8641	0.721	0.182
11	118	GGA	15	2.88	2.50	0.6525	0.631	0.045
21	147	GT	20	6.82	3.23	0.8535	0.762	0.129
24	271	TC	5	1.55	1.70	0.3529	0.298	0.184
28	127	CAA	28	6.19	3.19	0.8384	0.462	0.453
30	202	CT	29	4.04	2.81	0.7527	0.725	0.048
35	162	ACC	11	2.18	2.06	0.5416	0.515	0.053
40	308	CT	19	10.49	3.47	0.9047	0.806	0.122
41	143	GGA	11	3.71	2.69	0.7302	0.685	0.072
45	139	TCT	20	5.84	3.10	0.8287	0.683	0.194

Of the loci screened, primer 24 had the lower number of alleles (5), and primer 30 had the highest, 29. Allelic richness (A_R) was highest for locus 40 (3.47) and lowest for locus 24 (1.70) (Nei's gene diversity corrected for sample size (H_E) ranged between 0.3529 for locus 24 and 0.9047 for locus 40, and observed heterozygosity (H_O) ranged from 0.298 for locus 24 to 0.837 for locus 5. Apart from the low value for H_O, primer 28 also showed an extremely high value for F_{IS}, possibly a consequence of a large number of homozygotes.

4.3.2. Allelic patterns, Inbreeding, Selfing rate and Linkage Disequilibrium

The lowest level of individual inbreeding was found for *Pi*, 0.047, whereas the population with highest F_I, when assessed with this metric, was *Me* (0.150). *Me* was also the site at which more private alleles occurred (19), as assessed by GENALEX from the binary data matrix. *Mj* followed *Me*, with 13 private alleles. The lowest number of private alleles was found for *SA* (five); see Figure 28. *Me* was the population in which, on the basis of phenotype, both species seem to occur in sympatry, although Bayesian clustering data did not strongly (Q>0.90) assign any individual from *Me* to one of the genetic

clusters. *SA* was the population in which *O. fusca* does not appear to occur, and this was confirmed by genetic data (Table 23).

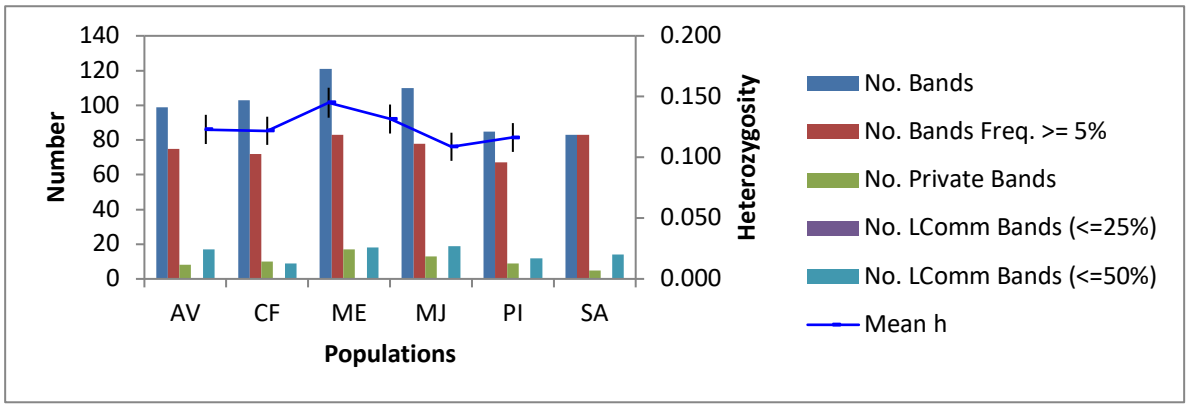


Figure 28. Allele patterns across populations. Data computed in Genalex from the binary data matrix, 163 samples, 13 loci.

4.3.3. Population differentiation

To assess population differentiation, values of mean *Rho* (Stace) were computed and are presented in Table 21, with values for Wright's F-statistics. Created for autotetraploids by Ronfort et al. (1998), this metric was not assessed per population but for each locus, amongst all the populations. Because of the frequency of double reduction in autopolyploids, segregation patterns are expected to vary among loci, and therefore *Rho* was not calculated for populations as a whole.

Table 21. Genetic differentiation indices calculated in SPAGEDI 163 individuals, from the genotypes matrix (C). Mean F_{ST}, R_{ST} – mean population differentiation; F_{IS}, R_{IS} – inbreeding coefficients for F and R statistics; Pairwise Rho – pairwised population differentiation (Ronfort et al., 1998), using the ANOVA framework.

	Mean F_{ST}	F_{IS}	R_{ST}	R_{IS}	Pairwise Rho
All loci	0.0431	0.1017	0.0588	0.1834	0.1359
01	0.0448	-0.1665	0.0179	-0.2721	0.3694
02	0.0293	0.3397	0.0267	0.2505	0.0502
05	0.0379	-0.0756	0.0242	-0.0967	0.2137
10	0.0616	0.1383	0.1384	0.1680	0.1488
11	0.0215	0.0273	0.0724	0.0313	0.1117
21	0.0188	0.1157	-0.0036	0.0838	0.0455
24	0.0267	0.1652	0.0665	0.1457	0.0594
28	0.0716	0.4188	0.0122	0.6890	0.1516
30	0.0394	0.0161	0.0236	0.0638	0.1754
35	0.0190	0.0382	0.0106	0.0129	0.0737
30	0.0658	0.0731	0.0983	-0.0032	0.2574
41	0.0375	0.0417	0.0116	0.1460	0.1398
45	0.0581	0.1531	0.1350	0.3174	0.1835
Mean	0.0432	0.1021	0.0594	0.1842	0.1165
SE	0.0054	0.0452	0.0222	0.0817	0.0146

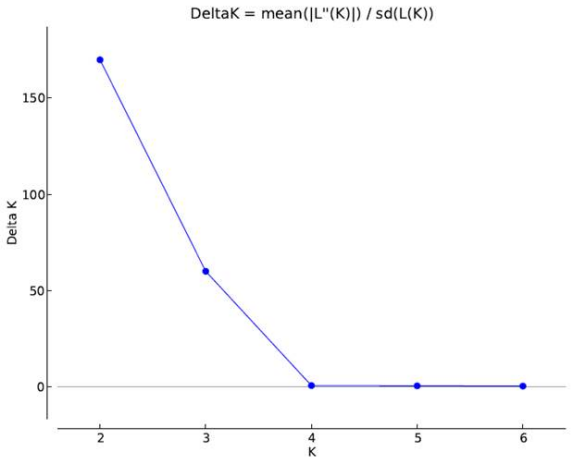
Global pairwise *Rho* coefficient (0.1359) indicated that there was significant differentiation between populations for all loci, as it significantly deviates from 0. When calculated per locus, it differed greatly from locus to locus, being lowest for locus 2 (0.0502) and locus 21 (0.0455) and highest for loci 1 (0.3694) and 30 (0.2754), that may be considered as moderate values.

When looking at AMOVA (Table 22), aiming to order to understand how genetic variability is portioned, results showed that most variation (79%) is contained between individuals within populations, and 21% is found among populations (P< 0.001). These results corroborate the idea of low genetic differentiation between populations (see below). Low genetic differentiation can also be associated with a high level of allele sharing between populations (25-50 % of common alleles) and the reduced number of private alleles, as shown in Figure 28.

Table 22. AMOVA table computed in Genalex.

Source of variation	d.f.	SS	MS	Variation	Φ - statistic
Among Pops	5	37.544	7.509	19 %	
Within Pops	157	157.535	1.003	81 %	0.195

Regarding the search for the number of genetic clusters using Bayesian clustering analysis, from both optimality criteria - the maximum value of ln P(D) and α provided by STRUCTURE and STRUCTURE HARVESTER (Earl and vonHoldt, 2012) - 2 was the number of genetic groups (K) within the dataset that gave optimal results (Figure 29).



K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	30	-12.315.793.333	1.266.255	—	—	—
2	30	-11.828.676.667	1.838.793	487.116.667	312.426.667	169.908.527
3	30	-11.653.986.667	3.577.394	174.690.000	214.480.000	59.954.259
4	30	-11.693.776.667	445.957.611	-39.790.000	242.403.333	0.543557
5	30	-11.491.163.333	268.476.212	202.613.333	107.730.000	0.401265
6	30	-11.396.280.000	52.747.390	94.883.333	14.456.667	0.274074
7	30	-11.315.853.333	90.976.771	80.426.667	—	—

Figure 29. Graphic plotted by STRUCTURE HARVESTER showing the variation of deltaK when testing several values for K (1-7) in Structure software. Below is the Evanno table provided by the same application (adapted).

The fact that both methods indicate the same number of genetic clusters, corresponding to the number of taxonomic entities considered here, allied to the proportion of plants from each population assigned to each cluster in Table 23, appears to indicate that the two species are still the main genetic groups and that introgression and hybridization events have not yet led to a genetic fragmentation of such groups. Table 23 illustrates the assignment of individuals to the genetic groups, after running Structure for K = 2, through the barplots with the distribution of admixture proportions (Q) for each

individual. Cluster 1 groups most individuals from *Mj*, *SA* (considered as populations of *O. dyrisI*) and *Me* (where both taxa occur in sympatry) in similar proportions, and cluster 2 groups most individuals from *AV* and *CF*, the populations phenotypically considered as pure populations of *O. fusca*. *Pi* was the population with the highest percentage of individuals assigned to different clusters: 51% to cluster 2 and 49% to cluster 1, all the range of Q values considered. Along with *Me*, this is a population in which both taxa occur. Considering only individuals strongly assigned to each cluster (Q>0.95), proportions of assignment in each population were as shown in Table 23.

Table 23. Numbers (#) and percentage (%) of individuals strongly assigned (Q>0.95) to each cluster. K = 2

		Cl.1 (> 0.95)		Cl.2 (> 0.95)	
Pop	**#**	**#**	**%**	**#**	**%**
AV	34	0	0	13	38.2
CF	30	0	0	10	33.3
Me	35	14	40	0	0
Mj	26	8	30.8	0	0
Pi	25	0	0	3	12
SA	13	5	38.5	0	0
Total	163	41		26	

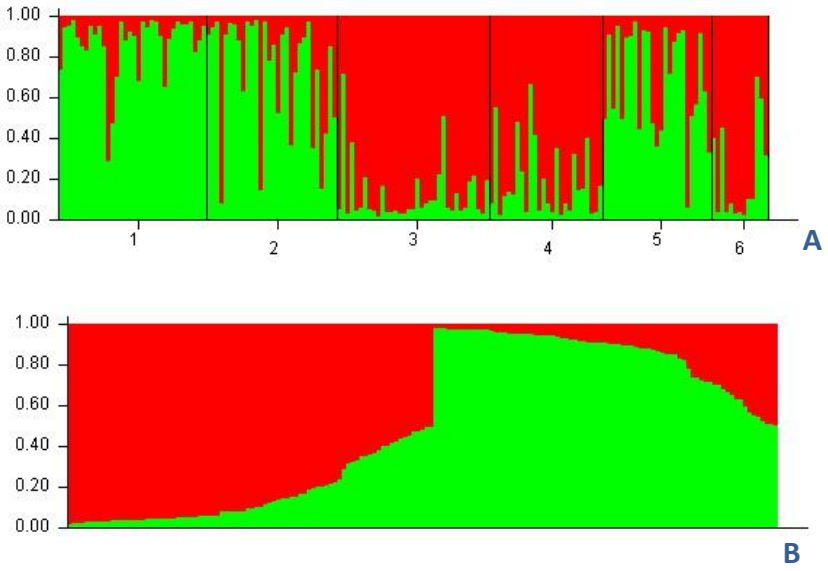


Figure 30. Individual admixture proportions (Q) given by Structure software for each individual of the six populations (K=2; 13 loci). Data plotted by original order (A) and by Q (B), for 163 individuals from all the populations sampled. Numbers 1 to 6 correspond to field populations: 1. AV; 2.CF; 3. Me; 4. Mj; 5.Pi; 6.SA

4.3.4. Gene flow and hybridization

Regarding the analysis of introgression using STRUCTURE, having chosen reference samples to be identified as parentals and considering two main clusters (K=2), all the populations seem to include individuals with mixed genotypes (Figure 30), rather than populations with only one cluster represented or populations with both clusters equally represented, as our initial hypotheses of one-taxon *versus* mixed-taxa populations might suggest.

Analysis of admixture proportions obtained this way show that almost half of the individuals, 44.8% (73 individuals), have mixed ancestry ($0.1 \leq Q \leq 0.9$). The population contributing the highest proportion was Pinheirinhos (*Pi*), with 17 individuals of admixed ancestry, corresponding to 68% of the individuals sampled in that population. Conversely, Serra de Santo António (*SA*) had the lowest number of introgressed plants (two; 23%). Following *Pi*, the second higher number of individuals with admixed ancestry was recorded for Casalinho Facho (CF), 60% (18).
Regarding hybrids, identified by Q values ranging from 0.35 to 0.65, they were identified in largest numbers in the *Pi* population, followed by *Mj* and *CF*, with four individuals each (0.28). *AV* was the population in which no hybrids were found.
Some of the results plotted in Table 23 were not previously expected, on the basis of field observations and morphometric data. Particularly from populations in which *O. dyris* is represented: *Mj* was phenotypically considered to consist of *O. dyris* alone, and the proportion of introgressed plants was even higher than the proportion of plants assigned to cluster 1. Also, in *SA*, despite the strong assignment to cluster 1, the proportion of hybrids was higher than of the remaining introgressed specimens. On the other hand, *CF* was not a population in which one would expect to have the second higher number of admixed plants, as it has been considered as a "pure" population of *O. fusca*.

Concerning genetic introgression, data suggest a genetic input from cluster 2 to cluster 1, as 50% (10 plants) of the introgressed individuals had admixture coefficients closer to the values of cluster 2, and only 15% (3 plants) were skewed towards cluster 1, considering the range of values $0.7<Q<0.9$.

A relevant aspect to consider is that, from the current results, many admixed plants occur in supposedly "pure" populations. Also, as mentioned before, efficiency and accuracy of hybridization and recent gene flow assessments should be further tested through simulation studies, simulating hybrids from individuals defined as pure parentals. At this stage, there appears to be a trend to genetic cluster 2 to be more represented in *O. fusca* and cluster 1 in *O. dyris*. If this is the case, *O. dyris* is the taxon which seems to be receiving genetic input from *O. fusca*, as genetic introgression seems to be occurring from cluster 2 to cluster 1, as mentioned above. The maintenance of this gene flow direction

through time will lead to a decrease of the genetic diversity of *O. dyris,* making this species more vulnerable. As we know from field observation and biogeographical data for both species, this is the one with a more restricted presence and lower habitat diversity. The persistence of such genetic dynamics could therefore be a threat to its future conservation.

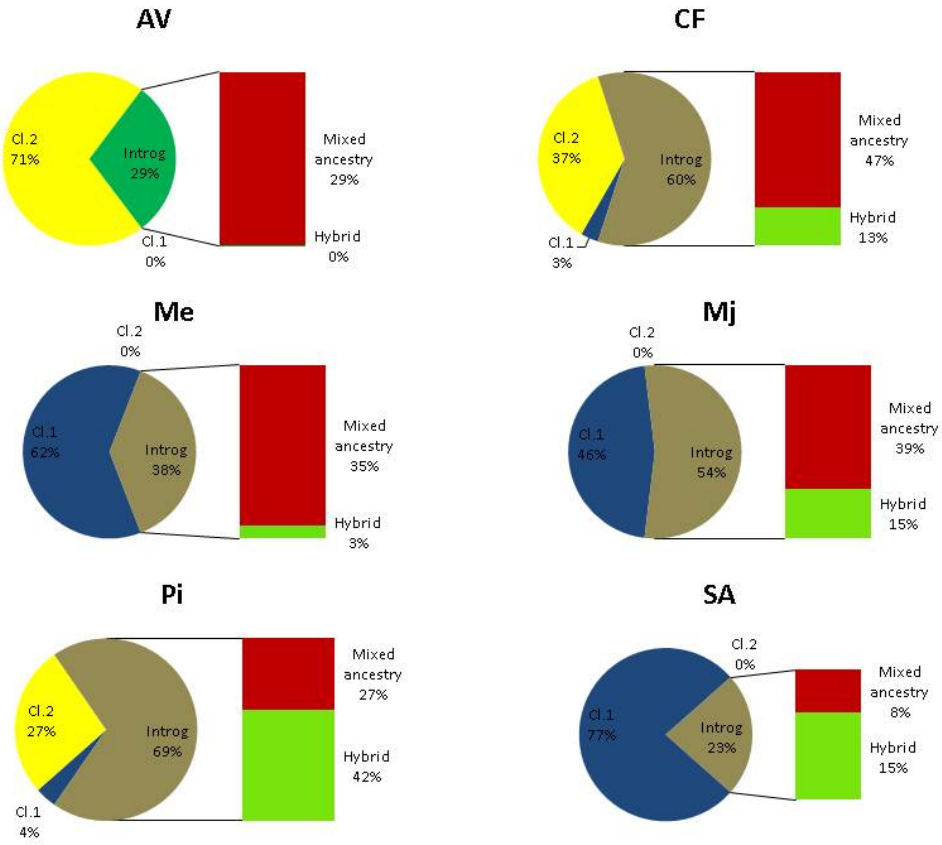


Figure 31. Admixture proportions per population, as given by STRUCTURE software, with K=2. Threshold values for mixed genotypes of 0.1 – 0.9. Threshold values for hybrids of 0.35-0.65. Cl. 1: cluster 1; Cl. 2: cluster 2; Cl. 3: cluster 3; Mix. gen.: mixed genotypes between clusters 1 and 2.

4.3.5. Multivariate statistics

In the principal component analysis (PCA) performed on Bruvo distance for all individuals of the genotypes data matrix (Figure 32), the two first eigen values explain 21% of the overall variability: the first axis approximately 12.0%, the second explaining approximately 10.0%. The graphical representation reveals no clear differentiation between sampling sites, but allows individuals grouping based on the taxa that predominates in each population to be inferred. *AV* and *CF* populations are composed of *O. fusca* and are located mainly in the lower part of the plot, *Mj* and *SA* are populations of *O. dyris* from which most individuals are placed in the upper part of the plot. *Pi* and *Me* are populations where both taxonomic entities occur, with several intermediate morphotypes.

Despite the distribution of these plants does not exhibit a clear trend, *Pi* individuals are scattered throughout the plot, although mostly in the lower part. *Me* specimens, the other mixed population, do not allow us to confirm the tendency of taxa-overlapping region in the centre of the plot.

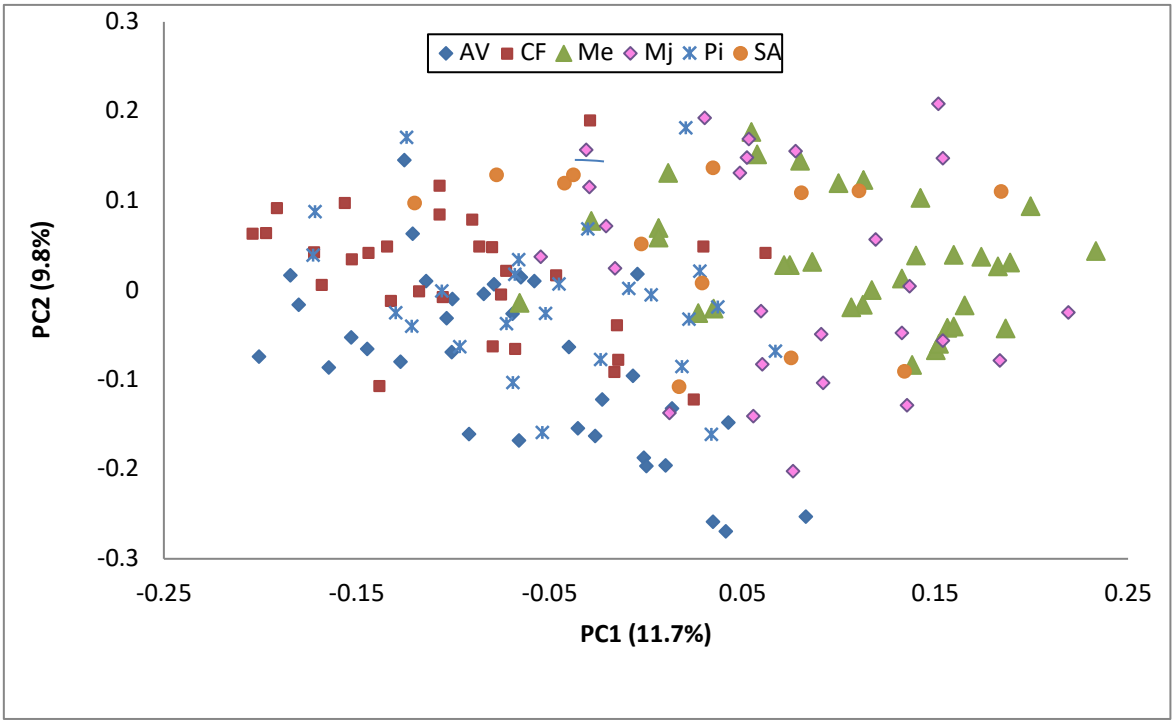


Figure 32. PCA performed on Bruvo distance obtained in Polysat, on the 163 individuals from the genotypes data matrix (C). Eigen values and coordinates x, y obtained in Poysat. Dots plotted in Excel. The first axis explains 12.0% of the total variation in the dataset, the second 10 %. Collored symbols are according to sampling populations: AV and CF are populations of *O. fusca*; Mj and SA are populations of *O. dyris;* Me and Pi are mixed populations.

4.3.6 Mantel Test

The Mantel test performed to compare morphometric and genetic distances matrices revealed a positive correlation, although weak (r = 0.116, p < 0.0001)

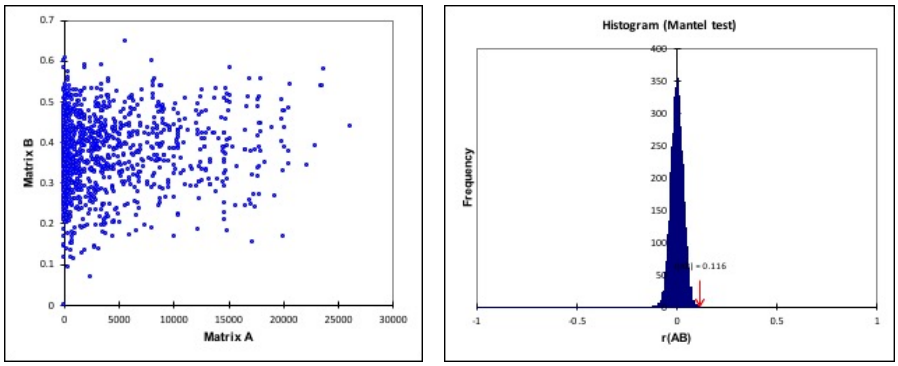


Figure 33. Mantel test performed on the distance matrices of the morphological and genetic datasets. A. Scatterplot of the cases of both matrices; B – Distribution of r (AB), estimated from 10000 permutations.

4.3.7. Bringing together seed set and the genetic data analysis

The hybrids group has slightly fewer (24.8%) and the *O. fusca* group 2 slightly more (31.4%) filled capsules. The *O. dyris* group 1 group showed an intermediate level of capsule filling (Figure 34). However, statistical comparison of the percentages of filled capsules per genetic group using the Kruskal-Wallis test confirmed that the three genetic groups show no significant differences in pollination success (p-value = 0.954).

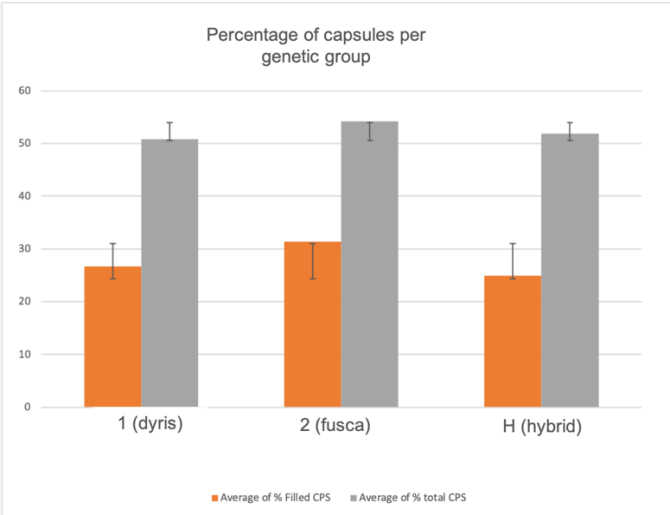


Figure 34. Percentage of capsules per genetic group (average values), dyris (1), fusca (2) and hybrid (H), measured as the number of inflated capsules per number of flowers in each plant. The left-hand side bars, orange, represent "Filled" capsules, the righthand side, in grey, the total capsules. Vertical lines in each bar correspond to standard deviation.

4.4. Conclusions

Regarding the initial hypothesis of having two or three genetic groups, from the current results we say that it falls between both scenarios, i.e., hybridization events are not so rare (42% of individuals in one of the populations) and gene flow even seems to be higher from cluster 2 to cluster 1, despite the fact that there are still two main genetic groups (A), introgression events having not yet led to their fragmentation; species do not keep their identity, as mentioned for one of the initial hypothetical scenarios. Introgression process does not seem to be equal in both senses, as one species (*O. fusca*)

appears to be acting as a genetic donor. Nothwithstanding, this hypothesis should be validated with evidence of the movement of potential pollinators among the taxa.
The comparison between the distance matrices of both genetic and morphometric dataset revealed a positive correlation, although weak. Indeed, the composition of the groups found with each analysis - chapter 2 and chapter 4 - is quite different, and the resulting placement of plants shows remarkable contrasts.

Contrary to what would be expected, no significant differences were found between genetic clusters regarding the percentages of pollination success, namely between hybrids and parentals (Figure 34). The parental species and hybrids seem to produce similar numbers of capsules, suggesting possible absence of post-zygotic barriers in the case of hybrids. This is consistent with the result of constant ploidy throughout the dataset, including both parental taxa and hybrids, as post-zygotic barriers are frequently attributed to cytological isolation.

5. GLOBAL DISCUSSION AND CONCLUSION

After analysing results achieved with the different datasets and in the different chapters, it is possible to have a more understanding perspective on hybridization and the affinities between these two entities. This opportunity was created in this study by sampling the same plants for genetic and phenotypic analysis, pollination success and cytological analysis. In the current section I discuss the conclusions from the chapters and from comparisons of the different datasets.

Comparisons of phenotype and genetics, with reference to hybridisation, have been carried out for diverse plant groups (Rieseberg et al., 2003, Shipunov et al., 2004, Stahlberg, 2009, Wilson and Hudson, 2011, Sedeek et al., 2014, Fogelqvist et al., 2015, Gögler et al., 2015, Lowe and Abbott, 2015), amongst others. Although there are different conclusions regarding the persistence and extent of reproductive barriers and the relations between datasets, these studies have demonstrated how the frequency of hybridisation occurrence amongst plants and the combination of genetic data with deep-field ecological and morphological studies can help to better understand the role of hybridisation phenomena.

Hybridisation can be followed by selection against hybrids, usually including the evolution of premating barriers, or trigger speciation, which may include reinforcement, i.e., the natural selection for assortative mating in parapatry or sympatry (Butlin, 1987). The incidence of reinforcement (sensu Butlin) in sympatric hybridisation areas can have marked effects in speciation (Abbott et al., 2013). It is believed to complete the speciation process between incipient species upon secondary contact by strengthening prezygotic barriers to prevent gene flow (Widmer et al., 2009). A study in the grass *Anthoxanthum odoratum* (Silvertown et al., 2005) showed evidence of reinforcement of a flowering time shift at the boundaries of the plots being studied. Also, a macroevolutionary analysis of the Cape Flora of South Africa, with a high level of pollination specialization, indicated that pollinator system shifts can occur upon secondary contact to prevent gene flow between incipient species that initially diverged as a consequence of adaptation to other ecological conditions (reviewed by Widmer et al. (2009). There are also recent works where shifts in the specific pollinator have been documented, novel pollinators allowing sexual reproduction to occur between hybrid plants. In *Narcissus*, natural hybrids recruited ants as novel pollinators, this shift having involved a combination of increased nectar secretion, the production of novel floral scent compounds, and possibly a breakdown of chemical defence against ants (Marques et al., 2015). The referred study provided evidence for the contribution of pollinator shifts to reproductive isolation between hybrids and their progenitors.

5.1. The understanding of pollinator-mediated floral evolution

When studying population-level flowering behaviour, it became apparent that, in the studied populations, floral isolation between the two taxa studied here (*Ophrys fusca* and *O. dyris*) was not complete; there is a short time span – at the beginning of April – when both can flower at the same time, although *Ophrys fusca* has been recorded to start flowering earlier, and *O. dyris* slightly later, as shown in previous studies (Lowe and Tyteca, 2012). According to the same records, flowering periods are longer in *O. fusca* than in *O. dyris*, and this may provide more opportunity for successful cross pollination. The first flowering records in each year have been obtained in mixed populations, from *O. fusca* individuals.

Percentages of pollination success obtained range from 3% in the *dyris* morphotype in 2013 to 37% in the *fusca* morphotype in 2014, with average values of 11.4% for *O. dyris* and of 23.4% for O. *fusca*.

In plants with such pollinator specificity as *Ophrys*, pollination success might be expected to be initially lower in hybrid plants, and pollinator shifts presumably more difficult to happen. In this study there were data describing pollination success and data describing the genetics, specifically whether plants were hybrids or not. Although the topic of pollinator shifts deserves attention and would likely be helpful in discussing the results of pollination success amongst hybrids, it was not not possible to include this in the scope of the current work. It is possible, however, to hypothesise that if pollinator specificity existed before, it has been or it is now being disrupted. This assumption is based both on the occurrence and distribution of intermediate morphotypes and on the record of the flowering peaks in the different years. Concerning intermediate morphotypes, their wide morphological range and dispersion (field observations), as well as, in some cases, the difficulty in their assignment to any of the morphotypes here considered, it is likely that both taxa can be visited by the same pollinator. On the other hand, average flowering peak by morphotype is not consistent between years (significant differences having been found in its analysis). Despite the likelihood of, at some extent, being biased by the time when sampling visits were made, it seems that different morphotypes respond differently to different conditions between years, this being relevant to discuss in terms of climate change.

Phenological events are a direct response to climate (Robbirt et al., 2011), which has been proved to affect them in different groups of plants (Sparks and Carey, 1995, Fitter and Fitter, 2002). A recent study from Hutchings et al. (2018) indicate that the sequence of peak flying date of male and female bees and peak of *Ophrys sphegodes* flowering date, has changed with a significant temperature warming recorded from 1659 to 2014. All three phenological events showed significant advancement because of global warming, significantly faster when the mean rate of warming was also higher (1970-2014). The time elapsed between each pair of phenological events also changed: there was a reduction in mean time between peak flying dates of males and females, and an increase between

male bee flying and peak orchid flowering. It is foreseen that such discrepancy of responses between orchids and their specialist pollinators might threaten the synchrony required for effective pollination by sexual deception (Hutchings et al., 2018, Robbirt et al., 2014).

For the five years when phenological events were recorded for the current study, the flowering peak of these plants does not seem to be globally anticipating. However, different responses were observed, also resulting in differences in the time intervals between the flowering peaks of each morphotype. The *fusca* morphotype is that which seems to show a trend to flower earlier, although statistical tests failed to resolve when morphotypes were considered as response variable. If this type of disruption is confirmed, it could represent an advantage for the *fusca* plants regarding the availability of pollinators, as the plants that flowers before the emergence of female insects are more likely to benefit from pseudocopulation by the males. Consequently, the actual small divergences in the rates of reproductive success between morphotypes are likely to be increased. Also, adaptive evolution might act positively on genetic variation to select for early flowering genotypes and select against later flowering genotypes. All these confer adaptative advantages on *O. fusca* plants in comparison with *O.dyris*, under the conditions surveyed in this study.

5.2. Are these separate entities? Is hybridization occurring between them?

Results from the morphological dataset (Chapter 2) revealed that there are still separate morphological groups that maintain some isolation. The number of groups found was either two or three; depending on the number of characters and plants used. The traits that contribute most to the separation of morphological groups, considering two or three, are qualitative ones.

Comparing the genetic and morphometric datasets revealed some inconsistencies. Exclusively in the morphological grouping assumed to represent *O. dyris* are the plants from Mj, and two plants from Pi. In the morphological grouping referred to as the *fusca* group are the plants from AV, CF and the remaining plants from Pi. The SA and Me plants are found in both the *fusca* and *dyris* groups, according to both the PCoA analyses, including all plants and including only the subset of plants for which there are genetic data. Previously assumed to represent *O. dyris*, SA is the population that includes the fewest plants of admixed ancestry, and the most genetically homogeneous, according to the hybridisation analysis described in Chapter 4. However, the few admixed individuals found are genetic hybrids. Another divergence between the results of the morphological and genetic analyses is apparent when one considers the populations initially assumed to be “pure” populations. Mj and CF populations represent two of these cases, but at the same time appear in second and third place regarding the number of genetic hybrids per population. On the other hand, the population Pi was

thought to be a "mixed" population, and this is confirmed by both morphological and genetic analyses, which revealed a higher number of hybrids. Such strong discrepancies between the results obtained from phenotypic and genotypic data within the *fusca* group have already been referred to by Francisco et al. (2015) and Bateman (2018c) in phylogenetic analyses, both having found a morphologically determined primitive position contrasting with a molecularly determined derived position for the *fusca* group, further exploration being needed.

Given the fact that all the populations included mixed genotypes (Q values between 0.1 and 0.9), found in 44.8% of the individuals, the initial hypothesis of "pure" (one taxon) versus mixed-taxa populations, when referring to genetic groups, ought to be refuted. Genetic hybrids were found in almost all the populations except AV. In Mj, a morphologically uniform population thought to consist of *O. dyris*, the proportion calculated for introgressed individuals was even higher than that for individuals assigned to the *dyris* cluster (cluster 1).

Genetic clusters do exist, but isolation between them is very permeable, gene flow occurring extensively. Hybrids have been detected, and introgression seems to be occurring from the *fusca* to the *dyris* cluster.

5.3. Considerations for Taxonomy

The maintenance of the taxonomic entities studied as different species will depend on the degree of circumscription in the concept used for species. The entities studied here are clearly distinct morphological identities, although interbreeding and gene flow occur between them. The results together show that divergence is not complete (morphological and genetic separation), gene flow and intermediate morphologies being very frequent. From one point of view, a combination of both the recognition of the average values regarding the diagnostic characters and habitat characteristics for each of the taxa would allow them to be identified individually. Bayesian cluster analysis of microsatellite data of Cotrim et al. (2016) identified four clusters within a set of samples from *O. fusca* and *O. dyris* populations in Portugal. All of the samples attributed to *O. dyris* were grouped in one cluster, not genetically isolated from the remaining three, the authors agreeing with taxonomic treatments which consider *dyris* as a subspecies of *fusca*, such as that of Aedo and Herrero (2005). One of the sampled populations in our study, SA, was found to be the most genetically isolated from the remaining populations. Morphologically plants resemble *O. dyris* with vigour traits in the upper extreme of the expression range. From the results of the microsatellites dataset, these are the plants suffering less hybridisation from the other cluster and with no *fusca* cluster presence having been

detected. A new species of *Ophrys* was described by Lowe and Tyteca (2012), *Ophrys lenae* M.R.Lowe & D.Tyteca, on the grounds of plants with this phenotype, the holotype having been collected in this region. To comment on its validity one would have to inspect other populations of *O. dyris* and quantify the genetic distance between them.

The data here produced do not allow us to take a decision on this subject or to decide if *O. dyris* deserves species status. We agree with Cotrim et al. (2016) that the *fusca* group should be addressed in a more global taxonomic treatment in Portugal and the Iberian Peninsula. Markers that sample larger parts of the genome, with the use of next-generation sequencing, would be useful in delimiting species within this group.

Notwithstanding, a large number of plants displaying intermediate morphotypes will not fall in any of the taxa and this approach would demand the taxonomic recognition of hybrids, so that every plant could be systematically classified. Despite we consider it as likely, it is still not known if the plants attributed to these different taxa share pollinators, occasionally or on a regular basis. This will be of great relevance for consideration in future studies.

5.4. Absence of cytological isolation

The same level of ploidy was found for all the plants tested, as well as the same DNA content, hybrids and parentals having the same ploidy (Chapter 3). A tetraploid cytotype has been found to occur across central Portugal, where these populations are located. This cytotype consistency contrasts with the diversity of phenotypes previously described (Chapter 2) and the different genetic groups (Chapter 4). Having confirmed the occurrence of recurrent hybridisation events (Structure estimates) and gene flow between genetic clusters, despite the persistence of two main clusters, and adding to the fact that both hybrids and parentals have proved to have the same cytotype, it was inferred that cytological isolation is not acting as a post-zygotic isolation barrier in the case of hybrids, and the hypothesis of the beginning of a homoploid hybrid speciation process has been raised. As mentioned in Chapter 4, hybridisation in the case of species of the same ploidy could have a range of outcomes, including the establishment of a hybrid swarm, the transfer of traits through introgressive hybridisation, and the origin of new homoploid hybrid species (Yakimowski and Rieseberg, 2014). As already mentioned by Paulus (1990) as a likely evolutionary mode in *Ophrys* in general, this might be the case with these taxa within this region; with such a starting point (considering the persistence of parentals and of two main genetic groups) these evolutionary processes can be studied in real time, representing a major advantage for the understanding of this unresolved group of *Ophrys.*

5.5. Is *Ophrys dyris* a hybrid?

The hypothesis of hybrid origin for *O. dyris* was raised by Devey *et al*. (2008). If this hypothesis holds true, the likely parents would be from sections *Pseudophrys* and *Ophrys* (*fuciflora* aggregate), based on the phylogenetic results obtained by the authors. Our current molecular dataset supports the idea of a hybrid origin for *O. dyris* (number and diversity of alleles, microsatellite peaks). To clarify this idea, plastid genes were also analysed in the scope of this study (data not publ.), which seem to indicate that, existing a hybrid origin, both *Ophrys lutea* (section *Ophrys*) and *O. fusca* could be parental species. In addition, Cotrim et al. (2016), using plastid haplotypes, reported introgressed individuals and hybridising populations between *fusca-lutea*, further supporting the close relationship between these species previously pointed out by Soliva et al. (2001). Once again, further investigations are needed, which should include more data, to allow this matter to be resolved.

The establishment of a species such as *O. dyris* might reflect the maintenance of viable populations for successive generations, or the rapid colonisation of new areas. Pre-zygotic isolating mechanisms, such as pollinator-mediated isolation, are of major importance for the establishment of such species (Petit 1999). The constancy of cytotypes, recurrent hybridisation events and gene flow between *O. dyris* and *O. fusca* suggest that isolation – at least from one of the putative parents – is not complete. To achieve a more comprehensive understanding of the evolution of this group including clarification of the origin of *O. dyris*, further analyses are needed.

5.6. Evolutionary Processes and Conservation

The outcome of the geneflow highlighted in this study, whether ultimately there is selection against hybrids or reinforcement that strengthens pre-zygotic barriers and halts gene flow, is still unknown, and the evolutionary consequences still remain to be addressed. The role played by such geneflow should be addressed in detail using larger datasets and other analytical methods. Nevertheless, considering the consistency of ploidy level throughout all the populations - including pure parental species and hybrids – the likely absence of chromosomal rearrangements following hybridization, and lack of reproductive and spatial isolation between the hybrids and the parental species (in some populations), it is plausible that the putative hybrid species in formation can be swamped through mating with the parental species, as hypothesised in the case of *Senecio* (Gross and Rieseberg, 2005, Gross, 2012). The same authors highlighted the importance of ecological selection in homoploid hybrid speciation. As in the cases of *Senecio* (James and Abbott, 2005) or *Helianthus* (Rieseberg et al.,

2003), parental species and original hybrid zones are still present in the native habitat of the *Ophrys* taxa studied here.

It is one opinion that the fate of the hybrids will strongly be influenced by local and environmental constraints (Lowe and Abbott, 2015), likely varying in different populations, depending whether there is sympatry or not between the different species. In the cases where the different species occupy different areas of the population (such as in the Pi population), assimilation of one of the species is not likely to occur, partly due to the patchy distribution of parental plants, that may limit hybridization to particular areas (Jacquemyn et al., 2012).

Furthermore, if genetic introgression is occurring from cluster 2 to cluster 1, as hypothesised in chapter 4, *O. dyris* is the taxon which would be receiving genetic input. The maintenance of the direction of this geneflow through time will lead to a decrease in the genetic identity of *O. dyris,* with it becoming swamped by *O. fusca*. As we know from field observation and biogeographical data, for both species, *O. dyris* is the one with a more restricted presence, greater vulnerability in terms of pollination efficiency (Chapter 2) and less variability in the expression of morphological traits, with a consequently reduced spectrum on which adaptive selection can act (Anderson and Stebbins, 1954). The persistence of such genetic dynamics could therefore be a threat to its future conservation.

If, on the one hand, conservation of evolutionary processes that generate biodiversity can be as important or even more so than conservation of taxonomic entities (Ennos, French et al. 2005; Ennos, Whitlock et al. 2012)**,** on the other hand, hybridisation can be a force driving populations or taxa towards extinction (Wolf et al., 2001). Conservation and management plans should thus consider whether hybridisation is occurring and which outcomes are being generated, evaluating the extinction risk, in the case of vulnerable species.

From one perspective, conservation measures should target ongoing processes as well as the named taxa. In the case of *Ophrys*, in which most of the taxa are young and closely related (Chase et al., 2003, Bateman et al., 2003), conservation of active processes allows continuous evolution and/or speciation in the face of disruptive factors such as climate change or habitat degradation. Notwithstanding this, due to the vulnerability here highlighted, conservation of *O. dyris* should be prioritised over that of *O. fusca*, the first taxon needing targeted conservation measures. If taxonomic assignment proves difficult, such as in many others taxonomically complex groups, the approach should follow a "Process-Based Species Action Plan", as developed by Ennos et al. (2012). In practical terms, future management measures should thus be aimed at conserving the taxa at sites with high environmental heterogeneity, as sympatric zones provide the stage for evolutionary processes in orchids (Cozzolino

et al., 2006) and allow natural selection on differentially adapted selfing lines derived from segregating hybrid swarms to act (Cowling and Pressey, 2001).

5.7. Future work

Future work should include the production of a comprehensive phylogenetic tree for section *Pseudophrys* of the genus *Ophrys*, including multiple accessions of all species in the section, particularly *O. dyris* and *O. fusca*, to confirm or refute the sister relationship between them, as the phylogenetic position of both species (*O. fusca* and *O. dyris*) also remains to be confirmed. Nevertheless, recent works have revealed an accelerated rate of evolution within the *fusca* lineage (Francisco et al., 2015, Bateman et al., 2018c).

As mentioned earlier in this chapter, more detailed analysis of the genetic dataset would be needed. For a more comprehensible study of hybridisation, the performance of artificial crosses and viability assessment of the resulting seeds would also be important. Comparing the rates of seed viability between both taxa the resulting hybrids would offer a more precise picture of post-zygotic isolation barriers.

Perhaps the most relevant work to do on this topic would be a study of pollination and pollinator behaviour, in the presence of hybrids and intermediate morphotypes. Is there a shift of pollinator in the case of hybrids? Do both parental species share the same pollinator? These are important questions still to be addressed.

From one perspective, and agreeing with Widmer *et al.* (2009), regions such as the one studied here – assuming some of the studied areas as hybrid zones - can serve as "evolutionary genomics laboratories for studying the genetic basis of multiple different reproductive barriers in situ".

It is important to look at this topic not as a taxonomic problem to solve – that will certainly not have a forthcoming or straightforward solution – but as a possibility for studying evolutionary processes in real time and from which insights should be taken to help to explain more global phenomena. In the current scenario of climatic change and uncertainty regarding the long-term conservation of these habitats, this is much more relevant then tackling classification issues.

5.8. Reflections on the work done

Fieldwork and data collection

This study describes the analyses of five data sets, morphological, phenological, reproductive, cytological and genetic. Three of these five data sets were collected in the field, these sites being visited throughout the flowering season in a field campaign over 5 years.

The collection of the phenological data on a regular basis for all the populations was particularly difficult, and some populations had missing data for some weeks. Field work might have been carried out in a different way. If it were to conducted again, I would prioritise complete data for a site over attempting to have some data for all sites. I would also consider the inclusion of other morphological characters, and persisting with some that have been dropped. For example, the colour of the lateral sepals was initially recorded, but its collection was abandoned due to difficulties in recording this character in the field. It has been shown to be relevant for the pollinator and the most homoplastic character in a tree of 14 analysed *Ophrys* taxa (Bateman et al., 2018c). A more systematic method for recording the colour in the field could have been established. More systematic recording of environmental variables, from the beginning, would have allowed us to conduct other types of analysis. These variables (number of nearest plants and distance to the nearest neighbour, other *Ophrys* taxa at the same site) were only consistently recorded when I started to use the Memento database for Android (2012), and some of them proved to be significant in influencing reproductive success in other works for *Ophrys* (here referred to as "pollination success) (Vandewoestijne et al., 2009). Incorporating these parameters would have enhanced this work, as environmental traits and density effects are thought to be relevant for pollination success, in both non-rewarding (e.g. (Johnson et al., 2003) and rewarding species of orchids (e.g. (Duffy and Stout, 2011).

At the outset of the work I explored means to record pollinator visits and identify pollinators. This proved to be very challenging for these sexually deceptive orchids, and it was abandoned from there onwards. Collaborations with researchers who can contribute technical solutions to such problems are needed. Some pictures and videos were made, but the insects captured in the images was not identified. However, confirming whether visitors are effective pollinators depends on observation of removal of pollinaria. Further work could have included in-field tests of selfing, and experimental hybridizations could have been carried out, such as those carried out by Jacquemyn *et al.* (2012). If artificial crosses had been performed, this would have allowed practical inspection of genuine hybrids and opened opportunities for further genetic analysis.

Lab work and data exploration

Laboratory work carried out included DNA extractions, microsatellite data collection (performed both at the University of Lisbon and Royal Botanic Gardens, Kew), the chromosome counting and genome size measurements (carried out at Royal Botanic Gardens, Kew). The two latter were sufficient to confirm tetraploidy of the study plants, and this was essential for the interpretation of the microsatellite data. The chromosome counts depended on the availability of actively growing roots, and plants collected in the field and grown on in Portugal proved difficult to establish. The cytological work reported in this thesis and published (Abreu et al., 2017) did not explore cytological methods of investigating hybrids, and they were beyond the scope of this thesis. That would have included research using more sophisticated cytogenetic methods such as Fluorescence *in situ* hybridisation - FISH (Devi et al., 2005), that have proved useful in providing insights into the origins of hybrids.

Since the outset of this thesis work there has been a revolution in the quantity of sequence data that is used to study plant evolution. New sequencing methods, specifically Next Generation Sequencing NGS have been used to study microevolution (Bateman et al., 2018c, Macaya-Sanz et al., 2016, Osborne et al., 2016). Nevertheless, microsatellites remain a method of choice for studies of hybridization because they lend themselves to the calculation of indices of heterozygosity (Putman and Carbone, 2014, Dufresne et al., 2014, Pfeiffer et al., 2011). Were I to develop the work again at this point I would still advocate the use microsatellites as molecular markers, potentially combining this with next-generation methodology.

Other laboratory work that was not prioritized included the assessment of seed viability. Seeds were collected and are presently stored in the seed bank at the Natural History Museum of Lisbon. Seed viability testing in orchids based on germination is not easy to implement, as germination may depend on the presence of mycorrhizal partners (Jacquemyn et al., 2009); assessments of viability are usually based in the tetrazolium (TZ) test, as an estimation of seeds germination through testing the seeds vigour by staining the living tissues (Patil and Dadlani, 2009)

Regarding the genetic analyses, availability of artificial, known hybrids (from artificial crosses) would have allowed further analysis. Methods are available to theoretically identify backcrossed and F1 and F2 individuals (Anderson and Thompson, 2002), and these could be explored. These methods depend on the fairly confident assignments of pure parentals, and they were already used for assignments for the Bayesian cluster analysis, based on previous Structure results and on examination of photographs of the specimens.

Phenological data can be explored in many ways, but most are relevant when different data sets are considered. When considered alone, long term phenological data sets, including those of orchids, have proven to be important in understanding the impact of climate change on flowering time (Hutchings, 2010). Regarding pollination success data, there is the opportunity to compare reproductive success of these plants with the values recorded in the literature for other orchids (Neiland and Wilcock, 1998, Ayasse et al., 2000, Vandewoestijne et al., 2009).

REFERENCE LIST

2013. *The Plant List* [Online]. Available: http://www.theplantlist.org/ [Accessed 22nd December 2020].

ABBOTT, R. J., ALBACH, D., ANSELL, S., ARNTZEN, J. D., BAIRD, S. J. E., BIERNE, N., BOUGHMAN, J., BRELSFORD, A., BUERKLE, C. A., BUGGS, R., BUTLIN, R. K., DIECKMANN, U. F., EROUKHMANOFF, F., GRILL, A., CAHAN, S. H., HERMANSEN, J. S., HEWITT, G., HUDSON, A. G., JIGGINS, C., JONES, J., KELLERO, B., MARCZEWSKI, T., MALLET, J., MARTINEZ-RODRIGUEZ, P., MÖST, M., MULLEN, S., NICHOLS, R., NOLTE, A. W., PARISOD, C., PFENNIG, K., RICE, A. M., RITCHIE, M. G., SEIFERT, B., SMADJA, C. M., STELKENS, R., SZYMURA, J. M., VAÏNÖLÄ, R., WOLF, J. B. W. & ZINNER, D. 2013. Hybridization and speciation. *Journal of Evolutionary Biology,* 26**,** 229–246.

ABREU, J. I. A., HAWKINS, J. A., COTRIM, H. C., FAY, M. F., HIDALGO, O. & PELLICER, J. 2017. Ophrys fusca and Ophrys dyris (Orchidaceae) - constancy of tetraploidy amongst populations in Central Portugal. *New Journal of Botany*.

ADDINSOFT 2018. XLSTAT 2018: Data Analysis and Statistical Solution for Microsoft Excel. 2018.2 ed. Paris, France: Kovach Computing Services.

AEDO, C. & HERRERO, A. 2005. Vol. XXI Smilacaceae-Orchidaceae. *In:* CASTROVIEJO, S. (ed.) *Flora Iberica: plantas vasculares de la Península Ibérica e Islas Baleares.* Real Jardín Botánico ed. Madrid.

AMICH, F., GARCÍA-BARRIUSO, M. & BERNARDOS, S. 2007. Poliploidy and speciation in the orchid flora of the Iberian Peninsula. *Botanica Helvetica,* 117**,** 143-157.

ANDERSON, E. 1948. HYBRIDIZATION OF THE HABITAT. *Evolution,* 2**,** 1-9.

ANDERSON, E. 1949. *Introgressive hybridization,* New York,, J. Wiley.

ANDERSON, E. & STEBBINS, G. L. 1954. Hybridization as an evolutionary stimulus. *Evolution,* 8**,** 378-388.

ANDERSON, E. C. & THOMPSON, E. A. 2002. A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. *Genetics,* 160**,** 1217.

ANDREAKIS, N., KOOISTRA, W. H. C. F. & PROCACCINI, G. 2009. High genetic diversity and connectivity in the polyploid invasive seaweed Asparagopsis taxiformis (Bonnemaisoniales) in the Mediterranean, explored with microsatellite alleles and multilocus genotypes. *Molecular Ecology,* 18**,** 212-226.

ARNOLD, M. L. 1992. Natural Hybridization as an Evolutionary Process. *Annual Review of Ecology and Systematics,* 23**,** 237-261.

ARNOLD, M. L. 1994. Natural Hybridization and Louisiana Irises: Defining a major factor in plant evolution. *BioScience,* 44**,** 141-147.

ARNOLD, M. L., BALLERINI, E. S. & BROTHERS, A. N. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana Irises. *Heredity (Edinb),* 108**,** 159-66.

ARNOLD, M. L. & KUNTE, K. 2017. *Adaptive Genetic Exchange: A Tangled History of Admixture and Evolutionary Innovation*.

ARNOLD, M. L., TANG, S., KNAPP, S. J. & MARTIN, N. H. 2010. Asymmetric introgressive hybridization among louisiana iris species. *Genes (Basel),* 1**,** 9-22.

ASCENSÃO, L., FRANCISCO, A., COTRIM, H. & PAIS, M. S. 2005. Comparative structure of the labellum in Ophrys fusca and O. lutea (Orchidaceae). *American Journal of Botany,* 92**,** 1059-1067.

AYASSE, M., SCHIESTL, F. P., PAULUS, H. F., LOFSTEDT, C., HANSSON, B., IBARRA, F. & FRANCKE, W. 2000. Evolution of reproductive strategies in the sexually deceptive orchid Ophrys sphegodes: How does flower-specific variation of odor signals influence reproductive success? *Evolution,* 54**,** 1995-2006.

AYASSE, M., STÖKL, J. & FRANCKE, W. 2011. Chemical ecology and pollinator-driven speciation in sexually deceptive orchids. *Phytochemistry,* 72**,** 1667-1677.

BAACK, E. J., WHITNEY, K. D. & RIESEBERG, L. H. 2005. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in Helianthus homoploid hybrid species. *The New phytologist,* 167**,** 623-630.

BARTON, N. H. 2001. The role of hybridization in evolution. *Mol Ecol,* 10**,** 551-68.

BARTON, N. H. & HEWITT, G. M. 1985. Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics,* 16**,** 113-148.

BASKETT, M. L. & GOMULKIEWICZ, R. 2011. Introgressive hybridization as a mechanism for species rescue. *Theoretical Ecology,* 4**,** 223-239.

BATEMAN, R. M., DEVEY, D. S., MALMGREN, S., BRADSHAW, E. & RUDALL, P. J. 2010. Conflicting species concepts underlie perennial taxonomic controversies in Ophrys. *Cahors Société Française d'Orchidophilie,* 8**,** 87–101.

BATEMAN, R. M., GUY, J. M., RUDALL, P. J., LEITCH, I., PELLICER, J. & LEITCH, A. R. 2018a. *Evolutionary and functional potential of ploidy increase within individual plants: somatic ploidy mapping of the complex labellum of sexually deceptive bee orchids*.

BATEMAN, R. M., HOLLINGSWORTH, P. M., PRESTON, J., YI-BO, L., PRIDGEON, A. M. & CHASE, M. W. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). *Botanical Journal of the Linnean Society,* 142**,** 1-40.

BATEMAN, R. M., MURPHY, A. R. M., HOLLINGSWORTH, P. M., HART, M. L., DENHOLM, I. & RUDALL, P. J. 2018b. Molecular and morphological phylogenetics of the digitate-tubered clade within subtribe Orchidinae s.s. (Orchidaceae: Orchideae). *Kew Bulletin,* 73**,** 54.

BATEMAN, R. M. & RUDALL, P. J. 2006. Evolutionary and morphometric implications of morphological variation among flowers within an inflorescence: A case-study using European orchids. *Annals of Botany,* 98**,** 975-993.

BATEMAN, R. M. & RUDALL, P. J. 2011. The life and death of a mythical British endemic, Orchis militaris L. var. tenuifrons P.D. Sell: why infraspecific taxonomy requires a field-based morphometric approach. *New Journal of Botany,* 1**,** 98-110.

BATEMAN, R. M., SRAMKÓ, G. & PAUN, O. 2018c. Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids. *Annals of Botany,* 121**,** 85-105.

BECKER, M. 2011. *Diagnosing collinearity in mixed models from lme4* [Online]. Available: https://hlplab.wordpress.com/2011/02/24/diagnosing-collinearity-in-lme4/ [Accessed 12/2020 2020].

BEHIE, A. M. & GROVES, C. P. 2015. *Taxonomic tapestries : the threads of evolutionary, behavioural and conservation research,* Acton, ACT, Australia, Australian National University Press.

BELSLEY, D. A. 1991. *Conditioning diagnostics : collinearity and weak data in regression*, New York (N.Y.) : Wiley.

BENNETT, M. & LEITCH, I. 2005. Genome Size Evolution in Plants.

BENNETZEN, J. L. 2000. Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology,* 42**,** 251-269.

BERNARDOS, S., AMICH, F. & GALLEGO, F. 2003. Karyological and taxonomic notes on Ophrys L. (Orchidoideae, Orchidaceae) from the Iberian Peninsula. *Botanical Journal of the Linnean Society,* 142**,** 395- 406.

BERNARDOS, S., CRESPI, A., DEL REY, F. & AMICH, F. 2005. The section Pseudophrys (Ophrys, Orchidaceae) in the Iberian Peninsula: a morphometric and molecular analysis. *Botanical Journal of the Linnean Society,* 148**,** 359-375.

BEVER, J. D. & FELBER, J. 1992. The theoretical population genetics of autopolyploidy. *Oxford Surveys in Evolutionary Biology.* Oxford, UK: Oxford University Press.

BHARGAVA, A. & FUENTES, F. F. 2010. Mutational dynamics of microsatellites. *Mol Biotechnol,* 44**,** 250-66.

BOUGEARD, S. & DRAY, S. 2018. Supervised Multiblock Analysis in R with the ade4 Package. *Journal of Statistical Software; Vol 1, Issue 1 (2018)*.

BRADSHAW, E., RUDALL, P. J., DEVEY, D. S., THOMAS, M. M., GLOVER, B. J. & BATEMAN, R. M. 2010. Comparative labellum micromorphology of the sexually deceptive temperate orchid genus Ophrys: diverse epidermal cell types and multiple origins of structural colour. *Botanical Journal of the Linnean Society,* 162**,** 504-540.

BREITKOPF, H., ONSTEIN, R. E., CAFASSO, D., SCHLÜTER, P. M. & COZZOLINO, S. 2014. Multiple shifts to different pollinators fuelled rapid diversification in sexually deceptive Ophrys orchids. *New Phytologist,* 207**,** 377-389.

BRENNAN, A. C., BARKER, D., HISCOCK, S. J. & ABBOTT, R. J. 2012. Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of Senecio squalidus (Asteraceae). *Heredity,* 108**,** 87-95.

BRIGGS, D. & WALTERS, S. M. 2016. *Plant variation and evolution,* Cambridge, Cambridge University Press.

BRUVO, R., MICHIELS, N. K., D'SOUZA, T. G. & SCHULENBURG, H. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology,* 13**,** 2101-2106.

BUERKLE, C. A. & RIESEBERG, L. H. 2001. Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution,* 55**,** 684-91.

BURGARELLA, C., LORENZO, Z., JABBOUR-ZAHAB, R., LUMARET, R., GUICHOUX, E., PETIT, R. J., SOTO, A. & GIL, L. 2009a. Detection of hybrids in nature: application to oaks (Quercus suber and Q. ilex). *Heredity,* 102**,** 442-452.

BURGARELLA, C., LORENZO, Z., JABBOUR-ZAHAB, R., LUMARET, R., GUICHOUX, E., PETIT, R. J., SOTO, A. & GIL, L. 2009b. Detection of hybrids in nature: application to oaks (Quercus suber and Q. ilex).

BURKE, J. M. & ARNOLD, M. L. 2001. Genetics and the fitness of hybrids. *Annu Rev Genet,* 35**,** 31-52.

BURTON, T. L. & HUSBAND, B. C. 2000. FITNESS DIFFERENCES AMONG DIPLOIDS, TETRAPLOIDS, AND THEIR TRIPLOID PROGENY IN CHAMERION ANGUSTIFOLIUM: MECHANISMS OF INVIABILITY AND IMPLICATIONS FOR POLYPLOID EVOLUTION. *Evolution,* 54**,** 1182-1191.

BUTLIN, R. 1987. Speciation by reinforcement. *Trends in Ecology & Evolution,* 2**,** 8-13.

CANESTRELLI, D., BISCONTI, R. & NASCETTI, G. 2014. Extensive unidirectional introgression between two salamander lineages of ancient divergence and its evolutionary implications. *Scientific Reports,* 4**,** 6516.

CARROLL, C. P. & BORRILL, M. 1965. Tetraploid hybrids from crosses between diploid and tetraploidDactylis and their significance. *Genetica,* 36**,** 65-82.

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

CHAPMAN, M. A., HISCOCK, S. J. & FILATOV, D. A. 2016. The genomic bases of morphological divergence and reproductive isolation driven by ecological speciation in Senecio (Asteraceae). *Journal of Evolutionary Biology,* 29**,** 98-113.

CHASE, M. W., CAMERON, K., BARRETT, R. & V FREUDENSTEIN, J. 2003. *DNA data and Orchidaceae systematics: a new phylogenetic classification*.

CHOOI, W. Y. 1971. COMPARISON OF THE DNA OF SIX VICIA SPECIES BY THE METHOD OF DNA-DNA HYBRIDIZATION. *Genetics,* 68**,** 213.

CLARK, L. V. & JASIENIUK, M. 2011. polysat: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources,* 11**,** 562-566.

COCKERHAM, C. C. 1969. Variance of Gene Frequencies. *Evolution,* 23**,** 72-84.

COCKERHAM, C. C. 1973. Analysis of gene frequencies of mates. *Genetics,* 94**,** 701-712.

COTRIM, H., MONTEIRO, F., SOUSA, E., PINTO, M. J. & FAY, M. F. 2016. Marked hybridization and introgression in Ophrys sect. Pseudophrys in the western Iberian Peninsula. *American Journal of Botany,* 103**,** 677-691.

COTRIM, H. C. 2007. Patterns of speciation of Portuguese populations of Ophrys (Orchidaceae)", funded by FCT and developed at FCT ed.: Unit of Biotechnology and Plant Molecular Biology, ICAT, University of Lisbon.

COTRIM, H. C., MONTEIRO, F., SOUSA, E. S., FAY, M. F., CHASE, M. W. & PAIS, M. S. 2009. Isolation and characterization of novel polymorphic nuclear microsatellite markers from Ophrys fusca (Orchidaceae) and cross-species amplification. *Conservation Genetics,* 10**,** 739-742.

COWLING, R. M. & PRESSEY, R. L. 2001. Rapid plant diversification: planning for an evolutionary

future. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 5452-5457.

COYNE, J. A. & ORR, H. A. 2004. *Speciation,* Sunderland, MA., USA, Sinauer Associates,.

COZZOLINO, S., NARDELLA, A. M., IMPAGLIAZZO, S., WIDMER, A. & LEXER, C. 2006. Hybridization and conservation of Mediterranean orchids: Should we protect the orchid hybrids or the orchid hybrid zones? *Biological Conservation,* 129**,** 14-23.

D'EMERICO, S., PIGNONE, D., BARTOLO, G., PULVIRENTI, S., TERRASI, C., STUTO, S. & SCRUGLI, A. 2005. Karyomorphology, heterochromatin patterns and evolution in the genus Ophrys (Orchids). *Botanical Journal of the Linnean Society,* 148**,** 87-99.

DARWIN, C. 1859. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*, London : John Murray, 1859.

DELFORGE, P. 2002. *Guía de las Orquideas de España y Europa, Norte de Africa y Próximo Oriente,* Barcelona, Lynx Editions.

DEVEY, D. S. 2007. *Systematics and conservation of Mediterranean insect-mimicking orchids (Ophrys).* PhD, University of Reading.

DEVEY, D. S., BATEMAN, R. M., FAY, M. F. & HAWKINS, J. A. 2008. Friends or Relatives? Phylogenetics and Species Delimitation in the Controversial European Orchid Genus *Ophrys*. *Annals of Botany,* 101**,** 1-18.

DEVI, J., KO, J. M. & SEO, B. B. 2005. FISH and GISH: Modern Cytogenetics techniques. *Indian Journal of Biotechnology,* 4**,** 307-315.

DOLEZEL, J., GREILHUBER, J. & SUDA, J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protocols,* 2**,** 2233-2244.

DOYLE, J. J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.,* 19**,** 11-15.

DUFFY, K. J. & STOUT, J. C. 2011. Effects of conspecific and heterospecific floral density on the pollination of two related rewarding orchids. *Plant Ecology,* 212**,** 1397-1406.

DUFRESNE, F., STIFT, M., VERGILINO, R. & MABLE, B. K. 2014. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology,* 23**,** 40-69.

DUKAS, R. & REAL, L. A. 1993. Effects of recent experience on foraging

decisions by bumble bees. *Oecologia,* 94**,** 244–246.

EARL, D. & VONHOLDT, B. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources,* 4**,** 359-361.

EBIHARA, A., ISHIKAWA, H., MATSUMOTO, S., LIN, S.-J., IWATSUKI, K., TAKAMIYA, M., WATANO, Y. & ITO, M. 2005. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the Vandenboschia radicans complex (Hymenophyllaceae) in Japan and adjacent areas. *American Journal of Botany,* 92**,** 1535-1547.

ENNOS, R. A., FRENCH, G. C. & HOLLINGSWORTH, P. M. 2005. Conserving taxonomic complexity. *TRENDS in Ecology and Evolution,* 20**,** 164-168.

ENNOS, R. A., WHITLOCK, R. A. J., FAY, M. F., JONES, B., NEAVES, L. E., PAYNE, R., TAYLOR, I. A. N., 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.

ESSELINK, G. D., NYBOM, H. & VOSMAN, B. 2004. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting—peak ratios) method. *Theoretical and Applied Genetics,* 109**,** 402-408.

EVANNO, G., REGNAUT, S. & GOUDET, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology,* 14**,** 2611-2620.

EXCOFFIER, L., SMOUSE, P. E. & QUATTRO, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,* 131**,** 479-491.

FAY, M. F. & CHASE, M. W. 2009. Orchid biology: from Linnaeus via Darwin to the 21st century. Preface. *Annals of Botany,* 104**,** 359-364.

FIELD, D. L., AYRE, D. J., WHELAN, R. J. & YOUNG, A. G. 2011. Patterns of hybridization and asymmetrical gene flow in hybrid zones of the rare Eucalyptus aggregata and common E. rubida. *Heredity,* 106**,** 841-853.

FITTER, A. H. & FITTER, R. S. 2002. Rapid changes in flowering time in British plants. *Science,* 296**,** 1689-91.

FOGELQVIST, J., VERKHOZINA, A. V., KATYSHEV, A. I., PUCHOLT, P., DIXELIUS, C., RÖNNBERG-WÄSTLJUNG, A. C., LASCOUX, M. & BERLIN, S. 2015. Genetic and morphological evidence for introgression between three species of willows. *BMC Evolutionary Biology,* 15**,** 1-10.

FRANCISCO, A. & ASCENSÃO, L. 2013. Structure of the Osmophore and Labellum Micromorphology in the Sexually Deceptive Orchids Ophrys Bombyliflora and Ophrys tenthredinifera (Orchidaceae). *International Journal of Plant Sciences,* 174**,** 619-636.

FRANCISCO, A., PORTO, M. & ASCENSÃO, L. 2015. Morphological phylogenetic analysis of Ophrys (Orchidaceae): insights from morpho-anatomical floral traits into the interspecific relationships in an unresolved clade. *Botanical Journal of the Linnean Society,* 179**,** 454-476.

GARCIA, S., LEITCH, I. J., ANADON-ROSELL, A., CANELA, M. A., GÁLVEZ, F., GARNATJE, T., GRAS, A., HIDALGO, O., JOHNSTON, E., MAS DE XAXARS, G., PELLICER, J., SILJAK-YAKOVLEV, S., VALLÈS, J., VITALES, D. & BENNETT, D. V. 2014. Recent updates and developments to plant genome size databases. *Nucleic Acids Research,* 42**,** D1159–D1166.

GARCÍA-BARRIUSO, M., BERNARDOS, S. & AMICH, F. 2010. Chromosomal evolution in Mediterranean species of Ophrys sect. Pseudophrys (Orchidaceae): An analysis of karyotypes and polyploidy. *Taxon,* 59**,** 525-537.

GARNIER, S., ROSS, N., RUDIS, B., SCIAINI, M. & SCHERE, C. 2018. viridis: Default Color Maps from 'matplotlib'. R package version 0.5.

GÖGLER, J., STÖKL, J., CORTIS, P., BEYRLE, H., BARONE LUMAGA, M. R., COZZOLINO, S. & AYASSE, M. 2015. Increased divergence in floral morphology strongly reduces gene flow in sympatric sexually deceptive orchids with the same pollinator. *Evolutionary Ecology,* 29**,** 703-717.

GOWER, J. C. 1971. A General Coefficient of Similarity and Some of Its Properties. *Biometrics,* 27**,** 857-871.

GREGORY, T. R. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev Camb Philos Soc,* 76**,** 65-101.

GREILHUBER, J. 1998. Intraspecific Variation in Genome Size: A Critical Reassessment. *Annals of Botany,* 82**,** 27-35.

GREILHUBER, J. & EHRENDORFER, F. 1975. Chromosome numbers and evolution inOphrys (Orchidaceae). *Plant Systematics and Evolution,* 124**,** 125-138.

GROSS, B. L. 2012. Genetic and phenotypic divergence of homoploid hybrid species from parental species. *Heredity,* 108**,** 157-158.

GROSS, B. L. & RIESEBERG, L. H. 2005. The ecological genetics of homoploid hybrid speciation. *The Journal of heredity,* 96**,** 241-252.

GUSTAFSON, J. P., LUKASZEWSKI, A. J. & BENNETT, M. D. 1983. Somatic deletion and redistribution of telomeric heterochromatin in the genus Secale and in Triticale. *Chromosoma,* 88**,** 293-298.

HAMILTON, J. A. & MILLER, J. M. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv Biol,* 30**,** 33-41.

HARDY, O. J. 2015. Population genetics of autopolyploids under a mixed mating model and the estimation of selfing rate. *Molecular Ecology Resources***,** n/a-n/a.

HARDY, O. J. & VEKEMANS, X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels 2: 618-620. *Molecular Ecology Notes,* 2**,** 618-620.

HARRISON, R. G. & LARSON, E. L. 2014. Hybridization, introgression, and the nature of species boundaries. *J Hered,* 105 Suppl 1**,** 795-809.

HARROWER, M. & BREWER, C. A. 2003. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal,* 40**,** 27-37.

HOPKINS, R. 2013. Reinforcement in plants. *New Phytol,* 197**,** 1095-103.

HOPKINS, R. & RAUSHER, M. D. 2011. Identification of two genes causing reinforcement in the Texas wildflower Phlox drummondii. *Nature,* 469**,** 411-4.

HOPKINS, R. & RAUSHER, M. D. 2012. Pollinator-mediated selection on flower color allele drives reinforcement. *Science,* 335**,** 1090-2.

HOWARD, D. J. 1986. A ZONE OF OVERLAP AND HYBRIDIZATION BETWEEN TWO GROUND CRICKET SPECIES. *Evolution,* 40**,** 34-43.

HUTCHINGS, M. J. 2010. The population biology of the early spider orchid Ophrys sphegodes Mill. III. Demography over three decades. *Journal of Ecology,* 98**,** 867-878.

HUTCHINGS, M. J., ROBBIRT, K. M., ROBERTS, D. L. & DAVY, A. J. 2018. Vulnerability of a specialized pollination mechanism to climate change revealed by a 356-year analysis. *Botanical Journal of the Linnean Society,* 186**,** 498-509.

JACQUEMYN, H., BRYS, R., HONNAY, O., ROLDAN-RUIZ, I., LIEVENS, B. & WIEGAND, T. 2011. Nonrandom spatial structuring of orchids in a hybrid zone of three Orchis species. *New Phytologist,* 193**,** 454-464.

JACQUEMYN, H., BRYS, R., HONNAY, O., ROLDAN-RUIZ, I., LIEVENS, B. & WIEGAND, T. 2012. Nonrandom spatial structuring of orchids in a hybrid zone of three Orchis species. *New Phytologist,* 193**,** 454-464.

JACQUEMYN, H., WIEGAND, T., VANDEPITTE, K., BRYS, R., ROLDÁN-RUIZ, I. & HONNAY, O. 2009. Multigenerational analysis of spatial structure in the terrestrial, food-deceptive orchid Orchis mascula. *Journal of Ecology,* 97**,** 206-216.

JAMES, J. K. & ABBOTT, R. J. 2005. Recent, allopatric, homoploid hybrid speciation: the origin of Senecio squalidus (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution,* 59**,** 2533-2547.

JERSÁKOVÁ, J., JOHNSON, S. D. & KINDLMANN, P. 2007. Mechanisms and evolution of deceptive pollination in orchids. *Biological Reviews,* 81**,** 219-235.

JERSAKOVA, J. & KINDLMANN, P. 2004. *Reproductive Success and Sex Variation in Nectarless and Rewarding Orchids*.

JOHNSON, S. D., PETER, C. I., NILSSON, L. A. & ÅGREN, J. 2003. Pollination success in a deceptive orchid is enhanced by co-occuring rewarding magnet plants. *Ecology,* 84**,** 2919-2927.

KAMVAR, Z. N., TABIMA, J. F. & GRÜNWALD, N. J. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ,* 2**,** e281.

KELLER, B., GANZ, R., MORA-CARRERA, E., NOWAK, M. D., THEODORIDIS, S., KOUTROUMPA, K. & CONTI, E. 2020. Asymmetries of reproductive isolation are reflected in directionalities of hybridization: integrative evidence on the complexity of species boundaries. *New Phytol*.

KIMURA, M. & OTA, T. 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proceedings of the National Academy of Sciences of the United States of America,* 72**,** 2761-2764.

KINDLMANN, P. & JERSÁKOVÁ, J. 2006. Effect of floral display on reproductive success in terrestrial orchids. *Folia Geobotanica,* 41**,** 47-60.

KLODA, J. M., DEAN, P. D., MADDREN, C., MACDONALD, D. W. & MAYES, S. 2008. Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British Restharrows (Ononis spinosa and Ononis repens). *Heredity (Edinb),* 100**,** 253-60.

KNIGHT, C. A. & ACKERLY, D. D. 2002. Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecology Letters,* 5**,** 66-76.

KNIGHT, C. A., MOLINARI, N. A. & PETROV, D. A. 2005. The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann Bot,* 95**,** 177-90.

KOVACH, W. L. 2013. MVSP - A MultiVariate Statistical Package for Windows, ver. 3.1. . Kovach Computing Services.

KULLENBERG, B. 1961. Studies in Ophrys pollination. *Zoologiska Bidrag från Uppsala,* 34**,** 1-340.

KUZNETSOVA, A., BROCKHOFF, P. B. & CHRISTENSEN, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software; Vol 1, Issue 13 (2017)*.

L MORJAN, C. & H RIESEBERG, L. 2004. *How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles*.

LAFON-PLACETTE, C., VALLEJO-MARIN, M., PARISOD, C., ABBOTT, R. J. & KOHLER, C. 2016. Current plant speciation research: unravelling the processes and mechanisms behind the evolution of reproductive isolation barriers. *New Phytol,* 209**,** 29-33.

LEWONTIN, R. C. & BIRCH, L. C. 1966. Hybridization as a source of variation for adaptation to new environments. *Evolution,* 20**,** 315-336.

LEXER, C., BUERKLE, C. A., JOSEPH, J. A., HEINZE, B. & FAY, M. F. 2007. Admixture in European Populus hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity (Edinb),* 98**,** 74-84.

LEXER, C., JOSEPH, J. A., VAN LOO, M., BARBARA, T., HEINZE, B., BARTHA, D., CASTIGLIONE, S., FAY, M. F. & BUERKLE, C. A. 2010. Genomic admixture analysis in European Populus spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics,* 186**,** 699-712.

LI, C., MESGARAN, M. B., ADES, P. K. & COUSENS, R. D. 2020. Inheritance of breeding system in Cakile (Brassicaceae) following hybridization: implications for plant invasions. *Ann Bot,* 125**,** 639-650.

LINDTKE, D., BUERKLE, C. A., BARBARA, T., HEINZE, B., CASTIGLIONE, S., BARTHA, D. & LEXER, C. 2012. Recombinant hybrids retain heterozygosity at many loci: new insights into the genomics of reproductive isolation in Populus. *Mol Ecol,* 21**,** 5042-58.

LINDTKE, D., GOMPERT, Z., LEXER, C. & BUERKLE, C. A. 2014. Unexpected ancestry of Populus seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular Ecology,* 23**,** 4316-4330.

LINNAEUS, C. V. 1753. *Species Plantarum,* Stockholm, 1753.

LOWE, A. J. & ABBOTT, R. J. 2004. Reproductive isolation of a new hybrid species, Senecio eboracensis Abbott & Lowe (Asteraceae). *Heredity (Edinb),* 92**,** 386-95.

LOWE, A. J. & ABBOTT, R. J. 2015. Hybrid swarms: catalysts for multiple evolutionary events in Senecio in the British Isles. *Plant Ecology & Diversity,* 8**,** 449-463.

LOWE, M. R. & TYTECA, D. 2012. Two new Ophrys species from Portugal. *Journal Europaischer Orchideen : Mitteilungsblatt des AHO Baden-Württemberg,* 44**,** 207-229.

LOWE, P., HARRIS, S. & ASHTON, P. 2007. *Ecological Genetics: design, analysis and application,* UK, Blackwell Publishing Ltd.

LYNCH, M. 1990. The similarity index and DNA fingerprinting. *Molecular Biology and Evolution,* 7**,** 478-484.

MABBERLEY, D. J. 2017. *Mabberley's Plant-book: A Portable Dictionary of Plants, their Classification and Uses,* Cambridge, Cambridge University Press.

MACAYA-SANZ, D., HEUERTZ, M., LINDTKE, D., GIOVANNI GIUSEPPE, V., LEXER, C. & GONZÁLEZ-MARTÍNEZ, S. 2016. *Causes and consequences of large clonal assemblies in a poplar hybrid zone*.

MALLET, J. 1995. A species definition for the modern synthesis. *Trends Ecol Evol,* 10**,** 294-9.

MALLET, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution,* 20**,** 229-237.

MALLET, J. 2007. Hybrid speciation. *Nature,* 446**,** 279-283.

MANTEL, N. 1967. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research,* 27**,** 209.

MANTEL, N. & VALAND, R. S. 1970. A Technique of Nonparametric Multivariate Analysis. *Biometrics,* 26**,** 547-558.

MARQUES, I., FELINER, G. N., DRAPER MUNT, D., MARTINS-LOUCAO, M. A. & AGUILAR, J. F. 2010. Unraveling cryptic reticulate relationships and the origin of orphan hybrid disjunct populations in Narcissus. *Evolution,* 64**,** 2353-68.

MARQUES, I., FUERTES AGUILAR, J., MARTINS-LOUÇAO, M. A., MOHARREK, F. & NIETO FELINER, G. 2017a. A three-genome five-gene comprehensive phylogeny of the bulbous genus Narcissus (Amaryllidaceae) challenges current classifications and reveals multiple hybridization events. *TAXON,* 66**,** 832-854.

MARQUES, I., JURGENS, A., AGUILAR, J. F. & FELINER, G. N. 2016. Convergent recruitment of new pollinators is triggered by independent hybridization events in Narcissus. *New Phytol,* 210**,** 731-42.

MARQUES, I., LOUREIRO, J., DRAPER, D., CASTRO, M. & CASTRO, S. 2017b. How much do we know about the frequency of hybridisation and polyploidy in the Mediterranean region? *Plant Biology,* 20**,** 21-37.

MARQUES, I., MONTGOMERY, S. A., BARKER, M. S., MACFARLANE, T. D., CONRAN, J. G., CATALÁN, P., RIESEBERG, L. H., RUDALL, P. J. & GRAHAM, S. W. 2015. Transcriptome-derived evidence supports recent polyploidization and a major phylogeographic division in Trithuria submersa (Hydatellaceae, Nymphaeales). *New Phytologist***,** n/a-n/a.

MARQUES, I., NIETO FELINER, G., MARTINS-LOUCAO, M. A. & FUERTES AGUILAR, J. 2012. Genome size and base composition variation in natural and experimental Narcissus (Amaryllidaceae) hybrids. *Ann Bot,* 109**,** 257-64.

MARQUES, I., ROSSELLO-GRAELL, A., DRAPER, D. & IRIONDO, J. M. 2007. Pollination patterns limit hybridization between two sympatric species of Narcissus (Amaryllidaceae). *Am J Bot,* 94**,** 1352-9.

MARTIN, N. H., BOUCK, A. C. & ARNOLD, M. L. 2006. Detecting Adaptive Trait Introgression Between Iris fulva and I. brevicaulis in Highly Selective Field Conditions. *Genetics,* 172**,** 2481.

MAYDEN, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. *In:* CLARIDGE, M. F., DAWAH, H. A. & WILSON, M. R. (eds.) *Species: The units of diversity.* Chapman & Hall.

MAYR, E. 1940. Speciation Phenomena in Birds. *The American Naturalist,* 74**,** 249-278.

MEIRMANS, P. G., LIU, S. & VAN TIENDEREN, P. H. 2018. The Analysis of Polyploid Genetic Data. *J Hered,* 109**,** 283-296.

MEIRMANS, P. G. & VAN TIENDEREN, P. H. 2013. The effects of inheritance in tetraploids on genetic diversity and population divergence. *Heredity,* 110**,** 131-137.

NEILAND, M. R. M. & WILCOCK, C. C. 1998. Fruit set, nectar reward, and rarity in the Orchidaceae. *American Journal of Botany,* 85**,** 1657-1671.

NEVADO, B., HARRIS, S. A., BEAUMONT, M. A. & HISCOCK, S. J. 2020. Rapid homoploid hybrid speciation in British gardens: The origin of Oxford ragwort (Senecio squalidus). *Molecular Ecology,* 29**,** 4221-4233.

NIETO FELINER, G., ALVAREZ, I., FUERTES-AGUILAR, J., HEUERTZ, M., MARQUES, I., MOHARREK, F., PINEIRO, R., RIINA, R., ROSSELLO, J. A., SOLTIS, P. S. & VILLA-MACHIO, I. 2017. Is homoploid hybrid speciation that rare[quest] An empiricist/'s view. *Heredity*.

OSBORNE, O. G., CHAPMAN, M. A., NEVADO, B. & FILATOV, D. A. 2016. Maintenance of Species Boundaries Despite Ongoing Gene Flow in Ragworts. *Genome Biology and Evolution,* 8**,** 1038-1047.

PATIL, V. N. & DADLANI, M. 2009. *Tetrazolium test for seed viability and vigour,* Oxfordshire, UK, CABI.

PAULUS, H. 2006. Deceived males - Pollination biology of the Mediterranean orchid genus Ophrys (Orchidaceae). *Journal Europäischer Orchideen,* 38**,** 303-353.

PAULUS, H. F. 2019. Speciation, pattern recognition and the maximization of pollination: general questions and answers given by the reproductive biology of the orchid genus Ophrys. *Journal of Comparative Physiology A,* 205**,** 285-300.

PAULUS, H. F. & GACK, C. 1990. POLLINATORS AS PREPOLLINATING ISOLATION FACTORS: EVOLUTION AND SPECIATION IN OPHRYS (ORCFFLDACEAE). *Israel Journal of Botany,* 39**,** 43-79.

PEAKALL, R. & SMOUSE, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics,* 28**,** 2537-2539.

PEAKALL, R. O. D. & SMOUSE, P. E. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes,* 6**,** 288-295.

PEDERSON, H. A. & FAURHOLDT, N. 2007. *Ophrys. The Bee Orchids of Europe,* Kew, Kew Publishing.

PELLEGRINO, G., BELLUSCI, F. & MUSACCHIO, A. 2010. *The effects of inflorescence size and flower position on female reproductive success in three deceptive orchids*.

PELLICER, J., CLERMONT, S., HOUSTON, L., RICH, T. C. G. & FAY, M. F. 2012. Cytotype diversity in the Sorbus complex (Rosaceae) in Britain: sorting out the puzzle. *Annals of Botany,* 110**,** 1185-93.

PETIT, C., BRETAGNOLLE, F. & FELBER, F. 1999. Evolutionary consequences of diploid-polyploid hybrid zones in wild species. *Trends in Ecology & Evolution,* 14**,** 306-311.

PFEIFFER, T., ROSCHANSKI, A. M., PANNELL, J. R., KORBECKA, G. & SCHNITTLER, M. 2011. Characterization of Microsatellite Loci and Reliable Genotyping in a Polyploid Plant, Mercurialis perennis (Euphorbiaceae). *Journal of Heredity,* 102**,** 479-488.

PICKUP, M., BRANDVAIN, Y., FRAISSE, C., YAKIMOWSKI, S., BARTON, N. H., DIXIT, T., LEXER, C., CEREGHETTI, E. & FIELD, D. L. 2019. Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. *New Phytol,* 224**,** 1035-1047.

PRIDGEON, A. M., BATEMAN, R. M., COX, A. V., HAPEMAN, J. R. & CHASE, M. W. 1997. Phylogenetic of subtribe Orchidinae (Orchidoideae, Orchidaceae) based on nuclear its sequences. 1. Intergeneric relationships and polyphyly of Orchis sensu lato. *Lindleyana,* 12**,** 89-109.

PRIDGEON, A. M., CRIBB, P. J., CHASE, M. W. & RASMUSSEN, F. N. 2001. Orchidoideae (Part one). *In:* PRIDGEON, A. M., CRIBB, P. J., CHASE, M. W. & RASMUSSEN, F. N. (eds.) *Genera Orchidacearum.* Second ed. United Kingdom: Oxford University Press.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics,* 155**,** 945-959.

PRITCHARD, J. K., WENA, X. & FALUSHB, D. 2010. Documentation for structure software: version 2.3.

PUTMAN, A. I. & CARBONE, I. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution,* 4**,** 4399-4428.

QUILODRAN, C. S., MONTOYA-BURGOS, J. I. & CURRAT, M. 2020. Harmonizing hybridization dissonance in conservation. *Commun Biol,* 3**,** 391.

RAJORA, O. P. & DANCIK, B. P. 1992. Genetic characterization and relationships of Populus alba, P. tremula, and P. x canescens, and their clones. *Theoretical and Applied Genetics,* 84**,** 291-298.

RIESEBERG, L. H. 1997. Hybrid Origins of Plant Species. *Annual Review of Ecology and Systematics,* 28**,** 359-389.

RIESEBERG, L. H., ELLSTRAND, N. C. & ARNOLD, M. 1993. What Can Molecular and Morphological Markers Tell Us About Plant Hybridization? *Critical Reviews in Plant Sciences,* 12**,** 213-241.

RIESEBERG, L. H., KIM, S. C., RANDELL, R. A., WHITNEY, K. D., GROSS, B. L., LEXER, C. & CLAY, K. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica,* 129**,** 149-65.

RIESEBERG, L. H., RAYMOND, O., ROSENTHAL, D. M., LAI, Z., LIVINGSTONE, K., NAKAZATO, T., DURPHY, J. L., SCHWARZBACH, A. E., DONOVAN, L. A. & LEXER, C. 2003. Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. *Science,* 301**,** 1211-1216.

RIESEBERG, L. H., WHITTON, J. & GARDNER, K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics,* 152**,** 713-727.

ROBBIRT, K. M., DAVY, A. J., HUTCHINGS, M. J. & ROBERTS, D. L. 2011. Validation of biological collections as a source of phenological data for use in climate change studies: a case study with the orchid Ophrys sphegodes. *Journal of Ecology,* 99**,** 235-241.

ROBBIRT, K. M., ROBERTS, D. L., HUTCHINGS, M. J. & DAVY, A. J. 2014. Potential disruption of pollination in a sexually deceptive orchid by climatic change. *Curr Biol,* 24**,** 2845-9.

RODRIGUEZ, D. J. 1996. A Model for the Establishment of Polyploidy in Plants. *The American Naturalist,* 147**,** 33-46.

RONFORT, J., JENCZEWSKI, E., BATAILLON, T. & ROUSSET, F. 1998. Analysis of population structure in autotetraploid species. *Genetics,* 150**,** 921-930.

SABARA, H. A., KRON, P. & HUSBAND, B. C. 2013. Cytotype coexistence leads to triploid hybrid production in a diploid–tetraploid contact zone of Chamerion angustifolium (Onagraceae). *American Journal of Botany,* 100**,** 962-970.

SAEFKEN, B., KNEIB, T., VAN WAVEREN, C.-S. & GREVEN, S. 2014. A unifying approach to the estimation of the conditional Akaike information in generalized linear mixed models. *Electron. J. Statist.,* 8**,** 201-225.

SCASCITELLI, M., WHITNEY, K. D., RANDELL, R. A., KING, M., BUERKLE, C. A. & RIESEBERG, L. H. 2010. Genome scan of hybridizing sunflowers from Texas (Helianthus annuus and H. debilis) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Mol Ecol,* 19**,** 521-41.

SCHIESTL, F. P. & AYASSE, M. 2002. Do changes in floral odor cause speciation in sexually deceptive orchids? *Plant Systematics and Evolution,* 234**,** 111-119.

SCHIESTL, F. P., AYASSE, M., PAULUS, H. F., ERDMANN, D. & FRANCKE, W. 1997. Variation of Floral Scent Emission and Postpollination Changes in Individual Flowers of Ophrys sphegodes Subsp. sphegodes. *Journal of Chemical Ecology,* 23**,** 2881-2895.

SCHIESTL, F. P., AYASSE, M., PAULUS, H. F., LÖFSTEDT, C., HANSSON, B. S., IBARRA, F. & FRANCKE, W. 1999. Orchid pollination by sexual swindle. *Nature,* 399**,** 421-421.

SCHIESTL, F. P., AYASSE, M., PAULUS, H. F., LÖFSTEDT, C., HANSSON, B. S., IBARRA, F. & FRANCKE, W. 2000. Sex pheromone mimicry in the early spider orchid (Ophrys sphegodes): patterns of hydrocarbons as the key mechanism for pollination by sexual deception. *Journal of Comparative Physiology A,* 186**,** 567-574.

SCHLÜTER, P. M., RUAS, P. M., KOHL, G., RUAS, C. F., STUESSY, T. F. & PAULUS, H. F. 2011. Evidence for progenitor-derivative speciation in sexually deceptive orchids. *Annals of Botany,* 108**,** 895-906.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology,* 18**,** 233-234.

SCHUMER, M., CUI, R., POWELL, D. L., ROSENTHAL, G. G. & ANDOLFATTO, P. 2016. Ancient hybridization and genomic stabilization in a swordtail fish. *Mol Ecol,* 25**,** 2661-79.

SCOPECE, G., MUSACCHIO, A., WIDMER, A. & COZZOLINO, S. 2007. Patterns of reproductive isolation in Mediterranean deceptive orchids. *Evolution,* 61**,** 2623-2642.

SEDEEK, K. E. M., SCOPECE, G., STAEDLER, Y. M., SCHÖNENBERGER, J., COZZOLINO, S., SCHIESTL, F. P. & SCHLÜTER, P. M. 2014. Genic rather than genome-wide differences between sexually deceptive Ophrys orchids with different pollinators. *Molecular Ecology,* 23**,** 6192-6205.

SEDEEK, KHALID E. M., WHITTLE, E., GUTHÖRL, D., GROSSNIKLAUS, U., SHANKLIN, J. & SCHLÜTER, PHILIPP M. 2016. Amino Acid Change in an Orchid Desaturase Enables Mimicry of the Pollinator's Sex Pheromone. *Current Biology,* 26**,** 1505-1511.

SHEEHAN, H., HERMANN, K. & KUHLEMEIER, C. 2012. Color and scent: how single genes influence pollinator attraction. *Cold Spring Harb Symp Quant Biol,* 77**,** 117-33.

SHIPUNOV, A. B., FAY, M. F., PILLON, Y., BATEMAN, R. M. & CHASE, M. W. 2004. Dactylorhiza (Orchidaceae) in European Russia: combined molecular and morphological analysis. *American Journal of Botany,* 91**,** 1419-1426.

SILVERTOWN, J., SERVAES, C., BISS, P. & MACLEOD, D. 2005. Reinforcement of reproductive isolation between adjacent populations in the Park Grass Experiment. *Heredity,* 95**,** 198.

SLATKIN, M. 1995. A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics,* 139**,** 457-462.

SOLIVA, M., KOCYAN, A. & WIDMER, A. 2001. Molecular phylogenetics of the sexually deceptive orchid genus Ophrys (Orchidaceae) based on nuclear and chloroplast DNA sequences. *Molecular Phylogenetics and Evolution,* 20**,** 78-88.

SOLIVA, M. & WIDMER, A. 2003. Gene flow across species boundaries in sympatric, sexually deceptive Ophrys (Orchidaceae) species. *Evolution,* 57**,** 2252-2261.

SOLTIS, P. & SOLTIS, D. 2000. The role of genetic and genomic attributes in the

success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America,* 97.

SOLTIS, P. S. & SOLTIS, D. E. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol,* 60**,** 561-88.

SPARKS, T. H. & CAREY, P. D. 1995. The Responses of Species to Climate Over Two Centuries: An Analysis of the Marsham Phenological Record, 1736-1947. *Journal of Ecology,* 83**,** 321-329.

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

STACE, C. A. E. 1975. *Hybridization and the Flora of the British Isles. 626 S., 10 Abb. Academic Press, London, New York, San Franzisco*, Wiley-Blackwell.

STAHLBERG, D. 2009. Habitat differentiation, hybridization and gene flow patterns in mixed populations of diploid and autotetraploid Dactylorhiza maculata s.l. (Orchidaceae). *Evolutionary Ecology,* 23**,** 295-328.

STEBBINS, G. L. 1971. *Chromosomal evolution in higher plants,* London, United Kingdom.

STEBBINS, G. L. & FERLAN, L. 1956. Population Variability, Hybridization, and Introgression in Some Species of Ophrys. *Evolution,* 10**,** 32-46.

STERN, W. L. 2008. Vegetative anatomy of subtribe Orchidinae (Orchidaceae). *Botanical Journal of the Linnean Society,* 124**,** 121-136.

STERN, W. L., GREGORY, M. & CUTLER, D. F. 2014. *Anatomy of the Monocotyledons: Orchidaceae*.

STÖKL, J., PAULUS, H., DAFNI, A., SCHULZ, C., FRANCKE, W. & AYASSE, M. 2005. Pollinator attracting odour signals in sexually deceptive orchids of the Ophrys fusca group. *Plant Systematics and Evolution,* 254**,** 105-120.

STOLTING, K. N., NIPPER, R., LINDTKE, D., CASEYS, C., WAEBER, S., CASTIGLIONE, S. & LEXER, C. 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Mol Ecol,* 22**,** 842-55.

STREINZER, M., ELLIS, T., PAULUS, H. F. & SPAETHE, J. 2010. Visual discrimination between two sexually deceptive Ophrys species by a bee pollinator. *Arthropod-Plant Interactions,* 4**,** 141-148.

STREINZER, M., PAULUS, H. F. & SPAETHE, J. 2009. Floral colour signal increases short-range detectability of a sexually deceptive orchid to its bee pollinator. *Journal of Experimental Biology,* 212**,** 1365.

SWARTS, N. D. & DIXON, K. W. 2009. Terrestrial orchid conservation in the age of extinction. *Annals of Botany,* 104**,** 543-556.

SYMONDS, V. V., SOLTIS, P. S. & SOLTIS, D. E. 2010. Dynamics of polyploid formation in Tragopogon (Asteraceae): recurrent formation, gene flow, and population structure. *Evolution,* 64**,** 1984-2003.

TAYLOR, S. A. & LARSON, E. L. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat Ecol Evol,* 3**,** 170-177.

TRAVNICEK, P., PONERT, J., URFUS, T., JERSAKOVA, J., VRANA, J., HRIBOVA, E., DOLEZEL, J. & SUDA, J. 2015. Challenges of flow-cytometric estimation of nuclear genome size in orchids, a plant group with both whole-genome and progressively partial endoreplication. *Cytometry A,* 87**,** 958-66.

TREMBLAY, R. L., ACKERMAN, J. D., ZIMMERMAN, J. K. & CALVO, R. N. 2005. Variation in sexual reproduction in orchids and its evolutionary consequences: a spasmodic journey to diversification. *Biological Journal of the Linnean Society,* 84**,** 1-54.

TYTECA, D. 1997. The Orchid Flora of Portugal. *Journal Europäicher Orchideen,* 29**,** 185-581.

VAHA, J. P. & PRIMMER, C. R. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology,* 15**,** 63-72.

VALLEJO-MARIN, M. & HISCOCK, S. J. 2016. Hybridization and hybrid speciation under global change. *New Phytol,* 211**,** 1170-87.

VANDEWOESTIJNE, S., ROIS, A. S., CAPERTA, A., BAGUETTE, M. & TYTECA, D. 2009. Effects of individual and population parameters on reproductive success in three sexually deceptive orchid species. *Plant Biology,* 11**,** 454-463.

VEREECKEN, N. J. & FRANCISCO, A. 2014. *Ophrys* pollination: from Darwin to the present day. *In:* EDENS-MEIER, R. & BERNHARDT, P. (eds.) *Darwin's Orchids - Then and Now.* Chicago: The University Chicago Press.

VIGNOLINI, S., DAVEY, M., BATEMAN, R., RUDALL, P., MOYROUD, E., TRATT, J., MALMGREN, S., STEINER, U. & J GLOVER, B. 2012a. The mirror crack'd: Both pigment and structure contribute to the glossy blue appearance of the mirror orchid, Ophrys speculum. 196.

VIGNOLINI, S., DAVEY, M. P., BATEMAN, R. M., RUDALL, P. J., MOYROUD, E., TRATT, J., MALMGREN, S., STEINER, U. & GLOVER, B. J. 2012b. The mirror crack'd: both pigment and structure contribute to the glossy blue appearance of the mirror orchid, Ophrys speculum. *New Phytologist,* 196**,** 1038-1047.

VONHOLDT, B. M., BRZESKI, K. E., WILCOVE, D. S. & RUTLEDGE, L. Y. 2018. Redefining the Role of Admixture and Genomics in Species Conservation. *Conservation Letters,* 11**,** e12371.

WAPLES, R. S. 2015. Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity,* 106**,** 1-19.

WEISING, K., ATKINSON, R. G. & GARDNER, R. C. 1995. Genomic fingerprinting by microsatellite-primed PCR: a critical evaluation. *PCR Methods Appl,* 4**,** 249-55.

WHITNEY, K. D., RANDELL, R. A. & RIESEBERG, L. H. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower Helianthus annuus. 187**,** 230-9.

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

WILCOCK, C. & NEILAND, R. 2002. Pollination failure in plants: why it happens and when it matters. *Trends in Plant Science,* 7**,** 270-277.

WILSON, Y. & HUDSON, A. 2011. The evolutionary history of Antirrhinum suggests that ancestral phenotype combinations survived repeated hybridizations. *The Plant Journal,* 66**,** 1032-1043.

WIRTH, L. R., GRAF, R., BRODBECK, S., REBER-FUNK, C., HOLDEREGGER, R. & LANDERGOTT, U. 2009. Fully informative microsatellite markers for the tetraploid plant Eritrichium nanum (Boraginaceae). *Molecular Ecology Resources,* 9**,** 312-314.

WOLF, A. A., ZAVALETA, E. S. & SELMANTS, P. C. 2017. Flowering phenology shifts in response to biodiversity loss. *Proceedings of the National Academy of Sciences,* 114**,** 3463.

WOLF, D. E., TAKEBAYASHI, N. & RIESEBERG, L. H. 2001. Predicting the Risk of Extinction through Hybridization. *Conservation Biology,* 15**,** 1039-1053.

WRIGHT, S. 1949. The genetical structure of populations. *Annals of Eugenics,* 15**,** 323-354.

XU, S., SCHLÜTER, P. M. & SCHIESTL, F. P. 2012. Pollinator-Driven Speciation in Sexually Deceptive Orchids. *International Journal of Ecology,* 2012.

YAKIMOWSKI, S. B. & RIESEBERG, L. H. 2014. The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany,* 101**,** 1247-1258.