Article:
Olofsson, J.K., Curran, E.V., Nyirenda, F. et al. (11 more authors) (2021) Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. Molecular Ecology. ISSN 0962-1083

https://doi.org/10.1111/mec.15871

Received: 25 October 2020 | Revised: 26 February 2021 | Accepted: 1 March 2021

DOI: 10.1111/mec.15871

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap

Jill K. Olofsson[1] | Emma V. Curran[1] | Florence Nyirenda[2] | Matheus E. Bianconi[1] | Luke T. Dunning[1] | Vanja Milenkovic[1] | Graciela Sotelo[1] | Oriane Hidalgo[3] | Robyn F. Powell[3] | Marjorie R. Lundgren[1] | Ilia J. Leitch[3] | Patrik Nosil[1] | Colin P. Osborne[1] | Pascal-Antoine Christin[1]

[1]Department of Animal and Plant Science, University of Sheffield, Sheffield, UK

[2]Department of Biological Sciences, University of Zambia, Lusaka, Zambia

[3]Royal Botanic Gardens, Kew, Richmond, Surrey, UK

Correspondence
Pascal-Antoine Christin, Department of Animal and Plant Science, University of Sheffield, Western Bank, Sheffield, UK.
Email: p.christin@sheffield.ac.uk

Present address
Jill K. Olofsson, Section for GeoGenetics, Globe Institute University of Copenhagen, Øster Voldgade 5-7, Copenhagen, DK-1350, Denmark
Oriane Hidalgo, Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Passeig del Migdia s.n., Barcelona, 08038, Spain
Marjorie R. Lundgren, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YW, UK
Patrik Nosil, Centre for Functional Ecology and Evolution, National Centre for Scientific Research, 1919 route de Mende, Montpellier, 34000, France

Funding information
Royal Society, Grant/Award Number: RGF\EA\181050 and URF\R\180022; Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/J004243/1; H2020 European Research Council, Grant/Award Number: ERC-2014-STG-638333; Medical Research Council, Grant/Award Number: MR/K001744/1; Natural Environment Research Council, Grant/Award Number: NE/M00208X/1 and R8/H10/56

Abstract

Geographical isolation facilitates the emergence of distinct phenotypes within a single species, but reproductive barriers or selection are needed to maintain the polymorphism after secondary contact. Here, we explore the processes that maintain intraspecific variation of C_4 photosynthesis, a complex trait that results from the combined action of multiple genes. The grass *Alloteropsis semialata* includes C_4 and non-C_4 populations, which have coexisted as a polyploid series for more than 1 million years in the miombo woodlands of Africa. Using population genomics, we show that there is genome-wide divergence for the photosynthetic types, but the current geographical distribution does not reflect a simple habitat displacement scenario as the genetic clusters overlap, being occasionally mixed within a given habitat. Despite evidence of recurrent introgression between non-C_4 and C_4 groups, in both diploids and polyploids, the distinct genetic lineages retain their identity, potentially because of selection against hybrids. Coupled with strong isolation by distance within each genetic group, this selection created a geographical mosaic of photosynthetic types. Diploid C_4 and non-C_4 types never grew together, and the C_4 type from mixed populations constantly belonged to the hexaploid lineage. By limiting reproductive interactions between photosynthetic types, the ploidy difference probably allows their co-occurrence, reinforcing the functional diversity within this species. Together, these factors enabled the persistence of divergent physiological traits of ecological importance within a single species despite gene flow and habitat overlap.

KEYWORDS
C_4 photosynthesis, hybridization, introgression, population genomics, secondary contact, selection

1 | INTRODUCTION

Geographical isolation leads to genetic divergence of populations and, over time, speciation (Bolnick & Fitzpatrick, 2007; Butlin et al., 2008; Jordan, 1905; Mayr, 1947; Templeton, 1981). Populations in allopatry may experience different local selection pressures, which can drive phenotypic divergence. Following the removal of geographical barriers, secondary contact between populations with different phenotypes allows gene flow to resume (Barton, 2001; Dong et al., 2020). If the different phenotypes are selectively equivalent, or one is advantageous over the whole species' range, secondary genetic exchanges will lead to a homogenization of the phenotype over time, potentially erasing some of the character states that evolved in isolation (Coyne & Orr, 2004). When populations with different locally adapted phenotypes come into secondary contact, the balance between gene flow and selection maintains the different phenotypes in their respective environments and can result in a stable hybrid zone (Abbott, 2017; Barton, 2001; Barton & Hewitt, 1985; Endler, 1977; Ingles & Biglione, 1952; Slatkin, 1973). These dynamics have been well studied in the case of simple traits controlled by a few genes and for quantitative traits (Gay et al., 2008; Mallet, 1986; Mallet et al., 1990; Nürnberger et al., 1995; Poelstra et al., 2014; Toews et al., 2016). Here, we test whether similar processes dictate the fate of complex physiological adaptations that emerge only when multiple genes are modified.

C_4 photosynthesis is a complex physiological trait that evolved over the ancestral C_3 type and increases plant productivity in warm and arid conditions (Atkinson et al., 2016; Ehleringer & Monson, 1993; Hatch, 1987; Sage et al., 2012). The C_4 phenotype results from the coordinated action of multiple enzymes in a suitable leaf anatomy (Hatch, 1987; Sage et al., 2012; Williams et al., 2013). Most C_4 lineages evolved 5–30 million years ago (Christin et al., 2011), and extant species generally include a single photosynthetic type. The clearest exception is the palaeotropical grass *Alloteropsis semialata*, which presents unequalled intraspecific photosynthetic diversity, with C_4 genotypes and non-C_4 populations with different amounts of C_4 cycle activity (Ellis, 1974,

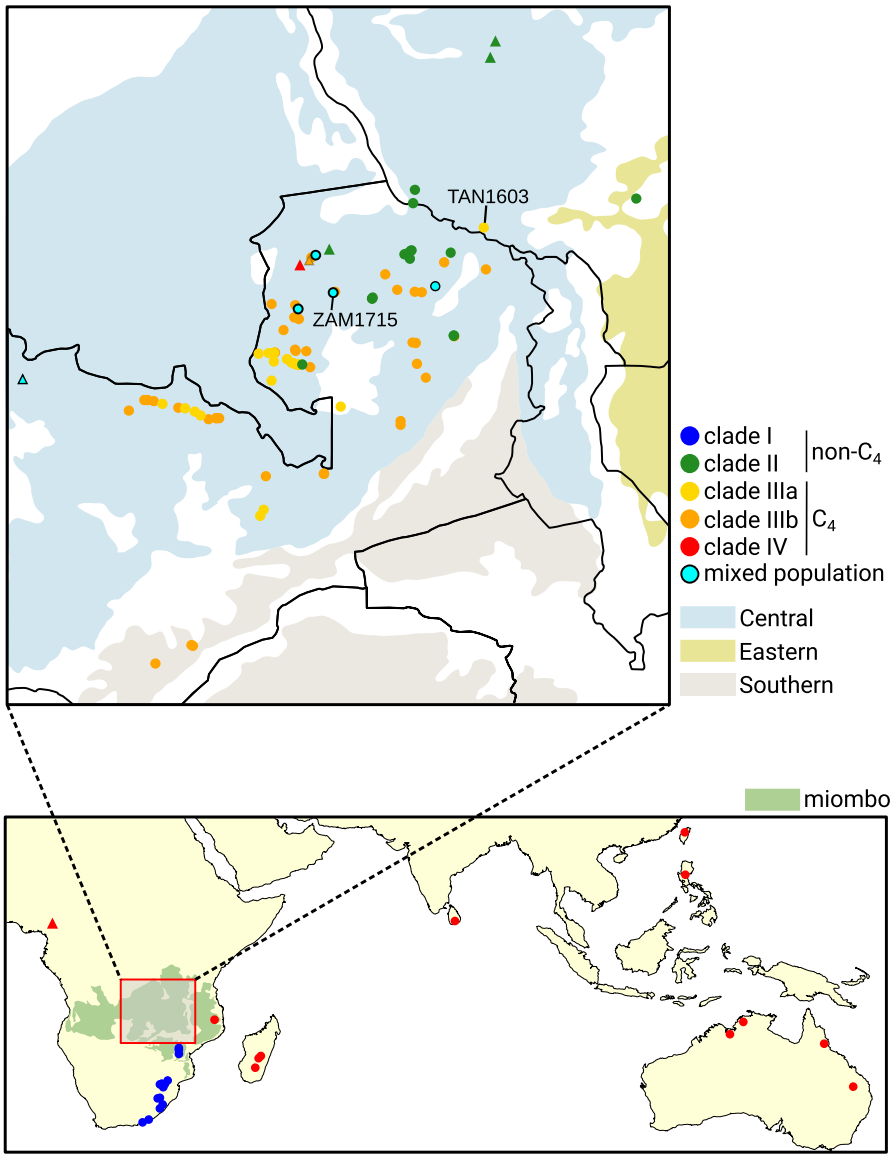

FIGURE 1 Distribution of sampled populations. Geographical locations of all sampled *Alloteropsis semialata* populations are shown on the large map (bottom panel), with the dense sampling of the Central Zambezian region shown in detail (upper panel). The full extent of miombo woodlands is shown on the bottom panel, with different miombo ecoregions on the upper panel based on Maquia et al., (2019). Populations are coloured based on their genetic lineage (see Figure S3). Populations with strong signs of admixture in the cluster analyses (Figure 4) are indicated with triangles, and the positions of the densely sampled population (ZAM1715; Figure S1) and one with strong signs of introgression between diploid C_4 and diploid non-C_4 (TAN1603; Figure 5) are indicated

1981; Frean et al., 1980; Lundgren et al., 2016). The species originated in the Central Zambezian miombo woodlands of Africa around 3 million years ago (Bianconi et al., 2020; Lundgren et al., 2015; Figure 1). Lineage isolation, potentially linked to episodic woodland contraction in the region during glacial cycles, was followed by the emergence of the C_4 phenotype, while migration during periods of re-expansion allowed secondary contacts between C_4 and non-C_4 groups (Bianconi et al., 2020; Olofsson et al., 2016), although whether hybrid zones were created remains unknown. The species later spread outside of its region of origin, with C_4 lineages colonizing the rest of Africa, Asia and Oceania, while a non-C_4 lineage migrated to southern Africa (Lundgren et al., 2015). Previous genome analyses have found evidence of admixture between C_4 and non-C_4 populations that happened recurrently over the past 2 million years (Bianconi et al., 2020; Olofsson et al., 2016), showing that reproductive barriers are at most partial. Despite their ability to hybridize, C_4 and non-C_4 lineages of *A. semialata* have coexisted in the Central Zambezian region for more than 1 million years (Bianconi et al., 2020). The factors allowing the persistence of these two photosynthetic types remain unknown.

Variation of ploidy levels has previously been reported within *A. semialata*. All the non-C_4 individuals analysed so far are diploid, while C_4 individuals can be di-, tetra-, hexa-, octo- or even dodecaploids (Bianconi et al., 2020; Ellis, 1981; Frean & Marks, 1988; Liebenberg & Fossey, 2001; Lundgren et al., 2015; Olofsson et al., 2016). Genome analyses have shown that polyploidization happened repeatedly in different parts of the species range, in some cases facilitating mixing of non-C_4 and C_4 genetic groups (Bianconi et al., 2020). In the Central Zambezian region, diploid and hexaploid C_4 have been reported alongside diploid non-C_4 (Bianconi et al., 2020). The effect of this variation of ploidy levels on gene flow dynamics across the region remains unexplored.

In this study, we analyse the photosynthetic types, geographical distribution, genome dynamics and ploidy levels of *A. semialata* populations to test for mechanisms responsible for the coexistence of C_4 and non-C_4 populations in Central Zambezian miombo woodlands. First, we characterize the geographical distribution of the different photosynthetic types to test the hypothesis that the photosynthetic innovation was accompanied by habitat displacement upon secondary contact (Brown & Wilson, 1956; Pfenning & Pfenning, 2010; Raabová et al., 2008; Schluter, 2000; Schluter & McPhail, 1992). Second, we establish the distribution of ploidy levels among photosynthetic types and genetic lineages to test the hypothesis that ploidy differences generate reproductive barriers (Moyle et al., 2004; Ramsey & Schemske, 1998). Third, we quantify the population structure and introgression events to test the hypothesis that secondary exchanges coupled with selection created one or more hybrid zones among groups of the same ploidy level (Barton, 2001; Barton & Hewitt, 1985; Endler, 1977; Ingles & Biglione, 1952). Fourth, we evaluate the patterns of differentiation across the genome to test the hypothesis that the photosynthetic types are maintained by C_4 genes acting as barrier loci (Barton & Bengtsson, 1986; Ravinet et al., 2017; Wolf & Ellegren, 2017), due to the breakdown of the C_4 apparatus in hybrids (Brown & Bouton, 1993; Oakly et al., 2014). Overall, our investigations indicate that different states for the C_4 complex trait are maintained among diploids because of low dispersal, while the presence of polyploids increases the overall diversity.

2 | MATERIAL AND METHODS

2.1 | Population sampling and determination of photosynthetic types

Populations of *Alloteropsis semialata* were located during consecutive field trips conducted in the rainy seasons between 2012 and 2019 (Table S1). Populations were visually identified while driving or found during stop-and-walk searches. At least one individual was sampled from all identified populations. At each collection point, GPS coordinates were recorded and leaf material from individuals growing at least 1 m apart were sampled and dried in silica gel. For some populations, live cuttings or seeds were also collected. Large populations where different photosynthetic types were suspected to co-occur after inspecting the vein architecture with a hand lens were sampled more intensively. In particular, for one population, ZAM1715, a total of 103 individuals were collected along parallel transects and individual-level GPS coordinates were recorded (Figure S1).

The photosynthetic type of each plant was determined using stable carbon isotopes. This method differentiates plants that grew using a C_4 photosynthetic cycle fixing the majority of their carbon with phosphoenolpyruvate carboxylase (PEPC) from non-C_4 plants that acquired most of their carbon without the C_4 cycle (non-C_4 plants; Bender, 1971; Stata et al., 2019). For some individuals, carbon isotope ratios were retrieved from previous studies (Table S1). For others, a total of 1–2 mg of dried leaf tissue was sampled and analysed using an ANCA GSL preparation module coupled to a 20–20 stable isotope analyser (PDZ Europa). The carbon isotopic ratio ($\delta^{13}C$, in ‰) was reported relative to the standard Pee Dee Belemnite (PDB). Individuals with a $\delta^{13}C$ value above −17‰ were considered to be C_4 whereas all other individuals were considered to be non-C_4, which can include plants without any C_4 activity (referred to as C_3 plants) and plants with a weak C_4 cycle activity (Stata et al., 2019). Non-C_4 individuals of *A. semialata* from the Central Zambezian region previously analysed used a weak C_4 cycle, while those from southern Africa were C_3 (Bianconi et al., 2020; Lundgren et al., 2016). The generality of this pattern cannot be confirmed based solely on carbon isotopes and we consequently refer collectively to these plants as non-C_4. If there was enough material left, the isotope measurements were repeated for individuals where the $\delta^{13}C$ values did not match other individuals of the population and genomic group (Table S1).

2.2 | Estimation of genome sizes

Genome sizes for some individuals were retrieved from the literature (Bianconi et al., 2020; Olofsson et al., 2016) and used to infer

ploidy levels based on known genome sizes of diploids and hexaploids of *A. semialata* (Olofsson et al., 2016). To supplement this data set, we estimated the genome sizes of additional populations from the Central Zambezian region. We first focused on individuals for which we had collected live cuttings and then used individuals grown from field-collected seeds. Finally, we added estimates based on silica gel-dried leaves for some populations in an effort to maximize the geographical and photosynthetic diversity of populations with genome size estimates (Table S1). The new genome sizes were estimated by flow cytometry using internal calibration standards, as described in Bianconi et al. (2020).

2.3 | Population-level RAD sequencing and genotyping

Genomic diversity was assessed using a reduced level sequencing approach. DNA was extracted from a 1- to 3-cm dried leaf fragment using the DNeasy Plant Mini Kit (Qiagen), following the manufacturer's protocol. Double-digested restriction-associated DNA sequencing (ddRADSeq) libraries were produced as previously outlined (Olofsson et al., 2019). Briefly, a barcoded and a common adaptor (Petersen et al., 2014) were ligated to ~200–700 ng DNA (7 μl DNA extract) double digested with *Eco*RI and *Mse*I. Standard Illumina sequencing primers were then added to the ligation products through PCR (polymerase chain reaction) amplification (16 cycles). Pools of barcode compatible libraries (93–96 libraries from the same or different projects) were gel size selected (300–600 bp) and purified using the QIAquick Gel Extraction Kit (Qiagen). Size-selected library pools were pair-end sequenced (125 bp) on one lane of an Illumina HiSeq2500 at the Edinburgh Genomics Centre, University of Edinburgh (UK), at the Centre for Genomic Research, Liverpool University (UK), or at the Sheffield Diagnostic Genetics Service (UK), following standard protocols.

Raw reads were trimmed, cleaned and demultiplexed as previously described (Olofsson et al., 2019). In short, the program TRIMMOMATIC tool kit (Bolger et al., 2014) was used to remove adaptor and primer sequences (ILLUMINACLIP option in palindrome mode). Using TRIMMOMATIC, low-quality bases (Q < 3) were further removed from the 5′ and 3′ ends, as were all bases with a low quality (Q < 15) in a four-base sliding window. The cleaned reads were demultiplexed and barcodes were removed using the process.RADtag.pl script from the program STACKS (Catchen et al., 2013). Reads mapping to the chloroplast were then removed using SAMTOOLS version 2.2.3 (Li, 2011; Li et al., 2009). Clean reads were mapped to the reference genome of *A. semialata* (ASEM_AUS1_v1.0; GenBank accession QPGU01000000; Dunning, Olofsson, et al., 2019) using the default settings for paired-end reads in BOWTIE2 version 2.2.3 (Langmead & Salzberg, 2012). Variants (single nucleotide polymorphisms [SNPs]) covered by 2–20 reads were called from uniquely mapped reads using a combination of SAMTOOLS and the consensus calling function in BCFTOOLS version 1.1.1 (Li, 2011) treating all samples as if they were diploids, because treating individuals as higher ploidy levels would be unreliable with this type and depth of sequencing (Stift et al., 2019). Finally, all variants were merged and filtered, only keeping individuals with less than 99% missing data and only keeping variants at least 1 kb apart on one of the nine chromosomes with a minor allele count of 56 (minor allele frequency of ~0.05) and less than 90% missing data.

2.4 | Phylogenetic analyses, population structure and tests for isolation by distance

A maximum likelihood phylogeny was inferred in RAXML version 8.2.11 (Stamatakis, 2014) for the nuclear SNP alignment using a GTR+G substitution model. Node support was evaluated with 100 rapid bootstraps.

Population structure was evaluated with a principal component analysis (PCA) implemented in the R package adegenet (Jombart, 2008; Jombart & Ahmed, 2011; R Core Team, 2018) and with a Bayesian admixture analysis implemented in NGSADMIX (Skotte et al., 2013). While some clustering analyses are robust to ploidy variation (Stift et al., 2019), PCAs tend to group individuals by ploidy level (Meirmans et al., 2018). These analyses were used here to describe the broad patterns of variation, with subsequent analyses repeated without suspected polyploids. In the PCA, missing data were coded as the mean of the allele frequency among individuals with data at each respective SNP, whereas they were coded as zero in the admixture analysis. The optimal number of population clusters in the admixture analysis was estimated using the Evanno et al., (2005) method, as implemented in CLUMPAK (Kopelman et al., 2015) from 10 independent runs for an increasing number of population clusters from one to 20. A separate clustering analysis was performed solely on individuals of the densely sampled population ZAM1715.

An analysis of molecular variance (AMOVA) among the five genomic groups (see Results) was performed on the variants using ARLEQUIN version 3.5.2.2 (Excoffier & Lischer, 2010). Populations where non-C_4 and C_4 individuals co-occurred (see Results) were split into two groups based on the assignment of individuals to nuclear clades. This analysis was repeated excluding the nuclear clade that contains mainly hexaploid samples (clade IIIb) as well as the known polyploid population from clade II (ZAM1958) and those from clade IV (Table S1).

Isolation by distance was evaluated by calculating pair-wise F_{ST} in VCFTOOLS version 0.1.14 (Danecek et al., 2011) between populations with at least three individuals. Populations from clade IV were not included as they span several continents and islands, making dispersal distances difficult to compare. Mixed populations were separated by nuclear clade, as above, and the polyploid population ZAM1958 was excluded, resulting in 96 populations (15 from clade I, 19 from clade II, 21 from clade IIIa and 41 from clade IIIb). Geographical distance (in km) between each pair of these populations was extracted from the GPS coordinates taking the curve of the Earth into account in R. The relationship between genetic and geographical distances was tested using Mantel tests. These

analyses were conducted independently for individuals belonging to the same or different genomic groups (see Results), in each case using the Spearman correlation and 9,999 permutations. For intragroup comparisons, rows and columns of the square, symmetrical distance matrix were shuffled simultaneously. The intergroup comparisons produced rectangular, nonsymmetrical matrices, and their columns and rows were shuffled independently. Slopes, intercepts and R^2 were estimated using linear regressions.

2.5 | Tests for introgression

Introgression between genetic clusters was tested with Patterson's *D*-statistics (ABBA-BABA statistics), using the software package DSUITE version 0.4, which counts site patterns using population allele frequencies (Malinsky et al., 2020). Based on the phylogenetic tree (see Results), clade I was used as the outgroup (O) and the three other groups (clades II, III and IV) were used in the P3, P2 and P1 positions, respectively. First, we combined all samples from the three clades that are thought to contain mainly diploid individuals: the non-C_4 clade II (excluding the hexaploid C_4 ZAM1958), the C_4 clade IIIa and the known diploid populations from C_4 clade IV. Second, we repeated this test with the C_4 clade IIIb that contains the hexaploids instead of clade IIIa. Third, we tested for gene flow between clades II (minus ZAM1958), IIIa and IIIb. Finally, we tested for gene flow between individual populations from clade II and individual populations from either clade IIIa or clade IIIb. For the last set of analyses, the Australian population from clade IV was always used in the P1 position as it the most distant from the Central Zambezian region, decreasing the chances of secondary gene flow.

2.6 | Genome scans

Genetic differentiation (F_{ST} and d_{xy}) was evaluated across the whole genome (nine chromosomes) of *A. semialata* between the two geographically overlapping non-C_4 and C_4 genomic groups containing diploids (groups II [excluding the hexaploid ZAM1958 population] and IIIa; see Results) as previously outlined (Olofsson et al., 2019). In short, F_{ST} was calculated in 500-kb sliding windows (10-kb slide) from all variants using VCFTOOLS. All variants with a minor allele count of 21 were included to allow for a minor allele frequency of approximately 0.05 among the included individuals (107 non-C_4 from clade II and 104 C_4 from clade IIIa). Absolute divergence (d_{xy}) was calculated in 500-kb sliding windows (10-kb slides) from allele frequencies obtained from VCFTOOLS using a combination of custom python, R and bash scripts as previously outlined (Olofsson et al., 2019). Null distributions of F_{ST} and d_{xy} were obtained according to Olofsson et al., (2019). Windows above the 99th percentile or below the 1st percentile of the F_{ST} and d_{xy} null distributions were considered outliers. Finally, the F_{ST} and d_{xy} in windows were plotted along the genome using the python library matplotlib (Hunter, 2007). The values of the two statistics were retrieved from the windows containing three genes with known differential expression between photosynthetic types in *A. semialata* (Dunning, Moreno-Villena, et al., 2019 and Dunning, Olofsson, et al., 2019) and were compared to the rest of the genome. To evaluate the influence of the general increase in differentiation and divergence around putative centromeres (Olofsson et al., 2019), we also compared differentiation of windows containing the three C_4 genes to those from the rest of the genome with a gene density ≥50/Mb (Dunning, Moreno-Villena, et al., 2019).

3 | RESULTS

3.1 | Different photosynthetic types can co-occur in the same location

Populations were densely sampled across Central Zambezian miombo woodlands (87 distinct locations within Zambia and southwest Tanzania; Figure 1), where most of the photosynthetic diversity in *Alloteropis semialata* is found (Bianconi et al., 2020; Lundgren et al., 2016). Other populations across the global species' range were added to the data set to provide comparison points (Figure 1).

Carbon isotope ratios can identify individuals that grew using the C_4 photosynthetic cycle (Bender, 1971). Values for a total of 593 individuals of *A. semialata* show a bimodal distribution, as previously reported (Lundgren et al., 2016; Figure S2, Table S1). All values above −17‰ were categorized as C_4, with all others referred to as non-C_4. The whole range of isotope values is observed within the Central Zambezian miombo woodlands, where C_4 and non-C_4 types geographically overlap (Figure 1). Most populations in this region are composed of individuals with a single photosynthetic type (either C_4 or non-C_4), but non-C_4 and C_4 carbon isotopes were detected in seven populations spread across Zambia (JKO2, JKO25, ZAM1503, ZAM1507, ZAM1715, ZAM1936 and ZAM1957; Table S1). In two of these (JKO2 and JKO25), the existence of different photosynthetic types (C_4 and non-C_4) was supported by a single individual, a lack of biological material prevented repetition of the isotopic measurements (Table S1), and these individuals were not genetically distinct from the other individuals in the same population (see below). These two populations were therefore not considered as mixed in this study. In the other five populations, repeated isotope measurements confirmed the existence of different photosynthetic phenotypes, with multiple individuals of each type sampled in four cases (Table S1). The coexistence of different photosynthetic types within five populations is thus unambiguous (populations indicated in Figure 1). Within one of these five populations (ZAM1715), sampling of 103 individuals confirmed that C_4 and non-C_4 types are spatially mixed (Figure S1; Table S1), and in some cases grow <3 cm apart. These distribution patterns show that the photosynthetic types overlap and can in some cases be found completely mixed.

3.2 | C_4 and non-C_4 types correspond to different genetic groups

All sampled individuals were assigned to previously delimited nuclear lineages (Bianconi et al., 2020; Olofsson et al., 2016) using a phylogenetic analysis inferred from 18,513 nuclear SNPs obtained with ddRADSeq for 566 individuals sampled from 87 populations within the Central Zambezian region and 36 populations (147 individuals) from the rest of the species range (Figure 1; Figures S3). As expected, the non-C_4 populations from southern Africa (Zimbabwe and South Africa) clustered with individuals from the previously identified clade I (Figure S3; Bianconi et al., 2020; Olofsson et al., 2016). For populations from the Central Zambezian region, most isotopically non-C_4 individuals were positioned within clade II, while the majority of C_4 individuals were placed within clade III (Figure 2a; Figures S3). The only exceptions are two individuals from a photosynthetic mixed population from Central Zambezia (ZAM1955-02 and ZAM1955-05) that are positioned at the base of clades III+IV, and a single C_4 population (ZAM1958-02) that is positioned at the base of clade II (Figure S3).

The first four axes of a PCA on nuclear markers explain a total of 39.81% of the variance in the data set and jointly identify five clusters corresponding to the three nuclear phylogenetic clades I, II and IV, plus clade III divided on the fourth component into two groups referred to as IIIa and IIIb (Figure 2; Figure S3). In the phylogenetic tree, clade IIIb forms a poorly resolved group that is paraphyletic with respect to the well-supported clade IIIa (Figure S3).

Based on our sampling, the nuclear clade IIIb is the most frequent in Zambia, but the distributions of clades II, IIIa and IIIb overlap geographically in the studied region (Figure 1). Our field records show that each group is found in both open grasslands and habitats with a tree cover (Figure S4). In one population with different photosynthetic types (ZAM1936), the non-C_4 individuals occupied an open grassland while the C_4 individuals were spread in the adjacent woodland, but in the four others (ZAM1503, ZAM1507, ZAM1715 and ZAM1957), C_4 and non-C_4 individuals appeared mixed within the same habitat. In all five populations with differences in photosynthetic types, the non-C_4 individuals belong to clade II and the C_4 individuals belong to clade IIIb (Table S1). This pattern was confirmed with a dense sampling of population ZAM1715 (Figure S1a).

3.3 | In mixed populations, photosynthetic types are associated with distinct ploidy levels

We report here genome size estimates for an additional 52 individuals from 26 populations, leading to a total of 92 individuals from 49 populations with genome sizes (Figure S5), of which 61 individuals from 28 populations are from the Central Zambezian region (Figure 3a). Within this region, most genome sizes are close to 2 Gb/2C (1.88–2.71 Gb, n = 42) or close to 6 Gb/2C (5.16–6.56 Gb, n = 15) (Figure 3b; Table S1), as previously reported for diploid and hexaploid *A. semialata* individuals, respectively (Bianconi et al., 2020; Olofsson et al., 2016). Four individuals with values close to 7 Gb/2C are possibly octoploids, but these values were estimated from silica gel-dried material, and analysis of fresh samples is needed to distinguish hexa- from octopolyploids (Figure 3; Table S1).

In the Central Zambezian region, all 28 non-C_4 individuals with genome size estimates are diploids (Figure 3; Figure S2, Table S1). All 13 analysed individuals from the C_4 clade IIIa are diploid, while all 16 from the paraphyletic clade IIIb and the two individuals from population ZAM1955 positions at the base of clades III+IV are probably hexaploid (Figure 3b). The only C_4 population assigned to clade II (ZAM1958) contains hexaploids, while the only population from clade IV from this region is diploid (Figure 3c). In the two populations with mixed photosynthetic types for which genome sizes are available, non-C_4 individuals are all diploid while C_4 individuals are all likely hexaploid (Figure 3c).

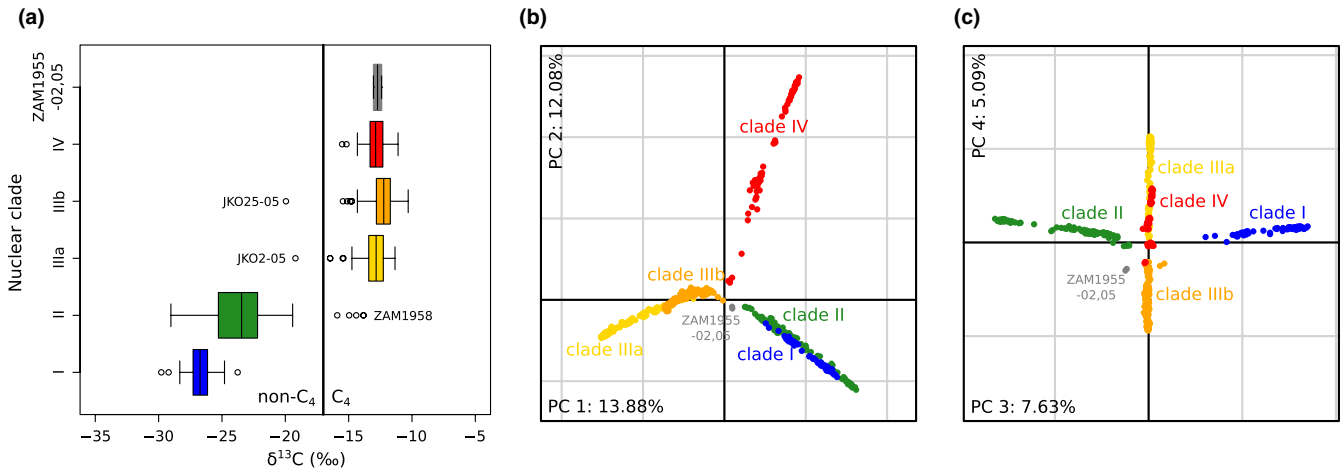

FIGURE 2 Photosynthetic and genetic diversity among nuclear groups. Nuclear groups within *Alloteropsis semialata* were identified based on phylogenetic analyses (Figure S3). (a) The distribution of carbon isotope ratios ($\delta^{13}C$) is shown for each nuclear group. A principal component analysis assessed the genetic variation, shown here along (b) the first two axes and (c) the third and four axes. Genetic groups are coloured as in Figure1

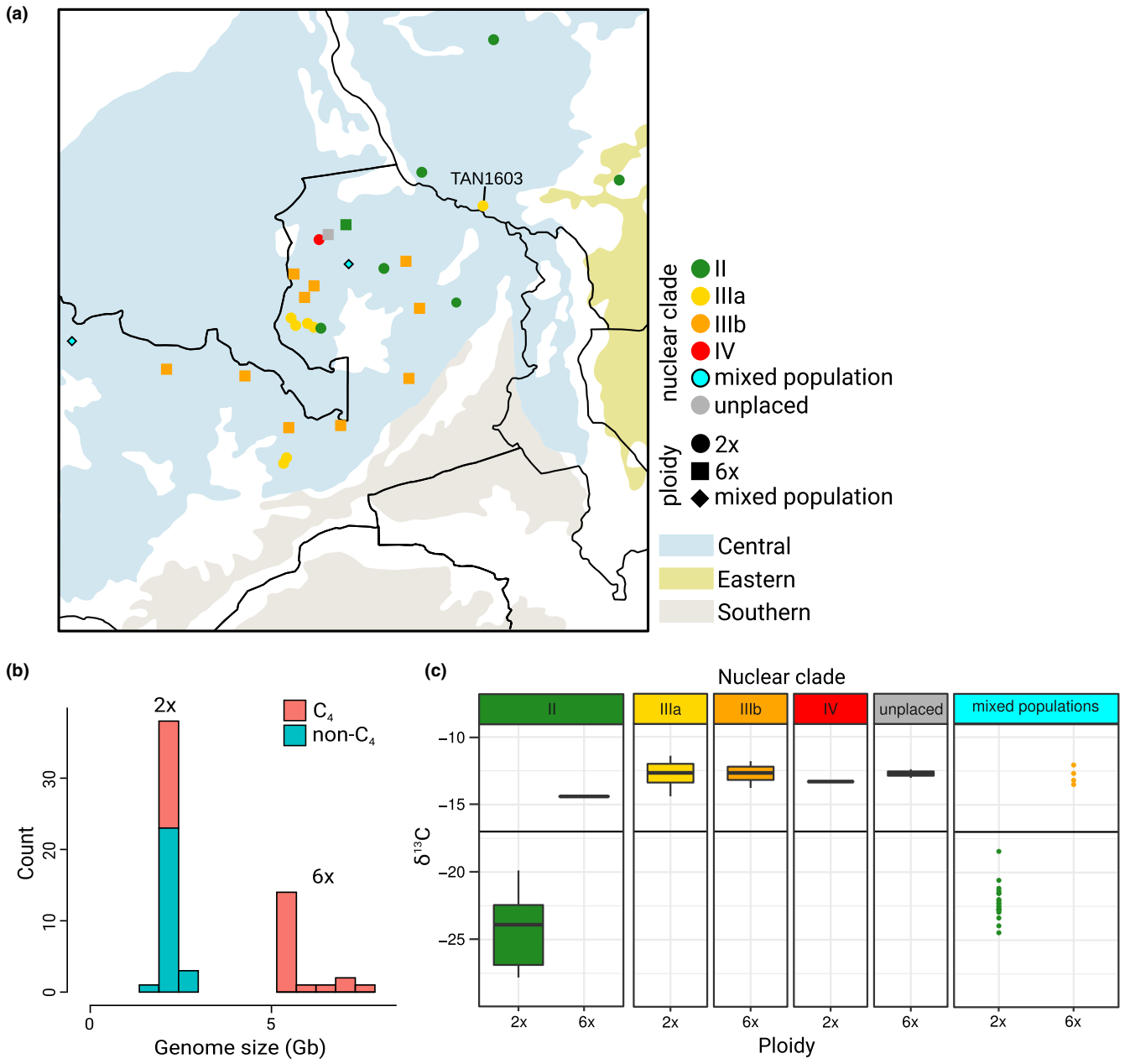

FIGURE 3 Ploidy levels and photosynthetic types. (a) Map corresponding to the one in Figure 1 showing the geographical location of populations from the Central Zambezian region for which at least one genome size was estimated. Point colours show the genetic groups (based on phylogenetic analyses; Figure S3), while symbols show the ploidy levels. (b) The distribution of estimated genome size (2C-values) is shown for 61 samples from the Central Zambezian region. (c) Carbon isotope ratios ($\delta^{13}C$) are shown for each lineage and ploidy level from the Central Zambezian populations

3.4 | Distinct genetic groups are maintained despite gene flow

The optimal number of population clusters estimated using Bayesian admixture analyses as implemented in NGSADMIX is three, which like the PCA based on nuclear markers splits the samples into groups corresponding to the non-C_4 clades I+II, the C_4 clade III, and the C_4 clade IV (Figure 4; Figure S6). A secondary optimum of six population clusters then sorts the samples into the five genomic groups identified in the PCA with clade IIIa splitting into two sub-groups (Figure 4; Figure S6). Most individuals are assigned to a single genomic group, but admixture between C_4 and non-C_4 clusters is clear in the population previously shown to contain dodecaploids (Bianconi et al., 2020), but also in some newly reported hexaploids from Zambia (Figure 4; Figure S7). Importantly, admixture between distinct photosynthetic types is also suggested in some diploid individuals (Figure 4; Figure S7), showing that it can occur without polyploidization. These admixed individuals are found in different parts of the range (Figure 1), and geographically close populations have different types of admixture proportions (Figure 4), which does not conform to typical hybrid zones.

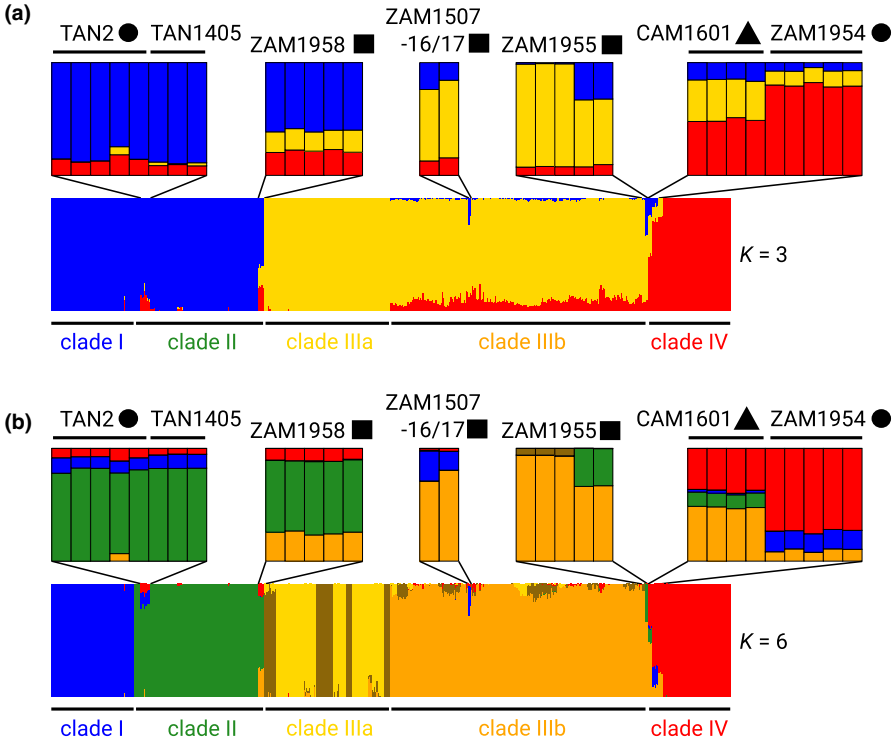

FIGURE 4 Genetic structure within *Alloteropsis semialata*. Results of an admixture analysis are shown for (a) three (K = 3) and (b) six clusters (K = 6). The nuclear groups are delimited at the bottom and coloured as in Figure 1. Each bar represents one individual. Samples with signs of admixture between different photosynthetic types are shown in detail on top, and the ploidy levels of samples within these populations are indicated with symbols next to the names; circle =2*x*, square =6*x* or 8*x*, triangle =12x

Patterson's *D*-statistics were used to test for gene flow among populations with distinct photosynthetic types (Figure 5). Using clade I as an outgroup, we found strong evidence for gene flow between the non-C_4 clade II and the C_4 clade IIIa (D = 0.09: p < .001) as well as between clade II and the C_4 clade IIIb (D = 0.05: p < .05; Figure 5). Introgression from clade II was stronger into clade IIIa, which contains diploids, than into clade IIIb, which contains hexaploids (D = 0.04; p < .005; Figure 5). A number of individual populations from clades II and IIIa showed signs of introgression when clade I was used as the outgroup with the Australian population from clade IV in the P1 position, and clade IIIa population TAN1603, which is located in a region where clade II is abundant (Figure 1), appeared especially introgressed (Figure 5). While some geographically close pairs of populations show no signs of introgression, all significant cases occurred among populations less than 500 km away (Figure 5). Similar patterns are observed with populations from clade IIIb (Figure 5), where many of the significant comparisons involve population ZAM1955, which includes admixed individuals (Figure 4).

AMOVAs show that the level of differentiation among the five genomic groups is high (47.2% of the total variation), with only 8.4% of the variation found among populations within each of the groups and 44.4% within populations. After excluding the groups known to contain polyploids (clade IIIb and population ZAM1958), these values changed to 57.3%, 14.5% and 28.2%, respectively. However, pairwise F_{ST} values between populations from different genomic groups only occasionally exceed 0.8. In many cases, the pairwise F_{ST} between populations from different genomic groups is below the values observed between geographically distant populations from the same genomic group (Figure 6; Figure S8). When considering the C_4 diploid clade IIIa, lower differentiation (F_{ST} as low as 0.4) is observed with the geographically overlapping non-C_4 clade II than with the geographically distant non-C_4 clade I (Figure S8). Together with the *D*-statistics and clustering analyses, these patterns support episodic genetic exchanges among the different photosynthetic types.

The connectivity among populations was evaluated by testing for isolation by distance, excluding clade IV, which is spread among multiple continents and islands (Figure 1), and the polyploid population ZAM1958. After correcting for multiple tests, the relationship between genetic differentiation and geographical distance is significant within each of the four clades I, II, IIIa and IIIb (Figure 6). However, the relationship is weaker within the non-C_4 group II, which might indicate the existence of undetected genetic subgroups. During field collections, we observed morphological variation among the non-C_4 from clade II, some having short, dense racemes while the others had long racemes similar to those of the C_4 plants. These two types were retrieved in the admixture analysis with seven clusters (Figure S6). When they were analysed separately, each showed evidence of isolation by distance (Figure S9). While some barriers might exist within group II, gene flow is therefore prevalent within the other genomic groups and mainly restricted by geography (Figure 6). When considering pairs of populations from different genomic groups, the pattern of isolation by distance is significant between the two non-C_4 clades I and II, and between the C_4 group IIIb that contains hexaploids and the non-C_4 group I as well as the C_4 group IIIa, which both contain diploids (Figure S8). In these cases, however, the intercept is high (0.67, 0.58 and 0.33, respectively), and the three other comparisons do not show evidence of isolation by distance (Figure S8). Similar conclusions are reached when considering the two subgroups from clade II with different inflorescence types (Figure S9). Gene flow between the genomic groups is therefore limited and in most cases it is independent of the geographical distance (Figure S8). These patterns, which mirror those reported in another grass species with marked

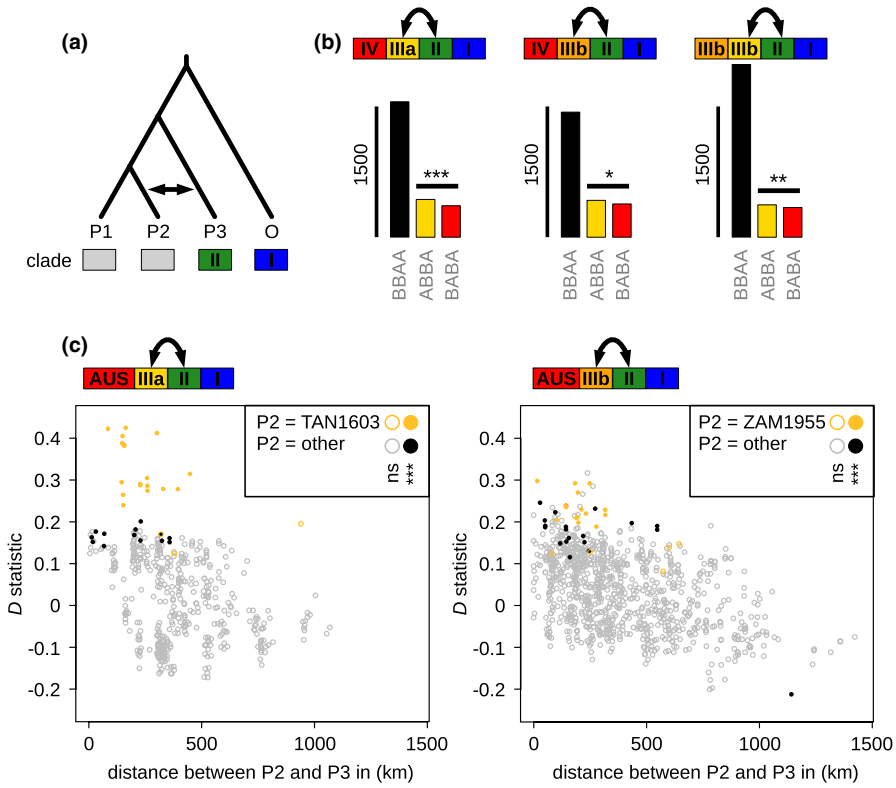

FIGURE 5 Evidence of introgression among C_4 and non-C_4 populations from the Central Zambezian region. (a) *D* statistics were used to test for gene flow between the non-C_4 clade II and one of the geographically overlapping C_4 clades. Generic positions are indicated at tips of the phylogeny, together with the groups considered in the tests (see Figure S3 for complete phylogeny). The arrow shows the introgression corresponding to a significant test. (b) The number of sites with different patterns is shown for three comparisons among four clades, shown at the top; BBAA sites present a derived variant shared by positions P1 and P2, as expected based on the species trees; ABBA sites present a derived variant shared by P2 and P3; BABA sites present a derived variant shared by P1 and P3. Significance levels are shown with asterisks (*p < .05, **p < .005, ***p < .001), with the suggested introgression indicated with an arrow at the top. (c,d) For all tests where P1 is the Australian group from clade IV, P2 is one population from clade IIIa (c) or from clade IIIb (d), P3 is one population of clade II, and the outgroup (O) is represented by clade I, the *D*-statistic is plotted against the geographical distance between the populations in positions P2 and P3. Comparisons with population TAN1603 (c) or population ZAM1955 (d) in the P2 position are coloured differently. Significant values after correction for multiple tests are indicated with filled circles

ecotypes (Grabowski et al., 2014), argue against clear hybrid zones, and show that the main genomic groups generally retain their identity despite close proximity and episodes of gene exchange.

3.5 | Genetic differentiation is widespread across the genome

The differentiation between the overlapping diploid clades II (excluding the polyploid ZAM1958) and IIIa (non-C_4 and C_4, respectively) was evaluated with genome scans (F_{ST} and d_{xy}). Genomic divergence is generally increased in regions with low gene content and high levels of transposable elements (Figure S10), a pattern attributable to reduced recombination around putative centromeres (Cruickshank & Hahn, 2014; Roesti et al., 2013), which was previously observed in *A. semialata* (Olofsson et al., 2019). Outside of these regions, numerous windows of low divergence suggest rampant introgression, while peaks of high differentiation are scattered across all chromosomes (Figure S10). Importantly, these peaks are not associated with genes known to be involved in C_4 photosynthesis. In particular, the three genes that play a role in the C_4 cycle and were shown to be differentially expressed among photosynthetic types of *A. semialata* (Dunning, Moreno-Villena, et al., 2019) showed patterns of diversity and differentiation similar to the rest of the genome, even after excluding regions with low gene density (Figure 7). These patterns indicate that divergence is driven by background selection or multiple regions scattered around the genome that do not encompass the genes previously identified to be involved in the photosynthetic differentiation of this species.

4 | DISCUSSION

4.1 | Photosynthetic types can share the same habitat

Based on phylogenomics studies conducted across the species' range, the different photosynthetic types of *Alloteropsis semialata* diverged during periods of geographical isolation, and thus the present-day

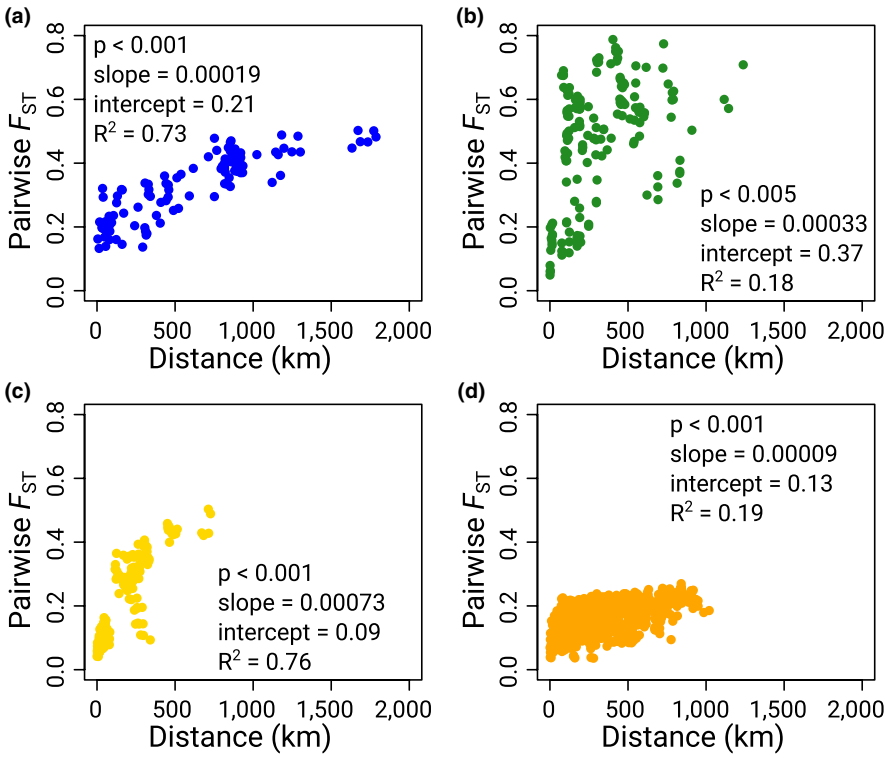

FIGURE 6 Patterns of isolation by distance within each group. For each of the four genetic groups within *Alloteropsis semialata*, pairwise genetic differentiation (F_{ST}) is plotted against geographical distance. (a) Non-C_4 clade I, (b) non-C_4 clade II, (c) C_4 clade IIIa and (d) C_4 clade IIIb. In each case, the *p*-value, obtained with a Mantel test and corrected for 10 tests (including intergroup analyses; Figure S8), is indicated. For significant relationships, slopes, intercepts and R^2 were estimated with linear regressions. Points are coloured as in Figure 1

overlapping ranges represent a secondary contact that occurred more than 1 million years ago (Bianconi et al., 2020; Olofsson et al., 2016). Across the entire species' range, the ecological niches of C_4 and non-C_4 *A. semialata* have been shown to overlap (Lundgren et al., 2015) and this is confirmed here by the co-occurrence of C_4 and non-C_4 individuals in five populations spread across Zambia (Figure 1; Figure S1), which adds to previously reported mixed populations in South Africa (Frean et al., 1980; Liebenberg & Fossey, 2001). These cases of co-occurrence show that the photosynthetic types are selectively equivalent, at least in some environments.

C_4 photosynthesis is advantageous mainly in warm, open habitats (Ehleringer et al., 1997; Ehleringer & Pearcy, 1983), and some Central Zambezian miombo woodlands might lie at the point where the C_4 and non-C_4 types are energetically equivalent. Alternatively, biotic factors might lead to niche differentiation within the same habitat, as reported in several animal systems (Comeault et al., 2015; Lindtke et al., 2017; Losey et al., 1997; Olendorf et al., 2006), thereby allowing the two types to exist in various environments. First, the two types might exploit different soil layers, with variation in rooting systems, which is affected by photosynthetic type (Wade et al., 2020), potentially allowing ecological partitioning of nutrient and water uptake (Fargione & Tilman, 2005; McKane et al., 2002; Phoenix et al., 2020). Second, mycorrhizal fungi might drive niche divergence in mixed populations, as previously reported in other systems (Osborne et al., 2018). Third, C_4 plants are generally less palatable and therefore preferentially avoided by herbivores (Boutton et al., 1978; Caswell et al., 1973). Such an advantage might counterbalance the energetic costs of the C_4 biochemical pathway, and allow C_4 populations to colonize habitats where non-C_4 individuals would otherwise be favoured. The coexistence of non-C_4 and C_4 *A. semialata* in mixed populations, nevertheless, demonstrates that they can grow in the same habitats. Our data therefore refute a complete habitat displacement of one of the photosynthetic types following secondary contact.

4.2 | Low dispersal probably prevents the homogenization of photosynthetic types among diploids

Previous genomic analyses found evidence of admixture between C_4 and non-C_4 plants in the Central Zambezian region (Bianconi et al., 2020; Olofsson et al., 2016), and the recurrent mixing of the genetic groups corresponding to the two photosynthetic types is confirmed here with our denser sampling. Strong admixture was detected in particular in polyploid individuals (Figure 4; Figure S7), mirroring other study systems (Fay et al., 2019; Grabowski et al., 2014; Novikova et al., 2020; Parisod et al., 2010; Schmickl & Koch, 2011). In addition, we obtained statistical evidence of admixture between diploid groups of C_4 and non-C_4 individuals (Figure 5), demonstrating that the distinct lineages are not completely incompatible. Despite these recurrent exchanges, the identity of each genomic group is maintained after more than 1 million years in the same region (Figures 2 and 4).

Despite overlapping ranges (Figure 1), the diploid non-C_4 and C_4 groups were never found in mixed populations (Figure 3). We hypothesize that the presence of diploids from the other photosynthetic type might interfere with reproduction. On the one hand, frequent pollination by the other type would decrease seed production if it results in unviable embryos or interferes with pollination

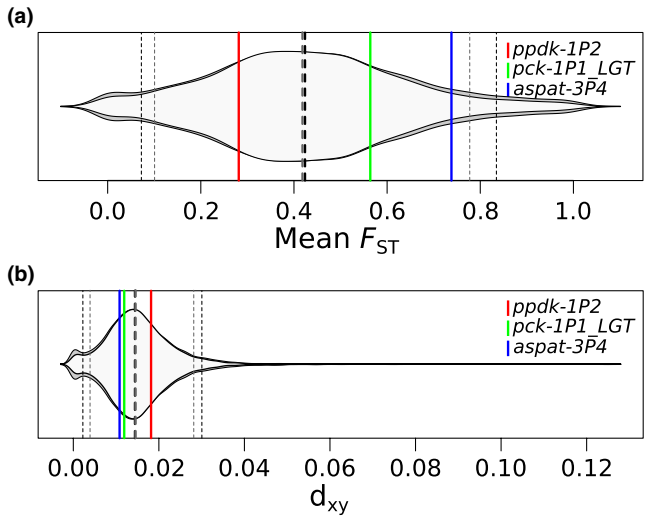

FIGURE 7 Distribution of genetic differentiation across the genome of *Alloteropsis semialata*. The genetic differentiation between the two geographically overlapping diploid genomic groups II (non-C_4) and IIIa (C_4) is shown using (a) F_{ST} and (b) d_{xy}. The beanplots show the distribution of values across nonoverlapping windows spanning the whole genome. Dark grey distributions represent all windows and light grey distributions represent high gene density (>50 genes/Mb) windows. Dashed vertical bars show the median (bold) and 5% and 95% percentiles (thin bars) with black representing all windows and grey all high gene density windows, and colour bars indicate the values for windows containing three genes previously shown to be upregulated for the C_4 cycle of *A. semialata* (see Dunning, Moreno-Villena, et al., 2019)

by the same type (e.g., Nishida et al., 2014). On the other hand, successful cross-fertilization would decrease fitness if the combination of genetic components disrupts the C_4 photosynthetic machinery through epistasis, as previously observed in crosses between photosynthetic types (in *Atriplex* and *Flaveria*; Brown & Bouton, 1993; Oakly et al., 2014). Our genome scans show that the divergence between C_4 and non-C_4 diploids is driven by a large number of regions scattered across chromosomes, which do not include the few genes with a known C_4 function that are differentially expressed between photosynthetic types of *A. semialata* (Figure 7; Dunning, Moreno-Villena, et al., 2019), but loci with a previously unknown function in the development or function of the C_4 trait might be responsible for hybrid depression. Alternatively, the hybrids might be at a disadvantage because of incompatibility of other dimensions of their phenotype, such as those involved in nutrient and water capture or herbivore and pathogen resistance.

Independent of the mechanism, both pre- and post-zygotic negative interactions would decrease the fitness of the photosynthetic type that is at low frequency in a mixed population and therefore has a high probability to reproduce with the alternative photosynthetic type (Lewis, 1961; Ray et al., 1979; Toll & Willis, 2018). In the absence of ecological differentiation, the photosynthetic type that first colonized and became abundant in a given area would therefore retain an advantage due to its higher frequency. While some introgression might still occur, newly arrived individuals would gradually disappear from each locality (Nishida et al., 2020; Ray et al., 1979; Templeton, 1981; Whitton et al., 2017), leading to the spatial sorting of photosynthetic types (Nishida et al., 2020). In the case of *A. semialata*, the low connectivity between diploid populations from the Central Zambezian region (i.e., strong isolation by distance; Figure 6) probably makes complete homogenization of photosynthetic types very slow and potentially impossible. We conclude that the low dispersal of diploids of *A. semialata* is key to the maintenance of distinct photosynthetic types within this region.

4.3 | Polyploidization enables co-occurrence of different photosynthetic types

Polyploids have arisen repeatedly in different parts of the range of *A. semialata* (Bianconi et al., 2020). In South Africa, C_4 polyploids co-occur with non-C_4 diploids (Frean et al., 1980; Liebenberg & Fossey, 2001) and we report here several populations spread across Zambia where non-C_4 diploids grow mixed with C_4 hexaploids (Figures 1 and 3). Some of the polyploid populations reported here present clear signs of admixture among photosynthetic types (e.g., ZAM1958; Figure 4). Polyploids might also facilitate gene flow among the two photosynthetic types, and tetraploidy in *Arabidopsis lyrata* has been shown to restore compatibility with diploids from the otherwise incompatible species *Arabidopsis arenosa* (Lafon-Placette et al., 2017). Future studies should test whether such a process occurs between diploids and hexaploids of *A. semialata*, but *D*-statistics indicate that gene flow occurs predominantly among the non-C_4 and C_4 clades that contain diploids (Figure 5). Based on analyses of organelle genomes, the polyploids from the C_4 clade IIIb emerged long after the split of the C_4 and non-C_4 lineages (Bianconi et al., 2020), and the initial divergence of photosynthetic types therefore occurred in a diploid context. However, our results indicate that ploidy differences might be required to allow the mixing of non-C_4 and C_4 individuals in the same population.

In all five populations from the Central Zambezian region with both C_4 and non-C_4 individuals, the C_4 plants belong to the nuclear group associated with hexaploids (Figures 1 and 3c). Similarly, in South Africa where C_4 and non-C_4 individuals occasionally form mixed populations, the C_4 individuals are polyploids while the non-C_4 individuals are diploid (Frean et al., 1980; Frean & Marks, 1988; Liebenberg & Fossey, 2001). Diploids might be disadvantaged in the presence of conspecific tetraploids because of asymmetrical gene flow and reproductive interference (Husband et al., 2002), but post-pollination barriers preventing gene flow between diploids and hexaploids have previously been reported in another system (Castro et al., 2011; Münzbergová et al., 2013). If cross-fertilization is reduced between diploid and hexaploid *A. semialata*, explaining the rarity of tetraploids (none detected out of 61 samples with genome sizes from the Central Zambezian region), the ploidy difference would allow long term coexistence of conspecifics with contrasted photosynthetic types. Similar conclusions were reached for ecotypes of the switchgrass, showing that ploidy generally helps maintain intraspecific functional diversity (Grabowski et al., 2014).

5 | CONCLUSIONS

In this study, we assessed the population genomics of the grass *Alloteropsis semialata* to determine how divergent photosynthetic types can have coexisted over more than 1 million years within a region of Africa. We showed that C_4 and non-C_4 individuals overlap geographically, sometimes occurring mixed within the same habitat, ruling out complete displacement upon secondary contact. We also found evidence of admixture and introgression between C_4 and non-C_4 populations. Groups with distinct photosynthetic types, however, behave as independent genomic entities, without classical hybrid zones. We suggest that selection against hybrids maintains the genetic groups, while low dispersal prevents their homogeneization. Polyploidy, which occurred repeatedly in the species, enables mixed populations, presumably because it strengthens prefertilization isolation. We conclude that low dispersal in the Central Zambezian region coupled with selection against hybrids and polyploidization explain the persistence of divergent photosynthetic types despite gene flow and a lack of habitat differentiation.

ACKNOWLEDGEMENTS

This work was funded by the European Research Council (grant ERC-2014-STG-638333) and the Natural Environment Research Council (grant NE/M00208X/1). M.E.B. is supported by the Royal Society (RGF\EA\181050) and P.A.C. is funded by a Royal Society University Research Fellowship (grant URF\R\180022). Sequencing was carried out by Sheffield Diagnostic Genetics Service, Sheffield Children's NHS Trust, Sheffield, the Centre for Genomic Research, Liverpool University, and Edinburgh Genomics, The University of Edinburgh. Edinburgh Genomics is partly supported through core grants from the NERC (R8/H10/56), MRC (MR/K001744/1) and BBSRC (BB/J004243/1).

AUTHOR CONTRIBUTIONS

J.K.O., P.N., C.P.O. and P.A.C. designed the study, J.K.O., F.N., E.V.C., M.E.B., L.T.D., M.R.L. and P.A.C. performed fieldwork, J.K.O. and G.S. produced the RADseq data, J.K.O. analysed the RADseq data, M.E.B., V.M. and M.R.L. produced and analysed the carbon isotope data, M.E.B., O.H., R.F.P. and I.L. produced and analysed the genome sizes, and J.K.O. and P.A.C. wrote the paper, with the help of all co-authors.

DATA AVAILABILITY STATEMENT

All raw reads are available in the short sequence archive under Accession nos. PRJNA560360 and PRJNA64872. Sample information is available in Table S1.

ORCID

Jill K. Olofsson https://orcid.org/0000-0002-9527-6573
Emma V. Curran https://orcid.org/0000-0002-1739-4603
Luke T. Dunning https://orcid.org/0000-0002-4776-9568
Graciela Sotelo https://orcid.org/0000-0002-0577-6655
Oriane Hidalgo https://orcid.org/0000-0002-1547-8627
Marjorie R. Lundgren https://orcid.org/0000-0002-2489-3646
Ilia J. Leitch https://orcid.org/0000-0002-3837-8186
Patrik Nosil https://orcid.org/0000-0002-8271-9005
Colin P. Osborne https://orcid.org/0000-0002-7423-3718
Pascal-Antoine Christin https://orcid.org/0000-0001-6292-8734

REFERENCES

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*, *55*, 238–258.

Atkinson, R. R. L., Mockford, E. J., Bennett, C., Christin, P.-A., Spriggs, E. L., Freckleton, R. P., Thompson, K., Rees, M., & Osborne, C. P. (2016). C_4 photosynthesis boosts growth by altering physiology, allocation and size. *Nature Plants*, *18*, 16038.

Barton, N. H. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*, 551–568.

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *56*, 357–376.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113–148.

Bender, M. M. (1971). Variations in the $^{13}C/^{12}C$ ratios of plants in relation to the pathway of photosynthetic carbon dioxide fixation. *Phytochemistry*, *10*, 1230–1244.

Bianconi, M. E., Dunning, L. T., Curran, E. V., Hidalgo, O., Powell, R. F., Mian, S., Christin, P. A. (2020). Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1938), 20201960. https://doi.org/10.1098/rspb.2020.1960

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.

Bolnock, D. I., & Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, *38*, 459–487.

Boutton, T. W., Cameron, G. N., & Smith, B. N. (1978). Insect herbivory on C_3 and C_4 grasses. *Oecologia*, *36*, 21–32.

Brown, H. R., & Bouton, J. H. (1993). Physiology and genetics of interspecific hybrids between photosynthetic types. *Annual Review of Plant Biology*, *44*, 435–456.

Brown, W. L., & Wilson, E. O. (1956). Character displacement. *Systematic Zoology*, *5*, 49–64.

Butlin, R. K., Galimdo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B*, *363*, 2997–3007.

Castro, S., Münzbergová, Z., Raabová, J., & Loureiro, J. (2011). Breeding barriers at a diploid-hexaploid contact zone in *Aster amellus*. *Evolutionary Ecology*, *25*, 795–814.

Caswell, H., Reed, F., Stephenson, S. N., & Werner, P. A. (1973). Photosynthetic pathways and selective herbivory: a hypothesis. *The American Naturalist*, *107*, 465–480.

Catchen, J., Hohenlohe, P. A., Basham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140.

Christin, P. A., Osborne, C. P., Sage, R. F., Arakaki, M., & Edwards, E. J. (2011). C_4 eudicots are not younger than C_4 monocots. *Journal of Experimental Botany*, *62*, 3171–3181.

Comeault, A. A., Flaxman, S. M., Riesch, R., Curran, E., Soria-Carrasco, V., Gompert, Z., Farkas, T. E., Muschick, M., Parchman, T. L., Schwander, T., Slate, J., & Nosil, P. (2015). Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. *Current Biology*, *25*, 1975–1981.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer & Associates.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

Dong, F., Hung, C. H., & Yang, X. J. (2020). Secondary contact after allopatric divergence explains avian speciation and high species diversity in the Himalayan-Hengduan Mountains. *Molecular Phylogenetics and Evolution*, *143*, 106671.

Dunning, L. T., Moreno-Villena, J. J., Lundgren, M. R., Dionora, J., Salazar, P., Adams, C., Nyirenda, F., Olofsson, J. K., Mapaura, A., Grundy, I. M., Kayombo, C. J., Dunning, L. A., Kentatchime, F., Ariyarathne, M., Yakandawala, D., Besnard, G., Quick, W. P., Bräutigam, A., Osborne, C. P., & Christin, P.-A. (2019). Key changes in gene expression identified for different stages of C_4 evolution in *Alloteropsis semialata*. *Journal of Experimental Botany*, *70*, 3255–3268.

Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., Dionora, J., Quick, W. P., Park, M., Bennetzen, J. L., Besnard, G., Nosil, P., Osborne, C. P., & Christin, P.-A. (2019). Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 4416–4425.

Ehleringer, J. R., Cerling, T. E., & Helliker, B. R. (1997). C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia*, *112*, 285–299.

Ehleringer, J. R., & Monson, R. K. (1993). Evolutionary and ecological aspects of photosynthetic pathway variation. *Annual Review of Ecology and Systematics*, *24*, 411–439.

Ehleringer, J., & Pearcy, R. W. (1983). Variation in quantum yield for CO_2 uptake among C_3 and C_4 plants. *Plant Physiology*, *73*, 555–559.

Ellis, R. P. (1974). The significance of the occurrence of both Kranz and non-Kranz leaf anatomy in the grass species *Alloteropsis semialata*. *South African Journal of Science*, *70*, 169–173.

Ellis, R. P. (1981). *Relevance of comparative leaf anatomy in taxonomic and functional research on the South African Poaceae (Unpublished DSc thesis)*. University of Pretoria, South Africa.

Endler, J. A. (1977). *Geographic variation, speciation, and clines (MPB-10), Volume 10*. Princeton University Press.

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., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Fargione, J., & Tilman, D. (2005). Niche differences in phenology and rooting depth promote coexistence with a dominant C_4 bunchgrass. *Oecologia*, *143*, 598–606.

Fay, J. C., Liu, P., Ong, G. T., Dunham, M. J., Cromie, G. A., Jeffery, E. W., & Dudley, A. M. A. (2019). Polyploid admixed origin of beer yeasts derived from European and Asian wine populations. *PLoS Biology*, *17*, e3000147.

Frean, M., Barrett, D., & (1980). Variability in leaf surface features and water efficiency utilisation in C_3 and C_4 forms of *Alloteropsis semialata* (R. Br.) hitchc. *Proceedings of the Annual Congresses of the Grassland Society of Southern Africa*, *15*, 99–103.

Frean, M. L., & Marks, E. (1988). Chromosome numbers of C_3 and C_4 variants within the species *Alloteropsis semialata* (R.Br.) Hitchc. (Poaceae). *Botanical Journal of the Linnean Society*, *97*, 255–259.

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*, *62*, 2789–2806.

Grabowski, P. P., Morris, G. P., Casler, M. D., & Borevitz, J. O. (2014). Population genomic variation reveals roles of history, adaptaion and ploidy in switchgrass. *Molecular Ecology*, *23*, 4059–4073.

Hatch, M. D. (1987). C_4 photosynthesis: A unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica Et Biophysica Acta – Bioenergetics*, *895*, 81–106.

Hunter, J. F. (2007). Mathplotlin: A 2D graphical environment. *Computing in Science & Engineering*, *9*, 90–95.

Husband, B. C., Schemske, D. W., Burton, T. L., & Goodwillie, C. (2002). Pollen competition as a unilateral reproductive barrier between sympatric diploid and tetraploid *Chamerion angustifolium*. *Proceedings of the Royal Society B*, *269*, 2565–2571.

Ingles, L. G., & Biglione, N. J. (1952). The contiguity of the ranges of two subspecies of pocket gophers. *Evolution*, *6*, 204–207.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071.

Jordan, D. S. (1905). The origin of species through isolation. *Science*, *22*, 545–562.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*, 1179–1191.

Lafon-Placette, C., Johannessen, I. M., Hornslien, K. S., Ali, M. F., Bjerkan, K. N., Bramsiepe, J., Glöckle, B. M., Rebernig, C. A., Brysting, A. K., Grini, P. E., & Köhler, C. (2017). Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proceedings of the National Academy of Sciences USA*, *114*(6), E1027–E1035.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie2. *Nature Methods*, *9*, 357–359.

Lewis, H. (1961). Experimental sympatric populations of *Clarkia*. *The American Naturalist*, *95*, 155–168.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*, 2987–2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

Liebenberg, E. J. L., & Fossey, A. (2001). Comparative cytogenetic investigation of the two subspecies of the grass *Alloteropsis semialata* (Poaceae). *Botanical Journal of the Linnean Society*, *137*, 243–248.

Lindtke, D., Lucek, K., Soria-Carrasco, V., Villoutreix, R., Farkas, T. E., Riesch, R., Dennis, S. R., Gompert, Z., & Nosil, P. (2017). Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology*, *2*, 6189–6205.

Losey, J. E., Harmon, J., Ballantyne, F., & Brown, C. (1997). A polymorphism maintained by opposite patterns of parasitism and predation. *Nature*, *388*, 269–272.

Lundgren, M. R., Besnard, G., Ripley, B. S., Lehmann, C. E. R., Chatelet, D. S., Kynast, R. G., Namaganda, M., Vorontsova, M. S., Hall, R. C., Elia, J., Osborne, C. P., & Christin, P.-A. (2015). Photosynthetic innovation broadens the niche within a single species. *Ecology Letters*, *18*, 1021–1029.

Lundgren, M. R., Christin, P. A., Esconar, E. G., Ripley, B. S., Besnard, G., Long, C. M., & Osborne, C. P. (2016). Evolutionary implications of C_3–C_4 intermediates in the grass *Alloteropsis semialata*. *Plant, Cell & Environment*, *39*, 1874–1885.

Malinsky, M., Marschiner, M., & Svardal, H. (2020). Dsuite – fast *D*-statsistics and related admixture evidence from VCF files. *Molecular Ecology Rescources*, *21*, 584–595.

Mallet, J. (1986). Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning colour clines. *Heredity*, *56*, 191–202.

Mallet, J., Barton, N., Gerardo, L. M., Jose, S. C., Manuel, M. M., & Eeley, H. (1990). Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics*, *124*, 921–936.

Maquia, I., Catarino, S., Pena, A. R., Brito, D. R. A., Ribeiro, N. S., Romeiras, M. M., & Ribeiro-Barros, A. I. (2019). Diversification of African tree legumes in miombo-mopane woodlands. *Plants*, *8*, 182.

Mayr, E. (1947). Ecological factors in speciation. *Evolution*, *1*, 263–288.

McKane, R. B., Johnson, L. C., Shaver, G. R., Nadelhoffer, K. J., Rastetter, E. B., Fry, B., Giblin, A. E., Kielland, K., Kwiatkowski, B. L., Laundre, J. A., & Murray, G. (2002). Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. *Nature*, *415*, 68–71.

Meirmans, P. G., Liu, S., & van Tienderen, P. H. (2018). The analysis of polyploid genetic data. *Journal of Heredity*, *2018*, 283–296.

Moyle, L. C., Olson, M. S., & Tiffin, P. (2004). Patterns of reproductive isolation in three angiosperm genera. *Evolution*, *58*, 1195–1208.

Münzbergová, Z., Šurinová, M., & Castro, S. (2013). Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales. *Heredity*, *110*, 123–130.

Nishida, S., Kanaoka, M. M., Hashimoto, K., Takakura, K. I., & Nishida, T. (2014). Pollen–pistil interactions in reproductive interference: Comparisons of heterospecific pollen tube growth from alien species between two native *Taraxacum* species. *Functional Ecology*, *28*, 450–457.

Nishida, S., Yakakura, K. I., Naiki, A., & Nishida, T. (2020). Habitat partitioning in native *Geranium* species through reproductive interference. *Annals of Botany*, *125*, 651–661.

Novikova, P. Y., Brennan, I. G., Booker, W., Mahony, M., Doughty, P., Lemmon, A. R., Moriarty Lemmon, E., Roberts, J. D., Yant, L., Van de Peer, Y., Keogh, J. S., & Donnellan, S. C. (2020). Polyploidy breaks speciation barriers in Australian burrowing frogs Neobatrachus. *PLoS Genetics*, *16*, e1008769.

Nürnberger, B., Barton, N., MacCallum, C., Gilchrist, J., & Appleby, M. (1995). Natural selection on quantitative traits in the *Bombina* hybrid zone. *Evolution*, *49*, 1224–1238.

Oakley, J. C., Sultmanis, S., Stinson, C. R., Sage, T. L., & Sage, R. F. (2014). Comparative studies of C_3 and C_4 *Atriplex* hybrids in the genomics era: Physiological assessments. *Journal of Experimental Botany*, *65*, 3637–3647.

Olendorf, R., Rodd, F. H., Punzalan, D., Houde, A. E., Hurt, C., Reznick, D. N., & Hughes, K. A. (2006). Frequency-dependent survival in natural guppy populations. *Nature*, *441*, 633–636.

Olofsson, J. K., Bianconi, M., Besnard, G., Dunning, L. T., Lundgren, M. R., Holota, H., Vorontsova, M. S., Hidalgo, O., Leitch, I. J., Nosil, P., Osborne, C. P., & Christin, P.-A. (2016). Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Molecular Ecology*, *25*, 6107–6123.

Olofsson, J. K., Dunning, L. T., Lundgren, M. R., Barton, H. J., Thompson, J., Cuff, N., Ariyarathne, M., Yakandawala, D., Sotelo, G., Zeng, K., Osborne, C. P., Nosil, P., & Christin, P.-A. (2019). Population-specific selection on standing variation generated by lateral gene transfers in a grass. *Current Biology*, *29*, 3921–3927.

Osborne, O. G., De-Karyne, R., Bidartondo, M. I., Hutton, I., Baker, W. J., Turnbull, C. G. N., & Savolainen, V. (2018). Arbuscular mycorrhizal fungi promote coexistence and niche divergence of sympatric palm species on a remote oceanic island. *New Phytologist*, *217*, 1254–1266.

Parisod, C., Holderegger, R., & Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytologist*, *186*, 5–17.

Petersen, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoeksta, H. E. (2014). Double digest RADseq: An inexpensive method for the *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*, e37135.

Pfenning, D. W., & Pfenning, K. S. (2010). Character displaement and the origins of diversity. *The American Naturalist*, *176*, S26–S44.

Phoenix, G. K., Johnson, D. A., Muddimer, S. P., Leake, J. R., & Cameron, D. D. (2020). Niche differentiation and plasticity in soil phosphorous acquisition among co-occurring plants. *Nature Plants*, *6*, 349–354.

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Muller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M. G., & Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*, 1410–1414.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raabová, J., Fischer, M., & Münzbergová, Z. (2008). Niche differentiation between diploid and hexaploid *Aster amellus*. *Oecologia*, *158*, 463–472.

Ramsey, J., & Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, *29*, 467–501.

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*, 1450–1477.

Ray, M. L., Ray, X. A., Dickinson, D. B., & Grossman, M. (1979). A model for the genetic modification of wild plant species. *Journal of Heredity*, *70*, 309–316.

Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome – patterns and consequences. *Molecular Ecology*, *22*, 3014–3027.

Sage, R. F., Sage, T. L., & Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology*, *63*, 19–47.

Schluter, D. (2000). Ecological character displacement in adaptive radiation. *The American Naturalist*, *156*, S4–S16.

Schluter, D., & McPhail, J. D. (1992). Ecological character displacement and speciation in sticklebacks. *The American Naturalist*, *140*, 85–108.

Schmickl, R., & Koch, M. A. (2011). Arabidopsis hybrid speciation processes. *Proceedings of the National Academy of Sciences*, *108*(34), 14192–14197.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*, 693–702.

Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics*, *75*, 733–756.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313.

Stata, M., Sage, T. L., & Sage, R. F. (2019). Mind the gap: The evolutionary engagement of the C_4 metabolic cycle in support of net carbon assimilation. *Current Opinion in Plant Biology*, *49*, 27–34.

Stift, M., Kolář, F., & Meirmans, P. G. (2019). STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity*, *123*, 429–441.

Templeton, A. R. (1981). Mechanisms of speciation – A population genetic approach. *Annual Review in Ecology and Systematics*, *12*, 23–48.

Toews, D. P., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, *26*, 2313–2318.

Toll, K., & Willis, J. H. (2018). Hybrid inviability and differential submergence tolerance drive habitat segregation between two congeneric monkeyflowers. *Ecology*, *99*, 2776–2786.

Wade, R. N., Seed, P., McLaren, E., Wood, E., Christin, P.-A., Thompson, K., Rees, M., & Osborne, C. P. (2020). The morphogenesis of fast growth in plants. *New Phytology*, *228*, 1306–1315.

Whitton, J., Sears, C. J., & Maddison, W. P. (2017). Co-occurrence of related asexual, but not sexual, lineages suggests that reproductive interference limits coexistence. *Proceedings of the Royal Society B*, *284*, 20171579.

Williams, B. P., Johnston, I. G., Covshoff, S., & Hibberd, J. M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *Elife*, *2*, e00961.

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Review Genetics*, *18*, 87–100.

SUPPORTING INFORMATION

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

How to cite this article: Olofsson JK, Curran EV, Nyirenda F, et al. Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. *Mol Ecol*. 2021;00:1–15. https://doi.org/10.1111/mec.15871