Article:
Sotelo, G., Gamboa, S., Dunning, L.T. et al. (2024) C4 photosynthesis provided an immediate demographic advantage to populations of the grass Alloteropsis semialata. New Phytologist, 242 (2). pp. 774-785. ISSN: 0028-646X

https://doi.org/10.1111/nph.19606

New Phytologist

Research

C_4 photosynthesis provided an immediate demographic advantage to populations of the grass *Alloteropsis semialata*

Graciela Sotelo[1], Sara Gamboa[1,2], Luke T. Dunning[3], Pascal-Antoine Christin[3] and Sara Varela[1]

[1]Universidade de Vigo, Departamento de Ecoloxía e Bioloxía Animal, 36310, Vigo, Spain; [2]Universidad Complutense de Madrid, 28040, Madrid, Spain; [3]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, S10 2TN, Sheffield, UK

Author for correspondence:
Graciela Sotelo
Email: graciela.sotelo@uvigo.gal, *gsotelo.fdz@gmail.com*

Received: *29 August 2023*
Accepted: *30 January 2024*

New Phytologist (2024)
doi: 10.1111/nph.19606

Key words: *Alloteropsis semialata*, C_4 photosynthesis, climatic niche, demographic history, key trait, paleoclimate reconstruction.

Summary

- C_4 photosynthesis is a key innovation in land plant evolution, but its immediate effects on population demography are unclear. We explore the early impact of the C_4 trait on the trajectories of C_4 and non-C_4 populations of the grass *Alloteropsis semialata*.
- We combine niche models projected into paleoclimate layers for the last 5 million years with demographic models based on genomic data.
- The initial split between C_4 and non-C_4 populations was followed by a larger expansion of the ancestral C_4 population, and further diversification led to the unparalleled expansion of descendant C_4 populations. Overall, C_4 populations spread over three continents and achieved the highest population growth, in agreement with a broader climatic niche that rendered a large potential range over time. The C_4 populations that remained in the region of origin, however, experienced lower population growth, rather consistent with local geographic constraints. Moreover, the posterior transfer of some C_4-related characters to non-C_4 counterparts might have facilitated the recent expansion of non-C_4 populations in the region of origin.
- Altogether, our findings support that C_4 photosynthesis provided an immediate demographic advantage to *A. semialata* populations, but its effect might be masked by geographic contingencies.

Introduction

Understanding how life copes with recurrent environmental shifts and how certain traits can confer key advantages for population success and/or spread into new areas are primary questions in evolutionary biology. In the long-term, climatic changes have been linked to macroevolutionary patterns, including taxonomic and biogeographic turnovers (e.g. Hewitt, 2004; Svenning *et al.*, 2015). At finer temporal and spatial scales, variation in temperature and precipitation regimes can alter the demographic dynamics of local populations and modify the adaptive value of specific traits (e.g. Parmesan & Yohe, 2003; Urban *et al.*, 2016). Nonetheless, addressing the immediate impact of the rise of a key trait on the fate of the populations is still a challenging topic (e.g. Miller *et al.*, 2023). First, major key traits evolved deep in the past, making it difficult to disentangle the effect of the emergence of the trait from that of further modifications accumulated through time. Second, fundamental traits rarely remain polymorphic at the intraspecific level, hindering the application of population approaches to capture the trait transition.

A remarkable exception is the evolution of C_4 photosynthesis in the grass *Alloteropsis semialata*. C_4 photosynthesis is a change in primary metabolism that represents a fundamental innovation not only in the evolution of land plants but of life on Earth (Sage, 2004), and *A. semialata* is the only known species where the trait is not fixed, encompassing C_3, C_4, and intermediate (type II C_3–C_4 intermediate *sensu* Edwards & Ku, 1987; hereafter C_3+C_4 *sensu* Dunning *et al.*, 2017) populations (Pereira *et al.*, 2023). C_4 plants have enhanced light-, water-, and nitrogen-use efficiency, outperforming the C_3 in conditions that reduce CO_2 availability in the leaf (Ehleringer *et al.*, 1997; Sage, 2004; Zhou *et al.*, 2018). The first C_4 plants appeared *c.* 30 million years ago (Ma) matching a global drop in atmospheric CO_2 (Christin *et al.*, 2008; Vicentini *et al.*, 2008), but C_4-dominated ecosystems expanded much later, *c.* 8–3 Ma, becoming prevalent in tropical and subtropical areas (Edwards *et al.*, 2010). C_4 photosynthesis convergently evolved many times in different lineages (Sinha & Kellogg, 1996; Sage *et al.*, 2011), and each lineage followed a particular trajectory depending on its specific evolutionary background and environmental context (Kadereit *et al.*, 2012; Spriggs *et al.*, 2014; Dunning *et al.*, 2017; Heyduk *et al.*, 2019; Bianconi *et al.*, 2020a,b). In the long-term, C_4 photosynthesis generally worked as a niche expander since most C_4 plants can tolerate a broader set of climatic conditions than non-C_4 plants (Aagesen *et al.*, 2016; Watcharamongkol *et al.*, 2018). However, whether/how the acquisition of C_4 photosynthesis was beneficial to populations in the first place remains unclear, and *A. semialata* offers an ideal system for tackling this question.

The photosynthetic types of *A. semialata* correspond to distinct genetic lineages that probably evolved during the Plio-Pleistocene (Lundgren *et al.*, 2015; Bianconi *et al.*, 2020a) from a common ancestor that used a weak C_4 cycle (Dunning *et al.*, 2017; Fig. 1). The diversification of the system has been studied attending to anatomy, physiology, ecology, and genomics (reviewed in Pereira *et al.*, 2023). Nonetheless, the demographic history of these lineages has not been explored yet. Understanding how their effective population sizes fluctuated through time and to what extent these fluctuations correlate with the C_4–non-C_4 split and with past climatic changes could provide direct insights into the role of the photosynthetic innovation in the relative success of the *A. semialata* lineages. Accessing that information is the main goal of this study.

The initial divergence between *A. semialata* C_4 and non-C_4 types most likely happened in the Central Zambezian miombo woodlands of Africa, where the species originated *c.* 3 Ma (Lundgren *et al.*, 2015; Bianconi *et al.*, 2020a). The C_3 lineage (clade I) later migrated to Southern Africa and a single C_4 lineage (clade IV) spread across Africa, Madagascar, Southeast Asia, and Oceania. The Central Zambezian region remained occupied by another C_4 lineage (clade III) and by $C_3 + C_4$ populations (clade II). The lineages evolved largely in isolation, but repeated episodes of genetic exchange might have contributed to the expansion of the different photosynthetic types (Lundgren *et al.*, 2015; Olofsson *et al.*, 2016, 2021; Bianconi *et al.*, 2020a). Currently, C_4 plants overlap with C_3 plants in Southern Africa and with $C_3 + C_4$ ones in the Central Zambezian region, but when they appear mixed (growing close to each other), the C_4 are polyploids and the non-C_4 are diploids, and this ploidy difference probably prevents gene flow between them (Olofsson *et al.*, 2021).

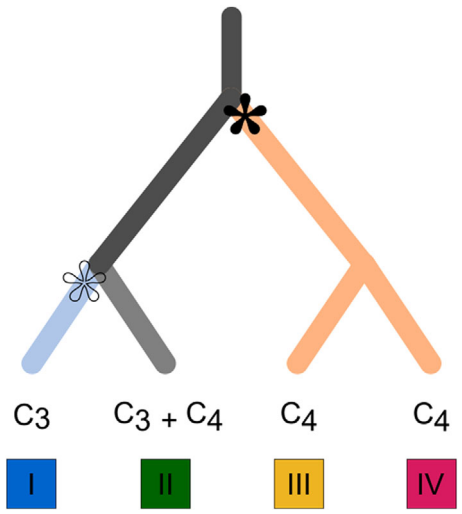

Fig. 1 Hypothesis of photosynthetic diversification in *Alloteropsis semialata* (following Dunning *et al.*, 2017) along the proposed nuclear phylogeny of the species (Bianconi *et al.*, 2020a). The most recent common ancestor likely used a weak C_4 cycle (similar to a $C_3 + C_4$ state, in dark grey), although different from that of the current intermediate lineage ($C_3 + C_4$, in light grey). The full C_4 cycle (in orange) was acquired after the initial split between non-C_4 and C_4 lineages (filled asterisk). A reversal to the ancestral C_3 cycle (in blue) occurred after the split between non-C_4 lineages (open asterisk). The tips of the tree represent the four main nuclear lineages currently recognized (I, II, III, and IV), with corresponding photosynthetic types indicated above.

Taking advantage of the uniqueness of this system, here we integrate demographic modelling with ecological niche modelling projected into paleoclimatic layers for the last 5 million years (Myr) aiming to test the following hypotheses: (1) major changes in effective population size occurred immediately after the divergence between C_4 and non-C_4 lineages; (2) C_4 lineages experienced larger increases in effective population size than non-C_4 lineages; (3) these larger demographic increases were due to a wider C_4 climatic niche resulting in a broader potential geographic range. All in all, this would tell us how the effective population size of the *A. semialata* lineages fluctuated through time and whether the C_4 trait acquisition was associated with immediate demographic advantages under changing climatic conditions.

Materials and Methods

Population sampling and data collection

Based on previous studies (Bianconi *et al.*, 2020a; Olofsson *et al.*, 2021), we used four well-defined groups of *A. semialata* (R. Br.) Hitchc. samples that represent the three photosynthetic types and the four main nuclear lineages described for the species. These samples are a subset of those studied by Olofsson *et al.* (2019, 2021). Specifically, we included C_3 samples from Zimbabwe and South Africa assigned to clade I, $C_3 + C_4$ samples from Zambia and Tanzania assigned to clade II, C_4 samples from Zambia and Tanzania assigned to clade IIIa, and C_4 samples from Australia that belong to clade IV. We excluded polyploids, such as C_4 samples from clade IIIb (which together with IIIa form clade III), given that polyploidy has been shown to affect population success in different ways (e.g. Monnahan *et al.*, 2019; Padilla-García *et al.*, 2023) and might confound our results. Briefly, these samples were collected during field trips between 2012 and 2019. Populations were spotted while driving or through stop-and-walk searches. GPS coordinates were recorded at each sampling site, and individuals were genotyped using a double-digested restriction-associated DNA sequencing (ddRADseq) approach, as described previously (Olofsson *et al.*, 2019, 2021). Both the genomic data and geographic coordinates used in this work were retrieved from the previous studies, accessible via the NCBI SRA projects PRJNA560360 and PRJNA649872. This dataset consisted of 475 individuals from 100 localities (Fig. 2; Supporting Information Table S1). For ecological niche modelling, we added 35 individuals that provide a more comprehensive coverage of the current geographic distribution of *A. semialata* (Fig. S1; Table S1). They were previously confirmed as diploids and assigned to photosynthetic type and nuclear lineage (Bianconi *et al.*, 2020a; Olofsson *et al.*, 2021).

Ecological niche modelling

Niche envelope and species distribution model We used mean annual temperature and total annual precipitation as predictive

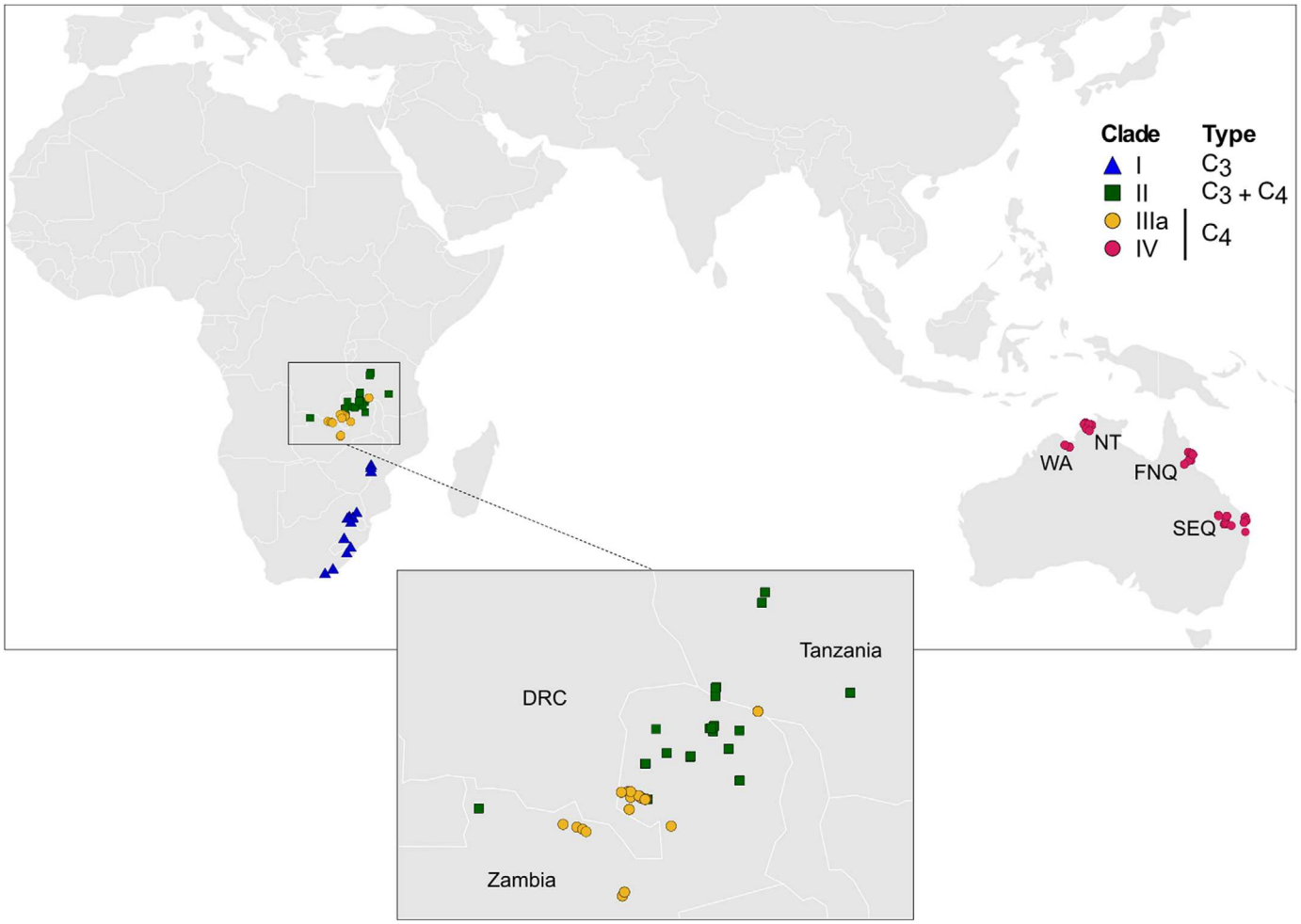

Fig. 2 Distribution of *Alloteropsis semialata* samples included in this study. Symbols correspond to photosynthetic types and colours, to clades. For the Australian samples, codes indicate geographic regions: WA, Western Australia; NT, Northern Territory; FNQ, Far North Queensland; and SEQ, South East Queensland. The Central Zambezian region, where clades II and IIIa overlap, is shown in more detail in the inset panel. DRC, Democratic Republic of Congo. See Supporting Information Fig. S1 for additional samples used for niche analyses.

variables for niche modelling. They have been identified as fundamental drivers of C_4 evolution (Edwards & Smith, 2010; Christin & Osborne, 2014), and the predictions on annual trends for past climatic scenarios are more robust to the underlying climatic model than seasonal or extreme trends (Varela *et al.*, 2015). We obtained preindustrial values for these variables from the high-resolution climate emulator PALEO-PGEM (Holden *et al.*, 2019), as illustrative of present-day conditions. We estimated the environmental niche of each clade using the classic 'climate-envelope model' Bioclim (Busby, 1991) implemented in the R package DISMO v.1.3-9 (Hijmans *et al.*, 2022). Bioclim employs presence-only data to define a multidimensional environmental space where species occur (Busby, 1991; Varela *et al.*, 2014). The resulting environmental space is represented as a box defined by the minimum and maximum values of all variables across the localities where species were sampled or reported. To mitigate the impact of outliers, we only considered those values within the 5th to 95th percentile range. The potential geographic range of each clade was then calculated as those areas where the values of temperature and precipitation fell within the estimated climatic requirements for the clade.

Niche overlap We quantified pairwise niche overlap using Schoener's *D* statistic (Schoener, 1968; Warren *et al.*, 2008) as implemented in the R package ENMTOOLS v.1.0.5 (Warren *et al.*, 2021). The *D* statistic ranges from 0 (no overlap) to 1 (full overlap). To assess the significance of the overlap, we applied equivalency and similarity tests under the null hypothesis that each pair of clades occupies the same environmental space. The tests were performed with 1000 permutations and a confidence level of 0.05. For the equivalency test, the null distribution was generated by pooling the occurrences (i.e. localities) of the corresponding pair of clades, randomizing these occurrences to get two new sets with the same number of observations as the originals, and calculating *D* for each permutation. The similarity tests were conducted in an analogous way but pooling the backgrounds of the clades instead of the occurrences, to account for the environmental conditions that are geographically available for the clades. The backgrounds were

built by adding a buffering circle of 200 km radius around each occurrence. For both tests, rejecting the null hypothesis indicated that niches overlap less than expected by chance.

Suitable areas through time We mapped the areas with suitable climatic conditions for each clade through time under the assumption of niche conservatism. We used PALEO-PGEM to extract monthly data of temperature (°C) and precipitation (mm) for the last 5 Myr with a spatial resolution of 0.5° and a temporal resolution of 1000 yr, and then, we used the R package LANDSCAPEMETRICS v.1.5.4 (Hesselbarth *et al.*, 2019) to sum the suitable areas at each 1000-yr layer under the equal-area projection 'Equal Earth' (Šavrič *et al.*, 2019). We calculated these areas (1) at the global scale, excluding the Americas where *A. semialata* has not been reported, (2) considering just Africa for every clade, and (3) considering just Australia for clade IV. Additionally, we estimated the areas that remained suitable for each clade over time, which could act as refugia. We assessed the extent of suitable areas through time and the number and/or connectivity of stable areas as potential predictors of the demographic success of the clades.

Demographic modelling

For demographic inference, we applied two composite likelihood methods based on the coalescent that use the site frequency spectrum (SFS) derived from genomic data: STAIRWAY PLOT v.2.1.1 (Liu & Fu, 2015, 2020), which estimates effective population size (N_e) trajectories for single populations without a predefined model, and FASTSIMCOAL2 v.2.709 (Excoffier *et al.*, 2013, 2021), which can evaluate more complex models specified by the user for one or more populations.

These methods assume that the genomic sites analysed (single nucleotide polymorphisms (SNPs)) are independent and evolve under neutrality. In this way, the demographic parameters can be translated into absolute units if a mutation rate and a generation time are provided and if the total length of the sequence analysed is known. Here, we used a mutation rate of 1×10^{-8} per site per generation based on available estimates for other grasses (maize and rice; Clark *et al.*, 2005; Jiao *et al.*, 2012; Yang *et al.*, 2015, 2017) and a tentative generation time of 5 yr based on glasshouse observations, while the sequence length was estimated as detailed below.

Another assumption of these methods is the absence of substructure in the data. The *A. semialata* clades behave as distinct genomic groups but not as panmictic units, since the populations within each clade (localities hereafter) show a strong pattern of isolation by distance (Olofsson *et al.*, 2021). This means that each clade better adjusts to a metapopulation, where localities exchange migrants depending on their geographic distance. Such structure can lead to spurious signatures of N_e change through time (Städler *et al.*, 2009; Mazet *et al.*, 2016; Maisano Delser *et al.*, 2019; Lesturgie *et al.*, 2022). Thus, we first assessed the N_e trajectories for each clade as a deme and for each locality as a smaller (near-) panmictic unit to verify that the results were consistent and similar at both scales. Then, we considered the N_e changes inferred for each clade to define models of their joint demographic history, to minimize potential biases due to unaccounted N_e changes in ancestral/descendant populations (Momigliano *et al.*, 2021).

Site frequency spectrum estimation We mapped clean ddRAD-seq reads for each individual to the reference genome of *A. semialata* (ASEM_AUS1_v1.0; GenBank accession QPGU01000000; Dunning *et al.*, 2019b) using BOWTIE2 v.2.4.4 (Langmead & Salzberg, 2012) with the default setting for paired-end reads. We retained properly paired reads with a unique mapping, and sorted and indexed them with SAMTOOLS v.1.10 (Li *et al.*, 2009).

To perform the analyses at the clade level, we derived folded SFSs from genotype calls. We generated a variant call format (VCF) file for each dataset (i.e. each single clade, pairs of clades, and the four clades together) as follows. We called sites with minimum mapping and quality scores of 20 using the *mpileup* and *call* functions from BCFTOOLS v.1.10.2 (Danecek *et al.*, 2021), and genotypes from biallelic SNPs aligning to the nuclear genome with VCFTOOLS v.0.1.16 (Danecek *et al.*, 2011). We treated a genotype as missing if its depth of coverage was below 7 in the corresponding individual (following recommendations as in Peterson *et al.*, 2012), and we discarded sites with > 70% missing data across individuals. We further discarded individual genotype calls with abnormally low or high coverage (< 0.5× or > 2× mean coverage for that individual), sites with heterozygosity excess across individuals, and individuals with > 70% missing data. We built a folded SFS from each VCF file with EASYSFS (https://github.com/isaacovercast/easySFS), projecting down the data to a smaller sample size to get rid of remaining missing values. We set the sample size per clade to 30 individuals (60 genomes), except when including the four clades where it was limited to 10 individuals (20 genomes) to not exceed the maximum SFS size that FASTSIMCOAL2 can handle.

We calculated the total sequence length as the sum of monomorphic and polymorphic (biallelic) sites in each SFS. To get the number of monomorphic sites in the data, we called all sites (not only variants) with BCFTOOLS and filtered them with VCFTOOLS by setting the minimum and maximum number of alleles to 1. We then adjusted the number of monomorphic sites in the SFS based on the proportion of monomorphic to polymorphic sites that passed the first filters on coverage (≥ 7) and missingness (< 70%), and the proportion of polymorphic sites in the original VCF file that were finally included in the SFS file.

To perform the analyses at the locality level, we estimated a folded SFS for each locality with at least four sampled individuals, from genotype likelihoods as implemented in ANGSD v.0.938 (Korneliussen *et al.*, 2014). The filtering strategy was analogous to that used for clades: We only considered sites mapping to the nuclear genome, with mapping and quality scores over 20, with no more than two alleles, and present in at least 70% of the samples. We evaluated the depth of coverage per site as the total sequencing depth across individuals per locality, setting minimum, and maximum values as half and twice the sum of mean coverage per individual. We also discarded sites with heterozygosity excess. Because the monomorphic sites were retained throughout the pipeline (no threshold was specified for '-SNP_pval' tag), the number of sites defining the total sequence

length was already included in the SFS file (without further adjustment).

One-population models To get an overview of N_e trajectories in each C_4 and non-C_4 clade/locality, we applied the multi-epoch model implemented in STAIRWAY PLOT 2, with default options: including singletons, using four numbers of random breakpoints, two thirds of the sites as training set and 200 replicate runs. The observed trajectories were contrasted with explicit models in FASTSIMCOAL2 (Methods S1, S2; Fig. S2a,b).

Four-population models To infer the joint demographic history of the four *A. semialata* clades, we defined four-population models in FASTSIMCOAL2. Based on the phylogenetic relationships inferred from the nuclear genomes (Bianconi *et al.*, 2020a), the models consisted of a first split of the most recent common ancestor of the four clades into a non-C_4 ancestral population and a C_4 ancestral population, which later split into clades I and II and clades IIIa and IV (respectively) at independent time points (Fig. S2c). Attending to results from one- and two-population models (Methods S1, S2), we included a sudden N_e change in each clade after the most recent splits. The models differed in the set of migration rates specified: The base model included migration between the non-C_4 and C_4 ancestral populations (IM1); next, we added migration between the Central Zambezian clades II and IIIa (IM2), then between the non-C_4 clades I and II (IM3), and last between the C_4 clades IIIa and IV (IM4). A fifth model included a sudden N_e change in non-C_4 and C_4 ancestral populations on top of the latter settings (IM4ac). All migration rates were specified as constant and allowed to be asymmetric. Each model was run 100 times, with 40 optimization cycles and 100 000 coalescent simulations per cycle. The models were compared with the Akaike information criterion, considering the run with the highest likelihood in each case.

To get confidence intervals for the parameter estimates under the best model, we used a nonparametric block-bootstrap approach. We divided the original dataset (i.e. VCF for the four clades together) into blocks of 1000 SNPs and generated 20 bootstrapped datasets by sampling with replacement 85 blocks to match the size of the original dataset (85 398 SNPs). For each bootstrap replicate, we obtained the SFS and run the demographic model as for the original dataset, and the estimates from the best run were used to calculate maximum and minimum values for each parameter.

Results

Ecological niche models

The environmental niche of the C_4 clade IV was found to be the broadest, nearly comprising the niches of all other clades except for C_3 clade I, while the niche of the C_4 clade IIIa appeared as the most restricted (Fig. S3a). All pairwise tests supported the nonequivalency of the niches, indicating that niches overlap less than expected by chance, although their backgrounds do not differ significantly (Table S2). The extensive niche of clade IV was reflected in the extent and distribution of predicted suitable areas for present-day conditions (Figs 3, S4): The predicted range largely resembles the putative distribution of *A. semialata* across Africa, Madagascar, Southeast Asia, and Oceania based on currently available records (Lundgren *et al.*, 2015, 2016). It also approaches the distribution of the C_4 type alone if polyploids were considered, as those reported in South Africa, which suggests that diploid and polyploid C_4 populations might share the same environmental niche. Further suitable areas were inferred along tropical and temperate regions of the Americas and around the Mediterranean basin in Southern Europe, although these regions likely remained inaccessible to *A. semialata* through its history. The predicted range of clade IV was mostly continuous within each continent, in contrast to the more fragmented ranges estimated for the other clades (Figs 3, S4).

The paleoclimate reconstructions revealed a cyclical variation in the extent of suitable areas for each clade over the last 5 Myr, which was in line with global changes in temperature during this period (fig. 2 in Holden *et al.*, 2019). Clade IV maintained the largest suitable areas both at the global scale and within Africa throughout the whole period. Within Australia, the suitable areas for clade IV were in the same order of magnitude as the areas of the other clades within Africa (Fig. S5; Table S3). In general, the climatic conditions became less favourable for *A. semialata* *c.* 2.5 Ma. Since then, despite the continuous fluctuations, the lower bound of suitable areas was higher than before for every clade. Within the last 2.5 Myr, maximum areas in Africa were estimated at *c.* 1.8, 1.5, 1.2, and 0.4 Ma for clades II, IV, IIIa, and I, respectively; and in Australia, *c.* 2.3 and 1.3 Ma for clade IV (Fig. S5; Table S3). Over time, no stable areas were identified for clade IIIa and only a few for clades I and II, while a network of suitable areas was identified for clade IV within each continent at every time step evaluated (Fig. S6; Table S4). The patterns observed across the paleoclimate reconstructions remained consistent when using the reduced dataset with clade IV samples restricted to Australia (data not shown).

Demographic models

One-population models The STAIRWAY PLOT analysis inferred an ancestral expansion for each clade (Figs 4a, S7), which started *c.* 300 thousand generations ago (kga) (*c.* 1.5 Ma) for clades II and IIIa and *c.* 150 kga (*c.* 0.8 Ma) for clades IV and I. Further steep increases in N_e started *c.* 50 kga (*c.* 0.3 Ma). Clade IV attained the largest N_e and experienced the largest increase with respect to the ancestral N_e (> 60-fold vs < 20-fold in the other three clades; Fig. 4a).

The analysis by locality detected the initial expansions but also posterior declines for multiple localities (Fig. 4b). These declines were supported by both the N_e trajectory and the shape of the normalized SFS (i.e. a reduction in singletons respect to the previous allele frequency category) in a few $C_3 + C_4$ localities from clade II and in most C_4 localities from clade IIIa and from clade IV along the easternmost region in Australia (South East Queensland; Figs S8, S9). Within clade IV, the timing of the demographic changes and the estimates of N_e towards the present

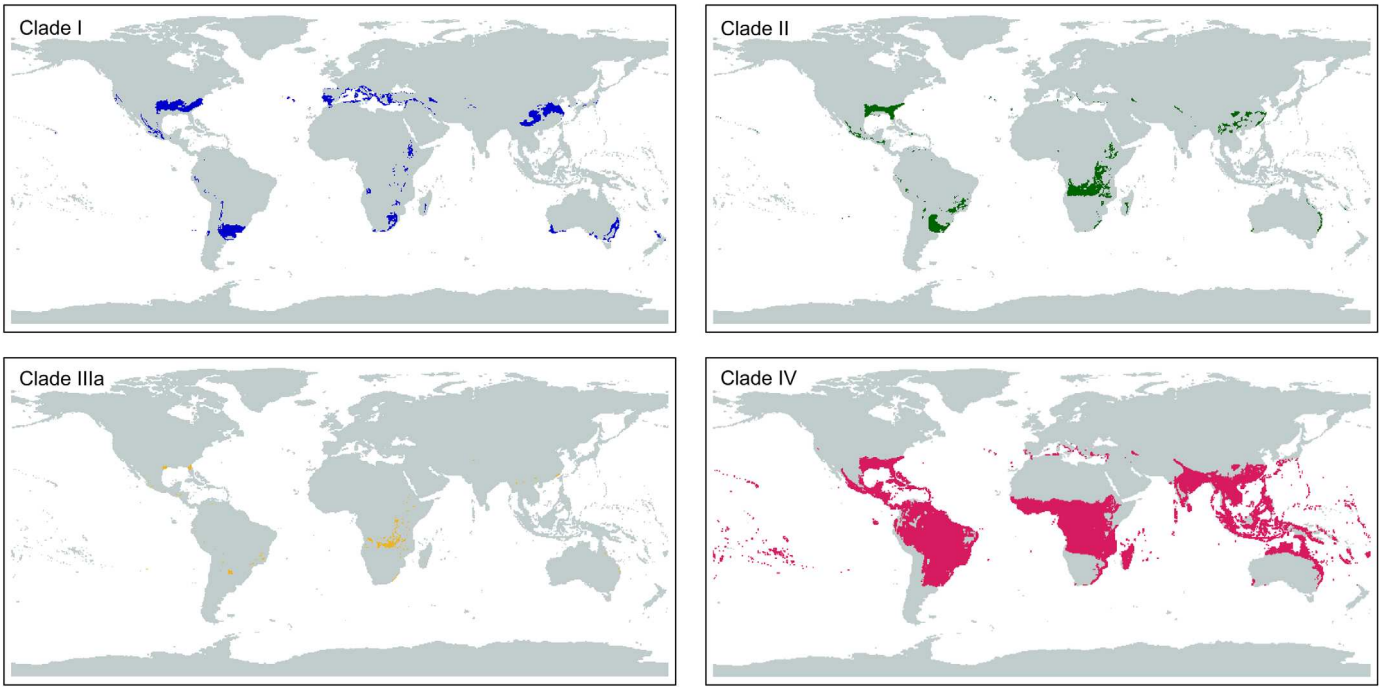

Fig. 3 Distribution of suitable areas estimated for each *Alloteropsis semialata* clade at the global scale for the present (preindustrial) time. Colours correspond to clades. See Supporting Information Fig. S4 for the distribution based on the dataset used for demographic analyses, with clade IV samples just from Australia.

showed a decreasing trend from northern to western and eastern localities, consistent with founder events following the species dispersal through Australia. Among all analysed samples, the largest recent N_e values corresponded to clade IV localities from the Northern Territory, in line with the clade analysis. Likewise, recent N_e values remained relatively large for almost all clade I localities, which did not show evidence of recent declines (Figs 4b, S8, S9).

Four localities from the Central Zambezian region in Africa exhibited distinct profiles compared with the other localities within the same clade: ZAM1716 and TAN2 from $C_3 + C_4$ clade II, and ZAM1505 and TAN1603 from C_4 clade IIIa (Fig. S8). ZAM1716 showed a further increase in N_e after the recent decline, and TAN2 displayed a longer period of constant N_e before the decline. ZAM1505 did not show signs of decline and TAN1603 presented a recent steep increase in N_e instead. Notably, TAN1603 currently extends over a large area (LTD, field observation) and was identified as particularly introgressed by Olofsson *et al.* (2021); thus, the large change in N_e could reflect an increase in both population size and gene flow from non-C_4 clades. Signatures of admixture were also detected in TAN2 (although its profile is quite the opposite) but not in ZAM1716 or ZAM1505 (Olofsson *et al.*, 2021), implying that other sources of heterogeneity cannot be ruled out (Notes S1).

In general, both STAIRWAY PLOT and FASTSIMCOAL2 models (Methods S1, S2; Fig. S10; Tables S5, S6) suggest that an expansion occurred in the ancestral populations of the clades, being compatible with the evolution of clades II and IIIa in the Central Zambezian region and the spread of clades I and IV out of this centre of origin, while some localities have likely experienced recent declines.

Four-population models The best supported scenario for the joint demographic history of the four *A. semialata* clades corresponded to the isolation with migration model IM4, which involved one sudden N_e change in each current clade and gene flow between both ancestral populations and current clades (Fig. 5; Table S7). The split of the most recent common ancestor into ancestral C_4 and non-C_4 populations was estimated to be *c.* 634 kga (*c.* 3 Ma) and was followed by the expansion of both populations, although the expansion was larger in the ancestral C_4 population. The split of non-C_4 clades I and II was estimated to be *c.* 228 kga (*c.* 1 Ma) and the split of C_4 clades IIIa and IV *c.* 66 kga (*c.* 0.3 Ma). A bottleneck was inferred at the origin of each clade, but it was stronger in the C_4 clades. Subsequent increases in N_e were inferred to roughly 140 kga (*c.* 0.7 Ma) for clade II, 72 kga (*c.* 0.4 Ma) for clade I, and 60 kga (*c.* 0.3 Ma) for clades IIIa and IV. The C_4 clades recovered N_e values close to their ancestral population shortly after they split, in relative terms, and overall clade IV experienced the largest increase in N_e towards the present. Besides, specifying further N_e changes in the ancestral C_4 and non-C_4 populations did not significantly improve the fit of the model (IM4ac, Table S7). Under the IM4ac model, the time estimates of ancestral N_e changes move towards the time of split between the ancestral populations under IM4, and the time of split moves towards 1 million generations ago, while all other estimates remain similar (Table S7).

Although all migration rates were low, the highest rate was estimated from the ancestral C_4 population to the ancestral

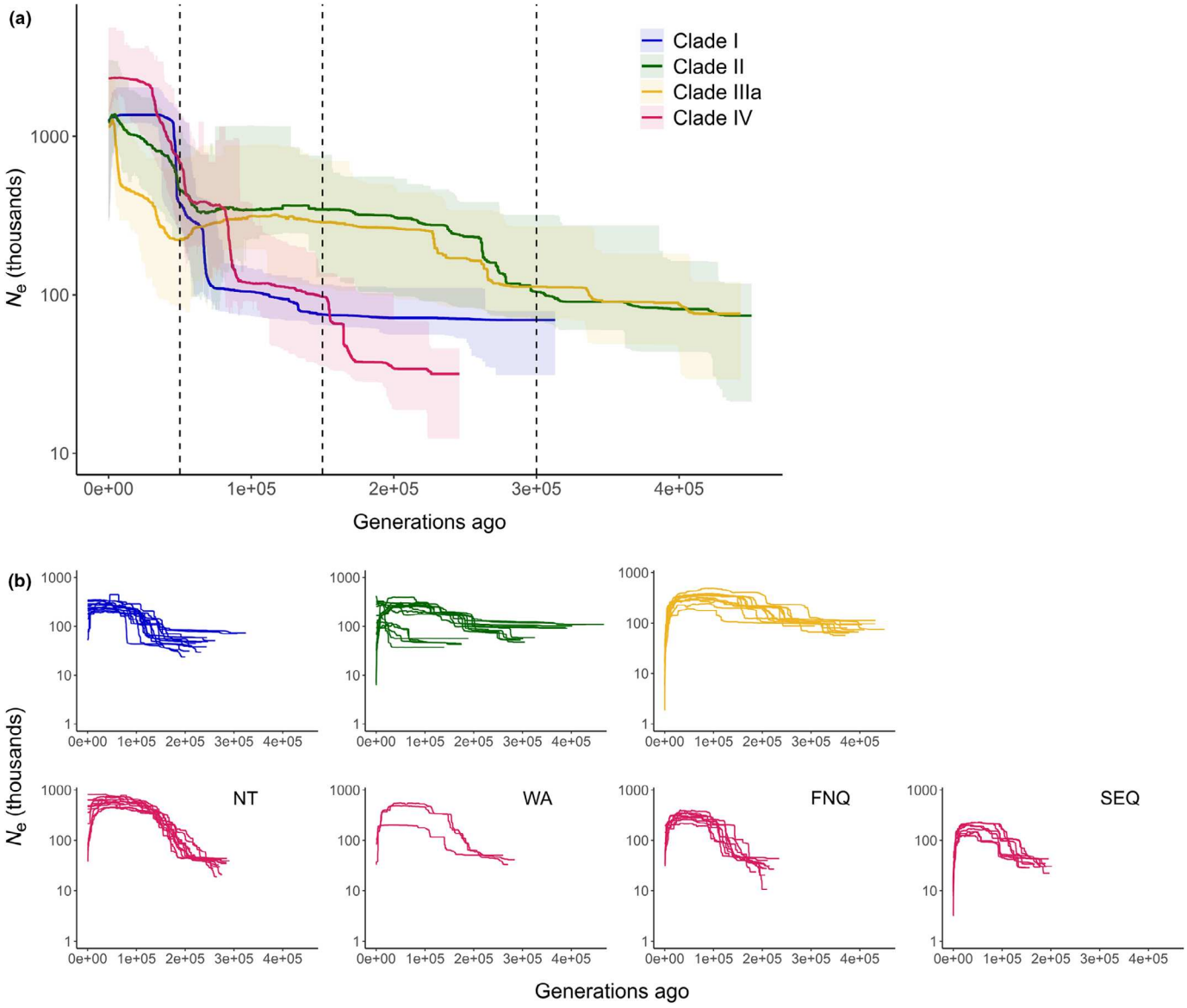

Fig. 4 Changes in effective population size (N_e) through time inferred with STAIRWAY PLOT 2 for *Alloteropsis semialata*. (a) Changes inferred by clade, based on folded site frequency spectrum (SFS) obtained after downsampling the data to 30 individuals per clade. Shaded areas represent 95% confidence intervals. Dotted vertical lines mark the approximate start of expansions (see main text). (b) Changes inferred for each locality with at least four sampled individuals. African localities are grouped by clade (upper panels) and Australian localities, by region (bottom panels: NT, Northern Territory; WA, Western Australia; FNQ, Far North Queensland; SEQ, South East Queensland). Four localities with profiles distinct from the others in the same clade are not shown: TAN2, ZAM1716 (clade II), and TAN1603, ZAM1505 (clade IIIa), but see Supporting Information Fig. S8 for separate plots with 95% confidence intervals for each locality. (a, b) Colours correspond to clades; *y*-axis is shown in log scale.

non-C_4 population (looking forward in time). This rate was one order of magnitude higher than migration from clade IIIa to clade II and from clade II to clade I, and two orders of magnitude higher than the rest (Fig. 5; Table S7). Therefore, our models support the ancient hybridization between C_4 and non-C_4 populations that was inferred from the incongruence between chloroplast and nuclear phylogenies and among nuclear gene trees (Bianconi *et al.*, 2020a; Raimondeau *et al.*, 2023), as well as the contribution of gene flow from C_4 to non-C_4 backgrounds to the diversification of the photosynthetic types proposed before (Olofsson *et al.*, 2016; Dunning *et al.*, 2017). In this regard, our models point to a scenario of constant migration between types (Methods S2; Tables S6, S7), but the considerably low estimates of migration rates (< 1 migrant - gene copy - per generation) and the geographic context would be rather compatible with recurrent episodes of genetic exchange (Bianconi *et al.*, 2020a; Olofsson *et al.*, 2021).

Discussion

Here, we combined niche models projected into paleoclimate layers spanning the last 5 Myr with demographic models based

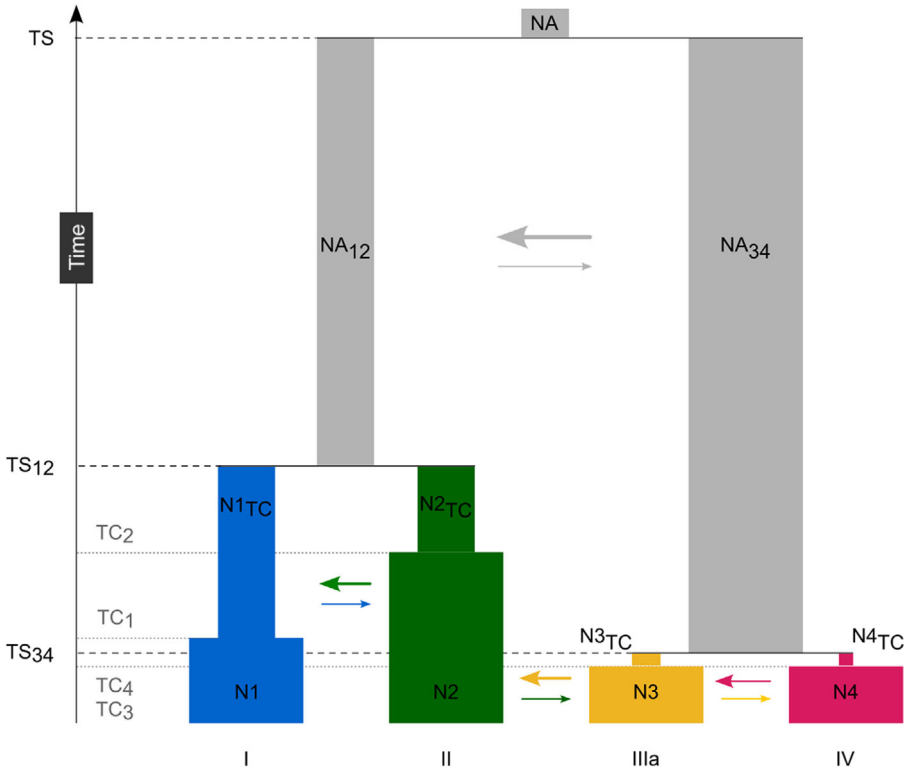

Fig. 5 Scheme of the joint demographic history of the four *Alloteropsis semialata* clades inferred with FASTSIMCOAL2 under the best fitted model (IM4, see Supporting Information Table S7 for parameter estimates and model comparisons). Results are based on the joint site frequency spectrum (SFS) obtained for a sample size of 10 individuals per clade. Column width is proportional to effective population size (N) and height, to time (T). Current effective population sizes: N1 – clade I, N2 – clade II, N3 – clade IIIa, N4 – clade IV. Effective population sizes of each clade before sudden expansions: $N1_{TC}$ to $N4_{TC}$, respectively. Ancestral effective population sizes: NA_{12} – ancestral non-C_4, NA_{34} – ancestral C_4, NA – most recent common ancestor. Time of split events: TS – between non-C_4 and C_4 ancestral populations, TS_{12} – between non-C_4 clades I and II, TS_{34} – between C_4 clades IIIa and IV. Time of sudden expansions for each clade: TC_1 to TC_4, respectively. Horizontal arrow size is proportional to migration rates between populations, forward in time. The vertical arrow represents time increasing from the present to the past.

on genomic data to capture the impact of the emergence of a key trait, photosynthetic innovation, on the fate of different populations of the grass *A. semialata*. We analysed a total of 510 samples that represent the three photosynthetic types (C_3, $C_3 + C_4$, and C_4) and the four main nuclear lineages (clades I, II, III, and IV) of this species (Fig. 1). We observed that niche breath, geographic range, and effective population size were generally coupled in these four clades. They were all successful in the sense that they occupied distinct niches and experienced demographic expansions. However, overall, the acquisition of C_4 photosynthesis was followed by a larger expansion of C_4 populations (Figs 4, 5; Table S7), linked to the ability of C_4 populations to occupy a broader climatic space (Figs 3, S3, S5).

The ancestral C_4 population underwent higher population growth than the ancestral non-C_4 in the region of origin after they split (Fig. 5). This points to an advantage brought by the C_4 trait, under otherwise equal conditions for both ancestors. Such an advantage is further reflected in the unparalleled expansion of C_4 clade IV (Fig. 4). It dispersed from Africa to Oceania within the last million years and attained the highest population growth among current clades. The rapid spread of clade IV is consistent with a large potential range over time derived from a wider climatic niche (Figs S3a, S5), together with long-distance dispersal events. These events probably facilitated the dispersal across the Indian Ocean since we did not identify potential land corridors from Africa to Asia, and transoceanic dispersal was also frequent in other angiosperm groups (de Queiroz, 2005; Linder *et al.*, 2018). The Australian localities are at the edge of the expansion range and accordingly show footprints of serial founder effects: population growth was more recent and lower as populations moved away from the northern region, the potential entry to the continent (Olofsson *et al.*, 2019), and recent declines were detected towards the easternmost localities (Figs 4b, S8, S9). Moreover, the Australian localities largely represent the climatic flexibility of the *A. semialata* C_4 type, as they span most of the temperature and precipitation conditions that the C_4 populations experience across the range (Figs S3b, S4).

By contrast, the other C_4 clade analysed, clade IIIa, showed the lowest population growth (Fig. 4) and the narrowest potential range over time (Fig. S5). It was almost confined to the region of origin in Central Zambezia. This region has a higher elevation than the surrounding lowlands and is associated with miombo forest, features that possibly acted as dispersal barriers (Bianconi *et al.*, 2020a; Olofsson *et al.*, 2021). The extent of the miombo forest also varied during the glacial cycles (Ivory *et al.*, 2018), and these fluctuations could be related to the recent declines observed in most clade IIIa localities (Fig. 4b). The trajectory of clade IIIa could thus result from geographic contingencies. This hypothesis is in line with the dependency of the C_4 trait effect on the particular context of each lineage (Christin & Osborne, 2014), as shown over the long-term evolution of C_4 grasses (Spriggs *et al.*, 2014; Aagesen *et al.*, 2016). The two *A. semialata* C_4 clades (III and IV) do not differ in photosynthetic performance, although the genetic variation underlying the C_4 pathway does vary among them (Lundgren *et al.*, 2016, 2019; Dunning *et al.*, 2019a,b). In this way, the success of clade IV could be driven by other adaptations not necessarily related to the C_4 pathway (see Christin & Osborne, 2014), such as enhanced dispersal linked to changes in seed morphology or germination control (see Linder *et al.*, 2018). However, despite the substantial

phenotypic diversity within *A. semialata*, so far, there is no evidence of consistent trait differences among the clades beyond the photosynthetic pathway (see Pereira *et al.*, 2023) that would support this hypothesis. Clades were found to be similar in terms of plant height, general morphology, flowering phenology, and seed size (Lundgren *et al.*, 2015), even though a comprehensive assessment of trait variation across the species range is still missing (see Pereira *et al.*, 2023). Besides, we note that we have not analysed the C_4 polyploids that prevail in Central Zambezia and belong to clade IIIb, the other component of the C_4 clade III (Olofsson *et al.*, 2021). Polyploidy may indeed improve establishment and ecological flexibility in grasses (Linder *et al.*, 2018), but the role of the polyploids in the evolution and success of the *A. semialata* types (and clades) is not understood yet and deserves further investigation.

The $C_3 + C_4$ clade II was also distributed in Central Zambezia, with local declines being compatible with the geographic constraints of the region. Nonetheless, clade II showed a larger population growth than clade IIIa (Fig. 4), and interestingly, its recent expansion (*c.* 0.7 Ma, Fig. 5; Table S7) did not coincide with any major change in geographic or climatic conditions (Fig. S5). This suggests that something happened to the $C_3 + C_4$ plants that made them more successful within the region they already inhabited. In this regard, previous studies demonstrated the transfer of genes encoding one of the core enzymes of the C_4 cycle, the phosphoenolpyruvate carboxykinase decarboxylase (PCK), from C_4 to $C_3 + C_4$ lineages after they diverged (Dunning *et al.*, 2017). The transfer of functional genes, between close or distant taxa, is currently recognized as an important source of adaptive variation for plant evolution (Christin *et al.*, 2012; Dunning *et al.*, 2019b; Olofsson *et al.*, 2019; Bianconi *et al.*, 2020b; Wickell & Li, 2020; Hibdige *et al.*, 2021; Raimondeau *et al.*, 2023). Therefore, it is tempting to suggest that the acquisition of this or any other C_4-related character might have promoted the demographic expansion of clade II in the absence of evident external triggers. Along these lines, our demographic models confirm that the diversification of the photosynthetic types occurred in the face of gene flow, between both ancestral populations and current clades, and that this preferentially happened from C_4 to non-C_4 groups (Fig. 5; Table S7).

Last, the C_3 clade I also underwent a rapid population growth (Fig. 4), consistent with a dispersal out of Central Zambezia as clade IV but into colder habitats. The niche of clade I was broader than that of clade II (Fig. S3), but its potential range remained smaller and patchier (Figs 3, S5). The range of clade I is currently limited to South African mountains, where the C_3 plants seem to be well-adapted as reflected in the local demographic trajectories (Fig. 4). The success of clade I was probably mediated by the development of freezing tolerance to cope with low-temperature extremes (Osborne *et al.*, 2008). In this respect, the evolution of cold tolerance across different C_3 lineages is considered another key factor in shaping the global distribution of grasses (Edwards & Smith, 2010; Linder *et al.*, 2018).

Overall, our inferences from the niche models projected into past climates and from the demographic models were concordant with each other, and with the spatial and temporal frame previously proposed for the photosynthetic diversification within *A. semialata* (Lundgren *et al.*, 2015; Olofsson *et al.*, 2016, 2019, 2021; Dunning *et al.*, 2017; Bianconi *et al.*, 2020a). Our approach provided new insights into the relative success of the clades after the acquisition of the C_4 trait, regardless of the limitations. We assumed niche conservatism to estimate the distribution of the clades over space and time. Yet, the possibility of niche evolution cannot be discarded, especially for clade IV giving its impressive range expansion. We also assumed the uncertainty on the mutation rate (1×10^{-8} per site per generation) and generation time (5 yr) used for demographic modelling. In our demographic models, besides, clade IV was only represented by Australian samples while the non-Australian ones would conform to 'ghost' populations not accounted for. This could mislead the demographic inferences to some extent (see Methods S2; Excoffier *et al.*, 2013; Maisano Delser *et al.*, 2019; Momigliano *et al.*, 2021). Despite all the caveats, altogether our results indicate that the *A. semialata* C_4 clades hold a history of larger effective population size than the non-C_4 clades. Furthermore, our results show that niche breath and predicted geographic extent are generally coupled with estimated effective population size in these clades, supporting an immediate beneficial effect of the C_4 trait on the demography of the populations driven by a broader climatic niche of the C_4 plants.

Conclusion

By integrating niche and demographic modelling, our study supports that C_4 photosynthesis was immediately beneficial to the *A. semialata* populations where it emerged. The C_4 trait probably conferred them the ability to thrive under a broader set of temperature and precipitation conditions, and thus to extend the geographic range and to increase the population size beyond other (non-C_4) populations in the face of recurrent environmental changes. Furthermore, the posterior transfer of some C_4-related characters to non-C_4 populations might have promoted the recent expansion of non-C_4 populations in the region of origin. Yet, the varying trajectories observed across C_4 localities and clades highlight that the effect of the photosynthetic trait might be masked by geographic contingencies. Overall, our insights into the demographic history of *A. semialata* C_4 and non-C_4 populations represent a step forward to disentangle how the emergence of a key trait can influence evolutionary trajectories over different timescales.

Acknowledgements

We thank Paolo Momigliano for advice on demographic analyses, and Lewis A. Jones and Sofía Galván for suggestions on data visualization. This study was funded by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement 947921) as part of the MAPAS project. Computational resources were provided by the Galicia Supercomputing Center (CESGA). Open access charge was covered by Universidade de Vigo and

Consorcio Interuniversitario do Sistema Universitario de Galicia (CISUG). SG was funded by the Ministry of Universities and the Next Generation European Union programme through a Margarita Salas Grant from Universidad Complutense de Madrid (CT31/21). LTD was supported by a Natural Environment Research Council Independent Research Fellowship (grant no.: NE/T011025/1).

Competing interests

None declared.

Author contributions

GS, P-AC and SV conceived and designed the study. GS performed the demographic analysis. SG performed the niche analysis. GS interpreted the results and wrote the manuscript with help from SG, LTD, P-AC and SV.

ORCID

Pascal-Antoine Christin https://orcid.org/0000-0001-6292-8734
Luke T. Dunning https://orcid.org/0000-0002-4776-9568
Sara Gamboa https://orcid.org/0000-0002-0829-3747
Graciela Sotelo https://orcid.org/0000-0002-0577-6655
Sara Varela https://orcid.org/0000-0002-5756-5737

Data availability

All ddRADseq data were retrieved from NCBI SRA BioProjects PRJNA560360 and PRJNA649872. The reference genome of *A. semialata* ASEM_AUS1_v1.0 was accessed through GenBank no. QPGU01000000. Sample information is provided in Table S1. The scripts used in this study are available from GitHub: https://github.com/MAPASlab/A_semialata_demography.

References

Aagesen L, Biganzoli F, Bena J, Godoy-Bürki AC, Reinheimer R, Zuloaga FO. 2016. Macro-climatic distribution limits show both niche expansion and niche specialization among C_4 panicoids. *PLoS ONE* **11**: e0151075.

Bianconi ME, Dunning LT, Curran EV, Hidalgo O, Powell RF, Mian S, Leitch IJ, Lundgren MR, Manzi S, Vorontsova MS *et al.* 2020a. Contrasted histories of organelle and nuclear genomes underlying physiological diversification in a grass species. *Proceedings of the Royal Society B: Biological Sciences* **287**: 20201960.

Bianconi ME, Hackel J, Vorontsova MS, Alberti A, Arthan W, Burke SV, Duvall MR, Kellogg EA, Lavergne S, McKain MR *et al.* 2020b. Continued adaptation of C_4 photosynthesis after an initial burst of changes in the Andropogoneae grasses. *Systematic Biology* **69**: 445–461.

Busby J. 1991. BIOCLIM – a bioclimate analysis and prediction system. *Plant Protection Quarterly* **6**: 8–9.

Christin P-A, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N. 2008. Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Current Biology* **18**: 37–43.

Christin P-A, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP. 2012. Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Current Biology* **22**: 445–449.

Christin P-A, Osborne CP. 2014. The evolutionary ecology of C_4 plants. *New Phytologist* **204**: 765–781.

Clark RM, Tavaré S, Doebley J. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Molecular Biology and Evolution* **22**: 2304–2312.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al.* 2011. The variant call format and VCFTOOLS. *Bioinformatics* **27**: 2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM *et al.* 2021. Twelve years of SAMTOOLS and BCFTOOLS. *GigaScience* **10**: giab008.

Dunning LT, Lundgren MR, Moreno-Villena JJ, Namaganda M, Edwards EJ, Nosil P, Osborne CP, Christin P. 2017. Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. *Evolution* **71**: 1541–1555.

Dunning LT, Moreno-Villena JJ, Lundgren MR, Dionora J, Salazar P, Adams C, Nyirenda F, Olofsson JK, Mapaura A, Grundy IM *et al.* 2019a. Key changes in gene expression identified for different stages of C_4 evolution in *Alloteropsis semialata*. *Journal of Experimental Botany* **70**: 3255–3268.

Dunning LT, Olofsson JK, Parisod C, Choudhury RR, Moreno-Villena JJ, Yang Y, Dionora J, Quick WP, Park M, Bennetzen JL *et al.* 2019b. Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences, USA* **116**: 4416–4425.

Edwards EJ, Osborne CP, Strömberg CAE, Smith SA, Bond WJ, Christin P-A, Cousins AB, Duvall MR, Fox DL, Freckleton RP *et al.* 2010. The origins of C_4 grasslands: integrating evolutionary and ecosystem science. *Science* **328**: 587–591.

Edwards EJ, Smith SA. 2010. Phylogenetic analyses reveal the shady history of C_4 grasses. *Proceedings of the National Academy of Sciences, USA* **107**: 2532–2537.

Edwards GE, Ku MSB. 1987. Biochemistry of C_3–C_4 intermediates. In: Hatch MD, Boardman NK, eds. *The biochemistry of plants: a comprehensive treatise.* Cambridge, MA, USA: Academic Press, 275–325.

Ehleringer JR, Cerling TE, Helliker BR. 1997. C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia* **112**: 285–299.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9**: e1003905.

Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. FASTSIMCOAL2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**: 4882–4885.

Hesselbarth MHK, Sciaini M, With KA, Wiegand K, Nowosad J. 2019. LANDSCAPEMETRICS: an open-source R tool to calculate landscape metrics. *Ecography* **42**: 1648–1657.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**: 183–195.

Heyduk K, Moreno-Villena JJ, Gilman IS, Christin P-A, Edwards EJ. 2019. The genetics of convergent evolution: insights from plant photosynthesis. *Nature Reviews. Genetics* **20**: 485–493.

Hibdige SGS, Raimondeau P, Christin P-A, Dunning LT. 2021. Widespread lateral gene transfer among grasses. *New Phytologist* **230**: 2474–2486.

Hijmans RJ, Phillips S, Leathwick J, Elith JL. 2022. *DISMO: species distribution modeling.* [WWW document] URL https://CRAN.R-project.org/package=dismo [accessed 30 October 2022].

Holden PB, Edwards NR, Rangel TF, Pereira EB, Tran GT, Wilkinson RD. 2019. PALEO-PGEM v.1.0: a statistical emulator of Pliocene–Pleistocene climate. *Geoscientific Model Development* **12**: 5137–5155.

Ivory SJ, Lézine A, Vincens A, Cohen AS. 2018. Waxing and waning of forests: Late Quaternary biogeography of southeast Africa. *Global Change Biology* **24**: 2939–2951.

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W *et al.* 2012. Genome-wide genetic changes during modern breeding of maize. *Nature Genetics* **44**: 812–815.

Kadereit G, Ackerly D, Pirie MD. 2012. A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). *Proceedings of the Royal Society B: Biological Sciences* **279**: 3304–3311.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**: 356.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with BOWTIE 2. *Nature Methods* **9**: 357–359.

Lesturgie P, Planes S, Mona S. 2022. Coalescence times, life history traits and conservation concerns: an example from four coastal shark species from the Indo-Pacific. *Molecular Ecology Resources* **22**: 554–566.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMTOOLS. *Bioinformatics* **25**: 2078–2079.

Linder HP, Lehmann CER, Archibald S, Osborne CP, Richardson DM. 2018. Global grass (Poaceae) success underpinned by traits facilitating colonization, persistence and habitat transformation. *Biological Reviews* **93**: 1125–1144.

Liu X, Fu Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nature Genetics* **47**: 555–559.

Liu X, Fu Y-X. 2020. STAIRWAY PLOT 2: demographic history inference with folded SNP frequency spectra. *Genome Biology* **21**: 280.

Lundgren MR, Besnard G, Ripley BS, Lehmann CER, Chatelet DS, Kynast RG, Namaganda M, Vorontsova MS, Hall RC, Elia J *et al.* 2015. Photosynthetic innovation broadens the niche within a single species. *Ecology Letters* **18**: 1021–1029.

Lundgren MR, Christin P-A, Escobar EG, Ripley BS, Besnard G, Long CM, Hattersley PW, Ellis RP, Leegood RC, Osborne CP. 2016. Evolutionary implications of C_3–C_4 intermediates in the grass *Alloteropsis semialata*. *Plant, Cell & Environment* **39**: 1874–1885.

Lundgren MR, Dunning LT, Olofsson JK, Moreno-Villena JJ, Bouvier JW, Sage TL, Khoshravesh R, Sultmanis S, Stata M, Ripley BS *et al.* 2019. C_4 anatomy can evolve via a single developmental change. *Ecology Letters* **22**: 302–312.

Maisano Delser P, Corrigan S, Duckett D, Suwalski A, Veuille M, Planes S, Naylor GJP, Mona S. 2019. Demographic inferences after a range expansion can be biased: the test case of the blacktip reef shark (*Carcharhinus melanopterus*). *Heredity* **122**: 759–769.

Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. 2016. On the importance of being structured: instantaneous coalescence rates and human evolution – lessons for ancestral population size inference? *Heredity* **116**: 362–371.

Miller AH, Stroud JT, Losos JB. 2023. The ecology and evolution of key innovations. *Trends in Ecology & Evolution* **38**: 122–131.

Momigliano P, Florin A-B, Merilä J. 2021. Biases in demographic modeling affect our understanding of recent divergence. *Molecular Biology and Evolution* **38**: 2967–2985.

Monnahan P, Kolář F, Baduel P, Sailer C, Koch J, Horvath R, Laenen B, Schmickl R, Paajanen P, Šrámková G *et al.* 2019. Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nature Ecology & Evolution* **3**: 457–468.

Olofsson JK, Bianconi M, Besnard G, Dunning LT, Lundgren MR, Holota H, Vorontsova MS, Hidalgo O, Leitch IJ, Nosil P *et al.* 2016. Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. *Molecular Ecology* **25**: 6107–6123.

Olofsson JK, Curran EV, Nyirenda F, Bianconi ME, Dunning LT, Milenkovic V, Sotelo G, Hidalgo O, Powell RF, Lundgren MR *et al.* 2021. Low dispersal and ploidy differences in a grass maintain photosynthetic diversity despite gene flow and habitat overlap. *Molecular Ecology* **30**: 2116–2130.

Olofsson JK, Dunning LT, Lundgren MR, Barton HJ, Thompson J, Cuff N, Ariyarathne M, Yakandawala D, Sotelo G, Zeng K *et al.* 2019. Population-specific selection on standing variation generated by lateral gene tansfers in a grass. *Current Biology* **29**: 3921–3927.

Osborne CP, Wythe EJ, Ibrahim DG, Gilbert ME, Ripley BS. 2008. Low temperature effects on leaf physiology and survivorship in the C_3 and C_4 subspecies of *Alloteropsis semialata*. *Journal of Experimental Botany* **59**: 1743–1754.

Padilla-García N, Šrámková G, Záveská E, Šlenker M, Clo J, Zeisek V, Lučanová M, Rurane I, Kolář F, Marhold K. 2023. The importance of considering the evolutionary history of polyploids when assessing climatic niche evolution. *Journal of Biogeography* **50**: 86–100.

Parmesan C, Yohe G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**: 37–42.

Pereira L, Bianconi ME, Osborne CP, Christin P-A, Dunning LT. 2023. *Alloteropsis semialata* as a study system for C_4 evolution in grasses. *Annals of Botany* **132**: 365–382.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**: e37135.

de Queiroz A. 2005. The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution* **20**: 68–73.

Raimondeau P, Bianconi ME, Pereira L, Parisod C, Christin P-A, Dunning LT. 2023. Lateral gene transfer generates accessory genes that accumulate at different rates within a grass lineage. *New Phytologist* **240**: 2072–2084.

Sage RF. 2004. The evolution of C_4 photosynthesis. *New Phytologist* **161**: 341–370.

Sage RF, Christin P-A, Edwards EJ. 2011. The C_4 plant lineages of planet Earth. *Journal of Experimental Botany* **62**: 3155–3169.

Šavrič B, Patterson T, Jenny B. 2019. The Equal Earth map projection. *International Journal of Geographical Information Science* **33**: 454–465.

Schoener TW. 1968. The Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology* **49**: 704–726.

Sinha NR, Kellogg EA. 1996. Parallelism and diversity in multiple origins of C_4 photosynthesis in the grass family. *American Journal of Botany* **83**: 1458–1470.

Spriggs EL, Christin P-A, Edwards EJ. 2014. C_4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS ONE* **9**: e97722.

Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* **182**: 205–216.

Svenning J-C, Eiserhardt WL, Normand S, Ordonez A, Sandel B. 2015. The influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annual Review of Ecology, Evolution, and Systematics* **46**: 551–572.

Urban MC, Bocedi G, Hendry AP, Mihoub J-B, Pe'er G, Singer A, Bridle JR, Crozier LG, De Meester L, Godsoe W *et al.* 2016. Improving the forecast for biodiversity under climate change. *Science* **353**: aad8466.

Varela S, Lima-Ribeiro MS, Terribile LC. 2015. A short guide to the climatic variables of the Last Glacial Maximum for biogeographers. *PLoS ONE* **10**: e0129037.

Varela S, Mateo RG, García-Valdés R, Fernández-González F. 2014. Macroecology and ecoinformatics: evaluating the accuracy of the ecological niche models calibrated with species occurrence data with biases and/or errors. *Ecosistemas* **23**: 46–53.

Vicentini A, Barber JC, Aliscioni SS, Giussani LM, Kellogg EA. 2008. The age of the grasses and clusters of origins of C_4 photosynthesis: clustered origins of C_4 photosynthesis. *Global Change Biology* **14**: 2963–2977.

Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**: 2868–2883.

Warren DL, Matzke NJ, Cardillo M, Baumgartner JB, Beaumont LJ, Turelli M, Glor RE, Huron NA, Simões M, Iglesias TL *et al.* 2021. ENMTOOLS 1.0: an R package for comparative ecological biogeography. *Ecography* **44**: 504–511.

Watcharamongkol T, Christin P-A, Osborne CP. 2018. C_4 photosynthesis evolved in warm climates but promoted migration to cooler ones. *Ecology Letters* **21**: 376–383.

Wickell DA, Li F-W. 2020. On the evolutionary significance of horizontal gene transfers in plants. *New Phytologist* **225**: 113–117.

Yang N, Xu X-W, Wang R-R, Peng W-L, Cai L, Song J-M, Li W, Luo X, Niu L, Wang Y *et al.* 2017. Contributions of *Zea mays* subspecies mexicana haplotypes to modern maize. *Nature Communications* **8**: 1874.

Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen J-Q, Hurst LD, Tian D. 2015. Parent–progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**: 463–467.

Zhou H, Helliker BR, Huber M, Dicks A, Akçay E. 2018. C_4 photosynthesis and climate through the lens of optimality. *Proceedings of the National Academy of Sciences, USA* **115**: 12057–12062.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Distribution of all *Alloteropsis semialata* samples analysed in this study.

Fig. S2 Scheme of demographic models evaluated with FASTSIMCOAL2.

Fig. S3 Environmental space of *Alloteropsis semialata* samples.

Fig. S4 Global distribution of suitable areas for each *Alloteropsis semialata* clade.

Fig. S5 Extent of suitable areas through time for each *Alloteropsis semialata* clade.

Fig. S6 Global distribution of stable areas for each *Alloteropsis semialata* clade.

Fig. S7 Normalized site frequency spectrum for each *Alloteropsis semialata* clade.

Fig. S8 Demographic trajectory for each *Alloteropsis semialata* locality.

Fig. S9 Normalized site frequency spectrum for each *Alloteropsis semialata* locality.

Fig. S10 Magnitude and timing of demographic changes across *Alloteropsis semialata* localities.

Methods S1 One-population demographic models with FASTSIMCOAL2.

Methods S2 Two-population demographic models with FASTSIMCOAL2.

Notes S1 Variation in demographic trajectories across *Alloteropsis semialata* localities.

Table S1 List of *Alloteropsis semialata* samples analysed in this study.

Table S2 Measure of niche overlap between *Alloteropsis semialata* clades.

Table S3 Extent of suitable areas through time for each *Alloteropsis semialata* clade.

Table S4 Number of stable areas for each *Alloteropsis semialata* clade.

Table S5 Estimates from one-population demographic models in FASTSIMCOAL2.

Table S6 Estimates from two-population demographic models in FASTSIMCOAL2.

Table S7 Estimates from four-population demographic models in FASTSIMCOAL2.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.