GBE

Genome Assemblies of the Daisy *Gorteria diffusa* Elucidate the Molecular-Genomic Basis of Pollination by Sexual Deception

Roman T. Kellenberger [1,2], Boris Delahaie [1,3,4], Joana I. Meier [5,6], Allan G. Ellis [7], Beverley J. Glover [1,*]

[1]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK
[2]Crop Genome Dynamics Group, Agroscope Changins, Nyon 1260, Switzerland
[3]CIRAD, UMR DIADE, Montpellier F-34398, France
[4]UMR DIADE, Université de Montpellier, CIRAD, IRD, Montpellier, France
[5]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK
[6]Tree of Life Programme, Wellcome Sanger Institute, Hinxton CB10 1SA, UK
[7]Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa

*Corresponding author: Email: bjg26@cam.ac.uk.

Accepted: February 28, 2026

Abstract

Plant sexual deception, the floral mimicry of female insects to attract mate-searching males for pollination, is a long-studied reproductive strategy with a poorly understood genomic basis. Here, we assembled the genomes of a sexually deceptive, a semi-deceptive, and a derived nondeceptive floral form of the South African daisy *Gorteria diffusa* to chromosome-scale and near-chromosome scale, respectively. We located several previously identified genes involved in the development of deceptive floral traits, including tandem duplications of *GdbHLH* and *GdMYBSG6* transcription factors regulating the complex coloration of sexually deceptive floral structures. Using additional genotyping-by-sequencing data of six *G. diffusa* floral forms, we further identified several large inverted genomic segments with a high fixation index (F_{ST}), which seem to play a role in maintaining the distinct identity of some floral forms in zones of secondary contact. Finally, genome synteny analyses revealed that the genome of the derived nondeceptive floral form is contracted and shows signs of recent genome-wide deletion of long terminal repeat retrotransposons. Our results provide insight into the genomic elements underlying plant sexual deception as well as some of the structural genomic differences between sexually deceptive and nondeceptive floral forms.

Key words: structural genomics, genome size evolution, chromosomal inversion, transposable element turnover, plant sexual deception, Asteraceae.

© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

Significance

Even though sexually deceptive flowers are well-known beyond the plant research community, the genomic basis of plant sexual deception is still poorly charted. Sparked by an unexpected difference in genome size, we conducted genomic comparisons between sexually deceptive and nondeceptive floral forms of a South African daisy. We detected several structural genomic patterns associated with plant sexual deception: A cluster of duplicated transcription factors that have been shown to regulate the coloration of sexually deceptive floral parts, large chromosomal inversions that contribute to phenotype stabilization of sexually deceptive and nondeceptive floral forms upon hybridization, and a genome-wide deletion of transposable elements coinciding with secondary loss of sexual deception. Our study highlights the flexibility of the genome and its role in adaptive increase and decrease of phenotypic complexity.

Introduction

Plant sexual deception, when flowers imitate cues of female insects to attract mate-searching males for pollination, is a highly specialized plant reproductive strategy that has been studied for decades (Correvon and Pouyanne 1916; Jersáková et al. 2006). Known cases of plant sexual deception mainly exist in American, Eurasian, African, and Australian orchid lineages, but also in the Asian iris *Iris paradoxa* and the South African daisy *Gorteria diffusa*, representing a classic example of convergent evolution (Vereecken et al. 2012; Peakall 2023). Sexually deceptive flowers are extremely elaborate and trick male pollinators into pseudocopulation by simultaneously imitating the morphology (de Jager and Peakall 2016), texture (Agren et al. 1983), and often the sexual pheromones (Schiestl et al. 1999; Bohman et al. 2016) of females. The floral traits involved are usually tightly integrated, and even small phenotypic changes can have far-reaching consequences such as a drastic reduction in pollination efficiency or even a switch to a different pollinator species (Xu and Schlüter 2015; Sedeek et al. 2016). Understanding the molecular-genomic basis of such complex fine-tuned interspecies interactions is a central goal of organismal biology, and several studies have identified individual genes and pathways underlying particular floral traits that contribute to pollinator mimicry. This includes anthocyanin pathway genes and *MYB-bHLH-WD40* (*MBW*) transcription factor (TF) complexes controlling floral coloration, *EXPANSIN* (*EXPA*) genes and *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) TFs altering flower morphology and development, and *STEAORYL-ACP DESATURASE* (*SAD*) and *β-KETOACYL-ACP SYNTHASE* (*KAS*) genes mediating pheromone production (Vignolini et al. 2012; Sedeek et al. 2013; Bohman et al. 2016; Xu et al. 2017; Wong et al. 2019; Kellenberger et al. 2023; Fattorini et al. 2024). However, due to the general lack of tools and resources for these nonmodel plants, the genomic underpinnings of plant sexual deception are largely unknown.

Variation in genome size and structure has repeatedly been linked to different plant reproductive strategies. Polyploidization can induce transitions from outcrossing to self-fertilization (Siopa et al. 2020; Dou et al. 2024), rearrangement of floral genes into supergene structures underlies sexual dimorphisms such as heterostyly (Li et al. 2016) or dioecy (Charlesworth and Charlesworth 1978), and processes such as genomic inversion (Wan et al. 2022) and duplication (Liang et al. 2023), gene degradation (Hoballah et al. 2007), and transposon activity (Hu et al. 2024) can lead to shifts in pollination mode. However, the generally large and repetitive orchid genomes (Leitch et al. 2009) have so far impeded genome-wide analyses of plant sexual deception. First insights from the recently published genome assembly of the bee-pollinated sexually deceptive orchid *Ophrys sphegodes* suggest that an expansion of transposable elements in combination with gene duplication and chromosome fusions provided the genomic basis for the extreme floral specialization and adaptive radiation of the *Ophrys* genus (Russo et al. 2024). Transcriptomic analyses of the sexually deceptive Australian orchid genus *Chiloglottis* further imply that fine-tuning of existing pigment and floral volatile pathways can promote phenotypic diversity of deceptive floral traits even if the genomic basis is largely conserved (Wong et al. 2019). While these studies provide invaluable insights into the genomic properties fueling the diversification of sexually deceptive clades, information concerning the initial shift to sexual deception in these lineages is limited. Pinpointing the molecular-genomic underpinnings of plant sexual deception ultimately requires genomic comparisons of closely related sexually deceptive and nondeceptive taxa.

The daisy *G. diffusa* Thunb. (Asteraceae) is an annual, self-incompatible herb from the winter-rainfall Succulent Karoo in South Africa. In the past 0.6 to 2.2 Ma, *G. diffusa* radiated into 15 parapatric floral forms (morphotypes), which hybridize in narrow zones of secondary contact (Ellis and Johnson 2012). All morphotypes are pollinated by the same bombyliid fly

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

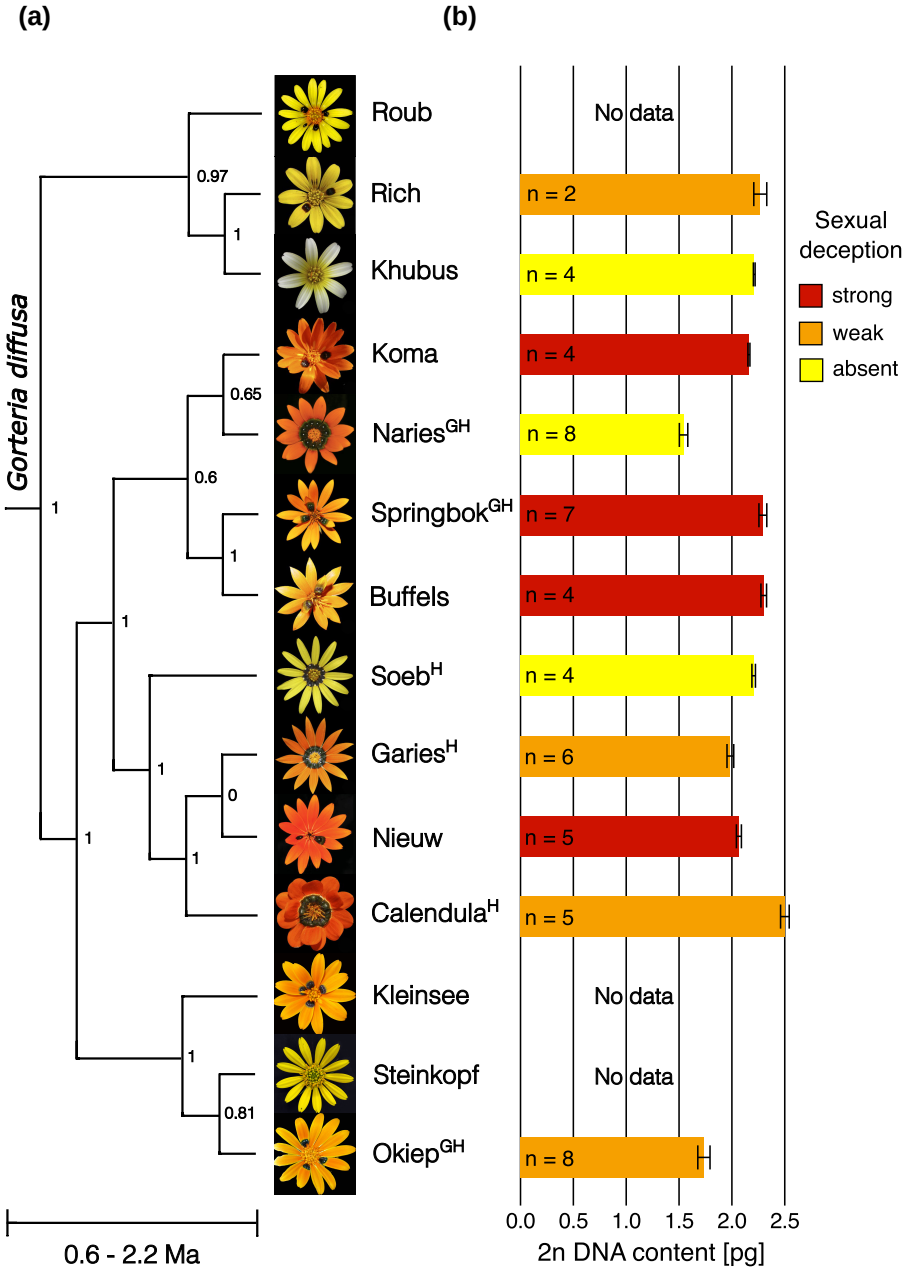


Fig. 1. Phylogeny, phenotype, sexual deception, and genome sizes of the *Gorteria diffusa* morphotypes. a) Floral phenotype and coalescent-based phylogeny of 14 *G. diffusa* morphotypes from Delahaie et al. (2022). Morphotypes labeled "G" were sequenced in this study, and morphotypes labeled "H" were used for the hybrid zone analysis. A 15th morphotype (*G. diffusa* "Oubees") was excluded from the analysis (all images taken by the authors). b) Degree of sexual deception and diploid genome size of 11 of the 14 assessed *G. diffusa* morphotypes. Ancestral trait reconstruction in Delahaie et al. (2022) showed that sexual deception evolved twice independently in the two central clades. Also, the derived nondeceptive Naries morphotype with small genome is nested within a clade of strongly deceptive morphotypes with large genomes. Sexual deception was experimentally determined in Ellis and Johnson (2010) and Ellis et al. (2014). Genome size was determined with flow cytometry (bars denote mean 2n DNA content in pg ± 1 SD, sample sizes for the flow cytometry analysis are indicated inside the bars).

Megapalpus capensis Wiedeman (Ellis and Johnson 2009 ; Delahaie et al. 2022) (Fig. 1a), but the reproductive strategy differs among morphotypes depending on the number, morphological complexity, and positioning of dark spots on the yellow to orange petal-like ray florets (Johnson and Midgley 1997). While morphotypes with simple petal spots arranged in nectar-guiding rings (de Jager et al. 2017) are visited for nectar and pollen, morphotypes with more complex, isolated petal spots imitate female *M. capensis* and elicit sexual responses

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

in *M. capensis* males ranging from inspection to pseudocopulation (Ellis and Johnson 2009, 2010; de Jager and Ellis 2012). The *G. diffusa* radiation thus consists of nondeceptive, semi-deceptive and deceptive morphotypes (Ellis et al. 2014) (Fig. 1a and b).

Previous studies have elucidated a large part of the developmental-genetic pathways underlying the three key traits of sexually deceptive *G. diffusa* petal spots, including dark green pigmentation, trichome-like cellular papillae, and isolated petal spot positioning (Thomas et al. 2009; Kellenberger et al. 2023; Fattorini et al. 2024). These traits seem to have evolved independently and sequentially, and only morphotypes in which all three petal spot traits are united are strongly sexually deceptive (weakly deceptive morphotypes lack either papillae or petal spot isolation) (Kellenberger et al. 2023). The recently resolved phylogenetic relationship of *G. diffusa* morphotypes further revealed that full sexual deception evolved twice within the *G. diffusa* complex and has been lost again in at least one case (Naries morphotype, see Fig. 1a and b) (Delahaie et al. 2022). The *G. diffusa* radiation is thus an ideal system in which to study the evolutionary origin and reversal of plant sexual deception.

Here, we aimed to elucidate the genomic basis of sexual deception in the *G. diffusa* radiation. We analyzed the genomes of the *G. diffusa* morphotypes by flow cytometry and found surprisingly large variation in genome size, suggesting the presence of structural genomic differences between morphotypes. We thus assembled the genome of the sexually deceptive *G. diffusa* Springbok morphotype to chromosome level, and the genomes of the semi-deceptive *G. diffusa* Okiep and derived nondeceptive *G. diffusa* Naries morphotypes to near-chromosome scale. We conducted genome comparisons as well as population-genomic analyses with genotyping-by-sequencing data of six morphotypes from four natural hybrid zones. Our analyses revealed genomic elements associated with plant sexual deception in *G. diffusa*, including clusters of genes and transcription factors underlying sexually deceptive traits, large-scale genomic inversions and genome-wide changes in transposable element content. Our results give an insight into the molecular-genomic basis of plant sexual deception in the *G. diffusa* system.

Results

Flow Cytometry Analysis of the *G. diffusa* Morphotypes

We estimated the genome size of 11 *G. diffusa* morphotypes from all major clades with flow cytometry (Fig. 1a and b). Results revealed surprisingly large variation in genome size, ranging from 2n = 1.5 to 2.5 pg (equivalent to ca. 1.5 to 2.5 Gbp), with no change in ploidy (2n = 10 chromosomes for all morphotypes) (Thomas 2009). Looking at the entire *G. diffusa* morphotype phylogeny, genome size does not correlate with the degree of sexual deception as determined previously through behavioral experiments with the pollinator *M. capensis* (Ellis and Johnson 2010; Ellis et al. 2014) (phylogenetic linear regression model, $n = 11$, $t = 0.976$, $P = 0.354$). However, the derived nondeceptive *G. diffusa* Naries morphotype turned out to have a ca. 33% smaller genome than its sexually deceptive sister morphotypes *G. diffusa* Springbok, Buffels and Koma in the same clade.

Chromosome Scale Assembly of the *G. diffusa* Springbok Genome

At 2n = 2.3 Gbp, the genome of the sexually deceptive *G. diffusa* Springbok morphotype is among the largest within the radiation, and a large part of previously published ecological, developmental, and genetic studies of the *G. diffusa* pollination system include this morphotype (Ellis and Johnson 2012; Kellenberger et al. 2023; Fattorini et al. 2024, <bibref rid="20">2009; Ellis and Johnson 2010; de Jager and Ellis 2012; Ellis et al. 2014; Delahaie et al. 2022). We extracted high-molecular weight DNA (HMW-DNA) from *G. diffusa* Springbok genotype 8 used in Kellenberger et al. (2023) and produced a total of 19.6 M PacBio CLR reads to an estimated coverage of 60× and 848.1 M Illumina PE 150 short reads to an estimated coverage of 100× (preceding the release of PacBio HiFi technology). An initial *k-mer* analysis estimated a high heterozygosity rate of 2.59% and 48.3% repetitive content. We used *canu* (Koren et al. 2017) to assemble the PacBio reads to 38,248 contigs with an N50 value of 249.1 Kbp (Fig. S1), conducted initial polishing with the PacBio reads, and purged haplotigs before submitting the assembly and leaf tissue to Dovetail Genomics LLC (USA) for omni-c scaffolding. After gap-filling, we closed residual gaps by polishing again with the PacBio reads, reduced sequencing errors by polishing with the Illumina reads, and manually joined two large scaffolds each representing one arm of chromosome 5 (subsequently called chromosome 5a and 5b). The final assembly contained five pseudomolecules corresponding to the five chromosomes of the haploid genome (Fig. 2a) plus 1,730 unplaced contigs. With a scaffold N50 of 148.6 Mbp and a total length of 963.0 Mbp, the assembled genome was slightly shorter than the 1n = 1.15 Gbp estimated by flow cytometry, likely due to the collapse of simple repeats during the assembly process (Koren et al. 2017). Nevertheless, 98.4% of Illumina reads could be mapped to the final assembly, and a BUSCO (Benchmarking

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

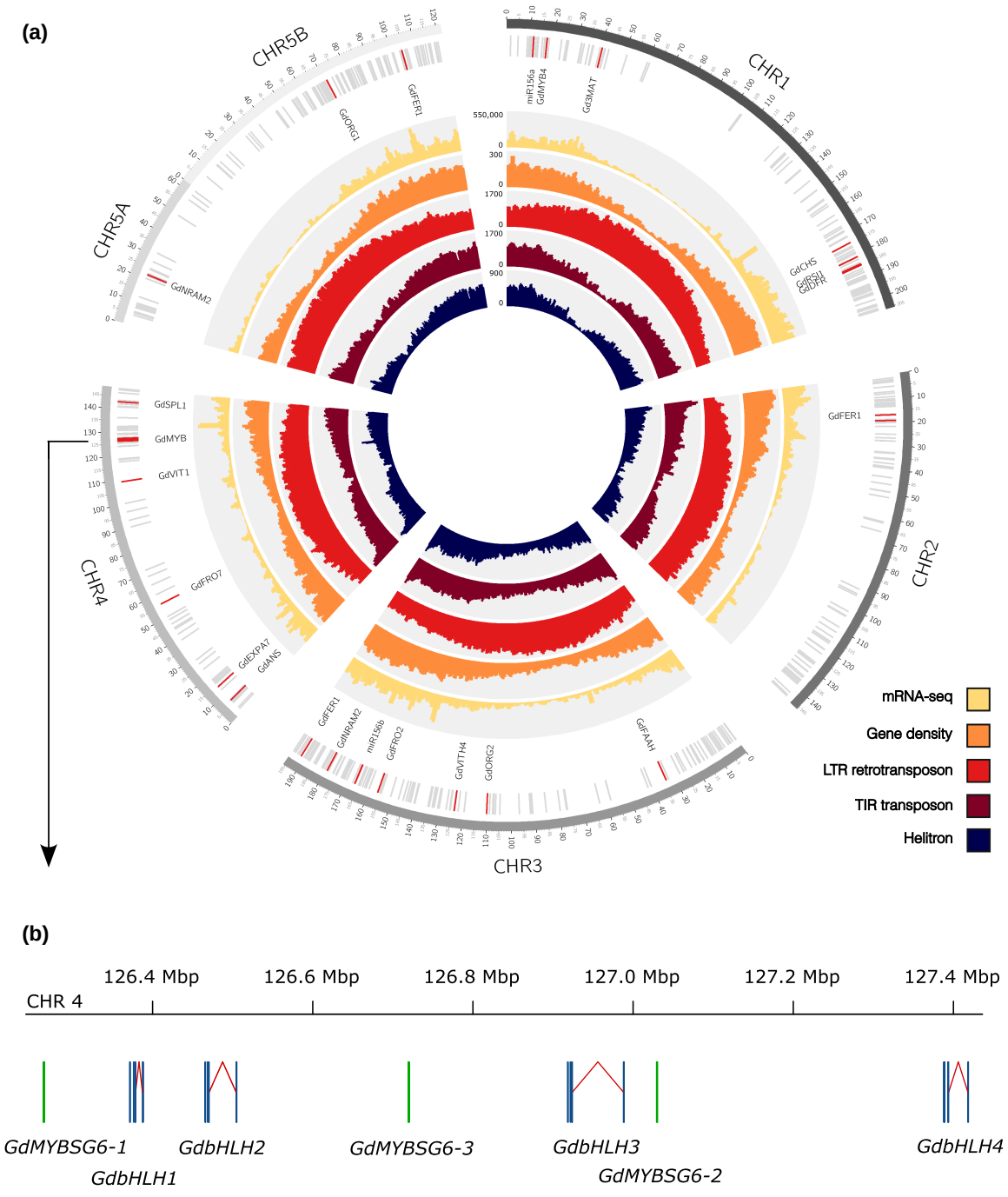


Fig. 2. Genome assembly of *Gorteria diffusa* Springbok with tandem duplication of *bHLH* and *MYB* transcription factors. a) Scaffolded G. *diffusa* Springbok genome assembly with five pseudochromosomes. Concentric plots depict (inwards): location of genes differentially expressed between spotted and unspotted Springbok petals with previously characterized genes labeled, mapping density of mRNA-seq reads from Springbok petals, gene density, density of long-terminal repeat (LTR) retrotransposons, density of tandem inverted repeat (TIR) transposons, helitron density (window size 2 Mbp with 1 Mbp overlap). b) Three *MYB* (*GdMYBSG6-1-3*) and four *bHLH* transcription factors (*GdbHLH1-4*) are clustered within a 1.2 Mbp region on chromosome 4, all encoded on the forward strand. These genes have been shown to regulate anthocyanin pigment production in the sexually deceptive *G. diffusa* petal spots (Fattorini et al. 2024). The *GdbHLH* genes have unusually long introns of up to 60 Kbp length (depicted as chevrons).

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

Universal Single-Copy Orthologs) analysis estimated a gene body completeness of 95.1%.

Location of Genetic Elements Underlying Plant Sexual Deception in the *G. diffusa* Springbok Genome

Using *BRAKER* and *EnTAP* with previously generated mRNA-seq data (Kellenberger et al. 2023) from the same *G. diffusa* Springbok genotype 8, we annotated a total of 62,513 gene structures in the Springbok genome, 41,631 with corresponding functional annotation. Matching annotations with previously characterized *G. diffusa* candidate genes (Kellenberger et al. 2023; Fattorini et al. 2024) revealed the location of several key genes underlying sexually deceptive petal spot traits (Fig. 2a). This includes the co-opted genes *EXPANSIN 7* (*GdEXPA7*) involved in petal spot cell elongation, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 1* (*GdSPL1*) (co)regulating petal spot positioning on the flower, and *VACUOLAR IRON TRANSPORTER 1* (*GdVIT1*), which plays a central role in iron-mediated petal spot coloration (Kellenberger et al. 2023). Other important petal spot genes from the iron homeostasis and anthocyanin pigmentation pathways, as well as the two precursor loci of the *miR156* microRNA that negatively controls *GdSPL1* expression (Kellenberger et al. 2023), were found as well.

In addition to the co-opted petal spot genes, we found an aggregation of three *MYB* (*GdMYBSG6-1-3*) and four *bHLH* (*GdbHLH1-4*) TF paralogs within a 1.2 Mbp region of chromosome 4 (Fig. 2b), all encoded on the forward strand. These TFs are homologous to previously characterized *MYB* and *bHLH* TFs which form flavonoid/anthocyanin regulating MYB-bHLH-WD40 (MBW) TF complexes (Yan et al. 2021). The interaction of GdMYBSG6 with GdbHLH has not yet been functionally tested, but previous studies on *G. diffusa* Springbok have shown that the expression of both *GdMYBSG6-1-3* and the *GdbHLH* cluster is highly upregulated in developing petal spots (Kellenberger et al. 2023 ; Fattorini et al. 2024). The upregulated MBW TFs activate the anthocyanin pathway genes *DIHYDROFLAVONOL 4-REDUCTASE* (*GdDFR*) and *ANTHOCYANIN 3-O-GLUCOSIDE-6′′-O-MALONYLTRANSFERASE* (*Gd3MAT*), which leads to localized, petal-spot specific accumulation of malonylated cyanidin pigment on the adaxial side of the ray florets (Fattorini et al. 2024). Cyanidin petal spot pigmentation is a synapomorphy of all *G. diffusa* morphotypes (Delahaie et al. 2022) and plays a key role in the mimicry of the female *M. capensis* exoskeleton in deceptive morphotypes (Thomas et al. 2009). A fourth, previously described, *MYB* paralog, *GdMYBSG6-4*, resides on chromosome 1 (Fig. 2a). This *MYB* copy is expressed throughout the ray florets and most likely regulates the accumulation of anthocyanin pigment on the abaxial side of the ray florets, which is unrelated to sexual deception (Fattorini et al. 2024). Within the anthocyanin TF complex on chromosome 4, *GdbHLH* and *GdMYBSG6* paralogs are clustered in alternate fashion, indicating at least two tandem duplications of an ancestral *bHLH* and *MYB* gene copy. We further noticed that intron seven of the *bHLH* gene copies is unusually long, separating the last two exons with one of them containing the *bHLH* motif by 8.5 Kbp (*GdbHLH1*), 33.7 Kbp (*GdbHLH2*), 60.7 Kbp (*GdbHLH3*), and 24.0 Kbp (*GdbHLH4*) from the rest of the coding sequence, with intron space mostly occupied by transposable elements. Overall, these results suggest that tandem duplication and subfunctionalization of *GdMYBSG6* and *GdbHLH* TFs underlie the cyanidin petal spot pigmentation present in all *G. diffusa* morphotypes, an important step in the evolution of sexually deceptive floral traits within the *G. diffusa* radiation.

Near-chromosome Scale Assemblies of the *G. diffusa* Naries and Okiep Genomes

Nested within a clade of three highly sexually deceptive *G. diffusa* morphotypes, Koma, Springbok, and Buffels, the Naries morphotype has the smallest genome of all examined morphotypes (2n = 1.5 Gbp) and has frequently been used in previous genetic and developmental analyses (Ellis and Johnson 2009; Ellis et al. 2014; Delahaie et al. 2022; Kellenberger et al. 2023). The Okiep morphotype is more distantly related to the Springbok and Naries morphotypes, intermediate in strength of sexual deception as well as in genome size (2n = 1.7 Gbp). We extracted HMW-DNA from *G. diffusa* Naries genotype 13 used in Kellenberger et al. (2023), as well as from the previously uncharacterized *G. diffusa* Okiep genotype 1, and generated a total of 9.0 and 8.5 M PacBio HiFi data to an estimated coverage of 20× each. Initial *k-mer* analyses estimated heterozygosity rates of 3.57% and 3.54% as well as 47.9% and 51.9% repetitive content for Naries and Okiep respectively. We used *hifiasm* (Cheng et al. 2021) to produce collapsed haploid assemblies consisting of 723 contigs for Naries (N50 = 83.7 Mbp, total length = 791.5 Mbp, BUSCO gene body completeness = 96.5%, Fig. S2) and 565 contigs for Okiep (N50 = 51.4 Mbp, total length = 959.3 Mbp, BUSCO gene body completeness = 96.7%, Fig. S3).

Association of Chromosomal Inversions with the Maintenance of Morphotype Identity Upon Secondary Contact

The *G. diffusa* morphotypes grow in parapatry and hybridize with neighboring morphotypes in zones of

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

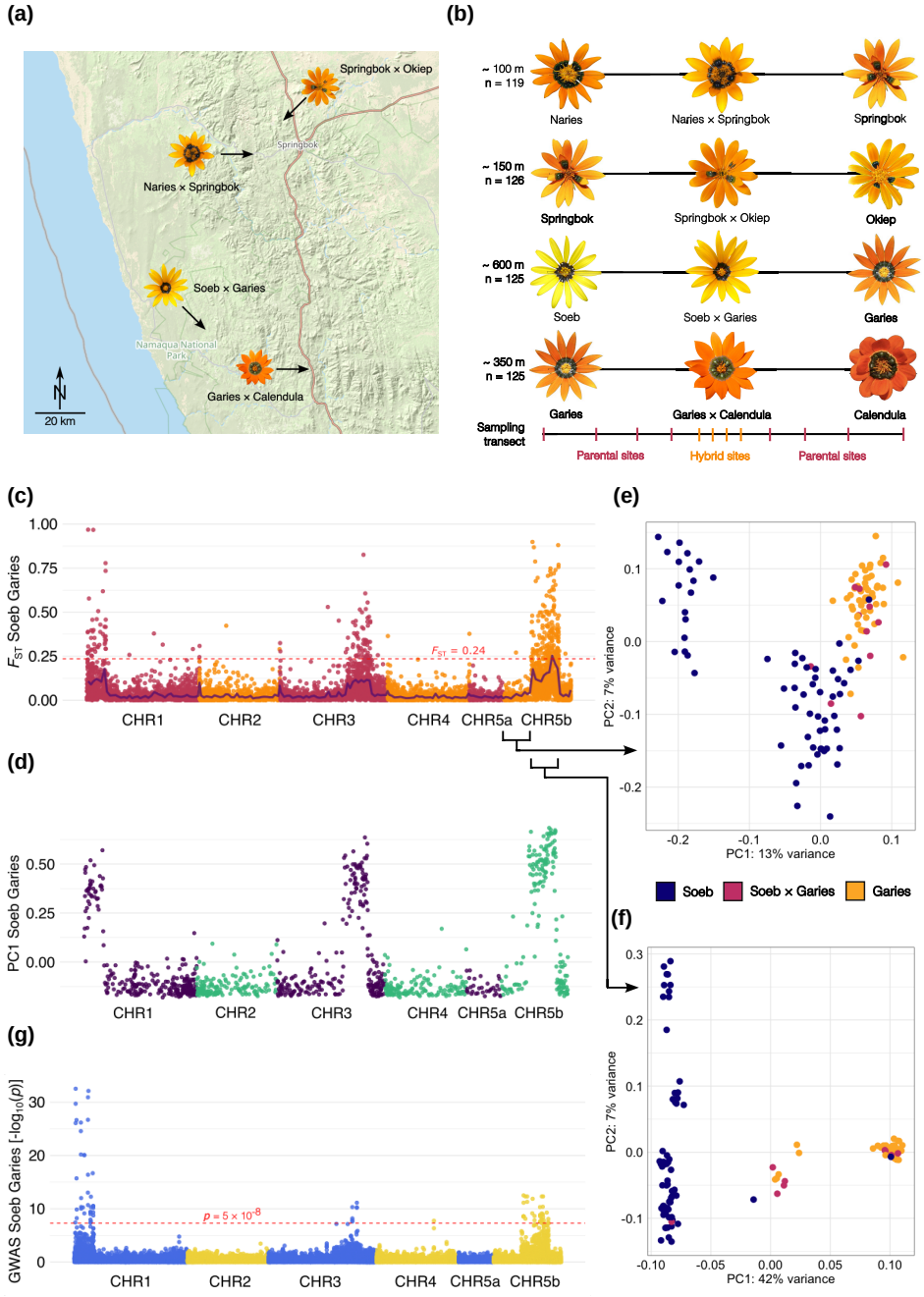


Fig. 3. Population genomic analyses of the *Gorteria diffusa* hybrid zones. a) Geographic location of the four investigated *G. diffusa* hybrid zones in South African Namaqualand. Map data from https://www.openstreetmap.org (2025). b) Sampling scheme of the GBS data. For each of the four *G. diffusa* hybrid zones, DNA and phenotypic data were collected from 12 plants each at 12 sampling sites (4 hybrid sites and 2 × 4 parental sites) situated along a transect of a few 100 m spanning the entire hybrid zone. Exclusion of some samples resulted in the sample sizes indicated for each hybrid zone (all images taken by the authors). c) F_{ST} scan in 5 Kbp (dots) and 5 Mbp windows (solid line) across the genome between *G. diffusa* Soeb and Garies. Islands of high F_{ST} occur on chromosomes 1, 3, and 5, with the one on chromosome 5 overlapping with the detected inversion between Springbok and Okiep. The F_{ST} outlier threshold of 0.24 corresponds to the 99th percentile of all SNPs in LD ≤ 0.4. d) Local PCA of Soeb and Garies in a sliding window of 100 SNPs across the genome. Large-scale heterogeneity in patterns of relatedness typical for chromosomal inversions occur at the genomic locations of the F_{ST} islands. e) Local PCA of Soeb and Garies within the un-inverted part of chromosome 5b adjacent to the inversion. Explanatory power of PC1 is low (13%) and individuals do not cluster well according to phenotype. f) Local PCA of Soeb and Garies within the inverted part of chromosome 5b. Explanatory power of PC1 is much higher (42%) and individuals cluster in three groups largely corresponding to the parental (homokaryotypic) and hybrid (heterokaryotypic) phenotypes. g) GWAS of petal color for Soeb (yellow petals) and Garies (orange petals). Petal color is strongly associated with the three inverted genomic segments (univariate linear mixed model with likelihood ratio test, $n = 125$, P-value threshold = 5×10^{-8}).

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

secondary contact (Ellis and Johnson 2009). These hybrid zones are extremely narrow with a width of only a few hundred meters, and virtually no hybrids are found outside their boundaries (Ellis and Johnson 2012). The factors that contain the spread of hybrids and prevent the introgression of genomic material from nondeceptive into sexually deceptive morphotypes thus contribute to the maintenance of sexual deception in *G. diffusa.* Among the three morphotypes for which we generated reference genomes, contact zones exist between the sexually deceptive Springbok and nondeceptive Naries, and between deceptive Springbok and weakly deceptive Okiep (Fig. 3a). Genome synteny analysis between Springbok and Naries and between Springbok and Okiep, respectively, revealed a large inversion of ca. 42 Mbp on chromosome 5b of the Okiep genome (Figs. S4 and S5). The inverted segment contains 3,939 gene structures, of which 2,721 have a functional annotation. Alignment of the combined Springbok and Naries reference transcriptome from (Kellenberger et al. 2023) to the Springbok assembly further showed that 54 genes differentially expressed between spotted and unspotted florets are located within the inversion (Table S1). We thus hypothesized that structural genomic differences influence the dynamics of some of the hybrid zones.

To test this, we produced Genotyping-By-Sequencing (GBS) data from a total of 490 *G. diffusa* individuals sampled in transects across four different hybrid zones between the morphotype pairs Naries-Springbok (NaSp), Springbok-Okiep (SpOk), Soeb-Garies (SoGa), and Garies-Calendula (GaCa). The additional SoGa and GaCa hybrid zones include three closely related morphotypes with different types of ring spots and background petal coloration generating little or no sexual deception (Fig. 3b). Including SoGa and CaGa thus allowed us to identify unique as well as common genomic features between hybrid zones both in presence and absence of strong sexual deception. We aligned the GBS data to the Springbok reference assembly; called, filtered, and phased a total of 104,281 single nucleotide polymorphisms (SNPs); and genetically confirmed the phenotype-based identification of pure and hybrid individuals with *fastSTRUCTURE* (Fig. S6). Global fixation index (F_{ST}) estimations from pure individuals were highest for SpOk ($F_{ST} = 0.13$), intermediate for NaSp ($F_{ST} = 0.11$) and GaCa ($F_{ST} = 0.08$), and lowest for SoGa ($F_{ST} = 0.06$). However, the degree of admixture within the corresponding hybrid zones seems to be mainly influenced by topographic heterogeneity. The overall proportion of hybrids relative to pure individuals is highest in the GaCa hybrid zone located on a flat plateau, with many more later-generation hybrids present than in the NaSp hybrid zone situated in a steep ravine (Fig. 3a and b and Fig. S3).

To identify regions of high genomic differentiation between morphotypes, we calculated F_{ST} values in 5 Kbp and 5 Mbp nonoverlapping windows across the genome for pure individuals from each hybrid zone. Since inverted chromosomal blocks in high linkage disequilibrium (LD) influence the identification of outlier windows, we pruned SNPs in LD > 0.4 and set the F_{ST} outlier threshold at the 99th percentile of the pruned F_{ST} value distribution. F_{ST} values for NaSp, SpOk, and GaCa showed high variation throughout the genome, and the inversion on chromosome 5 in SpOk was visible as a block of high F_{ST} (Fig. S7a to c). In contrast, F_{ST} values for SoGa were low throughout the genome except for three highly elevated F_{ST} islands on chromosomes 1, 3, and 5, with the one on chromosome 5 overlapping with the inversion detected in the Okiep assembly (Fig. 3c). To understand the impact of these F_{ST} islands on population structure of the hybrid zones, we conducted principal component analysis (PCA) on sliding windows of 100 SNPs across the genome for all individuals. While patterns of relatedness for NaSp were consistent across the genome (Fig. S8a), SpOk and SoGa showed large-scale heterogeneity at the genomic locations of the F_{ST} islands typical for genomic inversions (Mérot et al. 2021) (Fig. 3d and Fig. S8b). High genomic heterogeneity was also detected for GaCa in a narrower region of chromosome 3 downstream of the putative inversion in SoGa (Fig. S8c).

If the putatively inverted segments act as barrier loci between morphotypes, sequence polymorphisms within these regions should segregate according to floral phenotype. To test this, we performed local PCA for the identified F_{ST} island regions as well as for adjacent low F_{ST} regions of the same length (whenever possible) on the same chromosomes of the SpOk, SoGa, and GaCa hybrid zones. Assessing the first two principal components showed an increase of PC1 explanatory power within all four F_{ST} islands compared to the neighboring regions (Fig. 3e and f and Fig. S9a to d). Within the segments of high F_{ST}, individuals are separable in three groups along PC1 which largely correspond to the parental (homokaryotypic) and hybrid (heterokaryotypic) floral phenotypes and later-generation hybrids in GaCa cluster with their backcrossed parent morphotype (Fig. S9d). These findings were corroborated by an independently conducted local PCA in 10 Mbp sliding windows across the genome (Fig. S10a to d). A genome-wide association study (GWAS) for background petal color in the SoGa hybrid zone (yellow in Soeb, orange in Garies, and intermediate in hybrids, see Fig. 3b) further showed an association of petal color with a large number of associated SNPs within the three putatively inverted genomic

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

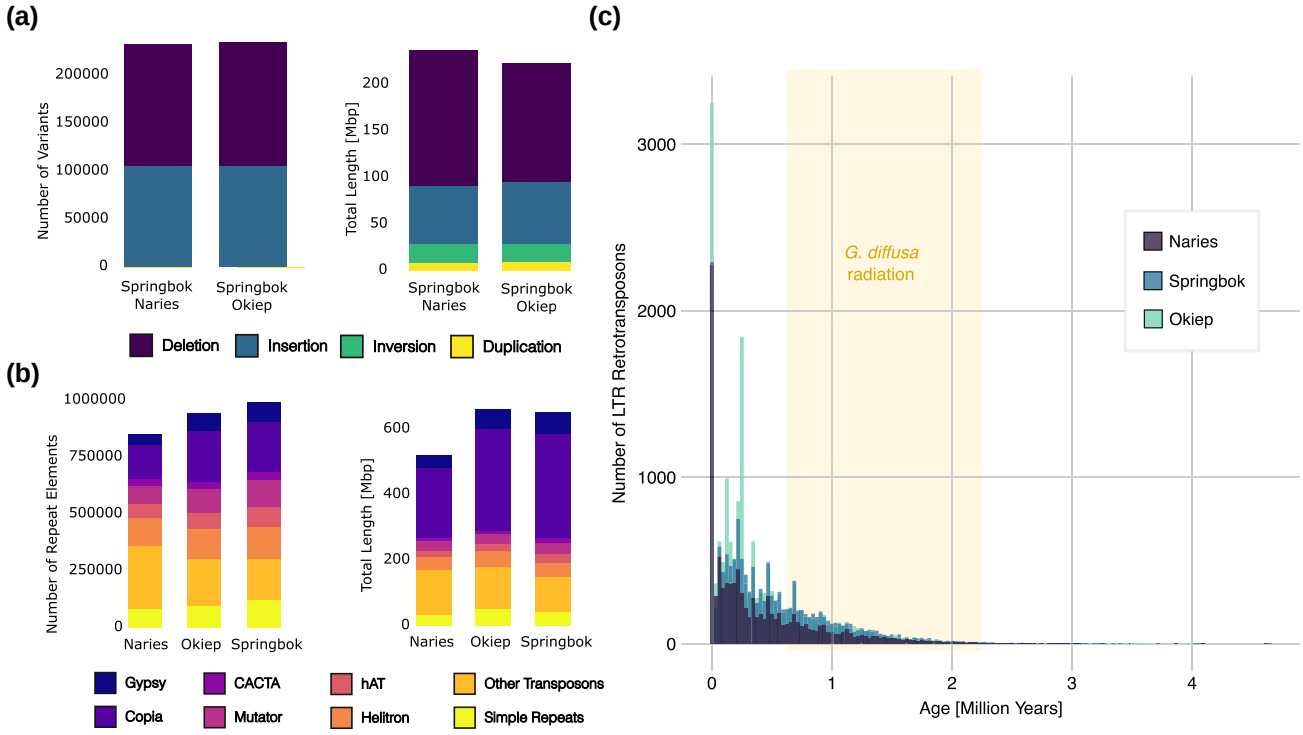


Fig. 4. Structural variant and transposable element analyses of *Gorteria diffusa* Springbok, Okiep, and Naries. a) Number (left plot) and cumulative length (right plot) of detected structural variants between the Springbok and Naries assemblies (left bar) and between the Springbok and Okiep assemblies (right bar). While the number of variants is similar in Naries and Okiep, the total length of deletions is longer, and the total length of insertions is shorter in Naries than in Okiep. b) Number (left plot) and cumulative length (right plot) of annotated repeat elements in the Springbok, Okiep, and Naries assemblies. While the total number of repeat elements increases gradually among the three assemblies, the total length of repeat elements is much shorter in the Naries assembly, mostly due to fewer Gypsy and Copia LTR retrotransposons. c) Number versus age of intact LTR retrotransposons (LTR-RT) in the Springbok, Okiep, and Naries assemblies. The Naries assembly contains fewer LTR-RT of all ages than the other two assemblies, except for the most recent LTR-RT insertions. In contrast, the abundance of LTR-RT is very similar in the Springbok and Okiep assemblies, except for several recent bursts of an unknown LTR-RT group in the Okiep assembly. The highlighted window denotes the estimated time of the initial *G. diffusa* radiation.

regions (Fig. 3g). Altogether, these findings suggest that some loci underlying key traits for morphotype identity are located in inverted genomic segments of high F_{ST}, contributing to the persistence of parental morphotypes in zones of secondary contact.

Genome Contraction and Deletion of Transposable Elements in the Derived Nondeceptive *G. diffusa* Naries Morphotype

To understand the pronounced difference in genome size between the closely related sexually deceptive *G. diffusa* Springbok and the derived nondeceptive *G. diffusa* Naries morphotypes, we performed synteny analysis between the two assemblies. This analysis did not reveal any chromosome-level deletion in the Naries genome (Fig. S4). The observed difference in genome size must thus be a cumulative result of many subchromosomal structural differences. Scanning for small-scale structural variation with *sniffles2* detected a total of 232,083 structural variants with a size of less than 1 Mbp between the Springbok and Naries assemblies (Fig. 4a). Among these are 127,466 deletions with a total length of 145.8 Mbp, 104,418 insertions with a total length of 62.2 Mbp, 118 inversions, 81 duplications, and 0 translocations (the terms "insertion" and "deletion" are relative to the Naries assembly since the ancestral state of indels is unknown). The cumulative length of all deletions surpasses the length of all insertions by 83.6 Mbp, amounting to 48.7% of the observed size difference of 171.6 Mbp between the two assemblies. This low number is likely a result of the rather strict filtering, excluding many variants with unclear break points and variants overlapping with assembly gaps. The amount of variants detected between the assemblies of the Springbok and the weakly deceptive Okiep morphotypes (total 234,233 structural variants, of which 128,943 deletions, 105,108 insertions, 117 inversions, 65 duplications, 0 translocations) was very similar to those between Springbok and Naries

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

(the terms "insertion" and "deletion" are relative to the Okiep assembly). However, the cumulative length of all deletions (127.2 Mbp) was shorter, and the length of all insertions (66.7 Mbp) was longer than for Springbok versus Naries, resulting in a lower total length difference of 60.5 Mbp.

The expansion and contraction of transposable elements (TEs) is a major driver of genome size dynamics in angiosperms (Lee and Kim 2014). To quantify the difference in TE load among the sequenced *G. diffusa* morphotypes, we de novo annotated TEs in each genome assembly with *EDTA* (Ou et al. 2019). We found 989,236 repetitive elements in the Springbok assembly, 942,569 elements in the Okiep assembly, and 849,317 elements in the Naries assembly (Fig. 4b). Despite the moderate difference in TE number between the assemblies, the cumulative length of all repetitive elements is much lower in the Naries assembly (520 Mbp, 65.7% of the genome) than in both the Okiep (661 Mbp, 68.9% of the genome) and Springbok assemblies (652 Mbp, 67.8% of the genome, slightly underestimated due to the deletion of simple repeats during the assembly of the PacBio CLR reads). The difference in total repeat length still amounts to 132.8 Mbp (77.4% of the observed genome size difference) between the Naries and Springbok assemblies and 141.3 Mbp (84.2% of the observed genome size difference) between the Naries and Okiep assemblies. A large part of this length difference can be attributed to Gypsy and Copia long terminal repeat retrotransposons (LTR-RTs), which are 45% and 31% less abundant in the Naries than in the Springbok assembly. Estimation of LTR-RT insertion times revealed that the Naries assembly contains ca. 40% to 45% fewer LTR-RTs of all ages than the Springbok and Okiep assemblies, except for the most recently inserted ones (Fig. 4c and Fig. S11). In contrast, LTR-RT load is very similar in the more distantly related Springbok and Okiep assemblies across all ages, except for several bursts of an unknown LTR-RT group in the Okiep genome within the last 0.2 M years. Considering that many of the assessed LTR-RTs inserted during or even before the time of the initial *G. diffusa* radiation (ca. 0.6 to 2.2 Ma), the constantly lower number of LTR-RTs across all but the most recent time frame in the Naries genome is likely an effect of recent genome-wide LTR-RT deletion rather than slower proliferation of these elements in just this lineage. Intersecting LTR-RT and gene annotations further showed that 1,788 of the Springbok LTR-RTs absent in Naries are inside or adjacent to annotated genes. Of these LTR-RTs, 267 overlap with genes differentially expressed between spotted, sexually deceptive petals and unspotted, nondeceptive petals of Springbok, including the previously characterized genes *GdVITH4* and *GdbHLH1* that are involved in petal spot pigmentation, suggesting a potential influence of LTR-RT excision on the expression of genes conferring sexual deception.

Discussion

With recent advances in long-read sequencing and analysis, it has become increasingly clear that structural genomic changes play a central role in the evolution of different organismic lifestyles (Kapheim et al. 2015; Fukushima et al. 2017; Segers et al. 2017; Cai et al. 2021; Looney et al. 2022). The comparatively well-understood genetic and ecological bases of plant reproductive strategies have allowed researchers to unravel some of the key associated genomic features (Li et al. 2016; Zhang et al. 2022; Liang et al. 2023). Plant sexual deception, however, has been largely excluded from these advances owing to the genomic complexity and the lack of model system characteristics of most sexually deceptive plant species. Our genomic analyses of sexually deceptive and nondeceptive *G. diffusa* morphotypes thus provide the first comparative insights at the intraspecies level into the genomics of plant sexual deception. Drawing parallels between *G. diffusa* and the only other fully sequenced sexually deceptive plant, *O. sphegodes* (Russo et al. 2024), allows the identification of some overarching patterns. In both taxa, tandem gene duplication and transposable element activity, which is known to influence gene expression dynamics (Dubin et al. 2018), seem to be associated with the highly elaborate and detailed floral patterning necessary to trick male pollinators into visiting the flowers. This is not surprising since these mechanisms are generally associated with increased variation and complexity of floral traits (Hoshino et al. 2001; Kuo et al. 2019). Our previous research has further shown that co-option of genes and pathways expressed in other developmental contexts has been central for the fine-tuning of floral traits toward sexual deception in *G. diffusa* (Kellenberger et al. 2023). Duplication and subsequent sub- and neofunctionalization of existing floral-related transcription factors such as *GbMYBSG6* and *GdbHLH* could thereby construct the regulatory network necessary for the integration of the newly co-opted genetic elements.

Although there are many similarities in the evolution of sexual deception between *Gorteria* and *Ophrys*, the evolutionary forces underlying further diversification of sexually deceptive structures are very different in the two taxa. While the *Ophrys* radiation was fueled by allomone-mediated pollinator shifts (Xu et al. 2012), all *G. diffusa* morphotypes are pollinated by the same fly species with no phylogeographic variation in behavior of male flies detected (de Jager and Ellis 2013). Previous studies suggest that the extant

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

morphotype diversity is a result of evolutionary tinkering with petal spot traits followed by pollinator-mediated integration of the most deceptive trait combinations (Ellis et al. 2014), perhaps influenced by learning patterns of male flies (de Jager and Ellis 2014). Irrespective of the initial diversifying processes, the apparent lack of reproductive barriers should theoretically lead to a quick erosion of morphotype identity upon secondary contact, but the narrow width of the contact zones suggests otherwise. The mechanisms maintaining morphotype identity may differ between contact zones, but our results indicate that genomic structure plays a role at least in some of them. Analogous to other plant breeding systems (Fishman et al. 2013; Hermann et al. 2013), the association of floral traits with our newly discovered chromosomal inversions implies reduced recombination between some genetic elements important for floral morphology and morphotype identity. The inversion on chromosome 5 was further detected twice in the Soeb and Okiep morphotype, albeit with slightly different inversion boundaries. Tracing the origin of these two inversions requires additional genomic data from sister morphotypes of Okiep, but depending on their evolutionary background they suggest a potential role of either segregation, genomic introgression, or parallel evolution of chromosomal inversions in morphotype diversification. However, the inverted segments we describe do not seem to overlap with the majority of the previously discovered genes and TFs involved in sexual deception. Also, the function of the 54 putative petal spot genes within the inversion on chromosome 5 has not yet been assessed, and the SNPs called from the GBS data are not dense enough to associate individual genes with floral trait differences between morphotypes. The significance of the detected inversions for the maintenance of sexually deceptive trait combinations thus remains to be determined.

A central question about evolutionary shifts of plant reproductive strategy is whether such transitions are reversible or not (Barrett 2013). Generally, highly specialized pollination systems such as plant sexual deception are expected to break down upon loss of the associated pollinator and either revert to a more generalized system or more commonly to self-pollination (Vereecken et al. 2012; Barrett 2013), as known from derived shelter mimicry in *Ophrys helenae* (Vereecken et al. 2012) and autogamy in *Ophrys apifera* (Darwin 1877; Fenster and Marten-Rodriguez 2007). Ancestral reconstruction of sexually deceptive floral traits showed that *G. diffusa* Naries lost the ability to produce papillae and isolated petal spots, reverting to nondeceptive flat ring spots as in other nondeceptive morphotypes from other clades (Delahaie et al. 2022). Unlike in the two *Ophrys* species, the evolutionary reversal of sexual deception in Naries was most likely not initiated by pollinator limitation and thus happened under very different selective pressure. Our results indicate that the strikingly reduced size of the Naries genome can largely be attributed to the lack of Gypsy and Copia LTR-RTs. The uniform distribution of missing LTR-RTs over all but the most recent insertion ages, some dating back to the most recent common ancestor of the *G. diffusa* morphotypes, points toward a recent genome-wide LTR-RT purging event in the Naries genome with many LTR-RTs absent in or near differentially expressed petal spot genes. Studies in other plant systems have shown that genome contraction and transposon excision can change the expression of floral genes such as MYB TFs that are homologous to known *G. diffusa* petal spot genes (Iida et al. 2004; Zou et al. 2023). Even though the function of most of the affected petal spot genes is yet unknown, it is thus plausible that genome contraction reverted petal spot papillae and positioning in Naries. The consequent elimination of sexually motivated visits from male flies, as well as continued or enhanced visits from female flies (all morphotypes produce nectar), may have resulted in the right combination of reproductive isolation and reproductive success to save the new morphotype from extinction.

Our study opens up many possibilities for further investigations, and we would like to point out four examples of particular interest. First, as noted earlier, complete sexual deception with pseudocopulation of male flies evolved twice in the *G. diffusa* radiation, in the clade of the Springbok morphotype and in the clade of the Nieuw morphotype (Delahaie et al. 2022) (Fig. 1a and b). Nieuw grows much further south than any other morphotype with no hybrid zone known to date. Genomic comparisons between these two clades would thus shed light on the molecular mechanisms enabling parallel evolution of plant sexual deception. Second, an extension of the genomic analyses to the other morphotypes of the Naries-Springbok clade, Buffels and Koma, would help to resolve outstanding questions on the genome size dynamics in the Naries morphotype, further elucidating the molecular mechanisms underlying the loss of plant sexual deception. Third, in vitro expression analyses of genes underlying sexually deceptive floral traits with nearby LTR-RTs edited in or out would allow a test of the possible connection between LTR-RT excision and loss of sexual deception in Naries. Fourth, fine-scale spatial expression analyses of the duplicated *GdMYBSG6* and *GdbHLH* TF paralogs would reveal which part of the petal spot is pigmented by which *GdMYBSG6–GdbHLH* combination. In addition, functional analyses of the *GdbHLH* copies would allow characterization of their molecular function beyond the regulation of anthocyanin production and

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

further illuminate the regulatory gene network behind the highly sophisticated sexually deceptive floral structures of *G. diffusa*.

Materials and Methods

Genome Size Estimation

Genome size was estimated using propidium iodide flow cytometry following the simplified two-step protocol of Doležel et al. (2007). Young leaves were harvested from 57 *G. diffusa* plants grown from field collected seed in 2009, representing 17 populations of 11 floral forms. Leaves were chopped in 0.5 mL of ice-cold Otto I buffer (0.1 M citric acid, 0.5% Tween 20) together with *Bellis perennis* L. (2C = 3.38 pg) as an internal reference standard that did not exhibit any signal overlap with any of the samples. The sample was then filtered through 42 µm nylon mesh and incubated for 10 min at room temperature before staining with a solution consisting of 1 mL Otto II buffer (0.4 M $Na_2HPO_4{\cdot}12H_2O$), propidium iodide and RNase IIA (both at final concentrations of 50 µg/mL), and β-mercaptoethanol (2 µL/mL). Samples were then run through a Partec CyFlow SL cytometer equipped with a diode-pumped solid state laser 532 nm (Cobolt Samba, 100 mW output power) which recorded fluorescence intensity of isolated nuclei (5,000 particles). Each sample was analyzed at least three times on different days. Correlation between estimated genome size and degree of sexual deception was analyzed in *R* v.4.3.2 with the package *phylolm* v.2.6.5 by fitting a phylogenetic linear regression model with mean genome size in pg per morphotype as response variable, degree of sexual deception (encoded as [0, 1, 2]) as dependent variable, and 100 bootstrap replicates. The model is based on the phylogenetic tree used for ancestral state reconstruction in (Delahaie et al. 2022).

Plant Cultivation

The *G. diffusa* plants were grown as outlined in Kellenberger et al. (2023). In brief, seed/bearing capitula of *G. diffusa* Springbok, Naries and Okiep were collected in South Africa (Springbok: 29° 41′ 37.1″ S, 17° 53′ 01.9″ E, Naries: 29° 44′ 52.2″ S, 17° 31′ 39.4″ E, Okiep: 29° 13′ 36.4″ S, 17° 40′ 20.5″ E), stratified with smoke primer (Cape Seed Primer, Fine Bush People, Cape Town, SA) for 24 h, and germinated in 9 × 9 cm pots with two-thirds compost (Levington's M3, Evergreen Garden Care, Nantgarw, Cardiff, United Kingdom) and one-third horticultural sand (Westland Horticulture, Dungannon, County Tyrone, United Kingdom). Plants were kept in a climate chamber (16 h illumination at 150 µmol/m^2/s, 20 °C, 60% relative humidity) with watering every 2 to 3 d and no fertilization and propagated monthly via cuttings.

HMW-DNA Extraction and Sequencing

To enable comparison across studies, the two focal *G. diffusa* genotypes previously characterized in Kellenberger et al. (2023), Springbok genotype 8 and Naries genotype 13, as well as the previously uncharacterized *G. diffusa* Okiep genotype 1, were chosen for sequencing. Young, developing leaves of the three *G. diffusa* genotypes were collected, divided into 80 mg samples, put into 2 mL Eppendorf tubes with a 4 mm glass ball added, and flash frozen in liquid N_2. Samples were ground frozen in a TissueLyser II (Qiagen, NL) and HMW-DNA was extracted individually from each sample according to the description for *G. diffusa* in Russo et al. (2022). Two samples per morphotype were pooled and DNA quantity was measured with a Qubit dsDNA BR Assay Kit on a Qubit 2.0 fluorometer (Thermo Fisher Scientific, United States), DNA purity was assessed on a NanoDrop 2000 (Thermo Fisher Scientific, United States), and DNA integrity was determined on a Femto Pulse system (Agilent Technologies, United States). The *G. diffusa* Springbok sample was sequenced on a PacBio Sequel 1 system (Pacific Biosciences of California, United States) to produce PacBio CLR reads to a depth of 40× and on an Illumina NovaSeq 6000 platform (Illumina, United States) to produce 150 bp paired-end Illumina reads to a depth of 100×. The *G. diffusa* Naries and Okiep samples were sequenced on a PacBio Sequel 2 system (Pacific Biosciences of California, United States) to produce PacBio HiFi reads to a depth of 20× each.

Assembly, Scaffolding and Annotation of the *G. diffusa* Springbok Genome

Prior to assembly, *k-mer* analysis was conducted with *kmc* v.3.0.0 and *GenomeScope* v.2.0 to estimate genome size, heterozygosity, and repeat content, and ploidy was confirmed with *SmudgePlot* v.0.2.3. Initial assembly of the PacBio CLR reads was conducted with *canu* v.2.0 using the parameters *genomeSize* = 2.3 g *correctedErrorRate* = 0.085 *corMhapSensitivity* = normal *corMhapFilterThreshold* = 0.0000000002 *corMhapOptions*="*–threshold* 0.80 *–num-hashes* 512 *–num-min-matches* 3 *–ordered-sketch-size* 1000 *–ordered-kmer-size* 14 *–min-olap-length* 2000 *–repeat-idf-scale* 50" *mhapMemory* = 60 g *mhapBlockSize* = 500 *ovlMerDistinct* = 0.975 *corOutCoverage* = 200 "*batOptions* = *-dg* 3 *-db* 3 *-dr* 1 *-ca* 500 *-cp* 50". The assembly was checked for contaminants with *BlobTools* v.2, assembly metrics were calculated with *assemblathon_stats.pl*, and gene body completeness was assessed with *BUSCO* v.4.0.6. The raw assembly was then long-read

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

polished three times with *gcpp* v.1.0.0 and the PacBio CLR reads until the *BUSCO* scores and mapping scores of the Illumina reads with *BWA-MEM* v.2.1 did not improve anymore. All contigs flagged as bubbles were removed, haplotypic duplication was purged from the polished assembly with *purge_dups* v.1.2.3 and minimap2 v.2.23 as aligner using the cutoffs *-l* 5 *-m* 67 *-u* 201, and the haploid assembly was sent along with 10 g flash frozen young leaves of *G. diffusa* Springbok to Dovetail Inc. (USA) for Omni-C scaffolding. Bubbles and purged scaffolds were added back to the scaffolded assembly, which was then gap-filled with *DENTIST* v.2.0.0 and long-read polished once again with *gcpp* v.2.0.0 and the PacBio CLR reads to fill further gaps. The Illumina reads were trimmed with the *BBDuk script* from *BBMap* v.38.69 with the parameters *ktrim* = r *k* = 23 *mink* = 11 *hdist* = 1 *tpe* = t *tbo* = t *qtrim* = r *trimq* = 28 *minlen* = 50 *trimpolyg* = 15 and then used for two rounds of short-read polishing with *NextPolish* v.1.3.1. After polishing, bubbles and purged contigs were again removed from the primary assembly. Scaffold 1 to 4 were renamed as chromosome1 to 4. Based on the results of the genome synteny analysis between the Springbok and Naries assembly (alignment to Naries scaffold ptg000011, see below and Fig. S4), scaffold 5 and 6, each representing one arm of chromosome 5, were later joined manually and renamed chromosome 5a and 5b. Repetitive elements were annotated with *EDTA* v.2.0.0 using the *–sensitive* parameter, and structural gene annotation was performed with *BRAKER* v.2.1.6 in RNA-seq mode. For this, the *EDTA* hard-masked assembly was converted to soft-masked, and the mRNA-seq reads of Springbok genotype 8 generated in Kellenberger et al.. (2023) were aligned to the Springbok assembly with *STAR* v.2.7.10a. *BRAKER* was then run using these files as input with the parameters *–AUGUSTUS_ab_initio, –softmasking –addUTR* = on. Functional gene annotation was done with the *EnTAP* v.0.10.8 pipeline. For this, asterisks (indicating stop codons) were removed from the *BRAKER* output file augustus.hints.aa, and the complete NCBI RefSeq, UniProt SwissProt, and Uniprot TrEMBL protein databases were downloaded. *EnTAP* was then run in *–runP* mode with these files as input, using both *EggNOG* and *InterProScan* with the *Pfam* and *Panther* databases for gene ontology analysis, and "Asteraceae" as relevant taxon for *Diamond* similarity search.

Assembly and Annotation of the *G. diffusa* Okiep and Naries Genomes

Prior to assembly, *k-mer* analysis was conducted with *kmc* v.3.0.0 and *GenomeScope* v.2.0 to estimate genome size, heterozygosity, and repeat content, and ploidy was confirmed with *SmudgePlot* v.0.2.3. The PacBio HiFi reads were assembled with *hifiasm* v.0.16.1 with default parameters and the in-built haplotig purging with *purge_dups*. The assemblies were checked for contaminants with *BlobTools* v.2, assembly metrics were calculated with *assemblathon_stats.pl*, and gene body completeness was assessed with *BUSCO* v.4.0.6. Repetitive elements were annotated with *EDTA* v.2.0.0 using the *–sensitive* parameter.

Genotyping-by-Sequencing (GBS) Data Generation

Four hybrid zones between pairs of six *G. diffusa* morphotypes were selected for GBS analysis: Naries-Springbok (29° 41′ 54.5″ S, 17° 42′ 43.6″ E), Springbok-Okiep (29° 34′ 28.7″ S 17° 49′ 25.7″ E), Soeb-Garies (30° 17′ 54.1″ S, 17° 31′ 23.3″ E), and Garies-Calendula (30° 26′ 24.0″ S, 17° 56′ 27.6″ E). An approximately linear transect extending from the hybrid zone center to adjacent sites of pure individuals from each parental morphotype was established through each hybrid zone. Along each transect, 12 sampling sites were established. Four sites were placed equidistantly across the zone containing putative hybrid phenotypes, and an additional four sites were placed in populations of parental phenotypes on either end, extending from the edge of the hybrid zone to approximately 10 km from it. At each site, open flowers of 12 *G. diffusa* plants were photographed, and young leaves were collected and desiccated in paper envelopes filled with silica gel. Dried leaf samples were ground into a fine powder using a TissueLyser II (Qiagen, NL). DNA was extracted from the powdered leaves following a modified cetyltrimethylammonium bromide (CTAB) protocol. Briefly, 500 μL of CTAB buffer was added to approximately 20 mg of leaf powder, vortexed, and incubated at 55 °C for 1 h. Subsequently, 5 μL of RNase A (New England Biolabs, United States) was added, and the samples were incubated for 30 min at 37 °C. Each sample was mixed with 500 μL of chloroform: isoamyl alcohol (24:1), vortexed, and centrifuged for 10 min at 12,000 × *g*. The aqueous phase was carefully collected, and its volume was estimated. An equal volume of chloroform: isoamyl alcohol (24:1) was added, followed by vortexing and centrifugation. The aqueous phase was once again retained. Based on the estimated aqueous phase volume, 0.08 volumes of ammonium acetate and 0.58 volumes of isopropanol were added. The samples were incubated at −20 °C for 1 h. After centrifuging for 15 min at 12,000 × *g*, the resulting pellet was washed sequentially with 70% (v/v) ethanol and 100% (v/v) ethanol. Finally, the DNA was resuspended in 20 μL of sterile water, and extracts were stored at

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

−20 °C. Some samples had to be discarded due to low DNA quality, resulting in total 119 samples for Naries-Springbok, 126 samples for Springbok-Okiep, 125 samples for Soeb-Garies, and 125 samples for Garies-Calendula.

The DNA samples were sent to the Beijing Genomics Institute (PRC) for GBS library preparation and sequencing. Library preparation followed the protocol of Elshire et al. (2011). The restriction enzyme *Ape*KI (G|CWGC) was selected for genome digestion due to its nature as a relatively rare cutter, which allowed for sufficient sequencing depth given the relatively large genome size of *G. diffusa*. Individuals were multiplexed using unique barcodes, and 150 bp paired-end reads were produced on six lanes of an Illumina HiSeq X Ten sequencer (Illumina, United States). A total of 96 individuals were multiplexed per lane.

SNP Calling, Admixture Proportion and F_{ST} Analysis

Raw GBS reads were trimmed with *Trim Galore!* v.0.6.6 and the settings *–paired -q* 20 *–fastqc –length* 90, demultiplexed with the *process_radtags* script from *stacks* v.2.54 and parameters *-e* apeKI *–inline_null -c -r -t* 90, and aligned to the Springbok assembly with *BWA-MEM2* v.2.1 and default parameters. Genotype likelihoods were generated and variants called with the *mpileup* and *call* functions of *BCFtools* v.1.15.1 and the settings *-d* 1000 *—ignore-RG.* Called variants were filtered with *VCFtools* v.0.1.15: All indels as well as SNPs with a minor allele frequency of less than 0.05, more than 20% missing data, phred quality of less than 20, depth smaller than 10, and greater than 200 were removed (*–remove-indels –maf* 0.05 *–max-missing* 0.8 *–minGQ* 20 *–minDP* 10 *–maxDP* 200). The filtered SNPs from the six pseudochromosomes were phased with *Beagle* v.5.4 and converted from vcf to fastSTRUCTURE format with *PGDSpider* v.2.1.1.5, and the ancestry of individuals was estimated separately for each hybrid zone with *fastSTRUCTURE* v. 1.0. Individuals with a *Q*-value ≥ 0.99999 for either morphotype (ie more than 99.999% of the SNPs originating from one morphotype) were considered pure. Global F_{ST} values were calculated from the filtered and phased SNPs with *VCFtools* v.0.1.15 for pure individuals of each hybrid zone, and sliding window F_{ST} values were calculated for each pseudochromosome with the script *popgenWindows.py* from *genomics_general* using the parameters *-w* 5000 *-m* 5 for 5 Kbp, and *-w* 5000000 *-m* 5 for 5 Mbp windows, respectively. Since inverted chromosomal blocks in high LD influence the identification of outlier windows, we did LD-based SNP pruning in *R* v.4.3.2 with the *R* library *SNPRelate* v.1.36.1 to remove all SNPs in LD > 0.4. Per SNP F_{ST} values were calculated from the pruned datasets with the script *popgenWindows.py* from *genomics_general* using the parameter *-w* 1, and the F_{ST} outlier threshold was set at the 99th percentile of the pruned F_{ST} value distribution.

Principal Component Analysis (PCA)

The filtered and phased GBS SNPs were subset into one vcf file per pseudochromosome and converted to bcf format with *BCFtools* v.1.15.1. Sliding window PCA was conducted for each pseudochromosome in *R* v.4.3.2 using the *R* library *lostruct* v.0.0.0.9000 with a window size of 100 SNPs.

For local PCA, approximate breakpoints of the inversions were determined based on sliding window F_{ST} values. The script *popgenWindows.py* from *genomics_general* was run with the parameters *-w* 50000 *-m* 5, and breakpoints were defined as boundaries of blocks of F_{ST} values exceeding the F_{ST} outlier threshold in the corresponding chromosomal regions (see above). Based on these inversion breakpoint estimations, the bcf files were split into inverted and adjacent uninverted segments of equal length with *BCFtools* v.1.15.1 and the setting *-r* CHR5b:48050000-91450000 for the inverted segment on SpOk CHR5b, *-r* CHR5b:5750000-49150000 for the un-inverted segment on SpOk CHR5b, *-r* CHR1:0-39700000 for the inverted segment on SoGa CHR1, *-r* CHR1:39700000-79400000 for the un-inverted segment on SoGa CHR1, *-r* CHR3:105200000-164900000 for the inverted segment on SoGa CHR3, *-r* CHR3:45500000-105200000 for the un-inverted segment on SoGa CHR3, *-r* CHR5:49100000-100000000 for the inverted segment on SoGa CHR5, *-r* CHR5:0-49100000 for the un-inverted segment on SoGa CHR5, *-r* CHR3:157500000-195750000 for the inverted segment on CaGa CHR3, and *-r* CHR3:119250000-157500000 for the un-inverted segment on CaGa CHR3. Local PCA was then conducted on these segments with *PLINK* v.1.90 and parameters *–pca –allow-extra-chr*.

For local sliding window PCA, the vcf files were filtered with *BCFtools* v.1.15.1 and parameters *-m* 2 *-M* 2 *-v* snps to keep only biallelic SNPs and subset into one vcf file per pseudochromosome. Local sliding window PCA was conducted with *WinPCA* v. 1.1. and parameters *-w* 10000000 *-i* 100000 corresponding to a window size of 10 Mbp and a step size of 100 Kbp.

Genome-wide Association Study (GWAS)

The filtered and phased GBS SNPs were converted from vcf to *PLINK* binary ped format with *PLINK* v.1.90. To test SNP association with petal color, the SNP dataset was subset to keep individuals from the SoGa hybrid zone

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

($n = 125$). Petal color was visually phenotyped from the standardized photographs (yellow = 0, yellow-orange = 1 and orange = 2) and added to the .fam files. *Gemma* v.0.98.5 with parameter *-gk* 1 was then used to compute the relatedness matrix, and a univariate MML model was fitted with parameter *-lmm* 4. Association between SNPs and petal color was assessed with a likelihood ratio test (*P*-value threshold 5×10^{-8}). GWAS of other hybrid zones was hampered by lack of floral traits unambiguously scorable from the photographs besides low SNP density and insufficient number of hybrids.

Genome Synteny and Structural Variation Analysis

For genome synteny analysis, the Okiep and Naries assemblies were aligned to the Springbok assembly with *minimap2* v2.23 and the parameter *-x* asm20. Dot plots were produced from the alignment files with *D-Genies* v.1.2.0. For structural variation analysis, the Okiep and Naries PacBio HiFi reads were aligned to the Springbok assembly with *minimap2* v.2.23 and the parameters *-ax* map-hifi *–MD*. *Sniffles2* v.2.0.6 was then run with default parameters using the alignment files as input. The identified structural variants were filtered with *Bcftools* v.1.15.1 similarly to (Weissensteiner et al. 2020). Structural variants larger than 1 Mbp, with a minor allele frequency of less than 0.3, with fewer than 5 supporting reads, imprecise breakpoints and/or overlap with assembly gaps were excluded (*-i* 'AF > 0.3 && IMPRECISE=0 && SVLEN<1000000 && SVLEN>-1000000 && SUPPORT>4').

To test for the presence of putative petal spot genes on the inverted segment of chromosome 5, the combined reference transcriptome of developing florets of Springbok and Naries from (Kellenberger et al. 2023) was aligned to the scaffolded Springbok assembly using *minimap2* v2.23 with parameters -ax splice:hq -uf –secondary=no. Transcripts aligning to the inversion (Scaffold_5:48050000-91450000) were compared to the list of transcripts differentially expressed between spotted and unspotted Springbok florets from (Kellenberger et al. 2023).

LTR-RT Insertion Analysis

To estimate LTR-RT age, the identity score (S) between the flanking 3′ and 5′-LTRs of intact LTR-RTs was extracted from the *EDTA* results. LTR-RT insertion time was then estimated as $(1–S)/(2 \times \mu)$, using the default nucleotide substitution rate of $\mu = 1.3 \times 10^{-8}$ substitutions per site per year (Ma and Bennetzen 2004). To identify LTR-RTs present in Springbok but absent in Naries, the Naries assembly was aligned to the Springbok assembly with *minimap2* v.2.23 and the parameter *-x* asm20, alignment gaps were extracted with *bedtools genomecov* and intersected with the *EDTA* output file of the Springbok assembly with *bedtools intersect -wa -u -f* 0.9 from *Bedtools* v.2.20.1 (ie at least 90% of the transposon has to be absent in Naries). The filtered *EDTA* file was then intersected with the *BRAKER* output with *bedtools intersect -wa -u* from *Bedtools* v.2.20.1 to find Springbok LTR-RTs within or adjacent to annotated genes (overlap ≥ 1 bp). To identify Springbok LTR-RTs in genes underlying sexual deception, transcripts differentially expressed between spotted and unspotted Springbok petals were obtained from (Kellenberger et al. 2023) and mapped to the Springbok assembly with *minimap2* v.2.23 and the parameter *-ax* splice. The output was then intersected with the filtered TE file using *bedtools intersect -wa -u* from *Bedtools* v.2.20.1 (overlap ≥ 1 bp).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Matthew Dorling and Emma Jackson for plant rearing and technical support; the late Prof. Jan Suda for flow cytometry work; Dr. Alessia Russo, Bettina Fischer, and Dr. Matthew Naish for help with HMW-DNA extraction; Dr. Etienne Bucher, Prof. Ian Henderson, Dr. Hamish Symington, Dr. Udhaya Ponraj, Dr. Róisín Fattorini, Farahnoz Khojayori, and Dr. Elinor Jax for helpful discussions; and the Northern Cape Department of Environment and Nature Conservation for issuing a collection permit to A.G.E. (FLORA 0057/2018).

Author Contributions

All authors conceived and designed this study. A.G.E. collected *G. diffusa* seeds. R.T.K. cultivated the *G. diffusa* plants. A.G.E. contributed the flow cytometry data. R.T.K. performed the HMW-DNA extraction, genome assemblies, and annotation. B.D., R.T.K., and A.G.E. did the hybrid zone phenotyping. B.D. conducted the hybrid zone DNA extraction and GBS data generation. R.T.K. performed the genetic and transposable element analyses. R.T.K. and J.I.M. did the genome synteny analyses. R.T.K., J.I.M., and B.D. conducted the hybrid zone analyses. R.T.K. wrote the original draft of the manuscript. All authors discussed and approved the final manuscript. B.J.G. and A.G.E. provided resources for this study. R.T.K., B.J.G., and A.G.E. acquired funding for this project.

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

Funding

This work was supported by a Swiss National Science Foundation (SNSF) Postdoc Mobility grant to R.T.K. (fellowship P400PB_194374), an Isaac Newton Trust grant to R.T.K. and B.J.G., a Natural Environment Research Council (NERC) grant to B.J.G. and A.G.E. (NE/P011764/1), and a Biotechnology and Biological Sciences Research Council (BBSRC) grant to B.J.G. (BB/V000314/1).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

All raw sequencing data (PacBio, Illumina, and GBS reads) and the assembled genomes are deposited on NCBI under BioProject number PRJNA1181622. The mRNA-seq data used for genome annotation is deposited on NCBI SRA under BioProject number PRJNA910570. All other datasets, including transposon and gene annotations, called genomic variants and phenotypic data from *G. diffusa* hybrid zones are deposited on Figshare under https://doi.org/10.6084/m9.figshare.27678051. All custom scripts are available from the lead contact upon request.

Literature Cited

Agren L, Kullenberg B, Sensenbaugh T. Congruences in pilosity between three species of *Ophrys* (Orchidaceae) and their hymenopteran pollinators. Nova Acta Regiae Soc Sci Upsal (Ser V:C). 1983:3:15–25.

Barrett SC. The evolution of plant reproductive systems: how often are transitions irreversible? Proc R Soc B Biol Sci. 2013:280:20130913. https://doi.org/10.1098/rspb.2013.0913.

Bohman B, Flematti GR, Barrow RA, Pichersky E, Peakall R. Pollination by sexual deception—it takes chemistry to work. Curr Opin Plant Biol. 2016:32:37–46. https://doi.org/10.1016/j.pbi.2016.06.004.

Cai L, et al. Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. (Rafflesiaceae). Curr Biol. 2021:31:1002–1011. https://doi.org/10.1016/j.cub.2020.12.045.

Charlesworth B, Charlesworth D. A model for the evolution of dioecy and gynodioecy. Am Nat. 1978:112:975–997. https://doi.org/10.1086/283342.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with *hifiasm*. Nat Methods. 2021:18:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Correvon H, Pouyanne A. Un curieux cas de mimétisme chez les Ophrydées. J Soc Nat Hortic. France. 1916:29–47.

Darwin C. The various contrivances by which orchids are fertilised by insects. John Murray; 1877.

de Jager ML, Ellis AG. Gender-specific pollinator preference for floral traits. Funct Ecol. 2012:26:1197–1204. https://doi.org/10.1111/j.1365-2435.2012.02028.x.

de Jager ML, Ellis AG. The influence of pollinator phylogeography and mate preference on floral divergence in a sexually deceptive daisy. Evolution. 2013:67:1706–1714. https://doi.org/10.1111/evo.12070.

de Jager ML, Ellis AG. Costs of deception and learned resistance in deceptive interactions. Proc R Soc B Biol Sci. 2014:281:20132861. https://doi.org/10.1098/rspb.2013.2861.

de Jager ML, Peakall R. Does morphology matter? An explicit assessment of floral morphology in sexual deception. Funct Ecol. 2016:30:537–546. https://doi.org/10.1111/1365-2435.12517.

de Jager ML, Willis-Jones E, Critchley S, Glover BJ. The impact of floral spot and ring markings on pollinator foraging dynamics. Evol Ecol. 2017:31:193–204. https://doi.org/10.1007/s10682-016-9852-5.

Delahaie B, et al. 2022. The phylogenetic history of the *Gorteria diffusa* radiation sheds light on the origins of plant sexual deception [preprint]. bioRxiv 521170. https://doi.org/10.1101/2022.12.22.521170

Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc. 2007:2:2233–2244. https://doi.org/10.1038/nprot.2007.310.

Dou S, et al. The self-compatibility is acquired after polyploidization: a case study of *Brassica napus* self-incompatible trilinear hybrid breeding system. New Phytol. 2024:241:1690–1707. https://doi.org/10.1111/nph.19451.

Dubin MJ, Scheid OM, Becker C. Transposons: a blessing curse. Curr Opin Plant Biol. 2018:42:23–29. https://doi.org/10.1016/j.pbi.2018.01.003.

Ellis AG, et al. Floral trait variation and integration as a function of sexual deception in *Gorteria diffusa*. Philos Trans R Soc B Biol Sci. 2014:369:20130563. https://doi.org/10.1098/rstb.2013.0563.

Ellis AG, Johnson SD. The evolution of floral variation without pollinator shifts in *Gorteria diffusa* (Asteraceae). Am J Bot. 2009:96:793–801. https://doi.org/10.3732/ajb.0800222.

Ellis AG, Johnson SD. Floral mimicry enhances pollen export: the evolution of pollination by sexual deceit outside of the Orchidaceae. Am Nat. 2010:176:E143–E151. https://doi.org/10.1086/656487.

Ellis AG, Johnson SD. Lack of floral constancy by bee fly pollinators: implications for ethological isolation in an African daisy. Behav Ecol. 2012:23:729–734. https://doi.org/10.1093/beheco/ars019.

Elshire RJ, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011:6:e19379. https://doi.org/10.1371/journal.pone.0019379.

Fattorini R, et al. Complex petal spot formation in the Beetle Daisy (*Gorteria diffusa*) relies on spot-specific accumulation of malonylated anthocyanin regulated by paralogous *GdMYBSG6* transcription factors. New Phytol. 2024:243:240–257. https://doi.org/10.1111/nph.19804.

Fenster CB, Marten-Rodriguez S. Reproductive assurance and the evolution of pollination specialization. Int J Plant Sci. 2007:168:215–228. https://doi.org/10.1086/509647.

Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP. Chromosomal rearrangements and the genetics of reproductive barriers *in Mimulus* (monkey flowers). Evolution. 2013:67:2547–2560. https://doi.org/10.1111/evo.12154.

Fukushima K, et al. Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. Nat Ecol Evol. 2017:1:0059. https://doi.org/10.1038/s41559-016-0059.

Hermann K, et al. Tight genetic linkage of prezygotic barrier loci creates a multifunctional speciation island in *Petunia*. Curr

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

Biol. 2013:23:873–877. https://doi.org/10.1016/j.cub.2013.03.069.
Hoballah ME, et al. Single gene–mediated shift in pollinator attraction in *Petunia*. Plant Cell. 2007:19:779–790. https://doi.org/10.1105/tpc.106.048694.
Hoshino A, Johzuka-Hisatomi Y, Iida S. Gene duplication and mobile genetic elements in the morning glories. Gene. 2001:265:1–10. https://doi.org/10.1016/S0378-1119(01)00357-2.
Hu J, et al. Transposable elements cause the loss of self-incompatibility in *Citrus*. Plant Biotechnol J. 2024:22:1113–1131. https://doi.org/10.1111/pbi.14250.
Iida S, Morita Y, Choi J-D, Park K-I, Hoshino A. Genetics and epigenetics in flower pigmentation associated with transposable elements in morning glories. Adv Biophys. 2004:38:141–159. https://doi.org/10.1016/S0065-227X(04)80136-9.
Jersáková J, Johnson SD, Kindlmann P. Mechanisms and evolution of deceptive pollination in orchids. Biol Rev Camb Philos Soc. 2006:81:219–235. https://doi.org/10.1017/S1464793105006986.
Johnson S, Midgley J. Fly pollination of *Gorteria diffusa* (Asteraceae), and a possible mimetic function for dark spots on the capitulum. Am J Bot. 1997:84:429–436. https://doi.org/10.2307/2446018.
Kapheim KM, et al. Genomic signatures of evolutionary transitions from solitary to group living. Science. 2015:348:1139–1143. https://doi.org/10.1126/science.aaa4788.
Kellenberger RT, et al. Multiple gene co-options underlie the rapid evolution of sexually deceptive flowers in *Gorteria diffusa*. Curr Biol. 2023:33:1502–1512. https://doi.org/10.1016/j.cub.2023.03.003.
Koren S, et al. *Canu*: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017:27:722–736. https://doi.org/10.1101/gr.215087.116.
Kuo Y-T, Chao Y-T, Chen W-C, Shih M-C, Chang S-B. Segmental and tandem chromosome duplications led to divergent evolution of the *chalcone synthase* gene family in *Phalaenopsis* orchids. Ann Bot. 2019:123:69–77. https://doi.org/10.1093/aob/mcy136.
Lee S-I, Kim N-S. Transposable elements and genome size variations in plants. Genomics Inform. 2014:12:87. https://doi.org/10.5808/GI.2014.12.3.87.
Leitch I, et al. Genome size diversity in orchids: consequences and evolution. Ann Bot. 2009:104:469–481. https://doi.org/10.1093/aob/mcp003.
Li J, et al. Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. Nat Plants. 2016:2:1–7. https://doi.org/10.1038/nplants.2016.188.
Liang M, et al. Taxon-specific, phased siRNAs underlie a speciation locus in monkeyflowers. Science. 2023:379:576–582. https://doi.org/10.1126/science.adf1323.
Looney B, et al. Evolutionary transition to the ectomycorrhizal habit in the genomes of a hyperdiverse lineage of mushroom-forming fungi. New Phytol. 2022:233:2294–2309. https://doi.org/10.1111/nph.17892.
Ma J, Bennetzen JL. Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci U S A. 2004:101:12404–-12410. https://doi.org/10.1073/pnas.0403715101.
Mérot C, et al. Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. Mol Biol Evol. 2021:38:3953–3971. https://doi.org/10.1093/molbev/msab143.
Ou S, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019:20:1–18. https://doi.org/10.1186/s13059-019-1905-y.
Peakall R. Pollination by sexual deception. Curr Biol. 2023:33:R489–R496. https://doi.org/10.1016/j.cub.2023.02.066.
Russo A, et al. Low-input high-molecular-weight DNA extraction for long-read sequencing from plants of diverse families. Front Plant Sci. 2022:13:1494. https://doi.org/10.3389/fpls.2022.883897.
Russo A, et al. Genome of the early spider-orchid *Ophrys sphegodes* provides insights into sexual deception and pollinator adaptation. Nat Commun. 2024:15:6308. https://doi.org/10.1038/s41467-024-50622-4.
Schiestl FP, et al. Orchid pollination by sexual swindle. Nature. 1999:399:421–421. https://doi.org/10.1038/20829.
Sedeek KE, et al. Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. PLoS One. 2013:8:e64621. https://doi.org/10.1371/journal.pone.0064621.
Sedeek KE, et al. Amino acid change in an orchid desaturase enables mimicry of the pollinator's sex pheromone. Curr Biol. 2016:26:1505–1511. https://doi.org/10.1016/j.cub.2016.04.018.
Segers FH, Kešnerová L, Kosoy M, Engel P. Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. ISME J. 2017:11:1232–1244. https://doi.org/10.1038/ismej.2016.201.
Siopa C, Dias MC, Castro M, Loureiro J, Castro S. Is selfing a reproductive assurance promoting polyploid establishment? Reduced fitness, leaky self-incompatibility and lower inbreeding depression in neotetraploids. Am J Bot. 2020:107:526–538. https://doi.org/10.1002/ajb2.1441.
Thomas MM. 2009. Evolution of the 'Beetle Daisy' petal spot. [PhD Thesis]. University of Cambridge.
Thomas MM, Rudall PJ, Ellis AG, Savolainen V, Glover BJ. Development of a complex floral trait: the pollinator-attracting petal spots of the beetle daisy, *Gorteria diffusa* (Asteraceae). Am J Bot. 2009:96:2184–2196. https://doi.org/10.3732/ajb.0900079.
Vereecken NJ, et al. Pre-adaptations and the evolution of pollination by sexual deception: Cope's rule of specialization revisited. Proc R Soc B Biol Sci. 2012:279:4786–4794. https://doi.org/10.1098/rspb.2012.1804.
Vignolini S, et al. The mirror crack'd: both pigment and structure contribute to the glossy blue appearance of the mirror orchid, *Ophrys speculum*. New Phytol. 2012:196:1038–1047. https://doi.org/10.1111/j.1469-8137.2012.04356.x.
Wan S, et al. A small chromosomal inversion mediated by MITE transposons confers cleistogamy in *Brassica napus*. Plant Physiol. 2022:190:1841–1853. https://doi.org/10.1093/plphys/kiac395.
Weissensteiner MH, et al. Discovery and population genomics of structural variation in a songbird genus. Nat Commun. 2020:11:3403. https://doi.org/10.1038/s41467-020-17195-4.
Wong DC, Amarasinghe R, Falara V, Pichersky E, Peakall R. Duplication and selection in *β-ketoacyl-ACP synthase* gene lineages in the sexually deceptive *Chiloglottis* (Orchidaceae). Ann Bot. 2019:123:1053–1066. https://doi.org/10.1093/aob/mcz013.
Xu H, et al. Complex sexual deception in an orchid is achieved by co-opting two independent biosynthetic pathways for pollinator attraction. Curr Biol. 2017:27:1867–1877. https://doi.org/10.1016/j.cub.2017.05.065.
Xu S, Schlüter PM. Modeling the two-locus architecture of divergent pollinator adaptation: how variation in SAD paralogs affects fitness and evolutionary divergence in sexually deceptive orchids. Ecol Evol. 2015:5:493–502. https://doi.org/10.1002/ece3.1378.
Xu S, Schlüter PM, Grossniklaus U, Schiestl FP. The genetic basis of pollinator adaptation in a sexually deceptive orchid. PLoS

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026

Genet. 2012:8:e1002889. https://doi.org/10.1371/journal.pgen.1002889.
Yan H, et al. MYB-mediated regulation of anthocyanin biosynthesis. Int J Mol Sci. 2021:22:3103. https://doi.org/10.3390/ijms22063103.
Zhang S, et al. Chromosome-scale assemblies of the male and female *Populus euphratica* genomes reveal the molecular basis of sex determination and sexual dimorphism. Commun Biol. 2022:5:1186. https://doi.org/10.1038/s42003-022-04145-7.
Zou Y, et al. Multi-integrated genomic data for *Passiflora foetida* provides insights into genome size evolution and floral development in *Passiflora*. Mol Hortic. 2023:3:27. https://doi.org/10.1186/s43897-023-00076-x.

Associate editor: Virginia Sanchez Puerta

Downloaded from academic.oup.com/gbe/article/18/3/evag065/8516517 by Bibliothek am Guisanplatz user on 09 November 2026