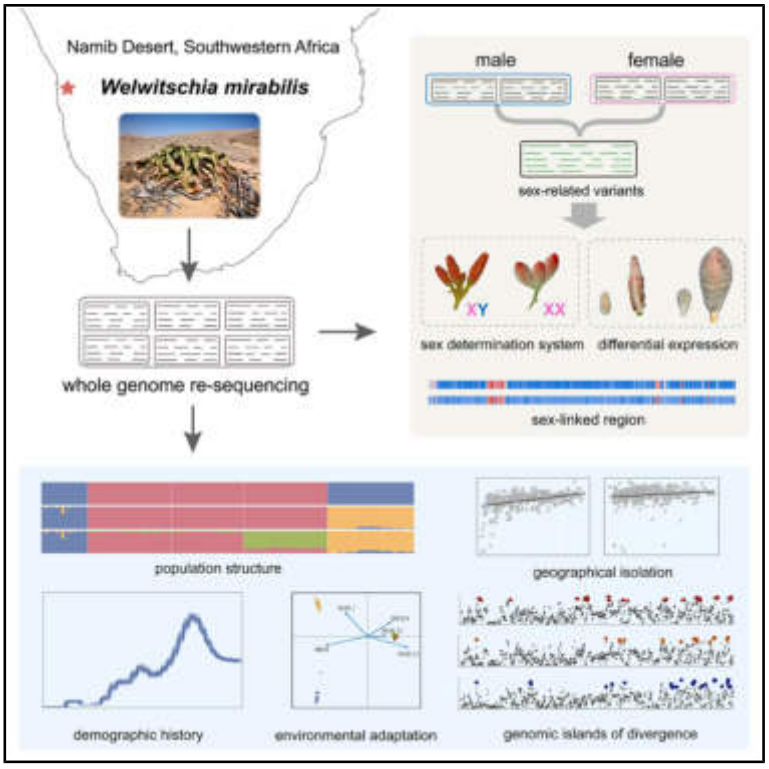
Cell Reports

Article

Genomic divergence and sex-linked region of *Welwitschia mirabilis* revealed by whole-genome re-sequencing

Graphical abstract



Highlights

- Significant genomic divergence exists among geographically close populations
- Ancient polymorphisms and divergence hitchhiking contribute to genomic islands
- *Welwitschia* has an XY-type sex system, with the sex-linked region on chromosome 6

Authors

Jing-Jing Sun (孙静静), Xiao-Mei Xia (夏小梅), Chang Liu (刘畅), ..., Jin-Hua Ran (冉进华), Qing-Feng Wang (王青锋), Xiao-Quan Wang (汪小全)

Correspondence

xiaoq_wang@ibcas.ac.cn

In brief

Sun et al. reveal the origin and maintenance of genomic divergence at a finer scale and explore the sex-determination mechanism in *Welwitschia mirabilis* through whole-genome re-sequencing of 130 individuals. This study provides insights into the evolutionary forces shaping genomic differentiation and the genetic basis of sex determination in gymnosperms.

Sun et al., 2026, Cell Reports *45*, 116836
January 27, 2026
https://doi.org/10.1016/j.celrep.2025.116836

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Genomic divergence and sex-linked region of *Welwitschia mirabilis* revealed by whole-genome re-sequencing

Jing-Jing Sun (孙静静),[1,2] Xiao-Mei Xia (夏小梅),[1,2] Chang Liu (刘畅),[1,2,3] Xi-Yu Wang (汪曦钰),[1,2,3] Sheng-Long Kan (阚胜龙),[1] Tao Wan (万涛),[4,5] Eugene Marais,[6] Gillian Maggs-Kölling,[6] Jin-Hua Ran (冉进华),[1,2,3] Qing-Feng Wang (王青锋),[4,5] and Xiao-Quan Wang (汪小全)[1,2,3,7,*]

[1]State Key Laboratory of Plant Diversity and Specialty Crops and Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[2]China National Botanical Garden, Beijing 100093, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]State Key Laboratory of Plant Diversity and Specialty Crops, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China
[5]Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China
[6]Gobabeb Namib Research Institute, Walvis Bay 13103, Namibia
[7]Lead contact
*Correspondence: xiaoq_wang@ibcas.ac.cn
https://doi.org/10.1016/j.celrep.2025.116836

SUMMARY

A central goal of population genomics is to unveil the evolutionary forces driving genomic divergence. Although genomic islands have been extensively studied in angiosperms, the patterns of genomic divergence remain poorly understood in gymnosperms. Here, by re-sequencing the whole genome of 130 individuals of *Welwitschia mirabilis*, we reveal significant genomic differentiation among its geographically close populations. Population genomic analyses suggest that geographical isolation and species characteristics play important roles in triggering and maintaining population divergence. We also identify multiple genomic islands of elevated divergence, which are primarily attributed to divergent sorting of ancient polymorphisms and divergence hitchhiking. Further genome-wide association analyses indicate that *Welwitschia* has an XY-type sex-determination system, with the sex-linked region located at 72.04–92.82 Mb on chromosome 6. Our findings are important for comprehending the evolutionary forces shaping genomic patterns of differentiation and the genetic basis of sex determination in gymnosperms.

INTRODUCTION

Understanding the patterns and processes of genomic divergence during speciation is a central issue in evolutionary biology.[1,2] The advances in high-throughput sequencing technologies have fundamentally changed our understanding of this issue.[2,3] Most studies have revealed a highly heterogeneous pattern of genomic divergence, with peaks of elevated differentiation widely spread over the genome.[4–8] Expanding upon the "genic model of speciation," the genomic regions with elevated divergence are resistant to gene flow,[2,3,9] most likely attributed to the presence of genes involved in local adaptation and/or reproductive isolation.[9–11] These regions, with divergence exceeding neutral expectations, are generally referred to as genomic islands, which may also arise in the absence of recent gene flow and can be formed by divergent sorting of ancient polymorphisms, recent selection at ecologically relevant loci, recent selective sweep, long-term linked selection, and/or genetic drift.[6,12–16] These processes are difficult to discriminate because they are not mutually exclusive and could cause similar divergence patterns in the genome.[3] Currently, the two measures, including F_{ST} (a relative measure of divergence) and D_{XY} (an absolute measure of divergence), are widely used to quantify genetic differentiation between populations, identify genomic islands of divergence, and distinguish evolutionary processes shaping genomic islands.[13,16,17] Although genomic patterns of divergence between diverging lineages have been frequently reported, almost all previous studies focused on angiosperms.[16,18–23] The origin and accumulation of genomic divergence in gymnosperms remain poorly understood.

Welwitschia mirabilis, the only extant species of the family Welwitschiaceae, belongs to the Gnetales, an ancient and enigmatic gymnosperm lineage that includes only two other genera, *Gnetum* (family Gnetaceae, ca. 40 species) and *Ephedra* (family Ephedraceae, ca. 54 species).[24,25] This lineage shows distinct morphological features, with members exhibiting unique reproductive structures and diverse ecological adaptations.[25] As a prime example, *W. mirabilis* is an extremely long-lived desert plant with a short, woody, unbranched stem and a massive woody concave crown bearing a single pair of opposite strap-shaped leaves that grow throughout the life of the plant.[24,25] This extraordinary plant is currently confined to the

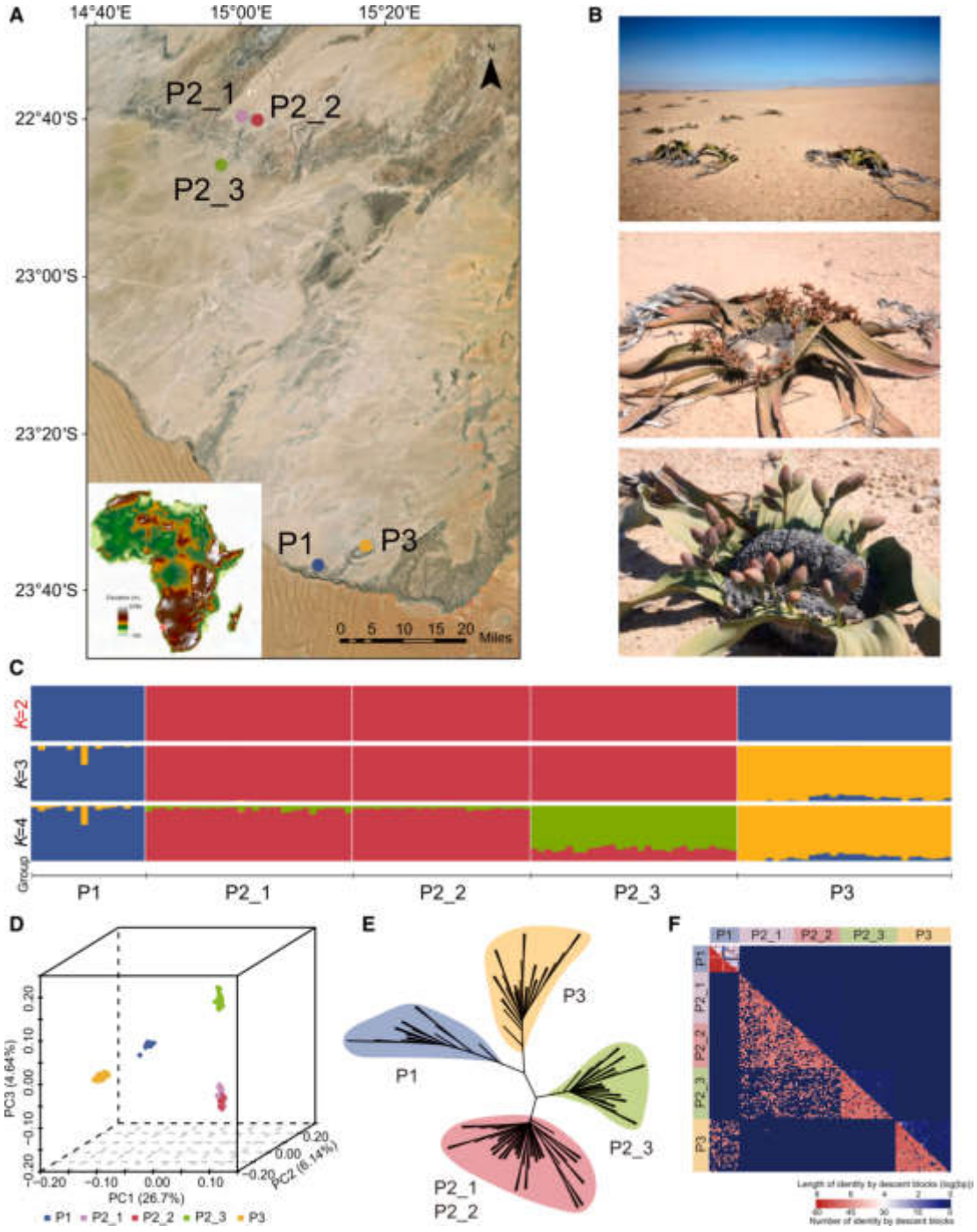


Figure 1. Sampling locations, habitat, morphology, and population genomic analysis of *Welwitschia*

(A) Geographical distribution of 5 *Welwitschia* populations. Color codes correspond to the results in the population genomic analysis. The map was downloaded and generated using ARCGIS 10.6 (http://www.esri.com/software/arcgis/arcgis-for-desktop).

(B) Coastal desert habitat (top) and male (middle) and female (bottom) individuals.

(C) Population genetic structure inferred by STRUCTURE. Each vertical bar represents an individual, with different colors representing the genetic ancestries. *K* represents the number of structure groups for each analysis, and the optimal *K* value is 2.

(legend continued on next page)

hyper-arid Namib desert in southwestern Africa, where its range is fragmented into isolated areas in a linear arrangement spanning a latitudinal range of over 1,000 km.[26] Within this range, *Welwitschia* occupies quite different environments, from temperate coasts to inland deserts, and has adapted to various habitat types, such as dry riverbeds, sandy plains, and rocky slopes.[26,27]

The morphological difference in the male strobili, allopatric distribution pattern, and molecular data all indicate the presence of two subspecies in *W. mirabilis*, with ssp. *namibiana* occurring in Namibia south of 18.7° and the typical ssp. *mirabilis* in Angola and the northwestern part of Namibia north of 18.7°.[26] Using six simple sequence repeat (SSR) markers, Jürgens et al.[26] have further revealed population differentiation in each of the two subspecies, with several different gene pools in isolated range fragments, which is consistent with a previous study on smaller-scale populations in the central Namib Desert using random amplified polymorphic DNA (RAPD) markers.[28] The significant genetic divergence among geographically isolated fragments of *W. mirabilis* can be largely attributed to the reduction or interruption of gene flow and may also be caused by environmental factors.[26] Intriguingly, this genetic differentiation persists even at a finer scale,[26,28] which prompts us to investigate the underlying mechanisms.

Moreover, *W. mirabilis* displays striking reproductive structures. It is functionally dioecious, with individuals producing either bisexual cones (fertile stamens and sterile ovules) or female cones (fertile ovules only).[25,29] The bisexual cones, which also occur in some species of *Gnetum* and *Ephedra*, differ significantly from the unisexual cones of other gymnosperms.[25,30,31] Although ongoing improvements in sequencing and analytical technologies provide new opportunities to reveal the molecular mechanisms of sex determination in gymnosperms, only preliminary insights have been gained for a few species, including *Ginkgo biloba*, *Cycas panzhihuaensis*, and *Taxus wallichiana*.[32–35] This limitation is particularly notable given the deep evolutionary divergence among major gymnosperm lineages.[34,36,37] Recent phylogenomic studies strongly support a sister relationship between Gnetales and Pinaceae.[34,37] Unlike the exclusively monoecious Pinaceae, the majority of Gnetales, including *Gnetum*, *Welwitschia*, and most *Ephedra* species, are dioecious.[25,38] Therefore, investigating whether the sex-determination system of *Welwitschia* differs from that of other studied gymnosperms is crucial for advancing our understanding of sex chromosome evolution in gymnosperms.

The chromosome-level genome sequence of *Welwitschia*, published in 2021, provides a foundational resource for evolutionary and genomic studies.[39] The assembled genome has a total size of 6.86 Gb, with a contig N50 of 1.48 Mb and a BUSCO completeness score of 83.47%, demonstrating the high quality of sequencing and assembly.[39] Based on this genome, it is particularly interesting to explore the mechanisms underlying the origin and maintenance of genetic variation among geographically closely related populations and to identify the sex-linked region (SLR) in *W. mirabilis*.

In this study, we re-sequenced the whole genomes of 130 individuals from five populations of *W. mirabilis* in the southernmost region of the Namib Desert. Firstly, we investigated population structure, gene flow, and the history of population divergence. Secondly, we evaluated the respective influence of environmental and geographical distances on inter-population genetic differentiation and pinpointed key environment-associated variants that drive local adaptation. Then, we identified differentiated genomic islands and assessed the contributions of different evolutionary processes in driving divergence. Finally, we tried to explore the sex-determination mechanism of *W. mirabilis* by identifying SLRs and differentially expressed genes (DEGs) between the sexes. Our study not only provides novel insights into the role of environmental and evolutionary processes in shaping the patterns of genetic variation in diverging lineages during speciation but is also important for comprehending the genetic basis of sex determination in gymnosperms.

RESULTS

Whole-genome re-sequencing and population genetic structure

In total, 13.27 Tb of sequencing data was generated across all individuals from the five populations (Figures 1A and 1B), with an average depth of 16.34× (Table S1). After mapping to the reference genome of *W. mirabilis*, the average mapping rate of the clean reads was 99.54% (Table S1).

Based on 3,442 independent genome-wide single-nucleotide polymorphisms (SNPs) from 129 individuals, the optimal number of genetic clusters in the STRUCTURE analysis was 2, with all sampled individuals clearly assigned to two specific groups (Figure 1C), one consisting of the 46 individuals from population (P)1 and P3 and the other comprising the 83 individuals from P2_1, P2_2, and P2_3. At $K = 3$, P1 and P3 stand out as distinct clusters, while at $K = 4$, P2_3 exhibits a distinct genetic component compared to P2_1 and P2_2. The results of principal-component analysis (PCA) were consistent with the STRUCTURE analysis. The first principal component (PC1; variance explained = 26.70%) separated P1 and P3 from the other three populations, while the second one (PC2; variance explained = 6.14%) separated P1 from P3. P2_3 and the other two populations were clearly differentiated by PC3 (variance explained = 4.64%), confirming them as genetically distinct (Figure 1D). The SNP-based unrooted phylogenetic tree included four distinct clades (Figure 1E). Individuals from P1, P2_3, and P3 formed separate clades, while P2_1 and P2_2 clustered together in a separate group. Moreover, the shared identical-by-descent haplotypes between P2_1 and P2_2 also indicated the common inheritance of many ancestral haplotypes between the two sides (Figure 1F). Given the nearly identical genetic background and negligible differentiation between P2_1

(D) Principal-component analysis based on genome-wide SNPs. The percentage of variation explained by each component is shown in parentheses.
(E) Unrooted neighbor-joining (NJ) tree based on SNP data.
(F) Estimated haplotype sharing between individuals. Heatmap colors represent the total length (above the diagonal) and total number (below the diagonal) of identity-by-descent blocks for each pairwise comparison.

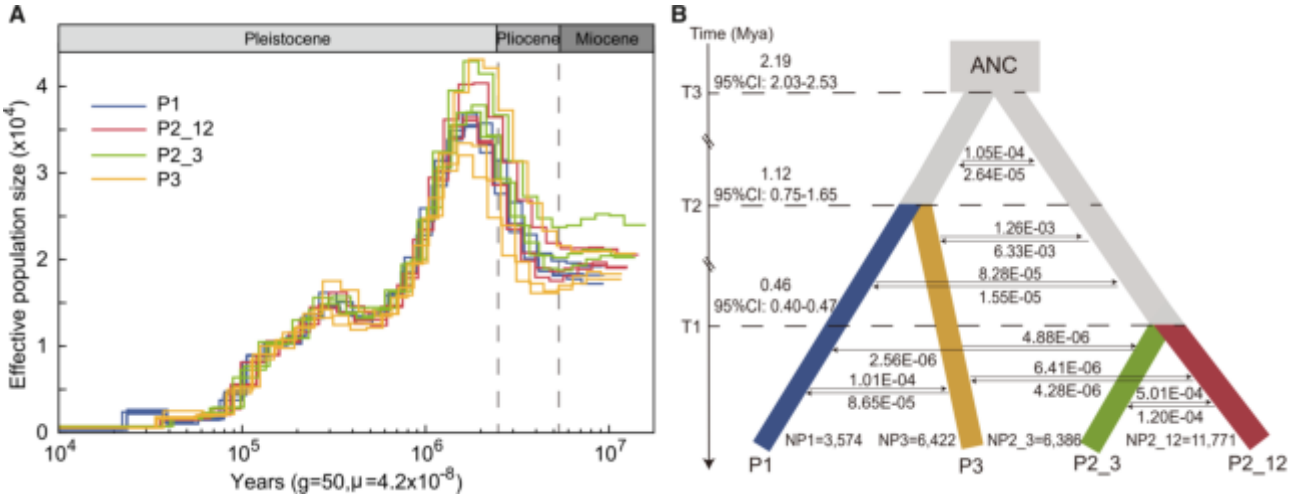


Figure 2. Demographic history of *Welwitschia*

(A) Changes in effective population size (*Ne*) through time inferred by the pairwise sequentially Markovian coalescent model. The timescale on the *x* axis is calculated assuming a mutation rate (μ) = 4.2×10^{-8} per site per generation and a generation time (g) = 50 years.

(B) Schematic of demographic scenario modeled using Fastsimcoal2. Estimates of gene flow between lineages are given in the migration fraction per generation. The point estimates and 95% confidence intervals of demographic parameters are shown in Table S3.

and P2_2, the two close populations were merged into a consolidated lineage, labeled as P2_12, in the following analyses.

Demographic history and gene flow

To explore the evolutionary history of the four lineages (P1, P2_12, P2_3, and P3), we first estimated the historical fluctuations in effective population size (*Ne*) using the pairwise sequential Markovian coalescence (PSMC) method. The results showed that all lineages experienced a dramatic decline in *Ne* approximately 1.8–0.5 million years ago (mya). Following this decline, a slight increase in *Ne* occurred between 0.45 and 0.3 mya, and a second decline started ~0.28 mya (Figure 2A). For P1, there was an expansion at 0.04 mya, followed by a sharp decrease at ~0.024 mya, showing the smallest historical *Ne*. Using the re-sequencing data, we further inferred divergence time and gene flow by the Fastsimcoal simulation analysis. The best-supported model (Figures 2B and S1; Tables S2 and S3) suggests that the common ancestor of P1 and P3 split from the common ancestor of P2_12 and P2_3 around 2.19 mya (95% confidence interval [CI]: 2.03–2.53 mya). P1 and P3 diverged 1.12 mya (95% CI: 0.75–1.65 mya), while P2_3 separated from P2_12 0.46 mya (95% CI: 0.40–0.47 mya). With increasing divergence time, gene flow between lineages gradually diminished. As indicated by BA3-SNPs analyses, we only detected a small number of contemporary migration events from P3 to P1 (0.033 ± 0.022) (Table S4). All of these findings suggest that both demographic history and rates of gene flow between lineages play important roles in shaping patterns of genomic divergence.

Effects of geographical isolation and environmental heterogeneity on genetic differentiation

To avoid potential correlation between geographical and environmental distances, we conducted partial Mantel tests to assess their respective contributions to genetic differentiation. The results showed significant patterns of both isolation by distance (IBD) and isolation by environment (IBE) in *Welwitschia* populations (Figure 3). Notably, geographical distance (partial Mantel's $r = 0.958$, $p < 0.001$) exerted a stronger influence on genetic differentiation than environmental distance (partial Mantel's $r = 0.621$, $p < 0.001$). We further conducted hierarchical partial Mantel tests across different geographical scales, and the results also showed a close correlation between genetic and geographical distances at finer scales (approximately 10 km; Figure S2). In addition, we evaluated the relative contributions of environmental factors in driving genetic differentiation. After filtering redundant variables, the final set consisted of 11 environmental variables, including six bioclimatic and five soil variables (Table S5; Figure S3). The result suggested that soil layers played a more important role in explaining genetic variation compared to climatic variables (Table S6). Among these, SOIL31 (topsoil sand fraction), SOIL1 (available water capacity [AWC] range), and SOIL21 (topsoil gypsum) emerged as the top three variables with the strongest impact.

Environmental adaptation-related variations

We used two complementary genotype-environment association (GEA) approaches, latent factor mixed model (LFMM) and redundancy analysis (RDA), to detect environment-associated genetic variants. The LFMM identified 1,528,526 SNPs that were significantly associated with one or more environmental variables. Notably, these SNPs were widely distributed across the genome and did not cluster in specific regions. For the RDA, we selected five low-correlated environmental predictors (Pearson's $r < 0.7$; Figure S4) and identified 292,211 environment-associated SNPs. The RDA results showed a clustering of individuals from different lineages consistent with the PCA, indicating the role of these environmental variables in driving local adaptation of *Welwitschia* (Figure 4).

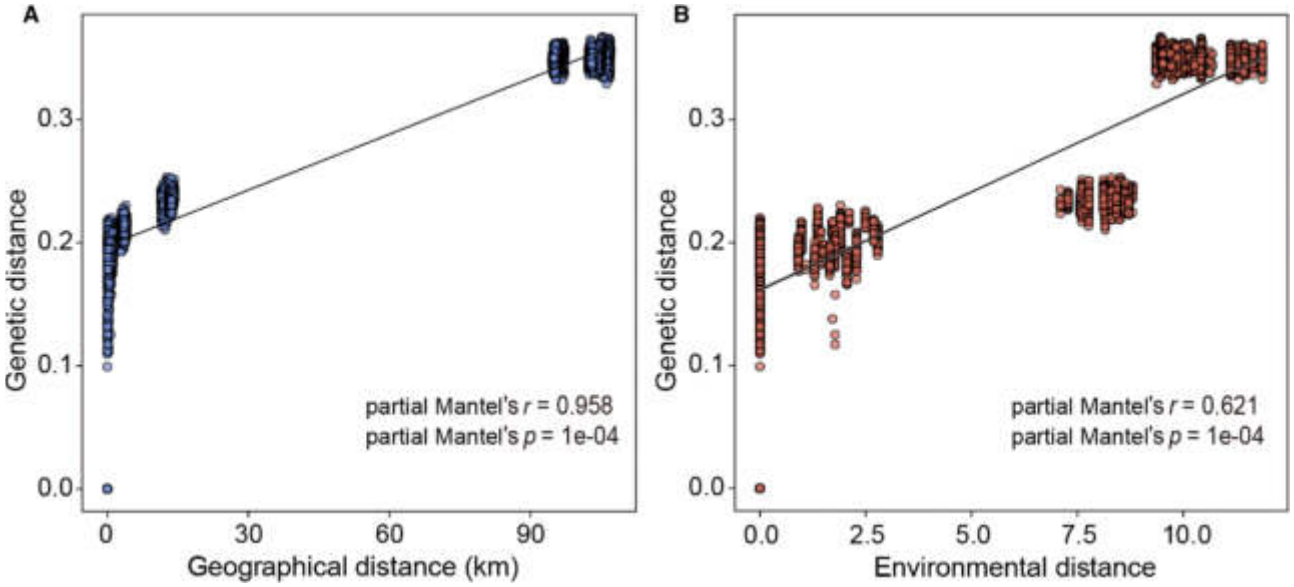


Figure 3. The relationship between genetic and geographical or environmental distances based on partial Mantel tests

(A) Isolation by distance (IBD) and (B) isolation by environment (IBE) for all individuals (n = 129). Genetic distances between all pairs of samples were measured using identity by state (IBS), which was calculated by PLINK v.1.90. Each partial Mantel test evaluated the independent effect of one factor (geographical or environmental distance) on genetic differentiation while controlling for the other.

A total of 162,996 non-redundant SNPs were obtained when the results of the two GEA analyses were combined. These shared variants were regarded as "core adaptive variants" for local adaptation, which were broadly distributed across the genome (Figure S5). Of the core adaptive variants, only 0.08% were non-synonymous, and 0.05% were synonymous mutations, with all remaining variants being non-coding (Table S7). These non-synonymous SNPs were associated with 86 genes. Gene Ontology (GO) enrichment analysis showed that these genes were significantly enriched in metabolic processes, cellular processes, and regulation of biological processes (Table S8). Finally, we found stronger genetic differentiation (F_{ST}) at these environment-associated SNPs compared to other SNPs (Figure S6), indicating that spatially varying selection has likely driven population differentiation at these local adaptive variants.

Population summary statistics

A set of genome-wide characteristics (π, Tajima's D, F_{ST}, D_{XY}, and D_f) was evaluated using an equal number of individuals from each lineage. Genetic diversity (π) ranged from 1.471 to 1.953 $\times$ 10^{-3} (Table S9), with P2_12 exhibiting the highest diversity and P1 the lowest. The positive Tajima's D values (1.411–1.682) may indicate a common historical demographic bottleneck event in *W. mirabilis* (Table S9). F_{ST} values between lineages ranged from 0.085 to 0.385 (Table S9). Consistent with expectations, F_{ST} values were notably higher between allopatric lineages than sympatric ones (Table S9). D_{XY} values showed the same pattern as F_{ST}, ranging from 2.111 to 2.914 $\times$ 10^{-3} (Table S9). Notably, the P1 lineage exhibited a higher degree of linkage disequilibrium (LD), a slower LD decay rate (Figure S7), and lower nucleotide diversity. These genomic patterns, together with the identity-by-descent analysis (Figure 1F), support the inference of a smaller *Ne* in P1, which may be primarily attributed to a long-term population bottleneck.

Genomic islands of divergence between lineages

The shape of the F_{ST} distribution varied between lineage pairs, which is consistent with the varying degrees of genetic differences between them (Table S9; Figure S8). Consistent with relatively recent population divergence, the F_{ST} values between P2_12 and P2_3 displayed an L-shaped distribution without any fixed differences (Table S9; Figure S8). Similarly, the F_{ST} values between P1 and P3 showed a comparable L-shaped distribution with only 127 fixed differences, which corresponds to a divergence scenario that includes contemporary gene flow (Tables S4 and S9; Figure S8).

We detected outlier windows for each pair of lineages based on their F_{ST} values and found that genetic differentiation between any two lineages was highly heterogeneous across the genome (Figure S9), irrespective of genetic and geographical distance. A total of 8,330 outlier windows with a size of 100 kb were identified (Table S10). Some of these outlier windows were shared between pairwise comparisons (Table S11), while there were no shared outlier windows among all lineages. Upon combining adjacent outlier windows, we obtained the final genomic islands and found that approximately half of these islands consisted of only one window (Table S10; Figure S10). In addition, these islands were characterized by reduced nucleotide diversity (π), more negative Tajima's D, and lower recombination rates, suggesting strong linked selection in these islands (Figure 5A; Table S12).

We further investigated the factors underlying the formation of genomic islands based on the D_{XY} values.[13,17,19] For all pairwise

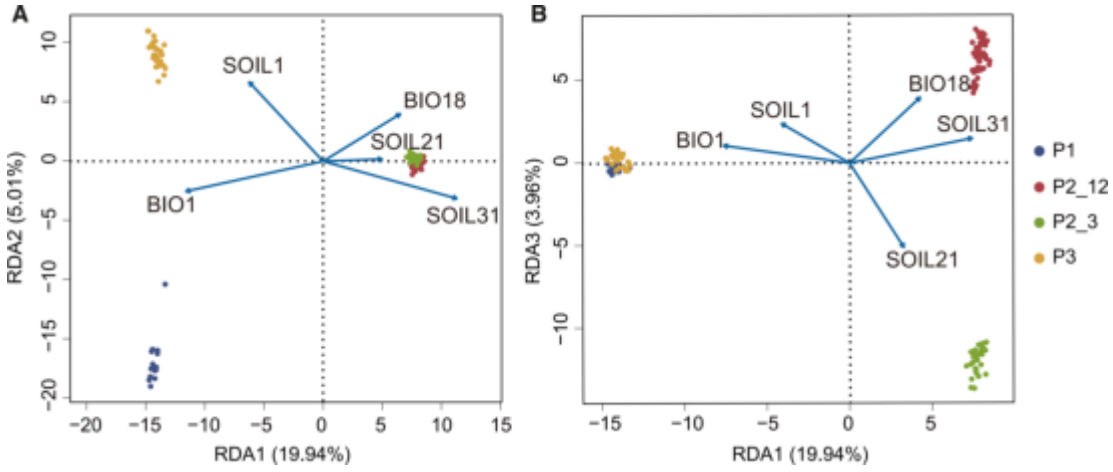


Figure 4. Redundancy analysis showing the relationships between the independent environmental parameters and the population structure of *Welwitschia*

(A) Redundancy analysis (RDA)1 and RDA2 axes and (B) RDA1 and RDA3 axes. Individuals are shown in colored points, and colors represent four lineages (P1, P2_12, P2_3, and P3). Blue vectors represent environmental variables (predictors).

comparisons, genomic islands exhibited significantly elevated D_{XY} compared with the genomic background (Table S12). This pattern was consistent in both allopatric and sympatric lineages, indicating that these diverged haplotypes in genomic islands likely originated from the ancestral populations. Although the number of genomic islands can be influenced by both gene flow and divergence time (Table S10), the widespread distribution of islands across the genome suggests that recent gene flow has not been a major factor in their formation (Figure S9). Instead, the strong positive correlation between D_{XY} and π implies the importance of ancestral polymorphism in island formation (Figure 5B). Additionally, we assessed whether environment-associated adaptive SNPs, identified by the LFMM and RDA, were enriched within the genomic islands. Strikingly, 30.2% (49,264/162,996) of these adaptive SNPs were located within 1,122 islands, showing a significant enrichment pattern (Fisher's exact test: odds ratio = 3.89, $p < 0.001$).

Identification of sex chromosome and SLRs

To elucidate the genetic mechanism of sex determination in *W. mirabilis*, we conducted genome-wide association studies (GWASs). In the SNP-based GWAS, we identified 184,297 SNPs that were significantly associated with sex phenotypes ($p < 1.06 \times 10^{-9}$; Figure 6A). Notably, 172,267 (93.47%) of these SNPs were clustered within specific regions on chromosome 6 (Figure S11A). To complement this approach, a k-mer-based GWAS was performed. We identified 898,946 k-mers that were successfully mapped to the reference genome, showing a significant association with sex ($p < 1.01 \times 10^{-11}$; Figure 6B). To minimize false positive signals caused by repetitive sequences and sequencing errors, we filtered out clusters with fewer than 20 k-mers in any 10 kb window, retaining 841,076 sex-related k-mers. Among these, 784,034 (93.22%) k-mers were located on chromosome 6 (Figures 6B and S11B), further supporting the pivotal role of this chromosome in sex determination. Additionally, chromosome 6 is the most differentiated between male and female *Welwitschia* genomes, with the highest fixation index (F_{ST}; Figure S12A) and the most differentiated nucleotide diversity (π; Figure S12B). All these results confirm that chromosome 6 is the sex chromosome of *W. mirabilis*. Furthermore, we identified SLRs, with the largest region spanning from 72.04 to 92.82 Mb (Figure 6C). Within these regions, F_{ST} values between males and females reach a maximum of 0.19 with an average of 0.11 (Figure S12A). Genotype analysis at these sex-associated loci revealed that 98.99% were homozygous in females, whereas 93.93% were heterozygous in males (Figure 6D), demonstrating that *Welwitschia* possesses an XY sex-determination system.

Transcriptome analysis of male and female cones

To further understand the genetic architecture of the SLRs in *W. mirabilis*, we analyzed the 157 genes within these regions, and 118 of them were functionally annotated (Table S13). Additionally, we identified 18 novel genes in the SLRs through transcriptome assembly, five of which were successfully annotated (Table S13). Among the 123 annotated genes, we identified several candidates potentially involved in reproductive development, such as *UFO* (*W. mirabilis.22100*) and *MSI1* (*MSTRG.16283*). Notably, no MADS-box genes were detected within the SLRs, although more than 30 MADS-box genes were identified elsewhere in the *Welwitschia* genome.

Moreover, we performed differential expression analysis between male and female cones at two developmental stages (M1 and M2 and F1 and F2; Figure S13). The result revealed widespread expression divergence between the sexes, with 4,000 significantly DEGs identified (Figures S14A and S14B; Table S14). Of these, 2,065 were upregulated in male cones and 1,935 in female cones (Figure S14A). GO enrichment analysis indicated that the genes upregulated in males were strongly associated with reproductive processes such as anther and stamen development, whereas those upregulated in females were significantly enriched in terms related to cell wall organization and secondary metabolism (Figure S14C). The transcriptional

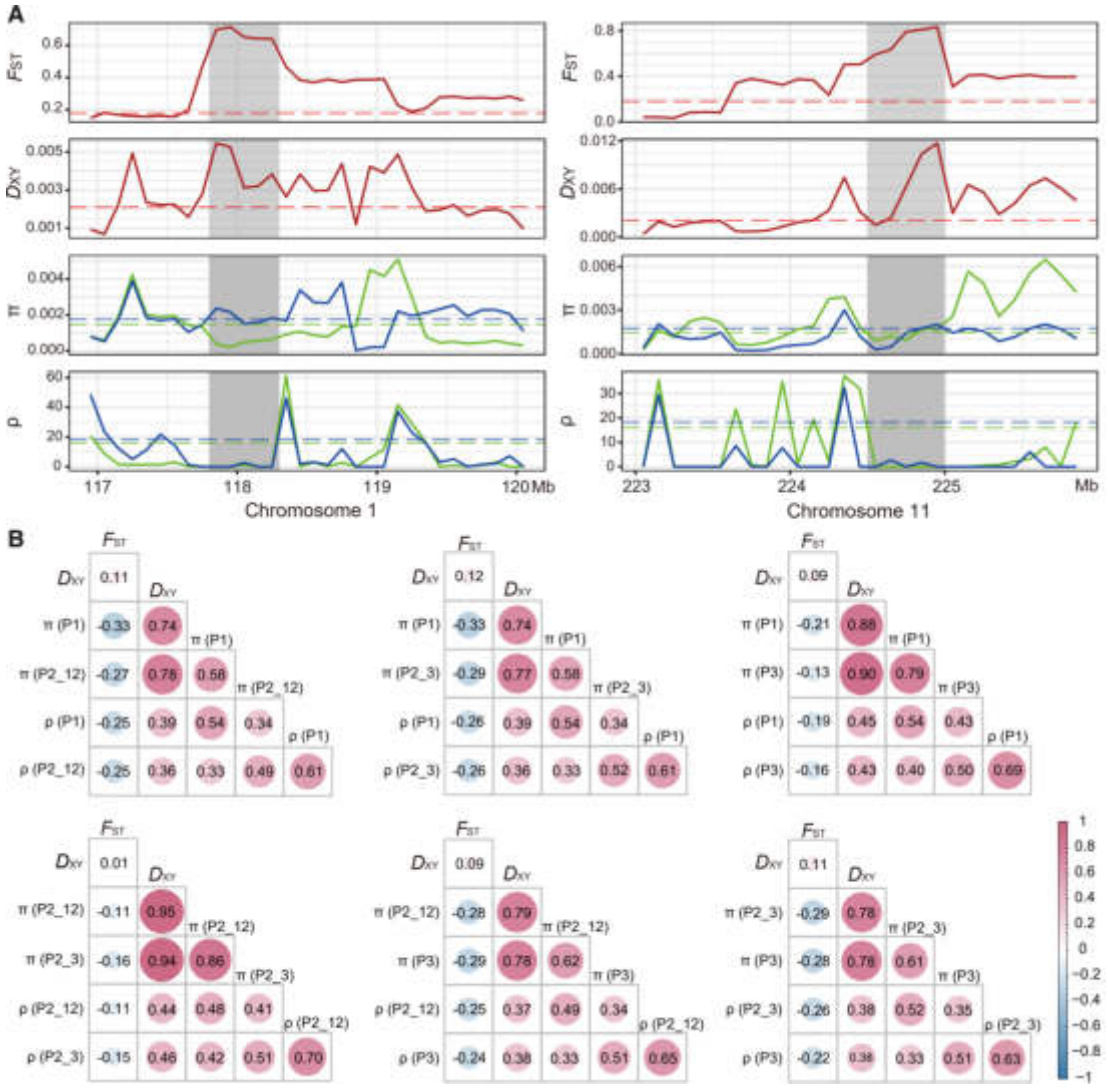


Figure 5. The heterogeneity of genomic divergence in four lineages

(A) The distribution of genomic parameters around two island examples (gray bar) between lineages P1 (green) and P3 (blue). The genetic divergence (F_{ST}), absolute divergence (D_{XY}), nucleotide diversity (π), and recombination rate (ρ) values are plotted using a 100 kb sliding window. Horizontal dashed lines represent the mean whole genome of corresponding values.

(B) Correlation of 100 kb window-based estimates of F_{ST}, D_{XY}, π, and ρ for all six pairwise comparisons. Pink circles indicate positive correlation, while blue ones indicate negative correlation. The color intensity and circle size are proportional to Spearman's correlation coefficient.

expression underlying male development was further highlighted by three key findings: first, we identified 454 male-biased genes and 98 female-biased genes based on their expression profiles (Figure S14A; Table S14); second, 10 MADS-box genes were significantly highly expressed in male cones, while only three in female cones (Figure S15); and third, the early stage of male cones (M1) had more upregulated DEGs than the later stage (M2) (Figure S14B). Together, these results suggest that male reproductive development involves more extensive and complex transcription processes, particularly during early stages.

We further analyzed the DEGs located within the SLRs. The result showed a remarkably small set of only 23 DEGs, with eight upregulated in male cones and 15 in female cones (Table S14). Among these, we identified one male-biased (*MSTRG.16683*) and one female-biased (*W. mirabilis.23857*) gene. Their specific genomic location and distinct expression patterns in cones make them noteworthy.

DISCUSSION

Origin and maintenance of genomic divergence in *W. mirabilis*

Elucidating the factors underlying the origin and maintenance of genetic variation among populations is a primary goal of population genomics. Unexpectedly, our present study reveals a great genomic divergence among five geographically closely

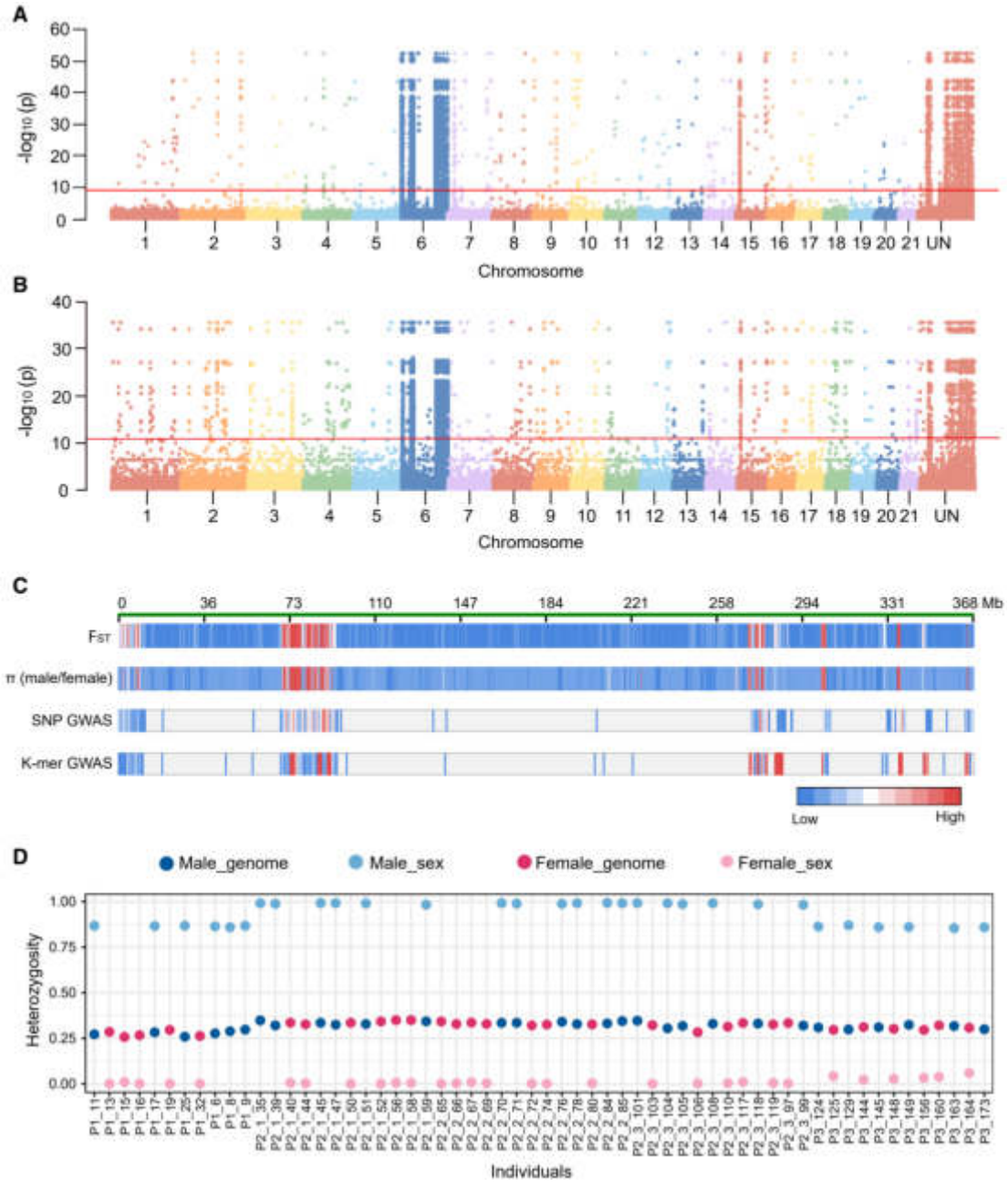


Figure 6. Sex-linked regions and sex-determining system of *W. mirabilis*

(A and B) Manhattan plot of SNP-based GWAS (A) and Manhattan plot of k-mer-based GWAS (B) of sex determination across 30 male and 30 female samples. The red horizontal line represents the Bonferroni-corrected threshold for genome-wide significance ($\alpha = 0.05$). *p* values were calculated from mixed linear model associations.

(C) Sex-linked regions on chromosome 6 were identified by integrating four lines of evidence: F_{ST}, π, and SNP-based and k-mer-based GWASs. The primary region spans 72.04–92.82 Mb (size = 20.78 Mb), with additional smaller regions detected.

(D) Heterozygosity of the genome level and 184,297 significant sex-associated SNPs in male and female samples.

related populations of *W. mirabilis* from the southernmost region of the Namib Desert (Figures 1C–1F), although a general strong population differentiation of this species was detected by previous studies using different molecular markers.[26,28]

Climate change and species traits have been implicated in shaping both past and current species distribution and genetic structure.[40–42] By integrating demographic history inference with paleoclimate evidence, we obtain a comprehensive

understanding of the evolutionary history of *Welwitschia*. During the Pliocene, a large-scale population expansion occurred in all lineages, which coincided with a period of relative global warmth, potentially indicating optimal growth conditions (Figure 2A).[43,44] Following this period, the intensification of the Benguela upwelling system and the onset of hyper-arid conditions likely triggered both a substantial decline in *Ne* and the divergence of lineages, which continued until ~0.5 mya (Figure 2).[45–47] Although this decline in *Ne* is consistent with a demographic bottleneck, an alternative explanation is the gradual divergence among lineages, which could generate spurious signals of population size change.[48–50] Therefore, the observed signal may reflect both demographic change and population subdivision. The population fragmentation led to a gradual decrease in gene flow, which further promoted genomic divergence among *Welwitschia* lineages. In addition, the finer-scale genetic structure is strongly influenced by species traits, such as the dispersal mode of pollen and seeds.[51] We found that there was limited contemporary gene flow among lineages, even between P2_12 and P2_3, which are distributed on opposite sides of the mountain and separated only by approximately 12 km (Table S4). This may suggest that insect pollination serves as the primary mode, and anemophily plays a minor role in *Welwitschia*. This inference is also supported by both field observations and pollen fossil records,[52–54] challenging the assumption that anemophily is the dominant mode of pollination in *W. mirabilis*.[29,55] Previous studies have suggested that pollination drops produced by both female and male cones of *Welwitschia* serve as rewards for insect pollinators.[25,53,56] Moreover, although the seeds of *Welwitschia* have wings that develop from the bracts, they are not windborne and generally fall under the mother plant due to their considerable weight, as confirmed by our field investigation.[28,55] The seed wings could increase buoyancy, aiding the dispersal by water, particularly during episodic rainfall.

In summary, the genomic divergence among populations of *Welwitschia* is primarily attributed to limited gene flow, a consequence of geographical isolation and species characteristics. The ongoing climate change significantly threatens the survival of *Welwitschia* populations, leading to increased adult mortality and smaller suitable habitats.[27,57–59] Under such circumstances, populations with smaller effective sizes are more vulnerable and prone to extinction in the future, e.g., P1. To address this challenge, conservation measures, such as artificial pollination or the establishment of supplemental seed banks, are needed to enhance the genetic connectivity and diversity.

Role of ancient polymorphisms in shaping genomic islands

A variety of evolutionary forces can shape genomic patterns of divergence during lineage differentiation.[13,16,22] Our study of *Welwitschia* shows that genetic divergence between lineages is highly heterogeneous, with genomic islands scattered across the genome (Figure S9). Several characteristics, including elevated F_{ST} and D_{XY} and reduced π, Tajima's *D*, and ρ, were observed in these islands (Figure 5A; Table S12), as reported in other plants.[6,7,16,60,61] The reduced π and ρ in genomic islands are generally associated with the hitchhiking effect, background selection, or selective sweeps, while elevated D_{XY} rejects the latter two possibilities.[13,17,62] Different from elevated F_{ST}, a characteristic of genomic islands, D_{XY} is elevated only in islands caused by differential gene flow between genomic regions or ancient balanced polymorphism but is unchanged or below average in recent ecological selection, ongoing background selection, or recurrent selective sweeps.[3,13,19] Therefore, in *Welwitschia*, the elevated D_{XY} suggests that differential gene flow between genomic regions or divergent sorting of ancient polymorphisms is the more likely primary driver of island formation, although environment-associated SNPs are significantly enriched in genomic islands. However, under the scenario of divergent selection with gene flow, the genomic islands tend to be clustered together,[63,64] which is inconsistent with our findings in *Welwitschia* as mentioned earlier. Our inference is further supported by the fact that the number of genomic islands decreases when gene change occurs between P1 and P3 (Tables S4 and S10), likely due to the homogenizing effect of gene flow. Thus, although the varying degrees of gene flow significantly influence genetic divergence,[63,65,66] the variation in gene flow is unlikely to be a major factor in generating genomic islands of divergence in *Welwitschia* lineages. Instead, we propose that a more plausible driver involves the divergent sorting of ancient polymorphisms. Evolutionary forces such as balancing selection can maintain highly differentiated haplotypes in an ancestral population. During population divergence, ecological selection for local adaptation or stronger genetic drift may lead to lineage sorting of those divergent haplotypes into the descendant populations unequally.[19,67,68] Together with divergence hitchhiking, these factors can result in an increase in both the number and size of genomic islands, ultimately generating the heterogeneous genomic patterns of divergence.

Genomic signatures of local adaptation

Local adaptation to diverse habitats is widespread in plant populations.[69] We found a strong association between genetic divergence and environmental variables, especially soil factors (Figure 3; Table S6). The important role of soil layers in explaining population genetic divergence is also supported by field observations and literature.[26] Actually, soil abiotic properties, such as nutritional status, water availability, and soil type, are the primary environmental factors that limit seedling establishment and drive soil-mediated local adaptation.[70–72]

In addition, we detected a wealth of genetic variants associated with environmental changes using the LFMM and RDA. Surprisingly, 99.87% of them were located in intergenic and intron regions (Table S7). This could indicate that regulatory elements located in non-coding regions exert a crucial effect on adaptive evolution by modifying gene expression.[73–75] Alternatively, it could be explained by the fact that the non-coding regions of the genome are generally less conserved than protein-encoding regions,[76,77] especially in recently diverged populations. Furthermore, we identified 86 environment-associated genes via non-synonymous SNPs. These genes are primarily enriched in metabolic process, cellular process, and regulation of biological process (Table S8), suggesting that they may play a key role in local adaptation and population divergence of *W. mirabilis*.

Sex-determination mechanism in *Welwitschia*

As an ancient, monotypic lineage in the enigmatic Gnetales,[37] *Welwitschia* is of significant value for studying plant sex-determination mechanisms. Our present study has successfully identified an XY system and a region from 72.04 to 92.82 Mb on chromosome 6 as the SLR in *W. mirabilis*. Nevertheless, we could not completely exclude the effect of several smaller regions on chromosome 6 in determining or aiding gender differentiation (Figure 6C). It is particularly intriguing that the MADS-box genes were not detected in the SLRs of *W. mirabilis*, given their crucial roles in sex determination for other gymnosperms.[32,34] A study on *C. panzhihuaensis* identified a MADS-box transcription factor within the male-specific region of the Y chromosome that is expressed exclusively in male cones.[34] Similarly, in *G. biloba*, most MADS-box genes are differentially expressed between the sexes, with four of them being specifically expressed in the staminate strobilus, implying an important role in sex determination.[32,78,79] These findings indicate a shared ancestral sex-determination mechanism controlled by MADS-box genes in the basal gymnosperms, including cycads and ginkgo.[34] The non-detection of MADS-box genes in the SLRs of *W. mirabilis* may suggest a unique evolutionary origin of sex chromosomes. This is plausible due to the deep evolutionary divergence and distinct reproductive organ features compared to those of cycads and ginkgo.[25,37] However, we cannot yet fully rule out the possibility that the higher proportion of repetitive sequences and severe recombination suppression in the SLR have impeded accurate assembly and annotation, potentially hindering the identification of such genes.[80–82] Nevertheless, the significant differential expression of MADS-box genes between male and female cones strongly implies that they are indeed involved in the development of reproductive organs of *W. mirabilis* (Figure S15). Our transcriptomic analysis identified only 23 DEGs within the SLRs, including two sexually biased genes. The male-biased gene (*MSTRG.16683*) is a homolog of *ENODL9*, which has been preliminarily linked to the shift from the vegetative to the reproductive stage and floral initiation via carbohydrate signaling in *Arabidopsis thaliana*, but further experimental evidence is required to support such a function.[83,84] The specific genomic location and sexually biased expression patterns suggest that these two genes may have an important role in sex differentiation, and it will be of interest to investigate whether they are involved in sex determination.

In addition, our transcriptomic analysis revealed distinct reproductive strategies of *Welwitschia*. Male development emerged as a more transcriptionally active and energy-consuming process, as evidenced by a greater number of male-biased genes, complex expression dynamics in early male development, and male-upregulated expression of key regulatory families such as MADS-box genes. In contrast, female-upregulated genes are primarily invested in structural stability and defensive functions, indicating a fundamental divergence in reproductive strategy between the sexes.

Limitations of the study

We acknowledge certain limitations in our study. Our efforts to assemble the male-specific region of the Y chromosome and identify sex-determining genes faced great challenges, primarily due to the limited number and length of the assembled contigs, which hindered gene prediction and annotation. The k-mer searching method widely used in angiosperms may be less effective for gymnosperms. Due to ultra-large genome size and high repetitive sequence content, previously assembled genomes of gymnosperms are relatively lower in quantity and quality compared to angiosperms.[85] The inherent characteristic of the SLR also makes it difficult to assemble using short-read sequencing technologies.[80–82] The successful assembly of the male-specific region in *Ginkgo* may have benefited from the sibling population samples and female reference genome used in analysis.[32] Although the cone samples in our study provide vital clues, the comprehensive understanding of sex determination will require sampling of more complete developmental stages, especially the primordium stage. However, *W. mirabilis* is endemic to the Namib Desert, making the collection of such material challenging. The cultivation of multiple individuals will be helpful to further studies on the mechanisms of sex determination in *Welwitschia*. Beyond this, a high-quality haplotype-resolved genome assembly is necessary for resolving these challenges and further elucidating sex chromosome evolution in *Welwitschia*.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xiao-Quan Wang (xiaoq_wang@ibcas.ac.cn).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- All raw re-sequencing and transcriptomic data in this study have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn/) under accession number PRJCA052889.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Wei-Ping Zhang (South China Botanical Garden, Chinese Academy of Sciences), Yu Cao (Huazhong Agricultural University), Hong-Yin Hu (Lanzhou University), and Lian Zhou (Institute of Botany, Chinese Academy of Sciences) for their assistance with data analyses. This study was supported by the Key Research Program of Frontier Sciences, CAS (QYZDJ-SSW-SMC027), and the Scientific Research Program of Sino-Africa Joint Research Center.

AUTHOR CONTRIBUTIONS

X.-Q.W. conceived and designed the study. S.-L.K., J.-H.R., X.-Q.W., E.M., and T.W. collected plant materials. J.-J.S. performed the experiments. J.-J.S., X.-M.X., C.L., X.-Y.W., and X.-Q.W. analyzed and interpreted the data. J.-J.S. and X.-Q.W. wrote the manuscript. X.-Q.W., Q.-F.W., and G.M.-K. coordinated the project. All authors read and approved the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Library construction and sequencing
  - Read mapping and variant calling
  - Population structure analysis
  - Inference of lineage demographic history and gene flow
  - Effects of geographical distances and heterogeneous environments on genetic structure
  - Identification of environment-associated genetic loci
  - Analyses of linkage disequilibrium, recombination, genomic diversity, and divergence
  - Identification of genomic islands of divergence
  - Identification of the sex-linked region
  - Gene annotation and differential expression analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116836.

Received: May 21, 2025
Revised: October 28, 2025
Accepted: December 9, 2025
Published: January 13, 2026

REFERENCES

1. Feder, J.L., Flaxman, S.M., Egan, S.P., Comeault, A.A., and Nosil, P. (2013). Geographic mode of speciation and genomic divergence. Annu. Rev. Ecol. Evol. Syst. *44*, 73–97. https://doi.org/10.1146/annurev-ecolsys-110512-135825.
2. Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.P., Bank, C., Brännström, A., et al. (2014). Genomics and the origin of species. Nat. Rev. Genet. *15*, 176–192. https://doi.org/10.1038/nrg3644.
3. Wolf, J.B.W., and Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. *18*, 87–100. https://doi.org/10.1038/nrg.2016.133.
4. Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J., and Turner, G.F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science *350*, 1493–1498. https://doi.org/10.1126/science.aac9927.
5. Kautt, A.F., Kratochwil, C.F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Härer, A., Hulsey, C.D., Franchini, P., et al. (2020). Contrasting signatures of genomic divergence during sympatric speciation. Nature *588*, 106–111. https://doi.org/10.1038/s41586-020-2845-0.
6. Hu, H., Yang, Y., Li, A., Zheng, Z., Zhang, J., and Liu, J. (2022). Genomic divergence of *Stellera chamaejasme* through local selection across the Qinghai-Tibet plateau and northern China. Mol. Ecol. *31*, 4782–4796. https://doi.org/10.1111/mec.16622.
7. Guo, W., Yang, Y., Zhang, X., Chen, J., Wu, S., Yang, J., Qu, K., He, L., Liu, M., Hu, H., et al. (2023). Genomic divergence between two sister *Medicago* species triggered by the quaternary climatic oscillations on the Qinghai-Tibet plateau and northern China. Mol. Ecol. *32*, 3118–3132. https://doi.org/10.1111/mec.16925.
8. L Rocha, J., Silva, P., Santos, N., Nakamura, M., Afonso, S., Qninba, A., Boratynski, Z., Sudmant, P.H., Brito, J.C., Nielsen, R., and Godinho, R. (2023). North African fox genomes show signatures of repeated introgression and adaptation to life in deserts. Nat. Ecol. Evol. *7*, 1267–1286. https://doi.org/10.1038/s41559-023-02094-w.
9. Wu, C.-I. (2001). The genic view of the process of speciation. J. Evol. Biol. *14*, 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x.
10. Feder, J.L., Egan, S.P., and Nosil, P. (2012). The genomics of speciation-with-gene-flow. Trends Genet. *28*, 342–350. https://doi.org/10.1016/j.tig.2012.03.009.
11. Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M.A.F., Mehlig, B., and Westram, A.M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. *30*, 1450–1477. https://doi.org/10.1111/jeb.13047.
12. Nosil, P., Funk, D.J., and Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. Mol. Ecol. *18*, 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x.
13. Han, F., Lamichhaney, S., Grant, B.R., Grant, P.R., Andersson, L., and Webster, M.T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res. *27*, 1004–1015. https://doi.org/10.1101/gr.212522.116.
14. Chen, X.Y., Zhou, B.F., Shi, Y., Liu, H., Liang, Y.Y., Ingvarsson, P.K., and Wang, B. (2024). Evolution of the correlated genomic variation landscape across a divergence continuum in the genus *Castanopsis*. Mol. Biol. Evol. *41*, msae191. https://doi.org/10.1093/molbev/msae191.
15. Peng, J.C., He, Z., and Zhang, Z.Q. (2025). Standing genetic variation and introgression shape the cryptic radiation of *Aquilegia* in the mountains of Southwest China. Commun. Biol. *8*, 684. https://doi.org/10.1038/s42003-025-08120-w.
16. Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H., Abbott, R.J., et al. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. Proc. Natl. Acad. Sci. USA *115*, E236–E243. https://doi.org/10.1073/pnas.1713288114.
17. Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. *23*, 3133–3157. https://doi.org/10.1111/mec.12796.
18. Stankowski, S., Chase, M.A., Fuiten, A.M., Rodrigues, M.F., Ralph, P.L., and Streisfeld, M.A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. PLoS Biol. *17*, e3000391. https://doi.org/10.1371/journal.pbio.3000391.
19. Wang, B., Mojica, J.P., Perera, N., Lee, C.R., Lovell, J.T., Sharma, A., Adam, C., Lipzen, A., Barry, K., Rokhsar, D.S., et al. (2019). Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. Genome Biol. *20*, 126. https://doi.org/10.1186/s13059-019-1729-9.
20. Choi, J.Y., Purugganan, M., and Stacy, E.A. (2020). Divergent selection and primary gene flow shape incipient speciation of a riparian tree on Hawaii Island. Mol. Biol. Evol. *37*, 695–710. https://doi.org/10.1093/molbev/msz259.
21. Choi, J.Y., Dai, X., Alam, O., Peng, J.Z., Rughani, P., Hickey, S., Harrington, E., Juul, S., Ayroles, J.F., Purugganan, M.D., and Stacy, E.A. (2021). Ancestral polymorphisms shape the adaptive radiation of *Metrosideros* across the Hawaiian Islands. Proc. Natl. Acad. Sci. USA *118*, e2023801118. https://doi.org/10.1073/pnas.2023801118.
22. Todesco, M., Owens, G.L., Bercovich, N., Légaré, J.S., Soudi, S., Burge, D.O., Huang, K., Ostevik, K.L., Drummond, E.B.M., Imerovski, I., et al. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature *584*, 602–607. https://doi.org/10.1038/s41586-020-2467-6.
23. Xu, W.Q., Ren, C.Q., Zhang, X.Y., Comes, H.P., Liu, X.H., Li, Y.G., Kettle, C.J., Jalonen, R., Gaisberger, H., Ma, Y.Z., and Qiu, Y.X. (2024). Genome

sequences and population genomics reveal climatic adaptation and genomic divergence between two closely related sweetgum species. Plant J. *118*, 1372–1387. https://doi.org/10.1111/tpj.16675.

24. Hooker, J.D. (1863). I. On *Welwitschia*, a new genus of Gnetaceae. Trans. Linn. Soc. Lond. *24*, 1–48. https://doi.org/10.1111/j.1096-3642.1863.tb00151.x.
25. Ickert-Bond, S.M., and Renner, S.S. (2016). The Gnetales: recent insights on their morphology, reproductive biology, chromosome numbers, biogeography, and divergence times. J. Syst. Evol. *54*, 1–16. https://doi.org/10.1111/jse.12190.
26. Jürgens, N., Oncken, I., Oldeland, J., Gunter, F., and Rudolph, B. (2021). *Welwitschia*: phylogeography of a living fossil, diversified within a desert refuge. Sci. Rep. *11*, 2385. https://doi.org/10.1038/s41598-021-81150-6.
27. Oldeland, J., Gunter, F., and Jürgens, N. (2022). Ecological niche models of *Welwitschia mirabilis* and its subspecies in the Namib desert. S. Afr. J. Bot. *148*, 210–217. https://doi.org/10.1016/j.sajb.2022.04.036.
28. Jacobson, K.M., and Lester, E. (2003). A first assessment of genetic variation in *Welwitschia mirabilis* Hook. J. Hered. *94*, 212–217. https://doi.org/10.1093/jhered/esg051.
29. Endress, P.K. (1996). Structure and function of female and bisexual organ complexes in Gnetales. Int. J. Plant Sci. *157*, S113–S125. https://doi.org/10.1086/297407.
30. Jörgensen, A., and Rydin, C. (2015). Reproductive morphology in the *Gnetum cuspidatum* group (Gnetales) and its implications for pollination biology in the Gnetales. Plant Ecol. Evol. *148*, 387–396. https://doi.org/10.5091/plecevo.2015.1142.
31. Haycraft, C.J., and Carmichael, J.S. (2001). Development of sterile ovules on bisexual cones of *Gnetum gnemon* (Gnetaceae). Am. J. Bot. *88*, 1326–1330. https://doi.org/10.2307/3558344.
32. Liao, Q., Du, R., Gou, J., Guo, L., Shen, H., Liu, H., Nguyen, J.K., Ming, R., Yin, T., Huang, S., and Yan, J. (2020). The genomic architecture of the sex-determining region and sex-related metabolic variation in *Ginkgo biloba*. Plant J. *104*, 1399–1409. https://doi.org/10.1111/tpj.15009.
33. Zhang, H., Zhang, R., Yang, X., Gu, K.J., Chen, W., Chang, Y., Xu, Q., Liu, Q., Qin, Y., Hong, X., et al. (2019). Recent origin of an XX/XY sex-determination system in the ancient plant lineage *Ginkgo biloba*. Preprint at bioRxiv. https://doi.org/10.1101/517946.
34. Liu, Y., Wang, S., Li, L., Yang, T., Dong, S., Wei, T., Wu, S., Liu, Y., Gong, Y., Feng, X., et al. (2022). The *Cycas* genome and the early evolution of seed plants. Nat. Plants *8*, 389–401. https://doi.org/10.1038/s41477-022-01129-7.
35. Li, Z., Zong, H., Liu, X., Wang, X., Liu, S., Jiao, X., Chen, X., Wu, H., Liu, Z., Wang, Z., et al. (2025). Phased high-quality genome of the gymnosperm Himalayan Yew assists in paclitaxel pathway exploration. GigaScience *14*, giaf026. https://doi.org/10.1093/gigascience/giaf026.
36. Wang, X.Q., and Ran, J.H. (2014). Evolution and biogeography of gymnosperms. Mol. Phylogenet. Evol. *75*, 24–40. https://doi.org/10.1016/j.ympev.2014.02.005.
37. Ran, J.H., Shen, T.T., Wang, M.M., and Wang, X.Q. (2018). Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. Proc. Biol. Sci. *285*, 20181012. https://doi.org/10.1098/rspb.2018.1012.
38. Fu, L.G., Yu, Y.F., and Riedl, H. (1999). Ephedraceae (Science Press; Missouri Botanical Garden Press).
39. Wan, T., Liu, Z., Leitch, I.J., Xin, H., Maggs-Kölling, G., Gong, Y., Li, Z., Marais, E., Liao, Y., Dai, C., et al. (2021). The *Welwitschia* genome reveals a unique biology underpinning extreme longevity in deserts. Nat. Commun. *12*, 4247. https://doi.org/10.1038/s41467-021-24528-4.
40. Chen, I.C., Hill, J.K., Ohlemüller, R., Roy, D.B., and Thomas, C.D. (2011). Rapid range shifts of species associated with high levels of climate warming. Science *333*, 1024–1026. https://doi.org/10.1126/science.1206432.
41. Antão, L.H., Weigel, B., Strona, G., Hällfors, M., Kaarlejärvi, E., Dallas, T., Opedal, Ø.H., Heliölä, J., Henttonen, H., Huitu, O., et al. (2022). Climate change reshuffles northern species within their niches. Nat. Clim. Chang. *12*, 587–592. https://doi.org/10.1038/s41558-022-01381-x.
42. Sunde, J., Franzén, M., Betzholtz, P.E., Francioli, Y., Pettersson, L.B., Pöyry, J., Ryrholm, N., and Forsman, A. (2023). Century-long butterfly range expansions in northern Europe depend on climate, land use and species traits. Commun. Biol. *6*, 601. https://doi.org/10.1038/s42003-023-04967-z.
43. Marlow, J.R., Lange, C.B., Wefer, G., and Rosell-Melé, A. (2000). Upwelling intensification as part of the Pliocene-Pleistocene climate transition. Science *290*, 2288–2291. https://doi.org/10.1126/science.290.5500.2288.
44. Ruggieri, E., Herbert, T., Lawrence, K.T., and Lawrence, C.E. (2009). Change point method for detecting regime shifts in paleoclimatic time series: application to δ^{18}O time series of the Plio-Pleistocene. Paleoceanography *24*, PA1204. https://doi.org/10.1029/2007pa001568.
45. deMenocal, P.B. (1995). Plio-Pleistocene African climate. Science *270*, 53–59. https://doi.org/10.1126/science.270.5233.53.
46. Lancaster, N. (2002). How dry was dry? - Late Pleistocene palaeoclimates in the Namib Desert. Quat. Sci. Rev. *21*, 769–782. https://doi.org/10.1016/S0277-3791(01)00126-3.
47. Senut, B., Pickford, M., and Ségalen, L. (2009). Neogene desertification of Africa. C. R. Geosci. *341*, 591–602. https://doi.org/10.1016/j.crte.2009.03.008.
48. Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B., and Beaumont, M.A. (2010). The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. Genetics *186*, 983–995. https://doi.org/10.1534/genetics.110.118661.
49. Mazet, O., Rodríguez, W., and Chikhi, L. (2015). Demographic inference using genetic data from a single individual: separating population size variation from population structure. Theor. Popul. Biol. *104*, 46–58. https://doi.org/10.1016/j.tpb.2015.06.003.
50. Bansal, J.K., and Nichols, R.A. (2025). Can genomic analysis actually estimate past population size? Trends Genet. *41*, 559–567. https://doi.org/10.1016/j.tig.2025.03.003.
51. Gamba, D., and Muchhala, N. (2020). Global patterns of population genetic differentiation in seed plants. Mol. Ecol. *29*, 3413–3428. https://doi.org/10.1111/mec.15575.
52. Pearson, H.H.W. (1909). IX. Further observations on *Welwitschia*. Phil. Trans. R. Soc. Lond. B *200*, 331–402. https://doi.org/10.1098/rstb.1909.0009.
53. Wetschnig, W., and Depisch, B. (1999). Pollination Biology of *Welwitschia mirabilis* HOOK. f. (*Welwitschiaceae, Gnetopsida*). Phyton-Ann. Rei. Bot. A. *39*, 167–183.
54. Hofmann, C.-C., Roberts, E.A., and Seyfullah, L.J. (2022). Diversity of the dispersed Gnetalean pollen record from the Lower Cretaceous Crato Formation, Brazil: Entomophily, harmomegathy and habitat heterogeneity. Cretac. Res. *129*, 105020. https://doi.org/10.1016/j.cretres.2021.105020.
55. Bornman, C.H. (1978). *Welwitschia*: Paradox of a Parched Paradise (C. Struik Publishers).
56. Prior, N., Little, S.A., Boyes, I., Griffith, P., Husby, C., Pirone-Davies, C., Stevenson, D.W., Tomlinson, P.B., and von Aderkas, P. (2019). Complex reproductive secretions occur in all extant gymnosperm lineages: a proteomic survey of gymnosperm pollination drops. Plant Reprod. *32*, 153–166. https://doi.org/10.1007/s00497-018-0348-z.
57. Bombi, P. (2018). Potential impacts of climate change on *Welwitschia mirabilis* populations in the Namib Desert, southern Africa. J. Arid Land *10*, 663–672. https://doi.org/10.1007/s40333-018-0067-1.
58. Bombi, P., Salvi, D., Shuuya, T., Vignoli, L., and Wassenaar, T. (2021). Very high extinction risk for *Welwitschia mirabilis* in the northern Namib

Desert. J. Arid Environ. *190*, 104529. https://doi.org/10.1016/j.jaridenv.2021.104529.

59. Bombi, P., Salvi, D., Shuuya, T., Vignoli, L., and Wassenaar, T. (2021). Climate change effects on desert ecosystems: A case study on the keystone species of the Namib Desert *Welwitschia mirabilis*. PLoS One *16*, e0259767. https://doi.org/10.1371/journal.pone.0259767.
60. Ke, F., Vasseur, L., Yi, H., Yang, L., Wei, X., Wang, B., and Kang, M. (2022). Gene flow, linked selection, and divergent sorting of ancient polymorphism shape genomic divergence landscape in a group of edaphic specialists. Mol. Ecol. *31*, 104–118. https://doi.org/10.1111/mec.16226.
61. Liu, Y., Yu, W., Wu, B., and Li, J. (2022). Patterns of genomic divergence in sympatric and allopatric speciation of three Mihoutao (*Actinidia*) species. Hortic. Res. *9*, uhac054. https://doi.org/10.1093/hr/uhac054.
62. Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Phil. Trans. R. Soc. Lond. B *367*, 451–460. https://doi.org/10.1098/rstb.2011.0260.
63. Andrew, R.L., and Rieseberg, L.H. (2013). Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. Evolution *67*, 2468–2482. https://doi.org/10.1111/evo.12106.
64. Turner, T.L., Hahn, M.W., and Nuzhdin, S.V. (2005). Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. *3*, e285. https://doi.org/10.1371/journal.pbio.0030285.
65. Garant, D., Forde, S.E., and Hendry, A.P. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. Funct. Ecol. *21*, 434–443. https://doi.org/10.1111/j.1365-2435.2006.01228.x.
66. Tusso, S., Nieuwenhuis, B.P.S., Weissensteiner, B., Immler, S., and Wolf, J.B.W. (2021). Experimental evolution of adaptive divergence under varying degrees of gene flow. Nat. Ecol. Evol. *5*, 338–349. https://doi.org/10.1038/s41559-020-01363-2.
67. Guerrero, R.F., and Hahn, M.W. (2017). Speciation as a sieve for ancestral polymorphism. Mol. Ecol. *26*, 5362–5368. https://doi.org/10.1111/mec.14290.
68. Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. *2*, e64. https://doi.org/10.1371/journal.pgen.0020064.
69. Leimu, R., and Fischer, M. (2008). A meta-analysis of local adaptation in plants. PLoS One *3*, e4010. https://doi.org/10.1371/journal.pone.0004010.
70. Pregitzer, C.C., Bailey, J.K., Hart, S.C., and Schweitzer, J.A. (2010). Soils as agents of selection: feedbacks between plants and soils alter seedling survival and performance. Evol. Ecol. *24*, 1045–1059. https://doi.org/10.1007/s10682-010-9363-8.
71. Remke, M.J., Hoang, T., Kolb, T., Gehring, C., Johnson, N.C., and Bowker, M.A. (2020). Familiar soil conditions help *Pinus ponderosa* seedlings cope with warming and drying climate. Restor. Ecol. *28*, S344–S354. https://doi.org/10.1111/rec.13144.
72. Zhang, Z., Zheng, J., Guang, Y., Zhao, G., Luo, X., Chen, D., Jia, C., and Hu, X. (2023). Soil mediated local adaptation at the early-life stages of *Stipa breviflora* is context dependent. Plant Soil *484*, 457–472. https://doi.org/10.1007/s11104-022-05814-6.
73. Wittkopp, P.J., and Kalay, G. (2011). *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69. https://doi.org/10.1038/nrg3095.
74. Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54. https://doi.org/10.1038/s41586-019-1517-4.
75. Ballinger, M.A., Mack, K.L., Durkin, S.M., Riddell, E.A., and Nachman, M.W. (2023). Environmentally robust *cis*-regulatory changes underlie rapid climatic adaptation. Proc. Natl. Acad. Sci. USA *120*, e2214614120. https://doi.org/10.1073/pnas.2214614120.
76. Shabalina, S.A., Ogurtsov, A.Y., Rogozin, I.B., Koonin, E.V., and Lipman, D.J. (2004). Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals. Nucleic Acids Res. *32*, 1774–1782. https://doi.org/10.1093/nar/gkh313.
77. Tatarinova, T.V., Chekalin, E., Nikolsky, Y., Bruskin, S., Chebotarov, D., McNally, K.L., and Alexandrov, N. (2016). Nucleotide diversity analysis highlights functionally important genomic regions. Sci. Rep. *6*, 35730. https://doi.org/10.1038/srep35730.
78. Zhou, P., Wang, Z., Li, Y., and Zhou, Q. (2023). Identification and expression of the *MADS-box* gene family in different versions of the *Ginkgo biloba* genome. Plants *12*, 3334. https://doi.org/10.3390/plants12183334.
79. Yang, K., Liu, Z., Chen, X., Zhou, X., Ye, J., Xu, F., Zhang, W., Liao, Y., Yang, X., and Wang, Q. (2022). Genome-wide identification and expression analysis of the MADS-Box family in *Ginkgo biloba*. Forests *13*, 1953. https://doi.org/10.3390/f13111953.
80. Hobza, R., Kubat, Z., Cegan, R., Jesionek, W., Vyskot, B., and Kejnovsky, E. (2015). Impact of repetitive DNA on sex chromosome evolution in plants. Chromosome Res. *23*, 561–570. https://doi.org/10.1007/s10577-015-9496-2.
81. Li, S.F., Zhang, G.J., Yuan, J.H., Deng, C.L., and Gao, W.J. (2016). Repetitive sequences and epigenetic modification: inseparable partners play important roles in the evolution of plant sex chromosomes. Planta *243*, 1083–1095. https://doi.org/10.1007/s00425-016-2485-7.
82. Carey, S., Yu, Q., and Harkess, A. (2021). The diversity of plant sex chromosomes highlighted through advances in genome sequencing. Genes *12*, 381. https://doi.org/10.3390/genes12030381.
83. Khan, J.A., Wang, Q., Sjölund, R.D., Schulz, A., and Thompson, G.A. (2007). An early nodulin-like protein accumulates in the sieve element plasma membrane of Arabidopsis. Plant Physiol. *143*, 1576–1589. https://doi.org/10.1104/pp.106.092296.
84. Denancé, N., Szurek, B., and Noël, L.D. (2014). Emerging functions of nodulin-like proteins in non-nodulating plant species. Plant Cell Physiol. *55*, 469–474. https://doi.org/10.1093/pcp/pct198.
85. Wan, T., Gong, Y., Liu, Z., Zhou, Y., Dai, C., and Wang, Q. (2022). Evolution of complex genome architecture in gymnosperms. GigaScience *11*, giac078. https://doi.org/10.1093/gigascience/giac078.
86. Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. *37*, 4302–4315. https://doi.org/10.1002/joc.5086.
87. Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T., Verelst, L., and Wiberg, D. (2008). Global Agro-Ecological Zones Assessment for Agriculture (IIASA and FAO).
88. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
89. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.
90. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
91. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
92. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
93. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.

94. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics *35*, 1786–1788. https://doi.org/10.1093/bioinformatics/bty875.

95. Felsenstein, J. (1989). PHYLIP - Phylogeny inference package (version 3.2). Cladistics *5*, 164–166. https://doi.org/10.1086/416571.

96. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190. https://doi.org/10.1371/journal.pgen.0020190.

97. Team, R.C. (2020). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

98. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959. https://doi.org/10.1093/genetics/155.2.945.

99. Earl, D.A., and vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7.

100. Browning, B.L., and Browning, S.R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics *194*, 459–471. https://doi.org/10.1534/genetics.113.150029.

101. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097. https://doi.org/10.1086/521987.

102. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496. https://doi.org/10.1038/nature10231.

103. Sanderson, M.J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics *19*, 301–302. https://doi.org/10.1093/bioinformatics/19.2.301.

104. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. PLoS Genet. *9*, e1003905. https://doi.org/10.1371/journal.pgen.1003905.

105. Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A., and Sousa, V.C. (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics *37*, 4882–4885. https://doi.org/10.1093/bioinformatics/btab468.

106. Marchi, N., Kapopoulou, A., and Excoffier, L. (2024). Demogenomic inference from spatially and temporally heterogeneous samples. Mol. Ecol. Resour. *24*, e13877. https://doi.org/10.1111/1755-0998.13877.

107. Wilson, G.A., and Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. Genetics *163*, 1177–1191. https://doi.org/10.1093/genetics/163.3.1177.

108. Mussmann, S.M., Douglas, M.R., Chafin, T.K., and Douglas, M.E. (2019). BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data. Methods Ecol. Evol. *10*, 1808–1813. https://doi.org/10.1111/2041-210x.13252.

109. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst. Biol. *67*, 901–904. https://doi.org/10.1093/sysbio/syy032.

110. Frichot, E., Schoville, S.D., Bouchard, G., and François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. Mol. Biol. Evol. *30*, 1687–1699. https://doi.org/10.1093/molbev/mst063.

111. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

112. Sun, Y., Lu, Z., Zhu, X., and Ma, H. (2020). Genomic basis of homoploid hybrid speciation within chestnut trees. Nat. Commun. *11*, 3375. https://doi.org/10.1038/s41467-020-17111-w.

113. Gao, F., Ming, C., Hu, W., and Li, H. (2016). New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. G3 *6*, 1563–1571. https://doi.org/10.1534/g3.116.028233.

114. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824. https://doi.org/10.1038/ng.2310.

115. Kokot, M., Długosz, M., and Deorowicz, S. (2017). KMC 3: counting and manipulating k-mer statistics. Bioinformatics *33*, 2759–2761. https://doi.org/10.1093/bioinformatics/btx304.

116. Voichek, Y., and Weigel, D. (2020). Identifying genetic variants underlying phenotypic variation in plants without complete genomes. Nat. Genet. *52*, 534–540. https://doi.org/10.1038/s41588-020-0612-7.

117. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

118. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360. https://doi.org/10.1038/nmeth.3317.

119. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

120. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

121. Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Rep. *15*, 8–15. https://doi.org/10.1007/BF02772108.

122. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498. https://doi.org/10.1038/ng.806.

123. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media).

124. Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

125. Nadachowska-Brzyska, K., Burri, R., Smeds, L., and Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. Mol. Ecol. *25*, 1058–1072. https://doi.org/10.1111/mec.13540.

126. Lichstein, J.W. (2007). Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecol. *188*, 117–131. https://doi.org/10.1007/s11258-006-9126-3.

127. van den Wollenberg, A.L. (1977). Redundancy analysis an alternative for canonical correlation analysis. Psychometrika *42*, 207–219. https://doi.org/10.1007/BF02294050.

128. Capblancq, T., and Forester, B.R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. Methods Ecol. Evol. *12*, 2298–2309. https://doi.org/10.1111/2041-210x.13722.

129. Irwin, D.E., Milá, B., Toews, D.P.L., Brelsford, A., Kenyon, H.L., Porter, A.N., Grossen, C., Delmore, K.E., Alcaide, M., and Irwin, J.H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. Mol. Ecol. *27*, 4839–4855. https://doi.org/10.1111/mec.14858.

130. Shang, H., Field, D.L., Paun, O., Rendón-Anaya, M., Hess, J., Vogl, C., Liu, J., Ingvarsson, P.K., Lexer, C., and Leroy, T. (2023). Drivers of genomic landscapes of differentiation across a *Populus* divergence gradient. Mol. Ecol. *32*, 4348–4361. https://doi.org/10.1111/mec.17034.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
BGI Optimal DNA Library Prep Kit	BGI, Shenzhen, China	LD00R96
TruseqTM RNA sample Prep Kit	Illumina	N/A
Deposited data		
Re-sequencing and transcriptomic data	This study	National Genomics Data Center (NGDC): PRJCA052889
WorldClim	Fick and Hijmans[86]	https://worldclim.org/data/worldclim21.html
Harmonized World Soil Database v1.2	Fischer et al.[87]	https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12
Experimental models: Organisms/strains		
Welwitschia mirabilis: leaf, cones	Collected from Namib Desert	N/A
Software and algorithms		
FastQC version 0.11.8	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Trimmomatic v0.32	Bolger et al.[88]	https://github.com/usadellab/Trimmomatic
BWA v0.7.17-r1188	Li[89]	https://github.com/lh3/bwa
SAMtools v.1.16.1	Li et al.[90]; Danecek et al.[91]	http://samtools.sourceforge.net
Genomic Analysis Toolkit v4.1.9.0	McKenna et al.[92]	https://gatk.broadinstitute.org/hc/en-us
PLINK v1.90	Chang et al.[93]	https://www.cog-genomics.org/plink2/
PopLDdecay v3.40	Zhang et al.[94]	https://github.com/BGI-shenzhen/PopLDdecay
PHYLIP v3.697	Felsenstein[95]	https://phylipweb.github.io/phylip/
EIGENSOFT v6.0.1	Patterson et al.[96]	https://github.com/argriffing/eigensoft
R v.4.4.1	R Core Team[97]	https://www.r-project.org/
STRUCTURE v2.3.4	Pritchard et al.[98]	http://www.stats.ox.ac.uk/~pritch/home.html
STRUCTURE HARVESTER v.0.6.94	Earl and vonHoldt[99]	http://taylor0.biology.ucla.edu/structureHarvester/
BEAGLE v5.4	Browning and Browning[100]; Browning and Browning[101]	https://faculty.washington.edu/browning/beagle/beagle.html
Pairwise Sequential Markovian Coalescence v0.6.5-r67	Li and Durbin[102]	http://github.com/lh3/psmc
r8s v1.81	Sanderson[103]	http://ginger.ucdavis.edu/r8s
Fastsimcoal2	Excoffier et al.[104]; Excoffier et al.[105]; and Marchi et al.[106]	http://cmpg.unibe.ch/software/fastsimcoal2/
BayesAss3-SNPs v1.1	Wilson and Rannala[107]; Mussmann et al.[108]	https://github.com/stevemussmann/BayesAss3-SNPs
Tracer v1.7.2	Rambaut et al.[109]	http://beast.community/tracer
Latent Factor Mixed Model	Frichot et al.[110]	https://bcm-uga.github.io/lfmm/
VCFtools v0.1.16	Danecek et al.[111]	http://vcftools.sourceforge.net
Python3 scripts	Sun et al.[112]	https://github.com/yongshuai-sun/hhs-omei
FastEPRR	Gao et al.[113]	http://www.picb.ac.cn/evolgen/
Perl script	Han et al.[13]	https://genome.cshlp.org/content/27/6/1004/suppl/DC1
GEMMA v.0.98.5	Zhou and Stephens[114]	https://github.com/genetics-statistics/GEMMA
KMC ver.3.2.4	Kokot et al.[115]	https://github.com/refresh-bio/KMC
kmerGWAS v.0.2-beta	Voichek and Weigel[116]	https://github.com/voichek/kmersGWAS
Bowtie v1.3.1	Langmead et al.[117]	https://bowtie-bio.sourceforge.net/
HISAT2 v2.0.5	Kim et al.[118]	https://daehwankimlab.github.io/hisat2/
StringTie v2.2.1	Pertea et al.[119]	https://ccb.jhu.edu/software/stringtie/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TransDecoder v5.5.0	N/A	https://github.com/TransDecoder/TransDecoder
DESeq2	Love et al.[120]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

In 2019, a total of 130 individuals of *W. mirabilis* were collected from five populations in the southernmost region of the Namib Desert (Figures 1A and 1B; Table S1). Geographical coordinates of the sampling locations were recorded using a global positioning system (GPS).

METHOD DETAILS

Library construction and sequencing

Genomic DNA was extracted from leaf tissue using the modified cetyltrimethylammonium bromide (CTAB) method and quantified using both a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) and a Qubit fluorometer (Invitrogen).[121] Paired-end libraries were prepared using BGI Optimal DNA Library Prep Kit (BGI, Shenzhen, China) and sequenced on the DNBSEQ-T7 platform. All individuals were sequenced to an average depth of 10×, with a read length of 150 bp and an insert size of 350 bp.

Read mapping and variant calling

For raw resequencing reads, we used FastQC version 0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to assess read quality and Trimmomatic v0.32 to remove adapters and low-quality reads.[88] After quality control, all clean reads from each individual were mapped to the *W. mirabilis* reference genome using the bwa-mem algorithm in BWA v0.7.17-r1188.[39,89] SAMtools v.1.16.1 was used to convert Sequence Alignment Map (SAM) files to Binary Alignment Map (BAM) format and sort the alignments.[90,91] Subsequently, according to the Genome Analysis Toolkit (GATK) Best Practices, GATK v4.1.9.0 was used to mark and remove polymerase chain reaction (PCR) duplicates, call SNPs for each individual, and combine SNPs across all individuals.[92,122,123]

To remove potential false positive sites and obtain high-quality genome-wide SNPs, we further performed the following procedures: (1) SNPs were filtered using GATK VariantFiltration with 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR >3.0 || MQRankSum <-12.5 || ReadPosRankSum <-8.0'; (2) only biallelic SNPs at least 5 bp away from any indels were retained; (3) SNPs with a quality score <30, read depth (DP) < 5 or genotype quality (GQ) < 20 were excluded; (4) variants with a missing rate >5% or minor allele frequency (MAF) < 0.05 were removed; (5) SNP clusters with more than 3 SNPs within 10 bp window were removed; (6) SNPs within 10kb extension region of sex-related SNPs were excluded (see below); and (7) individuals with very close relationships (genetic distance >0.9) were removed based on identity by state (IBS), which was calculated using PLINK v1.90 with the parameter "–distance ibs" (Figure S16).[93] Finally, a total of 129 individuals and 27,740,007 SNPs were retained. Linkage disequilibrium at the species level was calculated using PopLDdecay v3.40.[94] To obtain independent SNPs, these SNPs were thinned using a distance filter of interval >2Mb based on LD results (Figure S7), which yielded a 3,442 pruned SNP set that was used in the analyses of STRUCTURE, BayesAss3-SNPs and Fastsimcoal2.

Population structure analysis

To investigate the population structure of *W. mirabilis*, a neighbor-joining (NJ) tree was constructed based on the pairwise genetic distances between individuals using PHYLIP v3.697.[95] The NJ tree was visualized using iTOL (https://itol.embl.de/). For PCA, we used EIGENSOFT v6.0.1 to extract the top 10 principal components (PCs) and plotted the top three PCs using ggplot2 package (https://ggplot2.tidyverse.org/) in R.[96,97] STRUCTURE v2.3.4 was used to cluster individuals using the admixture model with correlated allele frequencies.[98] We carried out 10 independent runs for each value of *K* (from 1 to 6) with 1,000,000 burn-in steps and 10,000,000 Markov chain Monte Carlo (MCMC) steps. The optimal value of *K* was determined using STRUCTURE HARVESTER v.0.6.94 according to the delta *K* method.[99,124] For further insight into relationships among populations, we performed identity-by-descent analysis using BEAGLE v5.4 with the following parameters: window = 100,000; overlap = 10,000; ibdtrim = 100; ibdlod = 10.[100,101]

Inference of lineage demographic history and gene flow

Based on the results of population structure analysis, P2_1 and P2_2 were merged into P2_12, due to their nearly identical genetic background and negligible differentiation (see results). For the resulting four lineages (P1, P2_12, P2_3, and P3), we first used the PSMC v0.6.5-r67 to infer historical changes in effective population size (*Ne*).[102] As recommended, we used sequencing data with a mean genome coverage of ≥18, a per-site filter of ≥10 reads, and no more than 25% of missing data.[125] Three individuals were

selected from each lineage (Table S1). PSMC requires diploid consensus sequences, which were generated using the "pileup" command of the SAMtools software. We converted the scaled population parameters into *Ne* and years using a generation time of 50 years and a mutation rate of 4.2×10^{-8} substitutions per site per generation, which was estimated by r8s v1.81 based on the maximum-likelihood (ML) tree published in Ran et al.[37,103]

Additionally, we inferred the demographic history and divergence time using a coalescent simulation-based composite-likelihood method implemented in Fastsimcoal2.[104–106] For all possible pairs of the four lineages, the two-dimensional joint site frequency spectra (2D-SFS) was converted from a variant call format file using easySFS.py (https://github.com/isaacovercast/easySFS). Subsequently, we designed and simulated 16 evolutionary models to test different hypotheses for divergence order and time, the presence/absence of asymmetric gene flow between each lineage pair, and recent changes in *Ne* (Figure S1). The global maximum-likelihood parameter estimates for each model were obtained using 100 independent runs with 500,000 coalescence simulations and 100 conditional maximization algorithm cycles. The best model was identified through the maximum value of likelihoods and Akaike's information criterion (AIC). To obtain the 95% confidence interval of the best model, we generated 100 parametric bootstraps and estimated the parameters on each bootstrap replicate using the same settings as for the analyses of the original dataset.

To evaluate the strength and direction of contemporary gene flow among lineages, we used BayesAss3-SNPs v1.1 (BA3-SNPs) to estimate the proportion of immigrants in a given lineage based on Bayesian inference.[107,108] We first performed initial runs to determine optimal mixing parameters using a binary search algorithm in BA3-SNPs. Subsequently, we ran five independent analyses with 20 million MCMC generations, 2 million burn-in and sampling every 1000th iteration. Tracer v1.7.2 software was used to calculate the mean and standard deviation of gene flow across the five runs.[109] Migration rates were considered significant if 95% credible sets (mean migration rate ± 1.96 × mean standard deviation) did not overlap zero.

Effects of geographical distances and heterogeneous environments on genetic structure

To investigate the effects of geographical distances and heterogeneous environments in shaping genetic structure, we performed partial Mantel tests to detect IBD and IBE. We calculated genetic and geographical distances between all pairs of samples using PLINK v1.90 and R package geosphere (https://github.com/rspatial/geosphere), respectively.[93] Additionally, we downloaded 19 bioclimatic variables (1970–2000, 30 s resolution) and one elevation factor from WorldClim version 2.1 (https://worldclim.org/data/worldclim21.html), and 35 soil layers from Harmonized World Soil Database v1.2 (HWSD, https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12).[86,87] After removing variables displaying no variation or containing missing data, 33 environmental factors were retained and used to calculate the Euclidean distances between individual pairs (Table S5). The correlations between geographical/environmental distances and genetic distances were tested in the R package vegan (https://github.com/vegandevs/vegan) with 9,999 permutations.

To determine the relative importance of environmental factors, we conducted multiple regression on distance matrices (MRM).[126] We first clustered these environmental variables and removed highly redundant ones ($|r^2| > 0.7$) using R package Hmisc (https://hbiostat.org/r/hmisc) to avoid multicollinearity. Subsequently, we standardized each environmental variable using R package MuMIn (https://cran.r-project.org/package=MuMIn) to allow for the comparison of their relative importance. Finally, multiple matrix regression was performed using R package ecodist (https://rdocumentation.org/packages/ecodist) with 9,999 permutations.

Identification of environment-associated genetic loci

We used two different approaches to detect environment-associated genetic loci across the whole genome. Prior to analysis, we retained variants with a missing rate <2%, resulting in 8,301,475 SNPs, which were phased and imputed using BEAGLE v5.4 for the GEA analyses.[100,101] Firstly, LFMM was implemented to explore associations between allele frequencies and the 33 aforementioned environmental variables.[110] Based on the number of ancestry clusters inferred by STRUCTURE v2.3.4,[98] we ran LFMM with two latent factors to account for population structure in the genotype data. For each environmental variable, five independent MCMC runs were conducted using a burn-in of 5000 iterations followed by 10,000 iterations. We combined z-scores obtained from these five runs and re-adjusted the *p*-values using a false discovery rate (FDR) correction with a cutoff of 1%. Secondly, we performed RDA, which is one of the best-performing multivariate GEA approaches.[127,128] Based on the relative importance estimated by the MRM, five environmental variables (BIO1, BIO18, SOIL1, SOIL21, SOIL31) with low pairwise correlation ($|r| < 0.7$) were selected for the analysis. The RDA was conducted using the R package vegan (https://github.com/vegandevs/vegan). Significant environment-associated variants were identified as those having loadings in the tails of the distribution using a standard deviation cutoff of 3 on any RDA axes. We further defined the overlapped genetic loci identified by both approaches as "core adaptive loci" for local adaptation.

Analyses of linkage disequilibrium, recombination, genomic diversity, and divergence

To explore the genomic basis underlying heterogeneous differentiation patterns, we calculated a range of summary statistics (e.g., π, F_{ST}, D_{XY}, and ρ).[13,129,130] To minimize bias arising from unequal sample sizes, we randomly subsampled each lineage to 16 individuals for subsequent analyses of genomic characteristics and genomic islands (Table S1). The nucleotide diversity (π) and Tajima's *D* values for each lineage, as well as relative (F_{ST}) and absolute (D_{XY}) genetic divergence between lineages, were calculated using VCFtools v0.1.16 and python3 scripts with 100 kb non-overlapping sliding window.[111,112] Recombination rates (ρ) were estimated along each chromosome of each lineage using FastEPRR software with 100 kb sliding window strategy.[113] A Perl script was employed to calculate the fixed difference (*Df*), which was defined as the variants that showed one allele in all samples of one lineage

and meanwhile showed the other allele in the compared lineages.[13] In addition, we calculated the correlation coefficient (r^2) of pairwise SNPs to estimate linkage disequilibrium for each lineage using PopLDdecay v3.40.[94]

Identification of genomic islands of divergence

We identified genomic islands for each pair of lineages by combining empirical and permutation approaches, following the method of Ma et al.[16] First, we performed 5,000,000 permutations of SNPs across the entire genome to establish a null F_{ST} distribution for each of the 100 kb non-overlapping windows holding different numbers of variable sites. Then, the *p*-values were determined through a comparison of window estimates of F_{ST} with the relative null distribution and corrected for multiple testing using FDR adjustment. We defined putative outlier windows as those within the top 5% of the empirical F_{ST} distribution and with an FDR lower than 0.01. To minimize the false positive signals caused by variant calling and a limited number of loci, we excluded the very end window of each chromosome and the windows containing less than 50 SNPs. Finally, adjacent outlier windows were combined to form larger divergent islands and the numbers and sizes of these islands were compared across all lineage pairs. Moreover, we compared genomic islands and genome-wide background for differences in multiple summary statistics using the Mann-Whitney U test, including F_{ST}, D_{XY}, π, Tajima's *D* and ρ. In addition, we performed an enrichment analysis using Fisher's exact test to determine whether environment-associated SNPs were overrepresented in genomic islands.

Identification of the sex-linked region

To identify the sex-linked region in the *Welwitschia* genome, we conducted both SNP-based and k-mer-based GWAS on 30 male and 30 female individuals with sex treated as a binary phenotype (Table S1). For the SNP-based GWAS, variant calling and filtering were performed as mentioned above, resulting in 47.14 million SNPs. GWAS analysis was performed using the genome-wide efficient mixed model association (GEMMA) algorithm v.0.98.5 with a Bonferroni correction ($p < 0.05$).[114] The kinship matrix was generated using GEMMA and the first four components derived from PCA using PLINK v1.90 were included as random effects.[93] For the k-mer-based GWAS, all 31bp k-mers were extracted from clean reads and counted using K-Mer Counter (KMC) ver. 3.2.4.[115] The k-mer presence/absence matrix was then generated and filtered to retain k-mers with a MAF $\geq$0.05 and a minor allele count (MAC) $\geq$ 5. Subsequently, kmerGWAS v.0.2-beta was used to convert the filtered k-mers to PLINK binary format.[116] Finally, the retained k-mers were aligned to the reference genome using Bowtie v1.3.1 and used for GWAS in GEMMA with the same parameters as the SNP-based analysis.[117] To further characterize the sex-linked region, we also calculated F_{ST} and π between the sexes with a non-overlapping 20 kb window using VCFtools.[111] By integrating significant GWAS signals with genomic regions in the top 5% for both F_{ST} values and π ratios, we identified sex-linked regions. Additionally, we calculated the observed heterozygosity of each individual at both genome-wide level and sex-associated SNPs using VCFtools.[111]

Gene annotation and differential expression analysis

To further understand the molecular mechanisms underlying sex determination in *Welwitschia*, we evaluated and categorized all genes within sex-linked regions based on the annotations provided by the reference genome. In addition, we performed transcript assembly and functional annotation using transcriptomic data that were generated from 10 male and 14 female cone samples across different developmental stages (Figure S13; Table S15). RNA-seq was conducted on the DNBSEQ-T7 platform, resulting in approximately 6 Gb of raw reads per sample. After removing adapters and low-quality reads using Trimmomatic v0.32,[88] clean reads were mapped to the *W. mirabilis* genome using HISAT2 v2.0.5 and genome-aligned transcripts were assembled for each sample using StringTie v2.2.1.[118,119] The assembled novel transcripts were functionally annotated as follows: Open reading frames were predicted using TransDecoder v5.5.0 (https://github.com/TransDecoder/TransDecoder), and the predicted protein sequences were used as queries for BLASTP searches against the UniProt database based on sequence homology.

In addition, we calculated transcripts per million (TPM) values to identify differentially expressed genes between male and female cones. Differential expression analysis was then performed with the DESeq2 R package.[120] Genes with an adjusted *p* value <0.05 and $|\log_2(\text{FoldChange})| \geq 1$ were considered as significantly differentially expressed genes. Finally, GO enrichment analysis was conducted using Omics-Share (https://www.omicshare.com/). Moreover, sex-biased genes were defined as genes expressed at varying levels between male and female cones, based on their gene expression profiles. Male-biased genes were determined by $\log_2(\text{male/female}) \geq 5$, while female-biased expression was defined as $\log_2(\text{female/male}) \geq 5$.

QUANTIFICATION AND STATISTICAL ANALYSIS

The differences in genomic parameters between genomic islands and the genomic background were assessed using the Mann-Whitney U test. Enrichment of environment-associated SNPs within genomic islands was evaluated using Fisher's exact test. For gene enrichment analysis, the statistical significance was determined using the hypergeometric test, and *p*-values were adjusted for multiple testing using Benjamini-Hochberg method on the Omics-Share platform. For all tests, a *p*-value of <0.05 was considered significant. The statistical details can be found in the "method details" section and corresponding figure legends or table titles.