

Genomic insights into the karyotypic radiation of a narrow endemic holocentric plant *Carex helodes*

Inés Gómez-Ramos^{1*}, Rogelio Sánchez-Villegas², Ashwini V. Mohan³, Camille Cornet³
André Marques⁴, Enrique Maguilla², Santiago Martín-Bravo², Kay Lucek³, Marcial Escudero¹

¹Department of Plant Biology and Ecology, Faculty of Biology, University of Seville, Seville, Spain.

²Botany Area, Department of Molecular Biology and Biochemical Engineering, Universidad Pablo de Olavide, Seville, Spain.

³Biodiversity Genomics, Faculty of Sciences, University of Neuchâtel, Neuchâtel, Switzerland.

⁴Institute for Plant Breeding Research, Max Planck, Cologne, Germany.

*Corresponding author: ines.gomez.ramos98@gmail.com igomez2@us.es

Orcid authors:

Inés Gómez-Ramos: 0000-0002-6585-5478

Rogelio Sánchez-Villegas: 0000-0002-9744-0306

Ashwini V. Mohan: 0000-0001-9505-0665

Camille Cornet: 0000-0002-4545-5485

André Marques: 0000-0002-9567-2576

Santiago Martín-Bravo: 0000-0003-0626-0770

Kay Lucek: 0000-0002-2253-2556

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

26 Enrique Maguilla: 0000-0002-8245-3024

27 Marcial Escudero: 0000-0002-2541-5427

28

29 **Abstract:**

30 Holocentric chromosomes allow rapid genome changes through chromosomal rearrangements
31 such as fissions, fusions, inversions or translocations. The plant genus *Carex* shows one of the
32 highest rates of karyotypic evolution among holocentric organisms. We studied the genomic
33 patterns underlying chromosomal rearrangements in the karyotypic radiation of the narrow
34 endemic species *Carex helodes* ($2n = 68-75$).

35

36 Comparing genome assemblies of *C. helodes* from the two karyologically distinct extremes of
37 its European distribution, revealed a striking number of eight chromosomal rearrangements
38 including fusions, translocations and inversions. Genomic breakpoints are gene-poor and TE-
39 rich, corroborating findings in other species and suggesting common genomic characteristics
40 that facilitate the evolution and establishment of chromosomal rearrangements. We identified
41 a chromosomal inversion exhibiting patterns of purifying selection and enrichment in
42 functional genes that potentially mediate rearrangement tolerance. Conversely, another
43 inversion displayed elevated sequence divergence and enrichment in response to temperature
44 stress and phosphate limitation, matching key environmental variables that differ between the
45 study localities.

46

47 The establishment of chromosomal rearrangements along *Carex helodes* European
48 populations was likely driven by demographic bottlenecks and distinct genomic features at
49 breakpoints. Our findings provide preliminary evidence on the rearrangement role in

population differentiation either as reproductive barriers or as genomic islands of differentiation.

Introduction

The role of chromosomal rearrangements (CRs, encompassing fusions, fissions, inversions, duplications, insertions, deletions and translocations) in eukaryote evolution is a long standing debate [1–3]. While some authors highlight their importance [1,3,4], others consider them to be largely incidental [5–7]. In plants, dysploid CRs are more frequent and significant for evolution than expected compared to polyploidy [8–10], despite their potential deleteriousness e.g. altering gene function or expression [11–13] and causing difficulties for chromosome pairing during meiosis in heterokaryotypes [1,14]. Despite this potential deleteriousness, CRs offer an opportunity for speciation when they cause hybrid dysfunction, and therefore, act as a barrier to gene flow [1,3]. CRs may also contribute to speciation through recombination suppression [1,2,15], whereby recombination between chromosomes carrying different CRs is reduced, promoting divergence and decreasing introgression in those regions. If CRs encompass multiple co-adapted genes that are physically coupled, they can be considered as a single supergene [16]. Suppressed-recombination speciation is especially efficient when the rearranged regions contain genes that are key for both reproductive isolation and ecological adaptation [17].

The ability for underdominant CRs to establish within populations and/or propagate to new areas is influenced by population dynamics [1,18]. For example, within small populations where drift is strong compared to selection, a new CR is more likely to establish even under strong underdominance [19,20]. Also, partial reproductive barriers within a population may

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

75 allow for some genetic structure and for underdominant CRs to establish within

76 subpopulations. Inbreeding, vegetative propagation and long life cycles may similarly help

77 the establishment of new CRs within a population [21,22]. Some of these features can also

78 promote the colonization of new areas, where drift is higher and, therefore, the establishment

79 of new CRs could be easier. Heterogeneous environments can facilitate the persistence of new

80 CRs within a different ecological niche, as locally adapted alleles may arise and become fixed

81 within rearranged areas where recombination is suppressed [23,24].

82

83 Similar CRs often arise repeatedly over the evolutionary history of lineages [11,25–27]. This

84 recurrent aspect of CRs is reflected by the increased degree of conserved syntenic blocks in

85 taxonomic groups with high CR rates, where CRs are constrained to regions with certain

86 genomic features [25]. Such genomic hotspots for CRs have been reported for both plants

87 [25,28], Lepidoptera [26] and mammals [29,30]. For example, CR hotspots are enriched for

88 repetitive regions that are more prone to non-allelic homologous recombination (NAHR)

89 [31,32]. Breakpoints also tend to be outside of gene rich regions, likely because CRs could

90 break gene sequence or separate it from regulatory regions [25,33].

91

92 While the mechanisms underlying CR formation are conserved throughout eukaryotes

93 [34,35], their evolutionary impact largely varies between different clades. The importance of

94 underdominance for CRs has been suggested to be lower for species with holocentric

95 chromosomes, where kinetochore activity is distributed throughout each chromosome instead

96 of being located in one region [36–38]. Contrary to monocentric species, fissions and fusions

97 are expected to be more tolerated in holocentric organisms [39,40] leading to increased rates

98 of CR evolution with frequent meiotic irregularities [39,41,42]. Some holocentric clades have

99 distinct genomic features such as a more uniform GC content and gene distribution along

their chromosomes [43–46], which is otherwise associated with the distance to the centromere in monocentric organisms [47–49]. In summary, the principles of CR evolution may differ between holocentric and monocentric clades [50].

One of the best studied holocentric lineages is the genus *Carex* L. (Cyperaceae; commonly known as true sedges), exhibiting high rates of CRs [37,41,51–55], which has been suggested to underlie its striking species richness (ca. 2000 sp.) [37,41,52,56]. This has been confirmed by recent phylogenetic macroevolutionary studies [57,58]. Significant differences in karyotype evolution occur both across and within different *Carex* clades [57]. For example, within Spirostachyae [54], some species have a stable chromosome number and others showing large ranges of intraspecific chromosome number variation (*Carex laevigata* Smith $2n = 69 - 84$ or *C. helodes* Link $2n = 68 - 75$). In fact, *C. helodes* even shows chromosome number variation between individuals with no genetic differentiation at the Internal Transcribed Spacer (ITS) DNA region or microsatellites, suggesting a recent origin [59,60].

Here, we investigate the genomic basis of the remarkable karyotypic diversity observed across the European range of the Ibero-African narrow endemic sedge *Carex helodes* [60]. Firstly, we characterize CRs underlying changes in chromosome number and estimate the number of CRs through synteny analyses between two individuals sampled from two populations from opposite extremes of the species' European range. We then test the hypothesis that CRs are enriched for genomic features that either facilitate their emergence and/or limit their deleteriousness, by comparing the genomic characteristics of breakpoints and rearranged chromosomes with those of conserved regions. We further evaluate the impact of CRs on gene evolution in inverted regions and areas flanking breakpoints, where

recombination is expected to be suppressed [1,15]. Finally, the functions of these genes are assessed through gene ontology analysis to infer the potential impact of CRs.

Methods

(a) Study species

Carex helodes Link, a member of the *Carex* subgenus *Carex* in the Cyperaceae family [61], is a diploid, wind-pollinated, perennial herb with caespitose habit (without creeping rhizomes), rough upper leaf surface and high numbers of male (1–4[–7]) and androgynous spikes. It is endemic to southern Portugal, southwestern Spain and northern Morocco [62].

The *C. helodes* karyotype varies from $2n = 74$ in Morocco (exceptionally $2n = 75$ in Portugal) to $2n=68$ in Aznalcollar, Spain [54,60,62]. An intrapopulation cytological study demonstrated that karyotypical diversity was even high within populations, with four different chromosome numbers ($2n=68, 69, 70, 72$) and six different karyotypes [60] suggesting the presence of multiple CRs, with $2n=70=35^{II}$ being by far the most common karyotype. Despite this great karyotypic diversity, all individuals in the population were genetically identical based on 34 microsatellites that were variable for the species outside this population [60].

Former genetic studies on *C. helodes* [60,62] point to Portuguese populations as being ancestral and genetically diverse while Spanish populations evolved more recently and have a much lower genetic diversity, consistent with a west-to-east expansion from southern Portugal to southwestern Spain involving multiple genetic bottlenecks and drift [60].

(b) Plant material, library construction and sequencing

149

150 During spring 2019, we collected two individuals from two populations of *C. helodes*, located

151 at the west-eastern limits of the species distribution area within Europe. One individual was

152 sampled in Aznalcollar (Seville, Spain; 37° 34' N, 6° 21' W) and the other in Caldas de

153 Monchique (Algarve, Portugal; 37° 16' N, 8° 31' W), from now on referred to as AZN and

154 MON, respectively. The populations represent the two extremes of the karyotype gradient for

155 this species in Europe (figure 1a). The living plants were kept at the greenhouse of the

156 University Pablo de Olavide, Spain. Experimental crosses and chromosome counts of

157 parentals and F1 were performed (see electronic supplementary materials, Supplementary

158 Methods, section (a)).

159

160 Fresh leaves were harvested and flash frozen with liquid nitrogen before storage at -80°C.

161 High-molecular weight DNA extraction and long-read HiFi sequencing on a PacBio Revio

162 was outsourced to Novogene (Cambridge, UK). Hi-C was generated from fresh young leaf

163 tissue. Firstly, nuclei extraction was carried out using the CellLytic™ PN Isolation/Extraction

164 Kit from Merck. Then, DNA crosslinking, ligation, amplification, and library preparation was

165 done using the Arima Hi-C high-throughput kit and a modified protocol for plant tissue.

166 Finally, the prepared Hi-C libraries were paired-end sequenced on an Illumina Novaseq X

167 (Illumina Inc., San Diego, CA, USA).

168

169 (c) Genome assembly

170

171 HiFi data quality was assessed through NanoPlot v1.20.0 [63] and a k-mer histogram

172 performed with kmer-jellyfish 2.3.1 [64] and GenomeScope 1.0 [65]. The length, coverage

173 and quality of the long reads were considered adequate for both genomes.

1
2
3 174 Hi-Fi data was assembled using HifiASM 0.25.0 [66] obtaining contig level primary and
4
5 175 alternate assemblies. Hi-C data was used in this step to correct sequencing and assembly
6
7 176 errors. Duplicates were purged through a preliminary HifiASM purge and a further analysis
8
9
10 177 with purge_dups 1.2.6 [67]. Scaffolding was performed using Hi-C data by mapping the Hi-C
11
12 178 reads against each assembly following the Arima Hi-C mapping pipeline [68]. Then,
13
14 179 scaffolding was performed with YaHS 1.2a.2 [69] with later manual curation performed in
15
16 180 JuiceBox 2.15 [70] to correct misalignments between scaffolds. At each step of the assembly,
17
18 181 quality controls were performed addressing i) assembly contiguity with gfastats 1.3.11 [71],
19
20 182 ii) assembly completeness based on BUSCO genes from the embryophyta_odb10 and
21
22 183 poales_odb10 databases using CompleASM 0.2.27 [72] and iii) with haplotype specific
23
24 184 analysis based on kmer using Merquy 1.3 [73]. Contaminant scaffolds (e.g. organellar
25
26 185 sequences, microbial sequences, or short low-coverage fragments) were identified using Tiara
27
28 186 1.0.3 [74] and removed. Finally, telomeric regions were identified and visualized to evaluate
29
30 187 how many scaffolds were telomere-to-telomere using tidk v0.2.7 [75].
31
32
33
34
35
36
37

38 189 **(d) Gene annotation, synteny analysis and CRs identification**

39
40 190
41
42 191 Gene annotation was performed with Braker3 v3.0.8 [79], which combines two methods: an
43
44 192 homologous search using protein data from poales_odb10 [78] via the `-Busco_lineage` option
45
46 193 and an *ab initio* approach. Annotation completeness has been assessed with BUSCO genes
47
48 194 from the embryophyta_odb10 and poales_odb10 databases using CompleASM [72] using the
49
50 195 option *analyze*. Prior to gene annotation, repetitive elements were softmasked with EarlGrey
51
52 196 v4.4.5 [79] (further details are shown in electronic supplementary materials, Supplementary
53
54 197 methods, section (b)).
55
56
57
58
59
60

Synten analysis between the two genomes was performed with GENESPACE v.1.2.3 [81], using default parameters. Rearrangement breakpoints were defined as genomic regions which were not included in syntenic blocks. Rearrangements were manually explored with the Interactive Genomic Viewer [81] to ensure that there were no false rearrangements, e.g. due to assembly errors of repetitive regions close to telomeres [82]. CR validation was performed through two methods: reciprocal Hi-C cross-mapping and long read split end support (see supplementary materials, Supplementary methods). The possibility of heterozygosity for the detected inversions was also assessed (see supplementary materials, Supplementary methods, section (c)).

(e) Characterization of conserved versus rearranged genomic regions

Conserved (collinear) regions and breakpoints were compared for gene density, GC content, total repeat content and repetitive elements families. The most common tandem repeat was separately analysed, as it is a possible candidate for holocentromeres (Table S1 in [25]). In addition, comparisons for the same genomic features were made between collinear chromosomes and rearranged ones. Prior to the comparative analyses, the relationships between genomic variables were assessed through pairwise correlation with the R package ggcorrplot v1.4.1 [83] using Spearman's rank correlation coefficient.

Linear and generalized linear models were first performed with each genomic characteristic. Each genomic feature was defined within 100 kb windows throughout the whole genome for AZN and MON with the 'tileGenome' function from the GenomicRanges R package 1.34.1 [84]. The basic model formula was: Genomic variable ~ (RC + BP:RC)*DC, which was run for each genomic feature separately. The independent variables include: RC (does the

1
2
3 224 window belong to a rearranged chromosome?), BP (does the window include a rearrangement
4
5 225 breakpoint?) and DC (window's distance to the center of the chromosome). The variable BP
6
7
8 226 is nested within RC. Chromosome identity was included as a random factor for a genomic
9
10 227 characteristic model when it increased AIC and reduced residuals variance. Different R
11
12 228 packages were used to perform these models: *'lm'* and *'glm'* function from stats for linear
13
14 229 models, *'boxcox'* transformation from the package MASS 7.3-65 [85], *'glm.nb'* from the
15
16
17 230 MASS package for negative binomial GLMs, *'lmer'* function from lme4 1.1-37 [86] for linear
18
19 231 models with random effects and *'glmmTMB'* from glmmTMB 1.1.9 for GLM with random
20
21 232 effects [87]. The final model and the transformation used for each genomic characteristic are
22
23 233 shown in table S4 (see electronic supplementary material). Chromosome size was compared
24
25 234 between rearranged and collinear chromosomes, using Wilcoxon rank-sum and Student's t-
26
27
28 235 tests in R, ensuring that RC effects were not confounded by differences in chromosome size.
29
30
31 236
32
33 237 Breakpoint characterization was performed with *'permTest'* from the regioneR package 1.30
34
35 238 [88], which implements an enrichment analysis based on permutation tests through
36
37
38 239 randomizing genome areas. The robustness does not depend on the number of breakpoints
39
40 240 and is able to detect effects at finer scales. The number of overlaps (*'numOverlaps'*) was used
41
42 241 as the evaluation statistic, except for GC content whose evaluation statistic was defined as
43
44 242 *'meanInRegions'*. Randomization was performed with *'randomizeRegions'* from the same
45
46
47 243 package with 1000 permutations. Breakpoints were compared against regions within the same
48
49 244 chromosome and telomeres (defined as regions within 10kb of the chromosome ends) were
50
51 245 excluded by masking them. Another permutation was performed dividing chromosomes into
52
53 246 subtelomeric (10% most terminal sections of a chromosome) and interstitial sections.
54
55
56 247 Breakpoints were compared against regions of the same chromosome within the same
57
58 248 genomic section using the option *'universe'* in regioneR. Section limits were assessed based
59
60

on the average point where GC content stabilizes from the telomeres. The magnitude of the enrichment was quantified by dividing the observed value by the median of the permuted distribution (*fold enrichment*). For zero-inflated variables (>20% of zeros) pseudocounts were used for the fold enrichment calculation: (observed value +1) / (mean permuted value + 1).

(f) Divergence and selection in rearrangements

To assess patterns of molecular evolution associated with rearranged regions, sequence divergence between the two genomes was compared between rearranged and collinear regions of rearranged chromosomes. In this analysis the MON genome was used as reference, as is likely to be the ancestral population given its demographic history [60]. Because recombination can be suppressed within inversions and regions flanking other CRs [15,89,90], divergence—whether driven by genetic drift or selection—is expected to be higher within these regions. In this analysis rearranged regions included: (i) regions inside inversions, and (ii) regions flanking translocations and fission/fusion breakpoints (200 kb to each side of the non-collinear region). For the fusions, the rearranged regions were defined as 200 kb from the two chromosome ends (chr17 and chr19) and it was compared to the rest of the two chromosomes. DNA and amino acid sequences of genes located in rearranged chromosomes of the MON individual and their orthologous counterparts in AZN (obtained from Genespace) were aligned using MAFFT v7 [91]. Each coding sequence (CDS) was analyzed individually to maximize alignment accuracy. Alignments with <120 bp and/or >10% gaps were filtered out. Codon alignment distinguishing between synonymous and non-synonymous mutations was performed using pal2nal 14.1 [92]. Next, synonymous (dS) and non-synonymous (dN) substitution rates and their ratio ($\omega = dN/dS$) were estimated through the ‘*codeml*’ function implemented in PAML [93], under the evolutionary Model 0, which

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

assumes a constant ω throughout the tree. CDS with high (>2) dN or dS values were discarded, because when multiple substitutions occur in the same CDS, saturation occurs, compromising the rate's reliability. Finally, substitution rates and their ratios were compared between rearranged regions and the rest of the rearranged chromosome. The ratio ω is a measure of selective pressure indicating if there is neutral evolution ($\omega \approx 1$), positive ($\omega > 1$) or purifying selection ($\omega < 1$). Ratio ω comparisons tested if selective pressure was affected by the rearrangements. The statistical significance of the differences was evaluated through enrichment analyses performed with regioneR [88]. Regions of the same size as the rearranged areas were sampled across the rest of the rearranged chromosome after telomeres were masked (10 kb from each extreme of the chromosome). Aggregated dN, dS, and ω values were calculated over 1,000 iterations.

(g) Gene ontology within rearranged regions

Gene ontology (GO) enrichment analyses within rearranged regions (as defined before) were performed using topGO 2.50.0 [94] in R based on the annotated gene set obtained from Braker3 [79]. A custom gene-to-GO mapping was generated using eggNOG-mapper v2.1.12 [95,96]. Enrichment for the three primary Gene Ontology domains—Biological Process (BP), Cellular Component (CC), and Molecular Function (MF)—was assessed using classic Fisher's exact tests implemented in topGO. GO-annotated genes within rearranged regions were considered the foreground and the rest of GO annotated genes in the genome as the background. GO-enrichment test were performed for all CRs combined as well as separately per CR for the MON genome Redundant BP GO terms were clustered using GOSemSim

2.24.0 [97] based on the *Arabidopsis thaliana* annotation database (org.At.tair.db) for visualization.

Results:

(a) Genome assembly and annotation:

Carex helodes long read sequencing yielded 60 Gb of data from 8.33 million HiFi reads representing a coverage of 42.4x for Aznalcollar (AZN) and 72 Gb of data from 14.1 million reads for a coverage of 49.2x for Monchique (MON). Hi-C data accounted for 145 Gb from 392 million reads for AZN and 204 Gb from 569 million reads for MON.

Our final assemblies were 346.74 Mb in 35 chromosome-scale pseudomolecules with more than 99.98% of the bases in the assembly for AZN and 348.51 Mb in 36 chromosome-scale pseudomolecules with more than 99.77% of the bases in the assembly for MON. The assemblies had a scaffold N50 of 10.12 Mb and 9.69 Mb for AZN and MON, respectively. More than half of the chromosome-scale pseudomolecules had telomeric repeats at both chromosome ends (see electronic supplementary materials, figure S1). BUSCO analysis identified around 93 % of complete single-copy genes and 4 % of missing genes for both genomes when using the lineage Embryophyta as a reference (see electronic supplementary materials, table S1). However, when using the more complete gene set of the Poales lineage, the proportion of missing genes rose to 17 % for AZN and 18 % for MON, which is consistent with other high quality *Carex* genomes [98,99] as the reference gene set contains proteins unique to Poaceae. Parental genomes were highly homozygous (0.23% and 0.26% heterozygosity for AZN and MON, respectively, according to GenomeScope [65]), consistent

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

with other *Carex* genomes where clonal reproduction is not prevalent and repeat content is relatively high with 50.20% of repetitive DNA in AZN and 50.38% for MON.

Gene annotation yielded 30,745 genes for AZN and 31,224 for MON. Missing genes did not increase after gene annotation based on BUSCO analysis with Poales as the reference dataset and only by 2 % when Embryophyta is the reference used (see electronic supplementary material, table S1).

(b) Synteny analysis: A surprising number of rearrangements

The synteny analysis performed with Genespace showed three inversions (Inv_chr6, Inv_chr8 and Inv-chr32; rearrangements were named according to the more ancestral MON genome [60]), 4 translocations (Trans_chr5, Trans_chr9, Trans_chr16 and Trans_chr24) and one fusion (Fusion-chr17 and fusion-chr19), which results in the largest AZN chromosome (Figure 1c). These CRs were confirmed through crossed Hi-C contact maps (see electronic supplementary materials, figure S2). Long-read split-read support analyses were able to adequately predict breakends for most CR breakpoints in at least one of the genomes (see electronic supplementary materials, table S2). The inversion in chr32 was heterozygous in the MON genome in accordance to both Hi-C contact maps (see electronic supplementary materials, figure S3) and long-read split-read support (76 out of 130 support the inversion, see electronic supplementary material, table S2).

(c) Germination success and cytogenetics

F₁ and F₂ germination successes were 86% (obtained from one cross) and 60% (mean: 60%, sd: 0.28, number of crosses: 9), respectively. The cytogenetic experiments, performed on the same parentals as the synteny analysis, revealed chromosome numbers of $2n = 36^{II} = 72$ for the MON parent and $2n = 35^{II} = 70$ for the AZN parent, both displayed exclusively bivalent formations. In the five meioses analyzed for these two F₁ individuals, all showed the same configuration. Specifically, the F₁ individuals showed intermediate chromosome numbers ($2n = 34^{II} + 1^{III} = 71$) with 34 bivalents and one trivalent (figure 1b).

(d) Genomic differences between rearranged and conserved areas

Both genomes were divided in 100kb windows resulting in 3467 windows for AZN and 3486 windows for MON, with 18 windows containing rearrangement breakpoints for AZN and 15 for MON. Most genomic features varied among chromosomes, with the exception for tandem repeat clusters (TRCs) and the most common TRC (electronic supplementary material, table S4, figure S4). The best predictor for all genomic features, except gene, LINE and SINE content, was the position of the window within a chromosome (DC) (electronic supplementary material, table S4, figure S4). Overall TE proportion, TEs subtypes (except LINEs and SINEs) and GC content increased towards the telomeres, while unknown repeats followed the opposite pattern, as demonstrated by the negative estimate of the DC effect for unknown repeats. This negative correlation between unknown repeats and GC content is reflected in the autocorrelation analysis (see electronic supplementary materials, figure S6), the highest correlation between variables that do not derive from each other.

The genomic composition differed significantly between rearranged and conserved chromosomes (figure 2b, electronic supplementary material, table S4, figure S4): rearranged

1
2
3 371 chromosomes displayed less LINEs for both genomes and lower GC content for the MON
4
5 372 genome. These differences were not explained by chromosome size (AZN: P (t test)=0.29;
6
7 373 AZN: P (Wilcoxon)=0.36; MON: P (t test)=0.47; MON: P (Wilcoxon)=0.27). If the interaction
8
9 374 between RC and DC is considered, rearranged chromosomes have a lower TE density towards
10
11 375 the telomeres. LTRs and within those Ty3/Gypsy had also significant effects in the same
12
13 376 direction. However, this was only significant for AZN, while for MON only DNA
14
15 377 transposons were scarcer in rearranged chromosomes near the telomeres.
16
17 378
18
19 379 Breakpoints similarly differed from the rest of the genome (figure 2a, table 1, electronic
20
21 380 supplementary material, table S4-S5, figure S5) even when compared within the same
22
23 381 sections (interstitial and subtelomeric) in the same chromosome (table 1). However, the
24
25 382 magnitude of the differences varied between genomes, analysis and the breakpoint distance to
26
27 383 the chromosome center. Gene density for interstitial breakpoints was almost half compared to
28
29 384 similar regions, but breakpoints gene impoverishment was not notable within subtelomeric
30
31 385 sections (figure 2, table 1, see electronic supplementary materials, table S4). TE enrichment
32
33 386 near breakpoints was significant both within interstitial and subtelomeric sections for the
34
35 387 AZN genome, mainly because of LTR enrichment, particularly those from the Ty3/gypsy
36
37 388 subclass (figure 2, table 1, see electronic supplementary materials, table S4-5). TE enrichment
38
39 389 was less clear for the MON genome, where the effect was marginally significant (i.e. $0.1 > P$
40
41 > 0.05) and only within interstitial areas and was not detectable according to the linear model
42
43 390 analysis (see electronic supplementary materials, table S4). The most common TRC was
44
45 391 enriched in breakpoints located in the subtelomeric section for the MON genome, which
46
47 392 include breakpoints involved in fusion/fission. Unknown repeats, the main component of
48
49 393 repetitive elements (see electronic supplementary materials, figure S6), are scarcer near
50
51 394 breakpoints in the GLMs but not the permutation analyses where their effects were only
52
53
54
55
56
57
58
59
60

suggestive for only the AZN genome (table 1, see electronic supplementary materials, table S4-5). GC content varied in the breakpoint regions, with the direction of the effect differing between genomes, sections and analysis, likely because of correlations with other variables such as with gene density and unknown repeats (see electronic supplementary materials, figure S6).

(e) Divergence and evolution in rearrangements

Substitution rates and selection pressure ($\omega = d_N / d_S$) was calculated for the 6 (out of 8 rearrangements) that reached the minimum number of sites required for the analysis following strict coding sequence (CDS) filtering (see electronic supplementary material, table S3). Two inversions (Inv_chr6 and Inv_chr32) showed marginally significant signs of altered gene evolution (table 2) but in opposite directions. Inv_chr6 showed patterns consistent with purifying selection ($\omega = 0$; $P = 0.059$) with no non-synonymous mutations inside the inversion ($d_N = 0$; $P = 0.050$). In contrast, Inv_chr32 showed higher divergence both for synonymous ($d_S = 0.29$; $d_S P = 0.059$) and non-synonymous mutation rate ($d_N = 0.66$; $P = 0.050$) but with no apparent effect on selection efficacy (ω).

(f) Gene ontology across rearranged regions

The eggNOG-mapper successfully assigned functional orthology to 23,975 genes for MON (76.8% of the total annotated genes) with 453 GO annotated genes in regions affected by CRs (70.8% of all annotated genes within the CRs): 39 in Inv_chr 6 (65%), 184 in Inv_chr8 (68.4%), 83 in Inv_chr32 (85.6%), 34 in Trans_chr5 (85.5%), 24 in Trans_chr9 (88.9%), 21

1
2
3 419 in Trans_chr16 (56.8 %), 40 in Trans_chr24 (67.8%), 9 in Fus_chr17 (69.2%) and 19 in
4
5 420 Fus_chr19 (65.5%).
6
7
8 421
9
10 422 Gene Ontology enrichment analyses for all CRs combined and separately for each CR
11
12 423 showed that rearranged regions are enriched for a wide variety of cellular functions, with 188
13
14 424 enriched GOs for biological processes (BP), 51 for molecular functions (MF) and 47 for
15
16 425 cellular compartments (CC) (see electronic supplementary material, table S6). These GO
17
18 426 terms are mainly focused on transcription regulation, protein glycosylation, proteasome
19
20 427 activation and DNA topoisomerase (Figure 3; see electronic supplementary material, table
21
22 428 S6). These functions are enriched mainly in Inv_chr6, Inv_chr8 and Trans_chr16. However,
23
24 429 other CRs are specialised in other functions: Trans_chr5 regulate carbohydrate catabolism,
25
26 430 Trans_chr9 manage chloroplast transmembrane transport and oxidoreductase activity and
27
28 431 Trans_chr24 covers acquired resistance and Inv_chr32 is focused on response to cold and
29
30 432 phosphate starvation (see electronic supplementary material, table S6).
31
32
33
34
35
36
37
38
39
40
41
42

434 Discussion

435

436 (a) A colonization through founder events reveals a striking number of 437 rearrangements

438

439 West-to-east decreasing dysploidy across European *C. helodes* populations was reported by
440 Escudero and colleagues [60]. Using genomic tools, eight CRs (1 fusion, 3 inversions and 4
441 translocations) were detected between individuals from the two geographically extreme
442 populations of the species' European range (figure 1), a much higher number than expected

59

based on traditional cytogenetic analyses (figure 1; [61]). The number of CRs is striking considering the recent history of this species, as the estimated age of the crown node for the European populations of *C. helodes* is 40,000 years [61]. This would translate to at least one CR per 10,000 years [60] which is much higher than any previous estimations for angiosperms [100], even considering other species of the karyologically dynamic genus *Carex* [54,56,101].

Consistent with former inferences [60], the identified CRs could have been established through recurrent genetic bottlenecks during range expansion. This hypothesis is further supported by the fact that the assembled genomes show the only (MON) or the most common (AZN) euploid karyotype for each population. Accordingly, fusions and other cytologically detectable CRs between these two euploid chromosome numbers ($2n = 36^{II}$ and $2n = 35^{II}$) are less likely to occur within each population (figure 1*b*; [60]). During range expansion, even underdominant CRs can establish and proliferate [1,18]. A long life cycle [61] coupled with the possibility for asexual reproduction further promotes the establishment and persistence of underdominant CRs. Selfing capability, evidenced by our crossing experiment and the high homozygosity in natural populations [61] provide a plausible mechanism for this hypothesis: individuals with new CRs may produce seeds by selfing that colonize and establish in new areas [102]. However, our hypothesis for CRs that are not detectable by cytology, such as inversions or translocations, remains tentative until the fixed or polymorphic status of these CRs is confirmed in each population through population-level genomic data. Indeed, one inversion was heterozygotic for MON. Nevertheless, this conclusion remains consistent for CRs that are detectable through cytology such as chromosome fissions and fusions [61].

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Hybrid dysfunction led by CR accumulation has been reported in other *Carex* species [103,104]. While there is no strong reproductive isolation between our studied populations as revealed by F1s high germination success, reproductive isolation increases in the following generation as F2s reach a germination success of only 60% (significantly lower than in natural populations [109]). This increased hybrid dysfunction could be potentially explained by partial sterility of the F₁s due to chromosome missegregation during meiosis. In congruence with this, hybrid dysfunction was already reported in *C. scoparia* where germination success decreases as chromosome differences increase between parents [103]. However, multi-generational fitness measurements and more exhaustive cytogenetic studies would be necessary to clarify the extent and impact of hybrid dysfunction. The accumulation of CRs preceding genetic differentiation [60,62] may suggest that CRs are more likely to be drivers of genetic differentiation rather than a consequence of it.

(b) Gene-poor and TE-rich environments in the genomic landscape facilitate viable CRs

Both breakpoints and rearranged chromosomes showed distinct genomic characteristics from the rest of the genome. This is consistent with previous findings suggesting that rearrangements are restricted and constrained to genomic regions with particular genomic features [25,31–33], including *Carex* [25] and other holocentric organisms [26]. The observed reduced gene density near breakpoints compared to the rest of the genome (table 1) support the hypothesis that CR breakpoints lie in gene poor areas which reduce the potential deleteriousness of CRs [25,33], which has been observed for both mono [33,106] and holocentric organisms [25,26,107].

Transposable elements (TEs) were overall enriched near the breakpoints yet the magnitude of the enrichment depended on the analysed genome (table 1), within TEs the most enriched were LTRs, especially Ty3/Gypsy subclass (table 1, figure 2, see electronic supplementary material, table S4-S5). TEs enrichment appears to be more local than gene impoverishment as its effect was less notable in the GLM analysis (figure 2, see electronic supplementary material, table S4). The association between TEs and LTRs with CRs has been consistently reported across different organisms [25,26,112,113]. Evidence at both macro- [109] and microevolutionary (shown here) scales support the role of LINEs and LTR (Ty3/Gypsy) in chromosome instability, as rearranged chromosomes tend to be enriched in these repetitive elements. TEs can cause CRs through non-allelic homologous recombination (NAHR) or alternative transposition for DNA transposons [110,111]. However, CRs can also lead to TE accumulation through recombination suppression [116,118,119]. In addition, TE rich regions are also gene-poor [47,48], so the observed correlation could also reflect a higher likelihood of CRs to establish due to their lower deleteriousness. The available data does not allow us to disentangle which of these explanations for the association is responsible for the observed pattern.

In many monocentric species, gene density and GC content decrease toward the centromere, increase toward the telomeres, and subsequently decline sharply at chromosomal ends, exhibiting an inverse relationship with repeat content [47–49]. Other patterns have been reported in holocentric organisms [43–46]. The uniformity of genomic characteristics for holocentric plants was only suggested for gene density in our data as distance to the chromosome center is an important predictor for nearly all the genomic characteristics measured (see electronic supplementary material; figure S4, table S4).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(c) Contrasting evolutionary trajectories and functional signatures of chromosomal inversions

Among the observed CRs, the inversions on chromosomes 6 and 32 were the only ones with distinct gene evolution patterns compared to the rest of the chromosome (table 2), but in opposite directions. The inversion on chromosome 32 showed increased neutral and non-synonymous divergence and, therefore, it did not translate into difference in selection pressure (ω). Such increased divergence is consistent with a role of inversions as islands of genetic differentiation, likely due to recombination suppression and linkage disequilibrium within heterozygotic karyotypes [114]. Notably, this CR is the only one with heightened neutral divergence, pointing to it being likely older than the other CRs, further supporting the hypothesis of its appearance prior to the Spanish colonization as our Portuguese individual was heterozygous for this CR (see electronic supplementary materials, table S2, figure S3). In contrast, the inversion on chromosome 6 showed patterns consistent with purifying selection (table 2). This typically points to important functions being located within these inversion, so mutations are potentially deleterious for the plant. While the inversion capturing important genes by chance is perfectly possible, purifying selection can be detected in non-fundamental genes when they are in linkage disequilibrium with fundamental ones, as it happens with genes within an inversion [115,116].

Our GO enrichment analysis offers an exploratory insight into the potential functional implications of the identified inversions. Therefore, these results must be interpreted with caution due to the inherent limitations of cross-species GO transfer and incomplete genome annotations. Inv_chr6 is enriched in functions related to positive regulation of proteasome activity, which has been linked to increased tolerance for CRs [117,118]. Therefore it could

mitigate proteotoxic stress arising from gene dosage imbalances that may come from duplicated and/or eliminated parts in translocations [119–121]. Notably, nuclear proteasomes can also increase CR rate, as they degrade trapped topoisomerases exposing double strand breaks (DSBs) for repair [122,123], but also promote their repair through CR-prone mechanisms [124,125].

The inversion on chromosome 32 (Inv_chr32) is enriched for functions related to cold response and phosphate starvation, matching the geological and climatic contrast between both localities: Aznalcóllar features hot, dry summers and is located in the the Iberian Pyrite Belt with acidic, sulfurous soils which limits phosphorus availability [126]. Conversely, the mountainous Monchique is a high-precipitation refuge with milder temperatures and less sulfurous soils [127]. This preliminary finding opens the door for future experimental studies to confirm and explore the potential role of these inversions in population differentiation and local adaptation.

Final Remarks

The synteny analysis characterized a striking number of CRs between the two typical karyotypic races of our *C. helodes* populations, suggesting a high rearrangement rate even for the already kariologically dynamic genus *Carex*. CRs establishment was likely facilitated by genetic bottlenecks during the colonization of southwestern Spain. Breakpoints were gene-poor and TE-rich, characteristics that facilitate the occurrence and/or limit the CRs deleteriousness, likely accounting for some of the high number of CRs. Rearranged chromosomes as a whole also have distinct genomic characteristics, such as an increased LINE content, supporting their possible role in CRs as was previously reported for *Carex*

[109]. Two inversions showed suggestive signs of distinct gene evolution in opposite directions. The inversion on chromosome 6 shows a pattern that is consistent with early signs of purifying selection, aligning with its enrichment in proteasomal activity, related to CR tolerance. The inversion on chromosome 32 shows increased divergence between the two genomes and was enriched in functions related to response to temperature stress and phosphorus deficit, two environmental factors that differ between the two localities.

Ultimately, the accumulation of CRs and breakpoints with specific genomic features, suggest rapid intraspecific karyotypic evolution in *C. helodes*. These mechanisms likely serve as the foundational architecture for eventual differentiation, where chromosomal barriers effectively sequester potential adaptive loci facilitating differential genomic differentiation.

Reference

1. Rieseberg LH. 2001 Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358. (doi:10.1016/S0169-5347(01)02187-5)
2. Faria R, Navarro A. 2010 Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669. (doi:10.1016/j.tree.2010.07.008)
3. White MJD (Michael JD. 1978 *Modes of speciation*. San Francisco : W. H. Freeman. See <http://archive.org/details/modesofspeciation00whit>.
4. King M. 1993 *Species Evolution*. Cambridge University Press. See <https://www.cambridge.org/universitypress/subjects/life-sciences/evolutionary-biology/species-evolution-role-chromosome-change>.
5. Coyne JA, Orr HA. 1998 The evolutionary genetics of speciation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **353**, 287–305. (doi:10.1098/rstb.1998.0210)
6. Futuyma DJ, Mayer GC. 1980 Non-Allopatric Speciation in Animals. *Syst. Zool.* **29**, 254–271. (doi:10.2307/2412661)
7. Sites JW, Moritz C. 1987 Chromosomal Evolution and Speciation Revisited. *Syst. Zool.* **36**, 153–174. (doi:10.2307/2413266)
8. Benítez-Benítez C *et al.* 2025 Bridging micro and macroevolution: Insights from chromosomal dynamics in plants. *Front. Plant Sci.* **16**. (doi:10.3389/fpls.2025.1606450)

- 596 9. Escudero M *et al.* 2014 Karyotypic Changes through Dysploidy Persist Longer over
597 Evolutionary Time than Polyploid Changes. *PLOS ONE* **9**, e85266.
598 (doi:10.1371/journal.pone.0085266)
- 599 10. Sader MA, Amorim BS, Costa L, Souza G, Pedrosa-Harand A. 2019 The role of
600 chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). *Syst.*
601 *Biodivers.* **17**, 7–21. (doi:10.1080/14772000.2018.1546777)
- 602 11. Carvalho CMB, Lupski JR. 2016 Mechanisms underlying structural variant formation in
603 genomic disorders. *Nat. Rev. Genet.* **17**, 224–238. (doi:10.1038/nrg.2015.25)
- 604 12. Lupski JR. 1998 Genomic disorders: structural features of the genome can lead to DNA
605 rearrangements and human disease traits. *Trends Genet. TIG* **14**, 417–422.
606 (doi:10.1016/s0168-9525(98)01555-8)
- 607 13. Shaw CJ, Lupski JR. 2004 Implications of human genome architecture for
608 rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* **13 Spec**
609 **No 1**, R57-64. (doi:10.1093/hmg/ddh073)
- 610 14. Stathos A, Fishman L. 2014 Chromosomal rearrangements directly cause
611 underdominant F1 pollen sterility in *Mimulus lewisii*–*Mimulus cardinalis* hybrids.
612 *Evolution* **68**, 3109–3119. (doi:10.1111/evo.12503)
- 613 15. Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001 Chromosomal inversions and the
614 reproductive isolation of species. *Proc. Natl. Acad. Sci.* **98**, 12084–12088.
615 (doi:10.1073/pnas.221274498)
- 616 16. Lowry DB, Willis JH. 2010 A widespread chromosomal inversion polymorphism
617 contributes to a major life-history transition, local adaptation, and reproductive isolation.
- 618 17. Trickett AJ, Butlin RK. 1994 Recombination suppressors and the evolution of new
619 species. *Heredity* **73 (Pt 4)**, 339–345. (doi:10.1038/hdy.1994.180)
- 620 18. Hedrick PW. 1981 The Establishment of Chromosomal Variants. *Evolution* **35**, 322–
621 332. (doi:10.1111/j.1558-5646.1981.tb04890.x)
- 622 19. Bush GL, Case SM, Wilson AC, Patton JL. 1977 Rapid Speciation and Chromosomal
623 Evolution in Mammals. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3942–3946.
- 624 20. Wilson AC, Bush GL, Case SM, King MC. 1975 Social structuring of mammalian
625 populations and rate of chromosomal evolution. *Proc. Natl. Acad. Sci.* **72**, 5061–5065.
626 (doi:10.1073/pnas.72.12.5061)
- 627 21. Charlesworth D, Wright SI. 2001 Breeding systems and genome evolution. *Curr. Opin.*
628 *Genet. Dev.* **11**, 685–690. (doi:10.1016/S0959-437X(00)00254-9)
- 629 22. Potter S, Bragg JG, Blom MPK, Deakin JE, Kirkpatrick M, Eldridge MDB, Moritz C.
630 2017 Chromosomal Speciation in the Genomics Era: Disentangling Phylogenetic
631 Evolution of Rock-wallabies. *Front. Genet.* **8**. (doi:10.3389/fgene.2017.00010)
- 632 23. Charlesworth B. 1983 Models of the evolution of some genetic systems. *Proc. R. Soc.*
633 *Lond. B Biol. Sci.* **219**, 265–279. (doi:10.1098/rspb.1983.0074)

1
2
3 634 24. Guerrero RF, Kirkpatrick M. 2014 Local Adaptation and the Evolution of Chromosome
4 635 Fusions. *Evolution* **68**, 2747–2756. (doi:10.1111/evo.12481)
5
6 636 25. Escudero M, Marques A, Lucek K, Hipp AL. 2024 Genomic hotspots of chromosome
7 637 rearrangements explain conserved synteny despite high rates of chromosome evolution
8 638 in a holocentric lineage. *Mol. Ecol.* **33**, e17086. (doi:10.1111/mec.17086)
9
10 639 26. Höök L, Näsvall K, Vila R, Wiklund C, Backström N. 2023 High-density linkage maps
11 640 and chromosome level genome assemblies unveil direction and frequency of extensive
12 641 structural rearrangements in wood white butterflies (*Leptidea* spp.). *Chromosome Res.*
13 642 **31**, 2. (doi:10.1007/s10577-023-09713-z)
14
15 643 27. Mandáková T, Lysak MA. 2018 Post-polyploid diploidization and diversification
16 644 through dysploid changes. *Curr. Opin. Plant Biol.* **42**, 55–65.
17 645 (doi:10.1016/j.pbi.2018.03.001)
18
19 646 28. Li W, Challa GS, Zhu H, Wei W. 2016 Recurrence of Chromosome Rearrangements
20 647 and Reuse of DNA Breakpoints in the Evolution of the Triticeae Genomes. *G3*
21 648 *GenesGenomesGenetics* **6**, 3837–3847. (doi:10.1534/g3.116.035089)
22
23 649 29. Gokcumen O *et al.* 2011 Refinement of primate copy number variation hotspots
24 650 identifies candidate genomic regions evolving under positive selection. *Genome Biol.*
25 651 **12**, R52. (doi:10.1186/gb-2011-12-5-r52)
26
27 652 30. Perry GH *et al.* 2006 Hotspots for copy number variation in chimpanzees and humans.
28 653 *Proc. Natl. Acad. Sci.* **103**, 8006–8011. (doi:10.1073/pnas.0602318103)
29
30 654 31. Raskina O, Barber JC, Nevo E, Belyayev A. 2008 Repetitive DNA and chromosomal
31 655 rearrangements: speciation-related events in plant genomes. *Cytogenet. Genome Res.*
32 656 **120**, 351–357. (doi:10.1159/000121084)
33
34 657 32. Sutherland GR, Baker E, Richards RI. 1998 Fragile sites still breaking. *Trends Genet.*
35 658 *TIG* **14**, 501–506. (doi:10.1016/s0168-9525(98)01628-x)
36
37 659 33. Lin Y-L, Gokcumen O. 2019 Fine-Scale Characterization of Genomic Structural
38 660 Variation in the Human Genome Reveals Adaptive and Biomedically Relevant
39 661 Hotspots. *Genome Biol. Evol.* **11**, 1136–1151. (doi:10.1093/gbe/evz058)
40
41 662 34. Burssed B, Zamariolli M, Bellucco FT, Melaragno MI. 2022 Mechanisms of structural
42 663 chromosomal rearrangement formation. *Mol. Cytogenet.* **15**, 23. (doi:10.1186/s13039-
43 664 022-00600-6)
44
45 665 35. Schmidt C, Schindele P, Puchta H. 2019 From gene editing to genome engineering:
46 666 restructuring plant chromosomes via CRISPR/Cas. *aBIOTECH* **1**, 21–31.
47 667 (doi:10.1007/s42994-019-00002-0)
48
49 668 36. Bureš P, Zedek F, Marková M. 2013 Holocentric Chromosomes. In *Plant Genome*
50 669 *Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes* (eds
51 670 J Greilhuber, J Dolezel, JF Wendel), pp. 187–208. Vienna: Springer. (doi:10.1007/978-
52 671 3-7091-1160-4_12)
53
54
55
56
57
58
59
60

- 672 37. Hipp AL, Rothrock PE, Roalson EH. 2009 The Evolution of Chromosome
673 Arrangements in *Carex* (Cyperaceae). *Bot. Rev.* **75**, 96–109. (doi:10.1007/s12229-008-
674 9022-8)
- 675 38. Marques A *et al.* 2015 Holocentromeres in *Rhynchospora* are associated with genome-
676 wide centromere-specific repeat arrays interspersed among euchromatin. *Proc. Natl.*
677 *Acad. Sci. U. S. A.* **112**, 13633–13638. (doi:10.1073/pnas.1512255112)
- 678 39. Faulkner JS. 1972 Chromosome studies on *Carex* section *Acutae* in north-west Europe.
679 *Bot. J. Linn. Soc.* **65**, 271–301. (doi:10.1111/j.1095-8339.1972.tb00120.x)
- 680 40. Luceño M. 1993 Chromosome Studies on *Carex* L. Section *Mitratae* Kükenth.
681 (Cyperaceae) in the Iberian Peninsula. *Cytologia (Tokyo)* **58**, 321–330.
682 (doi:10.1508/CYTOLOGIA.58.321)
- 683 41. Davies EW. 1956 Cytology, Evolution and Origin of the Aneuploid Series in the Genus
684 *Carex*. *Hereditas* **42**, 349–365. (doi:10.1111/j.1601-5223.1956.tb03022.x)
- 685 42. Schmid B. 2008 Karyology and hybridization in the *Carex flava* complex in
686 Switzerland. *Feddes Repert.* **93**, 23–59. (doi:10.1002/fedr.19820930103)
- 687 43. Grbić M *et al.* 2011 The genome of *Tetranychus urticae* reveals herbivorous pest
688 adaptations. *Nature* **479**, 487–492. (doi:10.1038/nature10640)
- 689 44. Hofstatter PG *et al.* 2022 Repeat-based holocentromeres influence genome architecture
690 and karyotype evolution. *Cell* **185**, 3153–3168.e18. (doi:10.1016/j.cell.2022.06.045)
- 691 45. Mandrioli M, Azzoni P, Lombardo G, Manicardi GC. 2011 Composition and Epigenetic
692 Markers of Heterochromatin in the Aphid *Aphis nerii* (Hemiptera: Aphididae).
693 *Cytogenet. Genome Res.* **133**, 67–77. (doi:10.1159/000323510)
- 694 46. Mandrioli M, Carlo Manicardi G. 2012 Unlocking Holocentric Chromosomes: New
695 Perspectives from Comparative and Functional Genomics? *Curr. Genomics* **13**, 343–
696 349. (doi:10.2174/138920212801619250)
- 697 47. Bernardi G, Olofsson B, Filipinski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M,
698 Rodier F. 1985 The mosaic genome of warm-blooded vertebrates. *Science* **228**, 953–
699 958. (doi:10.1126/science.4001930)
- 700 48. Dujon B *et al.* 1994 Complete DNA sequence of yeast chromosome XI. *Nature* **369**,
701 371–378. (doi:10.1038/369371a0)
- 702 49. Mandrioli M, Manicardi GC. 2012 Unlocking Holocentric Chromosomes: New
703 Perspectives from Comparative and Functional Genomics? *Curr. Genomics* **13**, 343–
704 349. (doi:10.2174/138920212801619250)
- 705 50. Lucek K, Augustijnen H, Escudero M. 2022 A holocentric twist to chromosomal
706 speciation? *Trends Ecol. Evol.* **37**, 655–662. (doi:10.1016/j.tree.2022.04.002)
- 707 51. Tanaka N. 1939 Chromosome Studies in Cyperaceae, IV. *Cytologia (Tokyo)* **10**, 51–58.
708 (doi:10.1508/cytologia.10.51)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

709 52. Hipp AL. 2007 Nonuniform processes of chromosome evolution in sedges (Carex:
710 Cyperaceae). *Evol. Int. J. Org. Evol.* **61**, 2175–2194. (doi:10.1111/j.1558-
711 5646.2007.00183.x)

712 53. Roalson EH. 2008 A Synopsis of Chromosome Number Variation in the Cyperaceae.
713 *Bot. Rev.* **74**, 209–393.

714 54. Escudero M, Hipp AL, Luceño M. 2010 Karyotype stability and predictors of
715 chromosome number variation in sedges: A study in *Carex* section *Spirostachyae*
716 (Cyperaceae). *Mol. Phylogenet. Evol.* **57**, 353–363. (doi:10.1016/j.ympev.2010.07.009)

717 55. Escudero M, Hipp AL, Waterway MJ, Valente LM. 2012 Diversification rates and
718 chromosome evolution in the most diverse angiosperm genus of the temperate zone
719 (Carex, Cyperaceae). *Mol. Phylogenet. Evol.* **63**, 650–655.
720 (doi:10.1016/j.ympev.2012.02.005)

721 56. Hipp AL, Rothrock PE, Whitkus R, Weber JA. 2010 Chromosomes tell half of the story:
722 the correlation between karyotype rearrangements and genetic diversity in sedges, a
723 group with holocentric chromosomes. *Mol. Ecol.* **19**, 3124–3138. (doi:10.1111/j.1365-
724 294X.2010.04741.x)

725 57. Tribble CM, Márquez-Corro JI, May MR, Hipp AL, Escudero M, Zenil-Ferguson R.
726 2025 Macroevolutionary inference of complex modes of chromosomal speciation in a
727 cosmopolitan plant lineage. *New Phytol.* **245**, 2350–2361. (doi:10.1111/nph.20353)

728 58. Márquez-Corro JI, Martín-Bravo S, Jiménez-Mejías P, Hipp AL, Spalink D, Naczi RFC,
729 Roalson EH, Luceño M, Escudero M. 2021 Macroevolutionary insights into sedges (
730 *Carex* : Cyperaceae): The effects of rapid chromosome number evolution on lineage
731 diversification. *J. Syst. Evol.* **59**, 776–790. (doi:10.1111/jse.12730)

732 59. Escudero M, Valcárcel V, Vargas P, Luceño M. 2008 Evolution in *Carex* L. sect.
733 *Spirostachyae* (Cyperaceae): A molecular and cytogenetic approach. *Org. Divers. Evol.*
734 **7**, 271–291. (doi:10.1016/j.ode.2006.08.006)

735 60. Escudero M, Arroyo JM, Sánchez-Ramírez S, Jordano P. 2024 Founder events and
736 subsequent genetic bottlenecks underlie karyotype evolution in the Ibero-North African
737 endemic *Carex helodes*. *Ann. Bot.* **133**, 871–882. (doi:10.1093/aob/mcad087)

738 61. Villaverde T *et al.* 2020 A new classification of *Carex* (Cyperaceae) subgenera
739 supported by a HybSeq backbone phylogenetic tree. *Bot. J. Linn. Soc.* **194**, 141–163.
740 (doi:10.1093/botlinnean/boaa042)

741 62. Escudero M, Vargas P, Valcárcel V, Garcés M. 2008 Strait of Gibraltar: An effective
742 gene-flow barrier for wind-pollinated *Carex helodes* (Cyperaceae) as revealed by DNA
743 sequences, AFLP, and cytogenetic variation. *Am. J. Bot.* **95**, 745–55.
744 (doi:10.3732/ajb.2007342)

745 63. De Coster W, Rademakers R. 2023 NanoPack2: population-scale evaluation of long-
746 read sequencing data. *Bioinformatics* **39**, btad311. (doi:10.1093/bioinformatics/btad311)

64. Marçais G, Kingsford C. 2011 A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770. (doi:10.1093/bioinformatics/btr011)
65. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017 GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204. (doi:10.1093/bioinformatics/btx153)
66. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021 Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175. (doi:10.1038/s41592-020-01056-5)
67. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020 Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898. (doi:10.1093/bioinformatics/btaa025)
68. ArimaGenomics. 2023 mapping_pipeline/Arima_Mapping_UserGuide_A160156_v03.pdf at master · ArimaGenomics/mapping_pipeline. *GitHub*. See https://github.com/ArimaGenomics/mapping_pipeline/blob/master/Arima_Mapping_UserGuide_A160156_v03.pdf (accessed on 9 February 2026).
69. Zhou C, McCarthy SA, Durbin R. 2022 YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808. (doi:10.1093/bioinformatics/btac808)
70. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016 Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101. (doi:10.1016/j.cels.2015.07.012)
71. Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, Fedrigo O, Jarvis ED. 2022 Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics* **38**, 4214–4216. (doi:10.1093/bioinformatics/btac460)
72. Huang N, Li H. 2023 compleasm: a faster and more accurate reimplement of BUSCO. *Bioinformatics* **39**, btad595. (doi:10.1093/bioinformatics/btad595)
73. Rhie A, Walenz BP, Koren S, Phillippy AM. 2020 Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245. (doi:10.1186/s13059-020-02134-9)
74. Karlicki M, Antonowicz S, Karnkowska A. 2022 Tiara: deep learning-based classification system for eukaryotic sequences. *Bioinformatics* **38**, 344–350. (doi:10.1093/bioinformatics/btab672)
75. Brown MR, Manuel Gonzalez de La Rosa P, Blaxter M. 2025 tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics* **41**, btaf049. (doi:10.1093/bioinformatics/btaf049)
76. Luceño Garcés M. 1988 Notas cariológicas, III. *An. Jardín Botánico Madr.* **45**, 189–196.

1
2
3 786 77. Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021 BRAKER2: automatic
4 787 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a
5 788 protein database. *NAR Genomics Bioinforma.* **3**, lqaa108. (doi:10.1093/nargab/lqaa108)
6
7
8 789 78. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 BUSCO:
9 790 assessing genome assembly and annotation completeness with single-copy orthologs.
10 791 *Bioinformatics* **31**, 3210–3212. (doi:10.1093/bioinformatics/btv351)
11
12 792 79. Baril T, Galbraith J, Hayward A. 2024 Earl Grey: A Fully Automated User-Friendly
13 793 Transposable Element Annotation and Analysis Pipeline. *Mol. Biol. Evol.* **41**, msae068.
14 794 (doi:10.1093/molbev/msae068)
15
16 795 80. Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D,
17 796 Goodstein DM, Schmutz J. 2022 GENESPACE tracks regions of interest and gene copy
18 797 number variation across multiple genomes. *eLife* **11**, e78526. (doi:10.7554/eLife.78526)
19
20
21 798 81. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov
22 799 JP. 2011 Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26.
23 800 (doi:10.1038/nbt.1754)
24
25 801 82. Treangen TJ, Salzberg SL. 2011 Repetitive DNA and next-generation sequencing:
26 802 computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46.
27 803 (doi:10.1038/nrg3117)
28
29 804 83. Kassambara A, Patil I. 2023 ggcorrplot: Visualization of a Correlation Matrix using
30 805 ‘ggplot2’.
31
32
33 806 84. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT,
34 807 Carey VJ. 2013 Software for computing and annotating genomic ranges. *PLoS Comput.*
35 808 *Biol.* **9**, e1003118. (doi:10.1371/journal.pcbi.1003118)
36
37 809 85. Venables WN, Ripley BD. 2002 *Modern Applied Statistics with S*. New York, NY:
38 810 Springer. (doi:10.1007/978-0-387-21706-2)
39
40
41 811 86. Bates D, Mächler M, Bolker B, Walker S. 2015 Fitting Linear Mixed-Effects Models
42 812 Using lme4. *J. Stat. Softw.* **67**, 1–48. (doi:10.18637/jss.v067.i01)
43
44 813 87. McGillicuddy M, Popovic G, Bolker BM, Warton DI. 2025 Parsimoniously Fitting
45 814 Large Multivariate Random Effects in glmmTMB. *J. Stat. Softw.* **112**, 1–19.
46 815 (doi:10.18637/jss.v112.i01)
47
48 816 88. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. 2016
49 817 regioneR: an R/Bioconductor package for the association analysis of genomic regions
50 818 based on permutation tests. *Bioinforma. Oxf. Engl.* **32**, 289–291.
51 819 (doi:10.1093/bioinformatics/btv562)
52
53
54 820 89. Navarro A, Ruiz A. 1997 On the Fertility Effects of Pericentric Inversions. *Genetics*
55 821 **147**, 931–933. (doi:10.1093/genetics/147.2.931)
56
57 822 90. White MJD (Michael JD. 1973 *Animal cytology and evolution*. Cambridge, [Eng.]
58 823 University Press. See <http://archive.org/details/animalcytologyev0000whit>.

- 824 91. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7:
825 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
826 (doi:10.1093/molbev/mst010)
- 827 92. Suyama M, Torrents D, Bork P. 2006 PAL2NAL: robust conversion of protein sequence
828 alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–612.
829 (doi:10.1093/nar/gkl315)
- 830 93. Yang Z. 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol.*
831 *Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
- 832 94. Alexa A, Rahnenführer J, Marini F. 2022 topGO: Enrichment Analysis for Gene
833 Ontology. See <https://bioc.r-universe.dev/topGO> (accessed on 15 December 2025).
- 834 95. Huerta-Cepas J *et al.* 2019 eggNOG 5.0: a hierarchical, functionally and
835 phylogenetically annotated orthology resource based on 5090 organisms and 2502
836 viruses. *Nucleic Acids Res.* **47**, D309–D314. (doi:10.1093/nar/gky1085)
- 837 96. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021
838 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
839 Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829.
840 (doi:10.1093/molbev/msab293)
- 841 97. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. 2010 GOSemSim: an R package for
842 measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**,
843 976–978. (doi:10.1093/bioinformatics/btq064)
- 844 98. Márquez-Corro JJ, Christenhusz MJM, Fay MF. 2026 The genome sequence of the
845 starved wood-sedge, *Carex depauperata* Curtis ex Woodw. (Poales: Cyperaceae)
846 [version 1; peer review: 1 approved, 2 approved with reservations]. *Wellcome Open Res.*
847 **11**. (doi:10.12688/wellcomeopenres.26113.1)
- 848 99. Ruhsam M, Griffiths A. 2025 The genome sequence of Bottle Sedge, *Carex rostrata*
849 Stokes. *Wellcome Open Res.* **10**, 132. (doi:10.12688/wellcomeopenres.23874.1)
- 850 100. Ohbayashi K, Hodoki Y, Kunii H. 2012 Estimation of the Genetic Composition of a
851 Near-Threatened Tidal Marsh Plant, *Carex rugulosa*, in Japan. *Wetlands* **32**, 175–184.
852 (doi:10.1007/s13157-011-0266-4)
- 853 101. Escudero M, Lovit M, Brown BH, Hipp AL. 2019 Rapid plant speciation associated
854 with the last glacial period: reproductive isolation and genetic drift in sedges. *Bot. J.*
855 *Linn. Soc.* **190**, 303–314. (doi:10.1093/botlinnean/boz016)
- 856 102. Allesio Leck M, Schütz W. 2005 Regeneration of Cyperaceae, with particular reference
857 to seed ecology and seed banks. *Perspect. Plant Ecol. Evol. Syst.* **7**, 95–133.
858 (doi:10.1016/j.ppees.2005.05.001)
- 859 103. Escudero M, Hahn M, Brown BH, Lueders K, Hipp AL. 2016 Chromosomal
860 rearrangements in holocentric organisms lead to reproductive isolation by hybrid
861 dysfunction: The correlation between karyotype rearrangements and germination rates in
862 sedges. *Am. J. Bot.* **103**, 1529–1536. (doi:10.3732/ajb.1600051)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

104. Whitkus R. 1988 Experimental Hybridizations Among Chromosome Races of *Carex* pachystachya and the Related Species *C. macloviana* and *C. preslii* (Cyperaceae). *Syst. Bot.* **13**, 146–153. (doi:10.2307/2419251)

105. Narbona E, Delgado A, Encina F, Miguez M, Buide ML. 2013 Seed germination and seedling establishment of the rare *Carex helodes* Link depend on the proximity to water. *Aquat. Bot.* **110**, 55–60. (doi:10.1016/j.aquabot.2013.05.005)

106. Lovell JT *et al.* 2025 Comparative Analyses of Four Reference Genomes Reveal Exceptional Diversity and Weak Linked Selection in the Yellow Monkeyflower (*Mimulus guttatus*) Complex. *Mol. Ecol. Resour.* **25**, e70012. (doi:10.1111/1755-0998.70012)

107. Huang C *et al.* 2025 A comparative genomic analysis at the chromosomal-level reveals evolutionary patterns of aphid chromosomes. *Commun. Biol.* **8**, 427. (doi:10.1038/s42003-025-07851-0)

108. Ahola V *et al.* 2014 The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat. Commun.* **5**, 4737. (doi:10.1038/ncomms5737)

109. Cornet C, Mora P, Augustijnen H, Nguyen P, Escudero M, Lucek K. 2024 Holocentric repeat landscapes: From micro-evolutionary patterns to macro-evolutionary associations with karyotype evolution. *Mol. Ecol.* **33**, e17100. (doi:10.1111/mec.17100)

110. Dolgin ES, Charlesworth B. 2008 The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**, 2169–2177. (doi:10.1534/genetics.107.082743)

111. Gray YHM. 2000 It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **16**, 461–468. (doi:10.1016/S0168-9525(00)02104-1)

112. Bartolomé C, Maside X. 2004 The lack of recombination drives the fixation of transposable elements on the fourth chromosome of *Drosophila melanogaster*. *Genet. Res.* **83**, 91–100. (doi:10.1017/S0016672304006755)

113. Kent TV, Uzunović J, Wright SI. 2017 Coevolution between transposable elements and recombination. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160458. (doi:10.1098/rstb.2016.0458)

114. Ayala FJ, Coluzzi M. 2005 Chromosome speciation: Humans, *Drosophila*, and mosquitoes. *Proc. Natl. Acad. Sci.* **102**, 6535–6542. (doi:10.1073/pnas.0501847102)

115. Hill WG, Robertson A. 1968 Linkage disequilibrium in finite populations. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **38**, 226–231. (doi:10.1007/BF01245622)

116. Ohta T, Kimura M. 1969 Linkage disequilibrium due to random genetic drift. *Genet. Res.* **13**, 47–55. (doi:10.1017/S001667230000272X)

117. Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9161–9166. (doi:10.1073/pnas.151244298)
118. Wright SI, Foxe JP, DeRose-Wilson L, Kawabe A, Looseley M, Gaut BS, Charlesworth D. 2006 Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* **174**, 1421–1430. (doi:10.1534/genetics.106.062588)
119. Cuypers B *et al.* 2022 Four layer multi-omics reveals molecular responses to aneuploidy in *Leishmania*. *PLOS Pathog.* **18**, e1010848. (doi:10.1371/journal.ppat.1010848)
120. Muenzner J *et al.* 2024 Natural proteome diversity links aneuploidy tolerance to protein turnover. *Nature* **630**, 149–157. (doi:10.1038/s41586-024-07442-9)
121. Schukken KM, Sheltzer JM. 2022 Extensive protein dosage compensation in aneuploid human cancers. *Genome Res.* **32**, 1254–1270. (doi:10.1101/gr.276378.121)
122. Zhang A, Lyu YL, Lin C-P, Zhou N, Azarova AM, Wood LM, Liu LF. 2006 A protease pathway for the repair of topoisomerase II-DNA covalent complexes. *J. Biol. Chem.* **281**, 35997–36003. (doi:10.1074/jbc.M604149200)
123. Krogan NJ *et al.* 2004 Proteasome Involvement in the Repair of DNA Double-Strand Breaks. *Mol. Cell* **16**, 1027–1034. (doi:10.1016/j.molcel.2004.11.033)
124. Gudmundsdottir K, Lord CJ, Ashworth A. 2007 The proteasome is involved in determining differential utilization of double-strand break repair pathways. *Oncogene* **26**, 7601–7606. (doi:10.1038/sj.onc.1210579)
125. Gnugnoli M, Rinaldi C, Casari E, Pizzul P, Bonetti D, Longhese MP. 2024 Proteasome-mediated degradation of long-range nucleases negatively regulates resection of DNA double-strand breaks. *iScience* **27**, 110373. (doi:10.1016/j.isci.2024.110373)
126. Almodóvar GR, Sáez R, Pons JM, Maestre A, Toscano M, Pascual E. 1997 Geology and genesis of the Aznalcóllar massive sulphide deposits, Iberian Pyrite Belt, Spain. *Miner. Deposita* **33**, 111–136. (doi:10.1007/s001260050136)
127. Stanislawski D. 1962 The Monchique of Southern Portugal. *Geogr. Rev.* **52**, 37–55. (doi:10.2307/213004)
128. Sanz-Arnal M *et al.* 2025 A Comprehensive Database of Expert-Curated Occurrences for the Genus *Carex* L. (Cyperaceae). *Glob. Ecol. Biogeogr.* **34**, e70123. (doi:10.1111/geb.70123)

Acknowledgements

The authors wish to thank Ms Nafiseh Sargheini for helping and teaching with Hi-C extraction and Ms Alegria Montero for performing DNA extraction, Dr Ignacio Márquez-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Corro for helping with field work and caring for the plants in the greenhouse and Mr Diego Herrero Doblado continuing greenhouse work, performing the experimental crosses and estimating germination success.

Funding information

This work was funded by the Spanish Government and FEDER funds (EuropeanCommission) through projects PID2021-122715NB-I00 and PID2024-157198NB-I00 and by the Andalusian regional government (Spain) through project ProyExcel_00125. The latter also supported SM-B and RS-V. IG-R was supported by the Andalusian Regional Ministry of Economy, Knowledge, Business, and University (PREDOC_00632). KL was supported by the Swiss National Science Foundation grants PCEFP3_202869 and 220868 as well as the Fondation Pierre Mercier pour la Science and the Fondation du Jardin Botanique de Neuchâtel.

Figures captions

Figure 1: Geographic locations, meiotic configurations, and chromosomal synteny between *Carex helodes* parental genomes.

(a) Satellite image of southern Spain and Portugal and northern Morocco with *Carex helodes* locations indicated [60,62,128]. Those included in our study are indicated, including previously reported chromosome numbers [60,62]. (b) Meiotic configurations of *C. helodes* parentals, MON (Caldas de Monchique, Algarve, Portugal) and AZN (Aznalcollar, Spain), and their F1 hybrid. Aberrant configurations are indicated by arrows. (c) Synteny plot generated with GeneSpace showing conserved gene collinearity and rearrangements between the genome assemblies of the parentals MON and AZN. Chromosomes are scaled by gene

rank order Chromosomes with inversions are highlighted in yellow and interchromosomal rearrangements (translocations, fusions and fissions) in purple and green tones.

Figure 2: Comparison of genomic feature distributions between breakpoint and non-breakpoint regions, and across rearranged and collinear chromosomes.

Violin and jitter plots showing the distribution of genomic features in breakpoint regions versus the rest of the genome (*a*), and in rearranged versus conserved chromosomes (*b*), using a 100kb window distribution. Statistical significance is indicated with asterisks ($p < 0.1$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$), corresponding to the linear model comparisons. For each genomic variable, the number of elements per window is represented, except for GC content for which the proportion is represented. Abbreviations as follows: RE: Repetitive elements, TEs: Transposable Elements, LTRs: Long Terminal Repeat Retrotransposons and LINEs: Long Interspersed Nuclear Repeats.

Figure 3: Functional enrichment and semantic clustering of Gene Ontology (GO) terms for genomic rearrangements

Semantic clustering and enrichment of Gene Ontology (GO) terms for all rearrangements (*a*), inversion in chr 6 (*b*) and chr 32 (*c*) for MON. The three function types are represented from left to right: Molecular Function (MF), Cellular Component (CC) and Biological Process (BP). Molecular Function and CC are represented through scatter plots where the x-axis represents the number of annotated genes throughout the genome and y-axis the number of genes in the area affected by the rearrangement. BP is represented through semantic plots, visualizing it through the 2D semantic space via Multidimensional Scaling (MDS). Highly redundant terms were grouped using GOSemSim (Resnik measure based on *Arabidopsis*

1
2
3 986 *thaliana* annotations) and were labeled by the most representative term. In all the plots, point
4
5 987 size indicates enrichment factor and color represents $-\log_{10}(\text{p-value})$.
6
7
8 988

9
10 989 **Tables captions**

11
12 990 **Table 1: Enrichment permutation test by genomic region for the (a) MON and (b) AZN**
13
14 991 **genomes.** Breakpoints are compared against the rest of the rearranged chromosome within the
15
16 992 same genomic region. For each genomic feature, the table reports the z-score, observed value,
17
18 993 95% confidence intervals (CIs) fold enrichment and p-value. The following abbreviations
19
20 994 have been used: REs: Repetitive elements, TEs: Transposable Elements, LTRs: Long
21
22 995 Terminal Repeat Retrotransposons, DNA trans: DNA transposons; LINEs: Long Interspersed
23
24 996 Nuclear Elements; SINE: Short Interspersed Nuclear Elements and TRC: Tandem Repeat
25
26 997 Cluster.
27
28
29

30 998
31
32 999 **Table 2: Gene evolution permutation test inside areas affected by chromosomal**
33
34 1000 **rearrangements.** Synonymous (dS), nonsynonymous (dN) substitution rates and their ratio
35
36 1001 ($\omega = \text{dN}/\text{dS}$) are calculated through aggregated analyses, where substitutions and sites were
37
38 1002 summed across coding sequences (CDS) within each rearrangement and compared with areas
39
40 1003 of the same size permuted across the chromosome of the rearrangement.
41
42
43

44 1004
45
46 1005
47
48 1006
49
50 1007
51
52
53
54
55
56
57
58
59
60

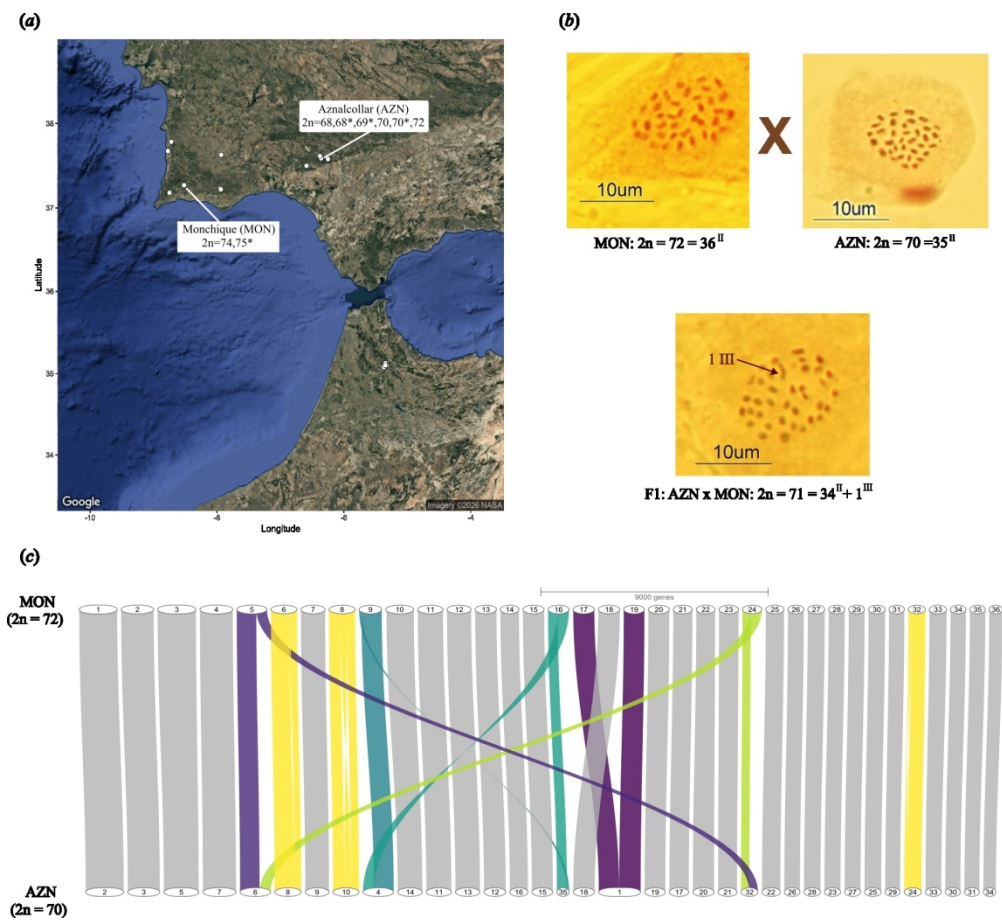


Figure 1: Geographic locations, meiotic configurations, and chromosomal synteny between *Carex helodes* parental genomes.

(a) Satellite image of southern Spain and Portugal and northern Morocco with *Carex helodes* locations indicated [60,62,128]. Those included in our study are indicated, including previously reported chromosome numbers [60,62]. (b) Meiotic configurations of *C. helodes* parents, MON (Caldas de Monchique, Algarve, Portugal) and AZN (Aznalcollar, Spain), and their F1 hybrid. Aberrant configurations are indicated by arrows.

(c) Synteny plot generated with GeneSpace showing conserved gene collinearity and rearrangements between the genome assemblies of the parents MON and AZN. Chromosomes are scaled by gene rank order. Chromosomes with inversions are highlighted in yellow and interchromosomal rearrangements (translocations, fusions and fissions) in purple and green tones.

168x152mm (300 x 300 DPI)

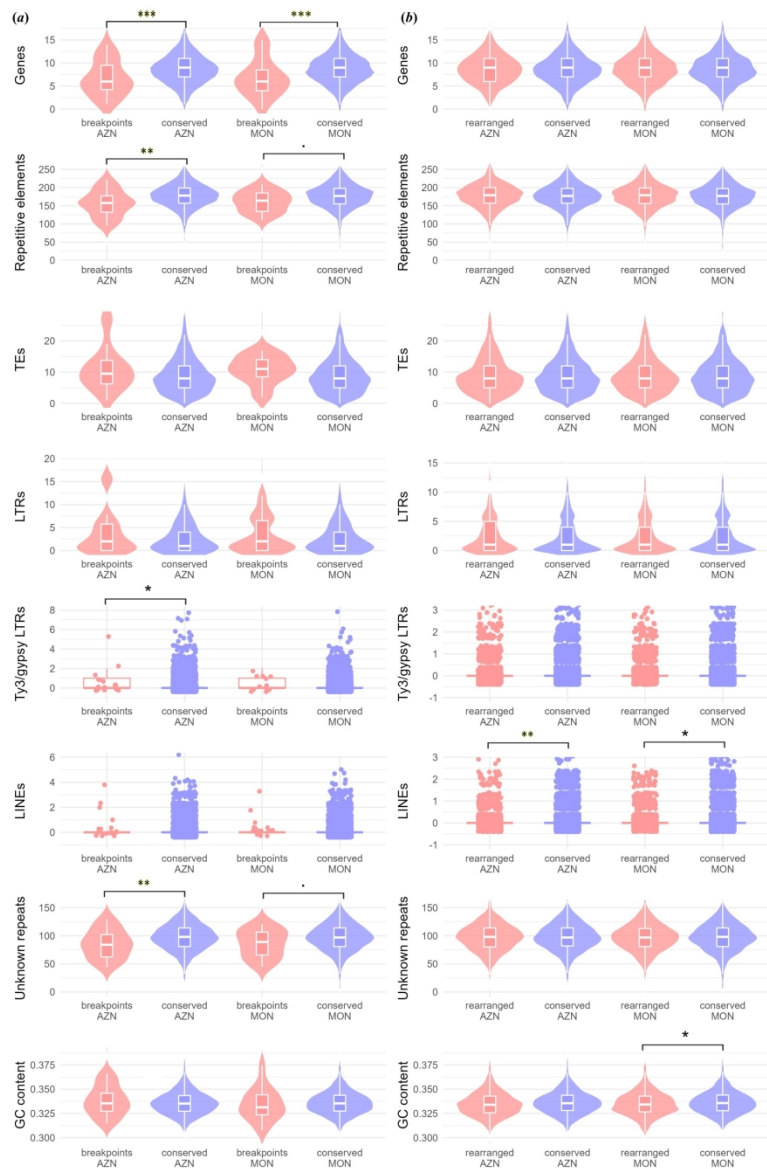


Figure 2: Comparison of genomic feature distributions between breakpoint and non-breakpoint regions, and across rearranged and collinear chromosomes. Violin and jitter plots showing the distribution of genomic features in breakpoint regions versus the rest of the genome (a), and in rearranged versus conserved chromosomes (b), using a 100kb window distribution. Statistical significance is indicated with asterisks (. $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$), corresponding to the linear model comparisons. For each genomic variable, the number of elements per window is represented, except for GC content for which the proportion is represented. Abbreviations as follows: RE: Repetitive elements, TEs: Transposable Elements, LTRs: Long Terminal Repeat Retrotransposons and LINEs: Long Interspersed Nuclear Repeats.

171x262mm (300 x 300 DPI)

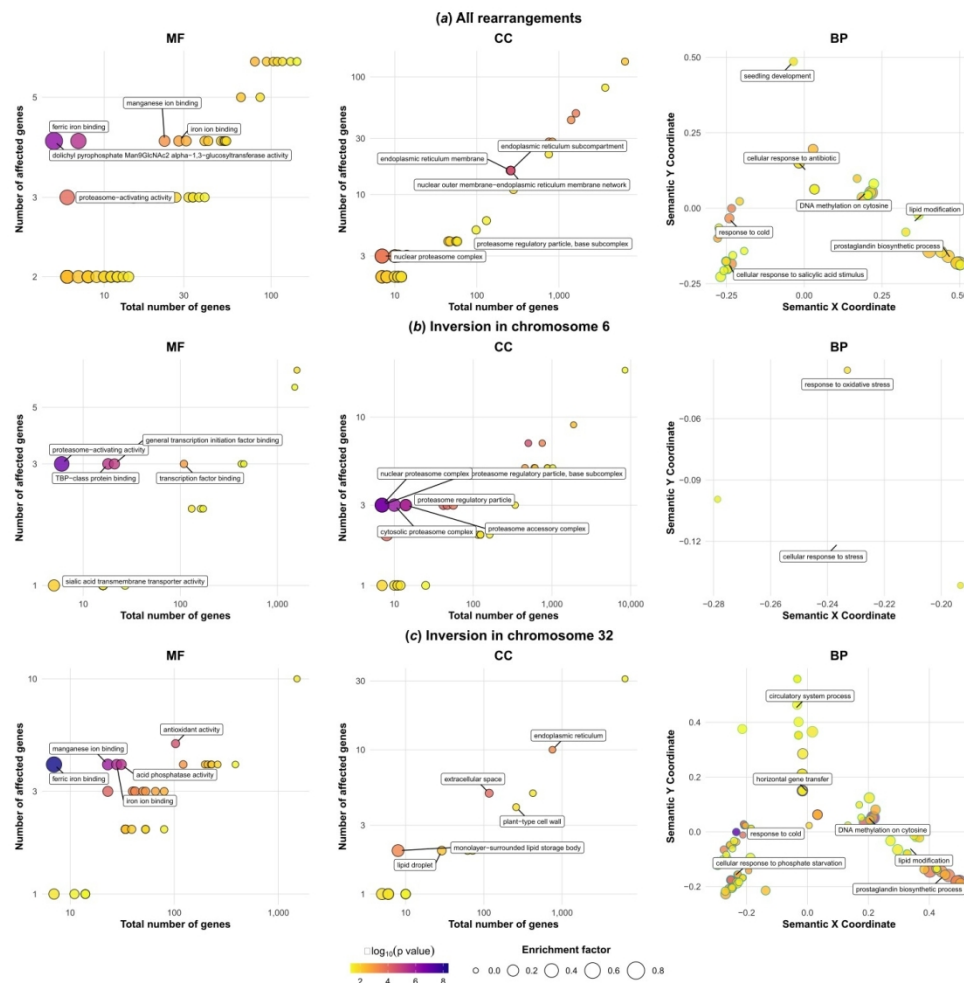


Figure 3: Functional enrichment and semantic clustering of Gene Ontology (GO) terms for genomic rearrangements

Semantic clustering and enrichment of Gene Ontology (GO) terms for all rearrangements (a), inversion in chr 6 (b) and chr 32 (c) for MON. The three function types are represented from left to right: Molecular Function (MF), Cellular Component (CC) and Biological Process (BP). Molecular Function and CC are represented through scatter plots where the x-axis represents the number of annotated genes throughout the genome and y-axis the number of genes in the area affected by the rearrangement. BP is represented through semantic plots, visualizing it through the 2D semantic space via Multidimensional Scaling (MDS). Highly redundant terms were grouped using GOSemSim (Resnik measure based on Arabidopsis thaliana annotations) and were labeled by the most representative term. In all the plots, point size indicates enrichment factor and color represents $-\log_{10}(\text{p-value})$.

407x412mm (96 x 96 DPI)

Table 1: Region-dependent permutation test for breakpoints enrichment in different genomic features for (a) MON and (b) AZN genomes.

Breakpoints are compared against the rest of the rearranged chromosome within the same genomic region. For each genomic feature, the table reports the z-score, observed value, 95% confidence intervals (CIs) fold enrichment and p-value. The following abbreviations have been used: REs: Repetitive elements, TEs: Transposable Elements, LTRs: Long Terminal Repeat Retrotransposons, DNA trans: DNA transposons; LINEs: Long Interspersed Nuclear Elements; SINE: Short Interspersed Nuclear Elements and TRC: Tandem Repeat Cluster.

a) MON genome						
Genomic feature	Region	z-value	Observed	95% CIs	Fold Enrichment	p-value
Genes	Interstitial	-2.27	22.00	[25, 50]	0.61	0.008
	Subtelomeric	-0.29	10.00	[5, 19]	0.91	0.481
REs	Interstitial	-0.42	686.00	[590, 822]	0.96	0.343
	Subtelomeric	-1.10	177.00	[142, 273.02]	0.83	0.133
TEs	Interstitial	1.51	53.00	[17, 60.02]	1.47	0.081
	Subtelomeric	-0.15	10.00	[2, 24]	1.00	0.518
LTRs	Interstitial	1.53	21.00	[1, 25]	2.10	0.091
	Subtelomeric	0.17	4.00	[0, 12]	1.67	0.390
Ty1/copia	Interstitial	0.98	6.00	[0, 9]	1.50	0.213
	Subtelomeric	-0.18	1.00	[0, 4]	1.00	0.660
Ty3/gypsy	Interstitial	0.54	2.00	[0, 5]	1.50	0.367
	Subtelomeric	-0.53	0.00	[0, 2]	1.00	0.723
DNA transp	Interstitial	0.88	3.00	[0, 5]	1.50	0.259
	Subtelomeric	0.70	1.00	[0, 2]	2.00	0.401
LINEs	Interstitial	-0.85	0.00	[0, 4]	0.50	0.444
	Subtelomeric	-0.49	0.00	[0, 2]	1.00	0.768
SINEs	Interstitial	-0.45	0.00	[0, 2]	1.00	0.760

	Subtelomeric	-0.27	0.00	[0, 1]	1.00	0.918
Unknown	Interstitial	-0.19	385.00	[305.97, 478]	0.98	0.432
	Subtelomeric	-0.82	98.00	[72.97, 163.02]	0.84	0.215
TRC	Interstitial	-0.65	0.00	[0, 3]	1.00	0.624
	Subtelomeric	2.05	1.00	[0, 1]	2.00	0.115
Most common TRC	Interstitial	-0.55	0.00	[0, 2]	1.00	0.710
	Subtelomeric	2.65	1.00	[0, 1]	2.00	0.086
GC content	Interstitial	-1.37	0.33	[0.32, 0.35]	0.98	0.083
	Subtelomeric	0.84	0.34	[0.32, 0.35]	1.02	0.192
b) AZN genome						
Genomic feature	Region	z-value	Observed	95% CIs	Fold Enrichment	p-value
Genes	Interstitial	-3.45	17.00	[28, 54]	0.42	0.00
	Subtelomeric	0.16	4.00	[0, 8]	1.33	0.49
REs	Interstitial	-1.46	690.00	[657.95, 900.02]	0.88	0.09
	Subtelomeric	-1.56	40.00	[28.98, 103.02]	0.57	0.06
TEs	Interstitial	2.79	73.00	[20, 64]	1.82	0.00
	Subtelomeric	2.42	12.00	[0, 12]	4.00	0.03
LTRs	Interstitial	2.83	31.00	[2, 28.02]	2.82	0.01
	Subtelomeric	3.45	8.00	[0, 7]	4.27	0.01
Ty1/copia	Interstitial	1.29	7.00	[1, 9]	1.75	0.16
	Subtelomeric	2.43	2.00	[0, 2]	2.15	0.07

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Ty3/gypsy	Interstitial	3.16	6.00	[0, 5]	2.91	0.02
	Subtelomeric	2.55	1.00	[0, 1]	1.82	0.09
DNA transp	Interstitial	-0.05	2.00	[0, 6]	1.00	0.67
	Subtelomeric	-0.40	0.00	[0, 1]	0.86	0.85
LINEs	Interstitial	-0.19	1.00	[0, 5]	0.89	0.63
	Subtelomeric	-0.30	0.00	[0, 2]	0.89	0.91
SINEs	Interstitial	-0.46	0.00	[0, 2]	0.73	0.73
	Subtelomeric	-0.17	0.00	[0, 1]	0.97	0.97
Unknown	Interstitial	-0.96	388.00	[344, 512.05]	0.90	0.17
	Subtelomeric	-1.74	17.00	[15.98, 66]	0.44	0.05
TRC	Interstitial	-0.59	0.00	[0, 2]	0.69	0.68
	Subtelomeric	-0.20	0.00	[0, 1]	0.95	0.96
Most common TRC	Interstitial	-0.57	0.00	[0, 2]	0.71	0.70
	Subtelomeric	-0.18	0.00	[0, 1]	0.96	0.97
GC content	Interstitial	1.50	0.35	[0.32, 0.35]	1.03	0.08
	Subtelomeric	4.32	0.44	[0.3, 0.39]	1.30	0.00

Table 2: Gene evolution permutation test inside areas affected by chromosomal rearrangements. Synonymous (dS), nonsynonymous (dN) substitution rates and their ratio ($\omega = \text{dN/dS}$) are calculated through aggregated analyses, were substitutions and sites were summed across coding sequences (CDS) within each rearrangement and compared with areas of the same size permuted across the chromosome of the rearrangement.

Rearran- gements	Observed (dN)	Observed (dS)	Observed (ω)	Predicted (dN)	Predicted (dS)	Predicted (ω)	95% CI (dN)	95% CI (dS)	95% CI (ω)	p-value (dN)	p-value (dS)	p-value (ω)
Inv_chr8	0.12	0.11	1.10	0.14	0.14	0.97	[0.03, 0.23]	[0.03, 0.24]	[0.38, 3.19]	0.287	0.277	0.208
Inv_chr6	0.00	0.01	0.01	0.10	0.05	0.85	[0, 0.77]	[0, 0.93]	[0, 3.49]	0.050	0.109	0.059
Inv_chr32	0.66	0.29	2.31	0.13	0.11	1.04	[0, 0.66]	[0.01, 0.29]	[0.1, 4.02]	0.050	0.059	0.327
Trans_chr16	0.04	0.04	0.94	0.19	0.10	0.99	[0, 1.66]	[0, 1.17]	[0.06, 5.14]	0.228	0.366	0.475
Trans_chr24	0.01	0.02	0.70	0.01	0.09	0.54	[0, 0.62]	[0, 0.63]	[0, 2.04]	0.436	0.317	0.683
Trans_chr5	0.03	0.05	0.51	0.03	0.10	0.68	[0, 0.76]	[0, 0.64]	[0.21, 1.61]	0.436	0.337	0.356