Centromere evolution in the fungal genus *Verticillium*

Michael F Seidl[1,2#], H Martin Kramer[2], David E Cook[2,3], Gabriel Lorencini Fiorin[2], Grardy CM van den Berg[2], Luigi Faino[2,4], and Bart PHJ Thomma[2,5#]

[1]Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, the Netherlands

[2]Laboratory of Phytopathology, Wageningen University, Wageningen, the Netherlands

[3]Plant Pathology, Kansas State University, Manhattan, United States of America

[4]Environmental Biology Department, Sapienza Università di Roma, Rome, Italy

[5]University of Cologne, Institute for Plant Sciences, Cluster of Excellence on Plant Sciences (CEPLAS), 50674 Cologne, Germany

[#]Corresponding authors: m.f.seidl@uu.nl and bart.thomma@wur.nl

Running title: Centromeres across the *Verticillium* genus

ABSTRACT

Centromeres are chromosomal regions that are crucial for chromosome segregation during mitosis and meiosis, and failed centromere formation can contribute to chromosomal anomalies. Despite this conserved function, centromeres differ significantly between and even within species. Thus far, systematic studies into the organization and evolution of fungal centromeres remain scarce. In this study, we identified the centromeres in each of the ten species of the fungal genus *Verticillium* and characterized their organization and evolution. Chromatin immunoprecipitation of the centromere-specific histone CenH3 (ChIP-seq) and chromatin conformation capture (Hi-C) followed by high-throughput sequencing identified eight conserved, large (~150 kb), AT-, and repeat-rich regional centromeres that are embedded in heterochromatin in the plant pathogen *V. dahliae*. Using Hi-C, we similarly identified repeat-rich centromeres in the other *Verticillium* species. Strikingly, a single repetitive element is strongly associated with centromeric regions in some but not all *Verticillium* species. Extensive chromosomal rearrangements occurred during *Verticillium* evolution, yet only a minority could be linked to centromeres, suggesting that centromeres played a minor role in chromosomal evolution. Nevertheless, the size and organization of centromeres differ considerably between species, and centromere size was found to correlate with the genome-wide repeat content. Overall, our study highlights the contribution of repetitive elements to the diversity and rapid evolution of centromeres within the fungal genus *Verticillium*.

IMPORTANCE

The genus *Verticillium* contains ten species of plant-associated fungi, some of which are notorious pathogens. *Verticillium* species evolved by frequent chromosomal rearrangements that contribute to genome plasticity. Centromeres are instrumental for separation of chromosomes during mitosis and meiosis, and failed centromere functionality can lead to chromosomal

anomalies. Here, we used a combination of experimental techniques to identify and characterize centromeres in each of the *Verticillium* species. Intriguingly, we could strongly associate a single repetitive element to the centromeres of some of the *Verticillium* species. The presence of this element in the centromeres coincides with increased centromere sizes and genome-wide repeat expansions. Collectively, our findings signify a role of repetitive elements in the function, organization and rapid evolution of centromeres in a set of closely related fungal species.

INTRODUCTION

Centromeres are crucial for reliable chromosome segregation during mitosis and meiosis. During this process, centromeres direct the assembly of the kinetochore, a multi-protein complex that facilitates attachment of spindle microtubules to chromatids (1-3). Failure in formation or maintenance of centromeres can lead to aneuploidy, i.e. changes in the number of chromosomes within a nucleus, and to chromosomal rearrangements (3-5). While these processes have been often associated with disease development (6), they can also provide genetic diversity that is beneficial for adaptation to novel or changing environments (7, 8). For example, aneuploidy in the budding yeast *Saccharomyces cerevisiae* can lead to increased fitness under selective conditions, such as the presence of antifungal drugs (9, 10). Thus, centromeric instability can contribute to adaptive genome evolution (11, 12).

Despite their conserved function, centromeres are among the most rapidly evolving genomic regions (13, 14) that are typically defined by their unusual (AT-rich) sequence composition, low gene and high repeat density, and heterochromatic nature (13, 15). Nevertheless, centromeres differ significantly in size, composition, and organization between species (13, 16). Centromeres in *S. cerevisiae* are only ~125 nucleotides long and are bound by a single nucleosome containing the centromere-specific histone 3 variant CenH3 (also called CENP-A or Cse4) (17-20). In contrast to these 'point centromeres', centromeres in many other fungi are more variable and larger, and have thus been referred to as 'regional centromeres' (15). For instance, in the opportunistically pathogenic yeast *Candida albicans*, the CenH3-bound 3-5 kb long centromeric DNA regions differ significantly between chromosomes, and rapidly diverged from closely related *Candida* species (21-23). Centromeres in the basidiomycete yeasts *Malassezia* are similar in size (3-5 kb) but contain a short AT-rich consensus sequence in multiple *Malassezia* species (11). In *Malassezia*, chromosomal rearrangements and karyotype changes are driven by centromeric loss through chromosomal breakage or by inactivation through sequence

diversification (11). Chromosomal rearrangements at centromeres have been similarly observed in the yeast *Candida parapsilosis*, suggesting that centromeres can be fragile and contribute to karyotype evolution (11, 12). CenH3-bound centromeric regions of the basidiomycete yeast *Cryptococcus neoformans* are relatively large, ranging from 30 to 65 kb, and are rich in Long Terminal Repeat (LTR)-type retrotransposons (16). Centromere sizes differ between *Cryptococcus* species as those lacking RNAi and DNA methylation have shorter centromeres, associated with the loss of full-length LTR retrotransposons at centromeric regions, suggesting that functional RNAi together with DNA methylation is required for centromere stability (16).

In filamentous fungi, centromeres have been most extensively studied in the saprophyte *Neurospora crassa* (15). In this species, centromeric regions are considerably larger than in yeasts (on average ~200 kb), and are characterized by AT-rich sequences that are degenerated remnants of transposable elements and sequence repeats that lack an overall consensus sequence (15, 24, 25). The increased AT-content and the degenerated nature of transposable elements in the genome of *N. crassa* are the result of a process called repeat-induced point mutation (RIP) (15, 26). RIP has been linked to the sexual cycle of ascomycetes and targets repetitive sequences by inducing C to T mutations, preferably at CpA di-nucleotides (26). The AT-rich centromeric regions are bound by CenH3 and enriched in the heterochromatin-specific histone modification histone 3 trimethylation of lysine 9 (H3K9me3) (25). Additionally, H3K9me3 and cytosine methylation occurs at the periphery of the centromeres (25). Alterations in H3K9me3 localization compromise centromeric localization, suggesting that the formation and location of heterochromatin, rather than the DNA sequence itself, is essential for function and localization of centromeres in *N. crassa* (15, 25). However, heterochromatin is not a hallmark for centromeres in all filamentous fungi. Centromeres in the fungal wheat pathogen *Zymoseptoria tritici* are shorter (~10 kb) and AT-poor, and their presence does not correlate with transposable elements nor with heterochromatin-specific histone modifications such as H3K9me3 or histone 3 trimethylation of

lysine 27 (H3K27me3) (27). Thus, even though centromeric function is highly conserved, fungal centromeres differ considerably in size, sequence composition, and organization.

Knowledge on centromeres has been impaired by their repetitive nature, which hampers their assembly and subsequent analyses (15, 28). However, recent advances in long-read sequencing technologies enables to study the constitution and evolution of centromeres (11, 16, 29-31). By using long-read sequencing technologies in combination with optical mapping, we previously generated gapless genome assemblies of two strains of the fungal plant pathogen *Verticillium dahliae* (32), whose genomes are characterized by genome rearrangements and the occurrence of lineage-specific (LS) regions (7, 8, 33-35) that are hypervariable between *V. dahliae* strains and contain genes with roles in adaptive evolution to plant hosts (7, 8, 33, 35). Repetitive elements within the LS regions display a distinct chromatin state when compared with other repetitive regions (36). The *Verticillium* genus consists of ten species that are all soil-borne and presumed asexual but have different life-styles (37). Nine of these species are haploid, while the species *Verticillium longisporum* is an allodiploid hybrid between a strain that is closely related to *V. dahliae* and an unknown *Verticillium* species (37-39). During the evolution of the different *Verticillium* species frequent chromosomal rearrangements occurred (8, 35, 40). Facilitated by the availability of high-quality genome assemblies of *V. dahliae* strains and of all other *Verticillium* species (32, 33, 40, 41), we here sought to identify and study the constitution and evolution of centromeres in the *Verticillium* genus.

RESULTS

CenH3-binding identifies large regional centromeres in *Verticillium dahliae*

Centromeres differ significantly between fungi, but most centromeres are functionally defined by nucleosomes containing CenH3 (1). To identify centromeres in *V. dahliae* strain JR2 by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq), we first identified the *V. dahliae* CenH3 ortholog (**Fig. S1a**) and generated transformants with N-terminally FLAG-tagged CenH3 (Table S1). To this end, the coding sequence for the FLAG-tagged CenH3 was inserted in locus behind the native *CenH3* promotor (**Figs. S1b-c**). We subsequently used anti-FLAG antibodies to purify FLAG-tagged CenH3-containing nucleosomes from two *V. dahliae* transformants (**Table S1a**) and sequenced the nucleosome-associated genomic DNA. Mapping of the sequencing reads to the *V. dahliae* strain JR2 genome assembly identified a single CenH3-enriched region per chromosome (**Fig. 1a; Fig. S1d-e**), while mapping of the sequencing reads derived from the WT strain did not reveal any CenH3-enriched region (**Fig. S1d-e**). The CenH3-enriched regions, designated as *Cen1-8*, range between ~94 and ~187 kb in size (**Fig. 1a**; **Table 1**). To corroborate these centromere sizes, we assessed centromere locations based on a previously generated optical map (32, 35) revealing no significant size differences (**Fig. S1f**). Thus, we conclude that CenH3-binding defines large regional centromeres in *V. dahliae* strain JR2.

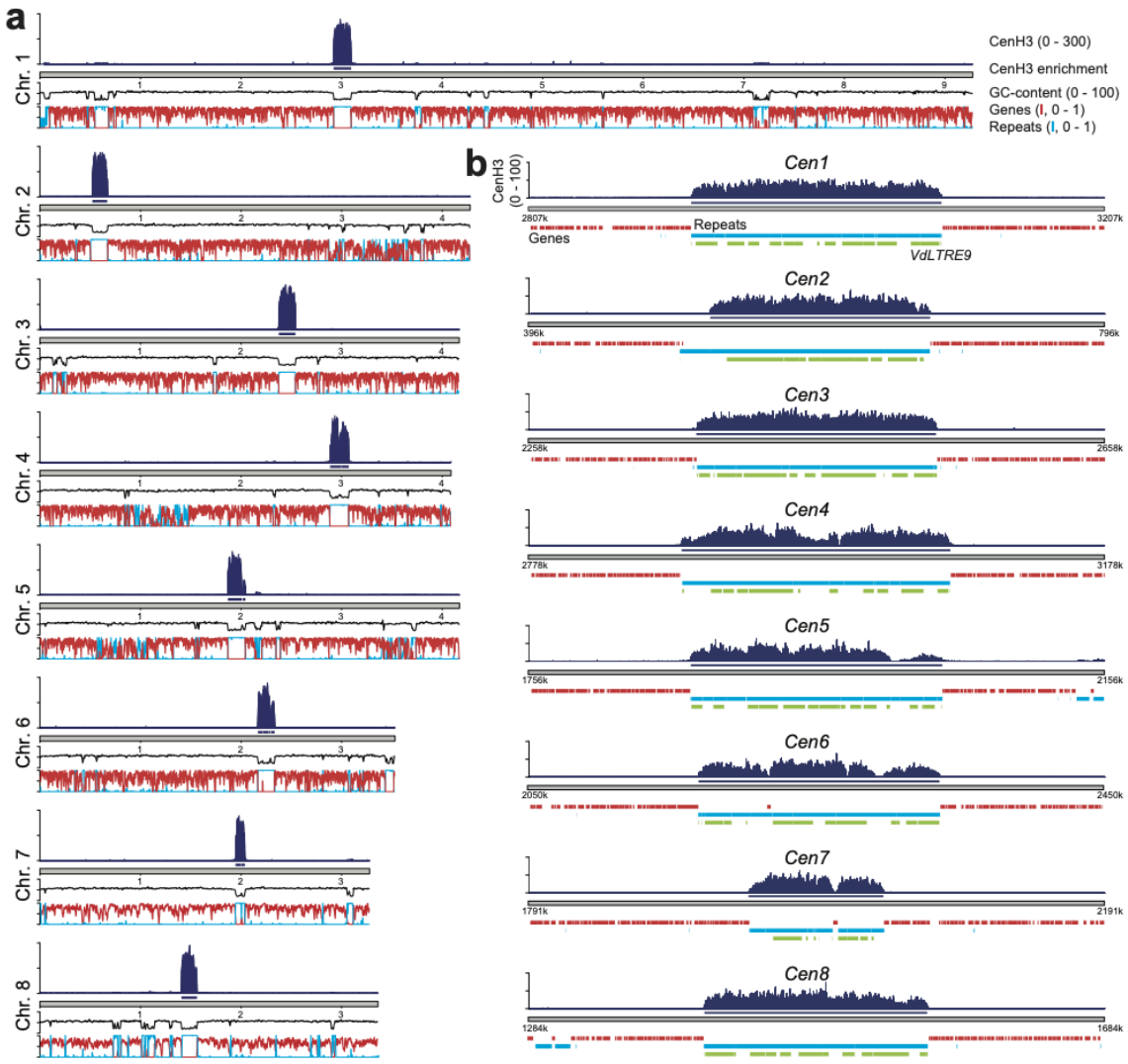


Figure 1 – CenH3-binding defines centromeres in *Verticillium dahliae* strain JR2. (a) Schematic overview of the chromosomes of *V. dahliae* strain JR2 showing the normalized CenH3 ChIP-seq read coverage (RPGC normalization in 1 kb bins with 3 kb smoothening), CenH3 enriched regions, GC-content, gene density (red line), and repeat density (blue line). (b) Magnification of a 400 kb region containing the centromere is shown for each of the eight chromosomes of *V. dahliae* strain JR2 (*Cen1-8*) depicting the CenH3 ChIP-seq read coverage (RPGC normalization in 10 bp bins with a 30 bp smoothening) and enrichment, as well as the presence of genes (red) and repetitive elements (blue). Regions carrying the centromere-specific long-terminal repeat element *VdLTRE9* are highlighted in green.

147 **Centromeres in *Verticillium dahliae* are repeat-rich and embedded in heterochromatin**

148 Centromeres are often characterized by increased AT-content, increased repeat density, and
149 depletion of protein coding genes (13, 15, 29). To characterize the centromeres in *V. dahliae*
150 strain JR2, we queried the eight chromosomes for the presence of large AT-rich, gene-sparse, and
151 repeat-rich regions. Seven of the eight chromosomes contain only a single large (>93 kb; average
152 size ~150 kb) AT-rich region (~74-78% versus ~46% genome-wide), nearly completely devoid of
153 protein-coding genes and enriched for repetitive sequences, that overlaps with the regions defined
154 by CenH3-binding (**Fig. 1a**; **Table 1**). In contrast, chromosome 1 contains three regions with
155 these characteristics (**Fig. 1a**; **Table 1**). However, only one of these overlaps with the centromeric
156 regions defined by CenH3-binding (**Fig. 1**).

157 Elevated AT-levels in repeat-rich regions are caused by RIP mutations in some
158 filamentous fungi (15, 25, 26, 42). Due to its presumably asexual nature (7), the occurrence of
159 RIP in *V. dahliae* is controversial (8, 43, 44), although signatures of RIP have previously been
160 reported in a subset of repeat-rich regions (36). We assessed the occurrence of RIP signatures in
161 centromeres using the composite RIP index (CRI) (45), which considers C to T mutations in the
162 CpA context. Intriguingly, genomic regions located at centromeres display significantly higher
163 CRI values than other genomic regions (e.g. genes or repetitive elements) (**Fig. 2a; Figs. S2,**
164 **S3a**), and thus RIP signatures at repetitive elements located at centromeres likely contribute to the
165 high AT-levels.

166 In most filamentous fungi and oomycetes, AT- and repeat-rich centromeres are embedded
167 in heterochromatin that is characterized by methylated DNA and by particular histone
168 modifications (H3K9me3 and H3K27me3) (13, 15, 16, 25, 30, 45). We recently determined
169 chromatin states in the genome of *V. dahliae* strain JR2 and revealed that repetitive sequences
170 outside of the LS regions display characteristics of heterochromatin (36). To define centromeric
171 chromatin states, we used previously generated bisulfite sequencing data to monitor DNA

172 methylation (mC) and ChIP-seq data to determine the distribution of the heterochromatic marks
173 H3K9me3 and H3K27me3 (36). To also determine the distribution of euchromatin, we performed
174 ChIP-seq with an antibody against the euchromatic mark di-methylation of lysine 4 of histone H3
175 (H3K4me2). We observed overall low genome-wide DNA methylation levels (36) (**Fig. 2a; Fig.**
176 **S2**), similar to the previously reported levels for *Aspergillus flavus* (46) and lower than for *N.*
177 *crassa* (47). Nevertheless, repetitive elements and centromeres show significantly higher DNA
178 methylation levels in all contexts when compared with genes (**Fig. 2b**). Methylation (in CG
179 context) at repetitive elements at centromeres is significantly higher than at repeats located along
180 the chromosomal arm, but not at sub-telomeric regions (**Fig. 2c**), and more methylation at
181 centromeres correlates with increased CRI (**Fig. 2a; Figs. S2, S3a**). DNA methylation co-
182 localizes with H3K9me3 at repeat-rich regions (36) (**Figs. 2a; Fig. S2**). H3K9me3 occurs
183 predominantly at repetitive elements localized at sub-telomeres and centromeres (**Figs. 2d-e;**
184 **Figs. S2, S3b**). In comparison, H3K4me2 and H3K27me3 are largely absent from centromeres
185 (**Figs. 2d-e; Fig. S3b**). Collectively, these observations indicate that centromeres of *V. dahliae*
186 display typical characteristics of constitutive heterochromatin.

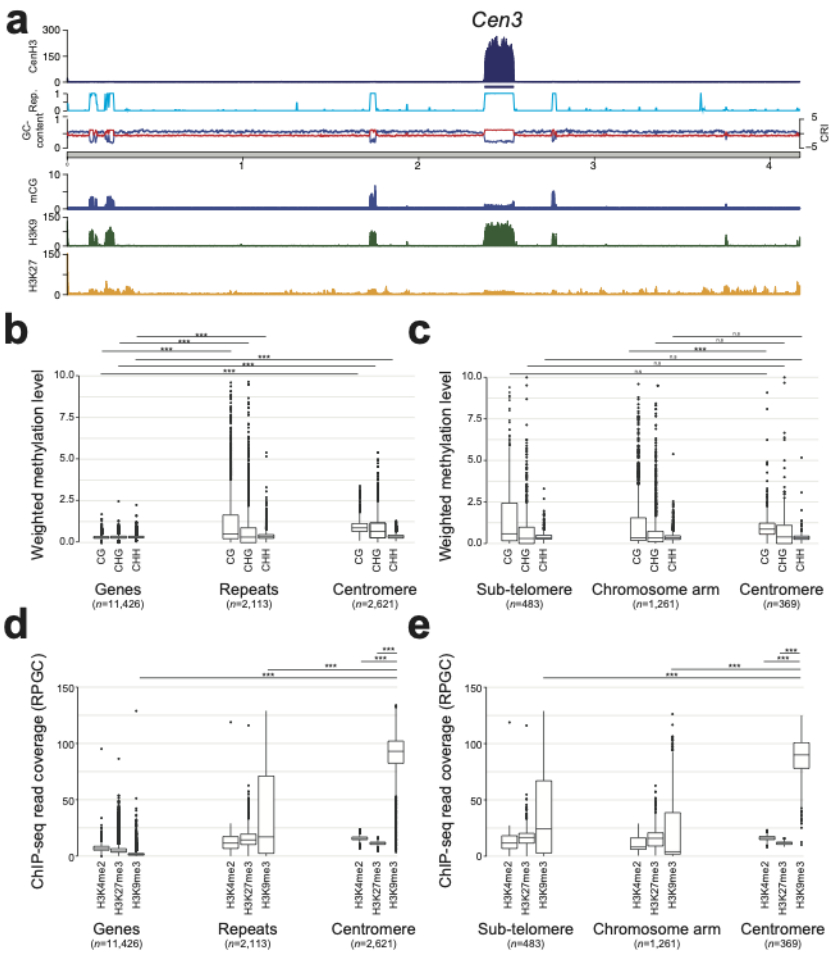


Figure 2 – Centromeres in *Verticillium dahliae* strain JR2 are embedded in heterochromatin. (a) Schematic overview of chromosome 3 of *V. dahliae* strain JR2, exemplifying the distribution of heterochromatin-associated chromatin modifications (mC, H3K9me3, and H3K27me3) in relation to the centromeres. The different lanes display the CenH3-FLAG ChIP-seq read coverage (RPGC normalization in 1 kb bins with 3 kb smoothening), the CenH3-FLAG enriched regions, the repeat-density, the GC-content, the CRI as well as the weighted cytosine methylation (all summarized in 5 kb windows with 500 bp slide), and the normalized H3K9me3 and H3K27me3 ChIP-seq read coverage (RPGC normalization in 1 kb bins with 3 kb smoothening). The schematic overview of all chromosomes is shown in Figure S2. (b) Box plots of weighted DNA methylation levels per genomic context (CG, CHG, or CHH) are summarized over genes, repetitive elements, or 5 kb genomic windows (500 bp slide) overlapping with the centromeric regions. (c) Weighted DNA methylation levels per genomic context (CG, CHG, or CHH) are summarized over repetitive elements that have been split based on their genomic location; sub-telomeres (within the first or last 10% of the chromosome), centromeres, or the remainder of the chromosome arm. (d) ChIP-seq read coverage (RPGC normalized; see (a)) for H3K4me2, H3K27m3, and H3K9me3 is summarized over genes, repetitive elements, or 5 kb windows (500 bp slide) overlapping with the centromeric regions. (e) ChIP-seq read coverage (RPGC normalized; see (a)) for H3K4me2, H3K27m3, and H3K9me3 is summarized over repetitive elements that have been split based on their genomic location; sub-telomeres (within the first or last 10% of the chromosome), centromeres, or the remainder of the

chromosomal arm. Statistical differences for the indicated comparisons were calculated using the one-sided non-parametric Mann-Whitney test; p-values < 0.001: ***.

A single repeat associates with *Verticillium dahliae* strain JR2 centromeres

Centromere identity and function is typically defined by CenH3-binding and not by specific DNA sequences, although various types of repetitive sequences, such as transposable elements, are commonly observed in centromeres of plants, animals, and fungi (13, 15, 48, 49). Unsurprisingly, CenH3-bound centromeres are repeat-rich in *V. dahliae* (**Fig. 1**). A detailed analysis of the eight centromeres revealed a near-complete (>96%) composition of repetitive elements belonging to only ten different repeat sub-families (**Figs. 1b, 3a; Table 1**), of which the majority shows similarity to LTR retrotransposons of the *Gypsy*- and *Copia*-like families (**Fig. 3a**). These elements show signs of RIP, are highly methylated, and non-transcribed (**Figs. S3c-e**), and thus likely inactive. Interestingly, a single LTR retrotransposon sub-family, previously designated *VdLTRE9* (8, 32), covers on average ~70% of the DNA sequences at the eight centromeres, ranging from 47% in *Cen7* to 83% in *Cen2* (**Fig. 3a; Table 1**). We scanned the genome for the localization of the ten repeat sub-families (**Fig. 3**). Intriguingly, although it is one of the most abundant repeats in the genome with 215 complete or partial matches, *VdLTR9* is associated to centromeres as 95% of the copies (204 out of 215; one-sided Fisher's exact test; multiple-testing corrected p-value 3e-106) occur at the eight centromeres, whereas only 5% of the copies are dispersed over the genome (**Fig. 3b-c**). The nine other repeat sub-families have additional matches that are located outside of the centromeres (**Figs. 1a; Figs. 3b-c)**, and only two of these repeats are significantly enriched and consistently present in all eight centromeres; 63% and 45% of the matches of these two sub-families occur at the centromeres (**Fig. 3c**). Collectively, these findings suggest that only the presence of *VdLTRE9* is strongly associated with centromeres in *V. dahliae* strain JR2.

VdLTRE9 displays similarity to LTR retrotransposons. The consensus sequence of *VdLTRE9* is ~7.3 kb long (the two LTR sequences are each ~200 bp long), and the individual matches share a high degree of sequence identity (~86%). Sequence similarity based TE-classifications using PASTEC (50) indicates that the consensus sequence displays remote similarity to *Gypsy*-like retrotransposons. Only ~25% of the *VdLTRE9* matches in the genome cover the entire (>97.5%) consensus sequence, but many of these are still fragmented as they occur as discontinuous copies. Furthermore, the *VdLTRE9* consensus sequence is AT-rich (~75% AT), which may be caused by RIP (**Fig. S3d**), indicating that *VdLTRE9* has significantly degenerated.

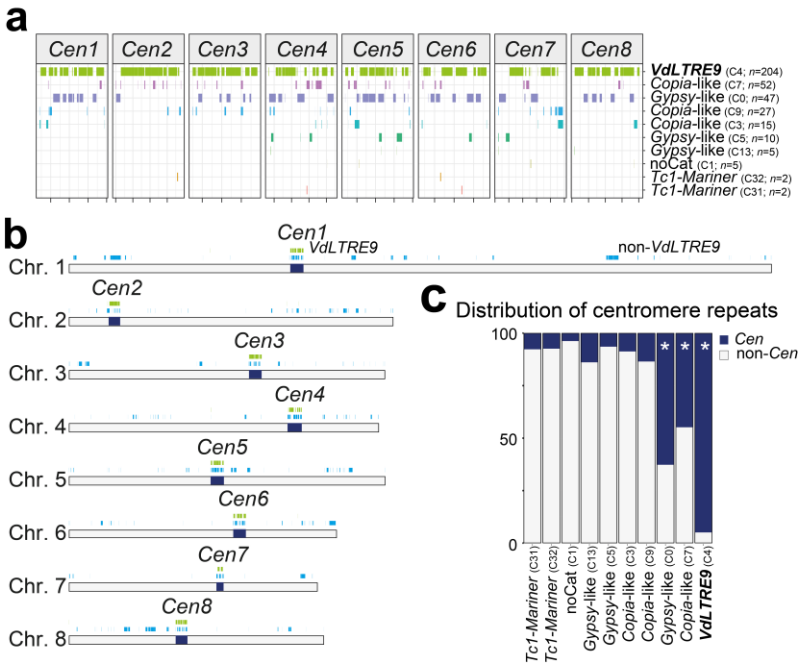


Figure 3 – A single repeat family associates with centromeres in *Verticillium dahliae* strain JR2. (a) The presence of different repeat sub-families is shown across the eight centromeres (*Cen1-8*), and the number of occurrences for each sub-family within the centromeres is indicated. The individual centromeres in the diagram are shown in equal scale. (b) Genome-wide distribution of the ten repeat sub-families occurring within the eight centromeres (*Cen1-8*; dark blue); the location of *VdLTRE9* is shown in green and the location of elements belonging to the other nine sub-repeat families (from panel (a)) is shown in light blue. (c) The distribution of different repeat sub-families in centromeres (*Cen*; dark blue) and across the genome (non-*Cen*; light grey). The enrichment of specific sub-families at centromeres was assessed using a one-sided Fisher's exact test. Significant enrichment (multiple-testing corrected p-value < 0.01) is denoted with an asterisk.

254

255 ***VdLTRE9* as hallmark of *Verticillium dahliae* centromeres**

256 To examine if *VdLTRE9* similarly occurs at centromeres in other *V. dahliae* strains, we made use
257 of the complete genome assembly of *V. dahliae* strain VdLs17 (8, 32, 35). The evolution of *V.*
258 *dahliae* is characterised by chromosomal rearrangements (8, 35) (**Figs. 4a; Figs. S4a-c**).
259 Nevertheless, synteny analyses between *V. dahliae* strains JR2 and VdLs17 revealed large regions
260 of co-linearity between chromosomes and identified significant sequence and synteny
261 conservation between the centromeres and their flanking regions (**Figs. 4b-c; Fig. S4a**),
262 suggesting that centromeric sequences and their locations are conserved. We queried the genome
263 of *V. dahliae* strain VdLs17 for the presence of *VdLTRE9* and identified a single region on each
264 chromosome, collectively containing 186 of the 207 (90%) complete or partial matches of
265 *VdLTRE9* in the genome (**Fig. 4d**) (one-sided Fisher's exact test; multiple-testing corrected p-
266 value 3e-146). These *VdLTR9*-rich regions are ~150 kb in size, AT-rich, gene-poor and repeat-
267 rich, and share similarity to the previously identified CenH3-bound and *VdLTRE9*-enriched
268 regions of *V. dahliae* strain JR2 (**Figs. 4b-c; Fig. S4d**), suggesting that these regions similarly
269 represent the centromeres of *V. dahliae* strain VdLs17.

270 Centromeres *N. crassa* and some other fungi co-localize within the nucleus (15, 51-55). This
271 co-localization can be experimentally determined using chromosome conformation capture (Hi-
272 C), which can identify centromeres by their increased inter-chromosomal contacts (55). To
273 confirm that Hi-C can be used to identify centromeres in *V. dahliae*, we first applied Hi-C to *V.*
274 *dahliae* strain JR2. As anticipated, we observed seven strong inter-chromosomal contacts for each
275 of the eight chromosomes (**Figs. 4e**). Importantly, the interacting regions overlap with the
276 CenH3-bound regions that we identified as centromeres (**Table S1b**), demonstrating that
277 centromeres in *V. dahliae* strain JR2 co-localize within the nucleus and supporting that Hi-C
278 reliably identifies centromeres (51, 52). We then applied Hi-C to *V. dahliae* strain VdLs17, and

similarly identified regions with strong inter-chromosomal contacts, one for each of the chromosomes (**Figs. 4f**). These regions overlap with the *VdLTRE9*-enriched regions (**Table S1b**), suggesting that these represent functional centromeres in *V. dahliae* strain VdLs17.

The two *V. dahliae* strains JR2 and VdLs17 are closely related and differ only by ~0.05% sequence diversity (8, 35). Thus, the conservation of *VdLTRE9* at centromeres could be driven by limited divergence between the two *V. dahliae* strains rather than representing a hallmark of *V. dahliae* centromeres. Therefore, we sought to determine centromeres in an additional *V. dahliae* strain with increased sequence diversity when compared with *V. dahliae* strains JR2 or VdLs17, namely strain CQ2 that displays ~1.05 percent sequence diversity (33). We previously obtained a long-read based genome assembly of this strain that encompasses 17 contigs (33). We generated Hi-C data for *V. dahliae* strain CQ2 and utilized intra-chromosomal contacts to assign the contigs into eight pseudo-chromosomes, leaving ~148 kb unplaced scaffolds (**Fig. 4g; Fig. S4e; Table S1c**). We subsequently identified a single region with seven strong inter-chromosomal contacts for each pseudo-chromosome that is significantly enriched for *VdLTRE9* (one-sided Fisher's exact test; multiple-testing corrected p-value 3.4e-166) (**Figs. 4d, g; Fig. S4e; Table S1b**). Synteny analyses between *V. dahliae* strains JR2 and CQ2 revealed that the eight *VdLTRE9*-rich regions and their flanking chromosomal regions are co-linear, suggesting that centromere locations are conserved between different *V. dahliae* strains (**Figs. 4; Figs. S4a-c, f**). With an average size of 165 kb, the centromeres of *V. dahliae* strain CQ2 are similar in size as the 144 kb and 157 kb average sizes in *V. dahliae* strains VdLs17 and JR2, respectively (**Table S1b**). The sizes of the corresponding (i.e. homologous) centromeres vary between the different *V. dahliae* strains. Yet, the consistent co-occurrence of the *VdLTRE9*-rich regions with the interaction data obtained by Hi-C throughout a selection of *V. dahliae* strains demonstrates that *VdLTRE9* is a hallmark of *V. dahliae* centromeres.

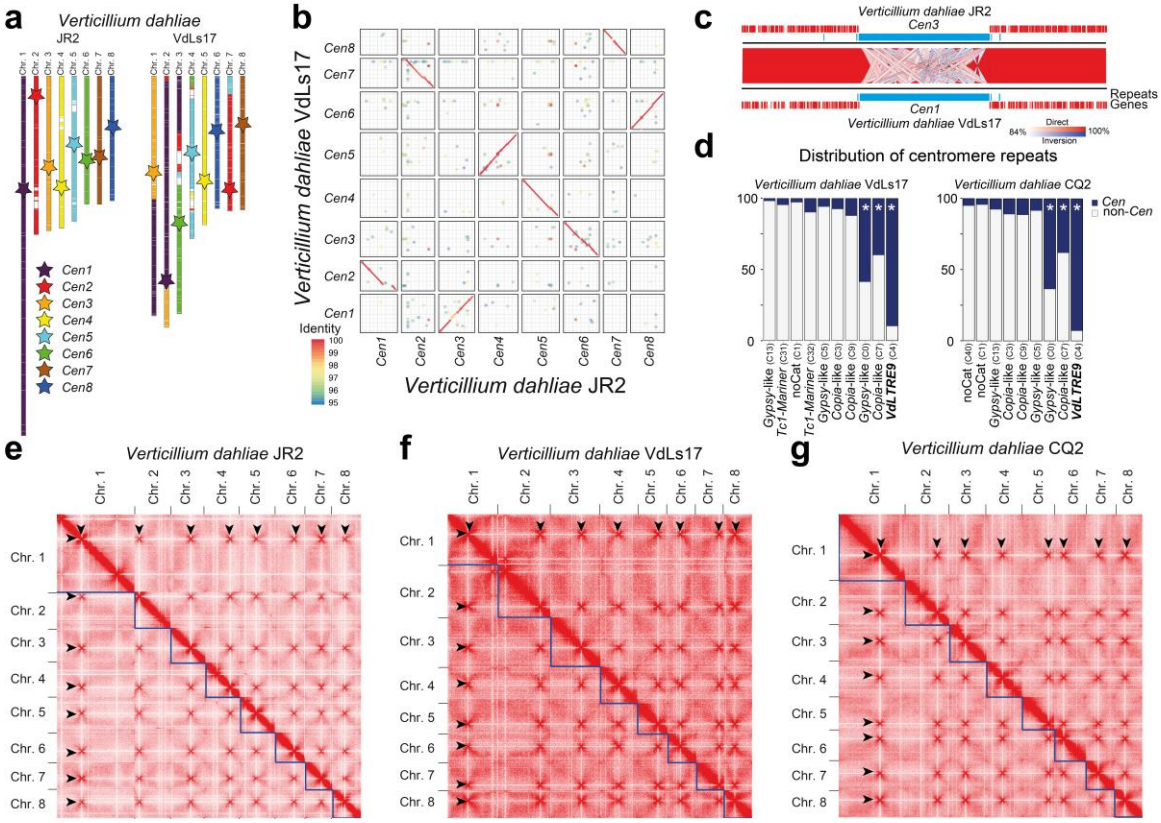


Figure 4 – Hi-C contact maps identify *VdLTRE9* as hallmark of centromeres in *Verticillium dahliae*. (a) Synteny analyses of the eight chromosomes of *V. dahliae* strains JR2 and VdLs17. Schematic overview of the eight chromosomes of *V. dahliae* strain JR2 (left) and the corresponding syntenic regions in *V. dahliae* strains VdLs17 (right). Approximate locations of centromeres are indicated by stars, and syntenic centromeres of *V. dahliae* strain VdLs17 are colored according to *Cen1-8* of *V. dahliae* strain JR2. (b) Sequence alignment of the centromeric regions ± 20 kb in *V. dahliae* strain JR2 and the corresponding regions in *V. dahliae* strains VdLs17 shown as dot-plot. For clarity, only alignments with >95% sequence identity are displayed. (c) Magnification of *Cen3* of *V. dahliae* strain JR2 and the syntenic *Cen1* of strain VdLs17. Synteny between regions is indicated by ribbons; entire centromeric regions *Cen1* and *Cen3* are syntenic and sequence similarity between individual *VdLTRE9* elements is visualized. The *Cen* regions ± 150 kb are shown as well as genes (red) and repeats (blue) are annotated within this region. (d) Distribution of different repeat families in centromeres (*Cen*; dark blue) and across the genome (non-*Cen*; light grey) for *V. dahliae* strains VdLs17 and CQ2. The enrichment of specific sub-families at centromeres was assessed using a one-sided Fisher's exact test. Significant enrichment (multiple-testing corrected p-value < 0.01) is denoted with an asterisk. (e-g) Hi-C contact matrix showing interaction frequencies between genomic regions in *Verticillium dahliae* strains JR2 (e), VdLs17 (f), and CQ2 (g). Regions of high inter-chromosomal interaction frequencies are indicative of centromeres and are highlighted by arrow heads. Interaction frequencies are summarized in 50 kb bins along the genome.

The evolution of *Verticillium* centromeres

In addition to *V. dahliae*, we previously generated genome assemblies of the eight haploid *Verticillium* species and the allodiploid *V. longisporum* (39, 40) (**Fig. 5a**) that ranged from 12 to 684 scaffolds (**Table S1c**). These ten *Verticillium* species have been traditionally separated over two distinct clades; Flavnonexudans and Flavexudans (**Fig. 5a**) (37). We generated Hi-C data to study the composition and evolution of centromeres in the different *Verticillium* species. By using intra-chromosomal interaction signals, we first assigned the vast majority of the previously assembled contigs into eight pseudo-chromosomes for each of the haploid *Verticillium* species and 16 pseudo-chromosomes for the diploid *V. longisporum*, leaving between 0.5 kb and 2,022 kb unassigned (**Fig. S5; Table S1c**). For most genome assemblies, the pseudo-chromosomes contain one or both telomeric repeats (**Table S1c)**, and thus we conclude that all *Verticillium* strains have eight chromosomes, and that this number doubled in *V. longisporum*. Based on the inter-chromosomal Hi-C interaction signals, we identified a single region with high inter-chromosomal contacts for each of the pseudo-chromosomes (**Fig. S5; Table S1d**), indicating that these are the centromeres in the different *Verticillium* species. The average centromere size in *Verticillium* is ~80 kb, yet we observed significant differences between the species (**Fig. 5b; Figs. S6a-b**). Centromeres within the Flavexudans clade are similarly sized and significantly smaller than the genus-wide average. By contrast, *V. dahliae* and *V. longisporum* centromeres are significantly larger.

We subsequently assessed whether *VdLTRE9* defines centromeres in the other *Verticillium* species besides *V. dahliae* as well. Interestingly, *VdLTRE9* is abundant at centromeres in the allodiploid *V. longisporum* and in *V. alfalfae*, but fewer (21) or no *VdLTRE9* copies were identified at centromeres in *V. nonalfalfae* and *V. nubilum*, respectively (**Fig. 5c; Fig. S6c-d**). Similarly, only few or no (partial) matches of *VdLTRE9* consensus could be identified in the genomes of the Flavexudans species (**Fig. 5c; Fig. S6-7; Table S1e**). Collectively, these

findings demonstrate that *VdLTRE9* is specific to Flavnonexudans species and has likely been recruited to the centromere only after the divergence of *V. nubilum* (**Fig. 5a; Fig. S6-7**).

Since *VdLTRE9* occurs only in few *Verticillium* species, we assessed to which extent other repetitive elements contribute to centromere organization. We analyzed the repeats identified by *de novo* repeat predictions for each of the *Verticillium* species. Centromeres in all species are AT- and repeat-rich (**Fig. S6a-b**), and some repeats occur in high frequency or nearly exclusively at centromeres in species that lack *VdLTRE9* (**Table S1e**). However, in contrast to *VdLTRE9*, these repeats cover only a minority (typically less than 10%) of the centromeres (**Table S1e**). Sequence similarity-based cluster analyses of the *de novo* repeat consensus sequences revealed that divergent repeat families contribute to *Verticillium* centromere organization (**Fig. S8**). Thus, in contrast to *VdLTRE9* in most Flavnonexudans species, we could not identify any additional repeat family as a hallmark of centromeres in other *Verticillium* species.

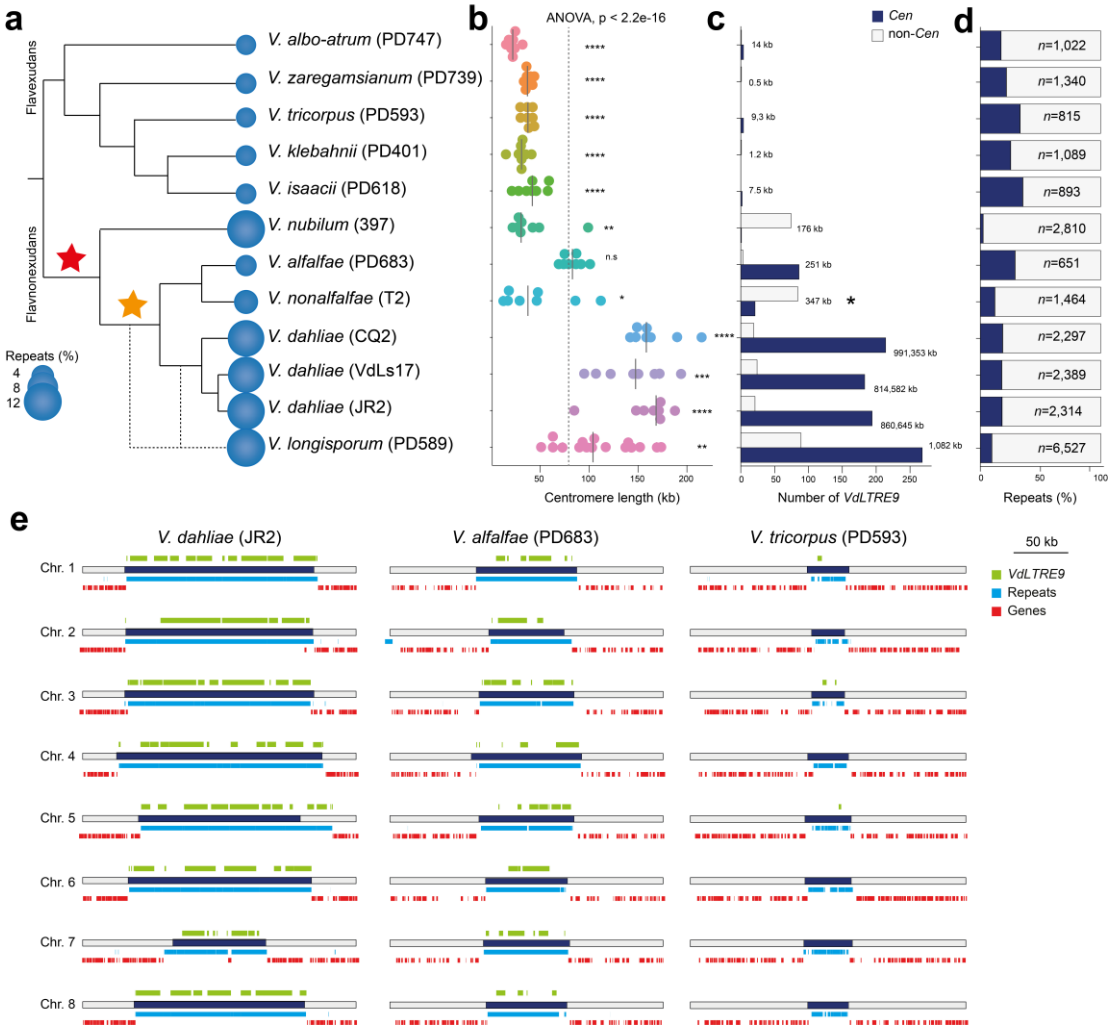


Figure 5 – Evolution of centromeres in the genus *Verticillium*. (a) Relationship of the ten members of the genus *Verticillium*. The predicted repeat content for each of the genomes is indicated (see Table S3 for details). The red star indicates the acquisition of *VdLTRE9* in the Flavnonexudans clade while the yellow star indicates the recruitment of *VdLTRE9* into centromeres. (b) Comparison of estimated centromere lengths (in kb) in the different *Verticillium* spp. Each dot represents a single centromere and the line represents the median size. (c) The number of (partial) *VdLTRE9* matches identified in centromeres (*Cen*; dark blue) and across the genome (non-*Cen*; light grey). The asterisk indicates the high number of *VdLTRE9* elements in unassigned contigs for *Verticillium nonalfalfae* strain T2 (see text for details). (d) Proportion of predicted repeat content localized at centromeres (*Cen*; dark blue) and across the genome (non-*Cen*; light grey). (e) Schematic overview of the eight centromeric regions (250 kb) in *Verticillium dahliae* strain JR2, and *Verticillium alfalfae* strain PD683 and *Verticillium tricorpus* stain PD593 as representatives for clade Flavnonexudans and clade Flavexudans, respectively. The centromeres are indicated by dark blue bars. The predicted genes (red) and repeats (light blue) are shown below each centromere, and location of (partial) *VdLTRE9* matches (light green) are shown above each centromere. Global statistical differences for the centromere sizes was calculated using one-way ANOVA, and differences for each species compared to the overall

mean were computed using unpaired T-tests; p-values < 0.0001: ****, p-values < 0.001: ***, p-values < 0.01: **, p-values < 0.05: *.

Centromeres contribute to *Verticillium* karyotype evolution

We previously used fragmented genome assemblies to identify chromosomal rearrangements during *Verticillium* evolution (8, 35, 40). We hypothesize that centromeres might have contributed to these chromosomal rearrangements. To identify genome rearrangements and to trace centromeres during *Verticillium* evolution, we used the pseudo-chromosomes of the haploid *Verticillium* species to reconstruct ancestral chromosomal configurations using AnChro (**Fig 6a**) (56). We reconstructed all potential ancestors that predominantly had eight chromosomes and ~8,000 genes (**Figs. S9a-b**), yet the number of ancestral chromosomes and genes varied when approaching the last common ancestor (**Figs. S9a-b**). By balancing the number of reconstructed chromosomes and genes, we identified a single most parsimonious ancestral genome with eight chromosomes and ~8,500 genes (**Fig. 6a; Fig. S9c**), except for the last common ancestor within the clade Flavexudans clade that had eight major chromosomes and two additional 'chromosomes' with only six and two genes (**Fig. S9d**). As these two smaller 'chromosomes' likely do not represent genuine chromosomes, we conclude that all of the ancestral genomes, similar to the extant haploid *Verticillium* genomes, had eight chromosomes (**Fig. 6a**). Confirming our previous report (40), we observed in total 198 chromosomal rearrangements (124 inversions and 74 translocations) (**Fig. 6a**). The number of chromosomal rearrangements is lower than previously recorded and we did not observe any chromosomal fusion or fission events, which is likely the result of the drastically improved genome assemblies, but the rearrangement signal on each branch is sufficient to nevertheless recapitulate the known *Verticillium* species phylogeny (**Fig. S9e**). Importantly, we observed 17 genomic rearrangements that occurred at, or in close proximity (within ~15 genes up or downstream) to, centromeres, both in extant *Verticillium* species as well as in the ancestors (**Fig. 6**). For example, at the branch from the last common ancestor (VA; **Fig. 6a**) to the ancestor of the clade Flavexudans (B1; **Fig. 6a**), two centromere-

410 associated translocations (between the ancestral chromosome 2 and 6) led to the formation of two
411 rearranged chromosomes. In total, we observed that five out of the eight ancestral centromeres
412 were associated with a chromosomal rearrangement at one point during evolution (**Fig. 6a**).
413 Nevertheless, comparisons of protein-coding genes that flank centromeres show that these are
414 syntenic in most extant species. Similarly, none of the recent chromosomal rearrangements
415 observed between *V. dahliae* strains is associated with centromeres (**Figs. 4a-b, 6a**). Thus, while
416 chromosomal rearrangements involving centromeres occurred during evolution, they do not
417 account for the majority of the karyotype variation between extant *Verticillium* species.

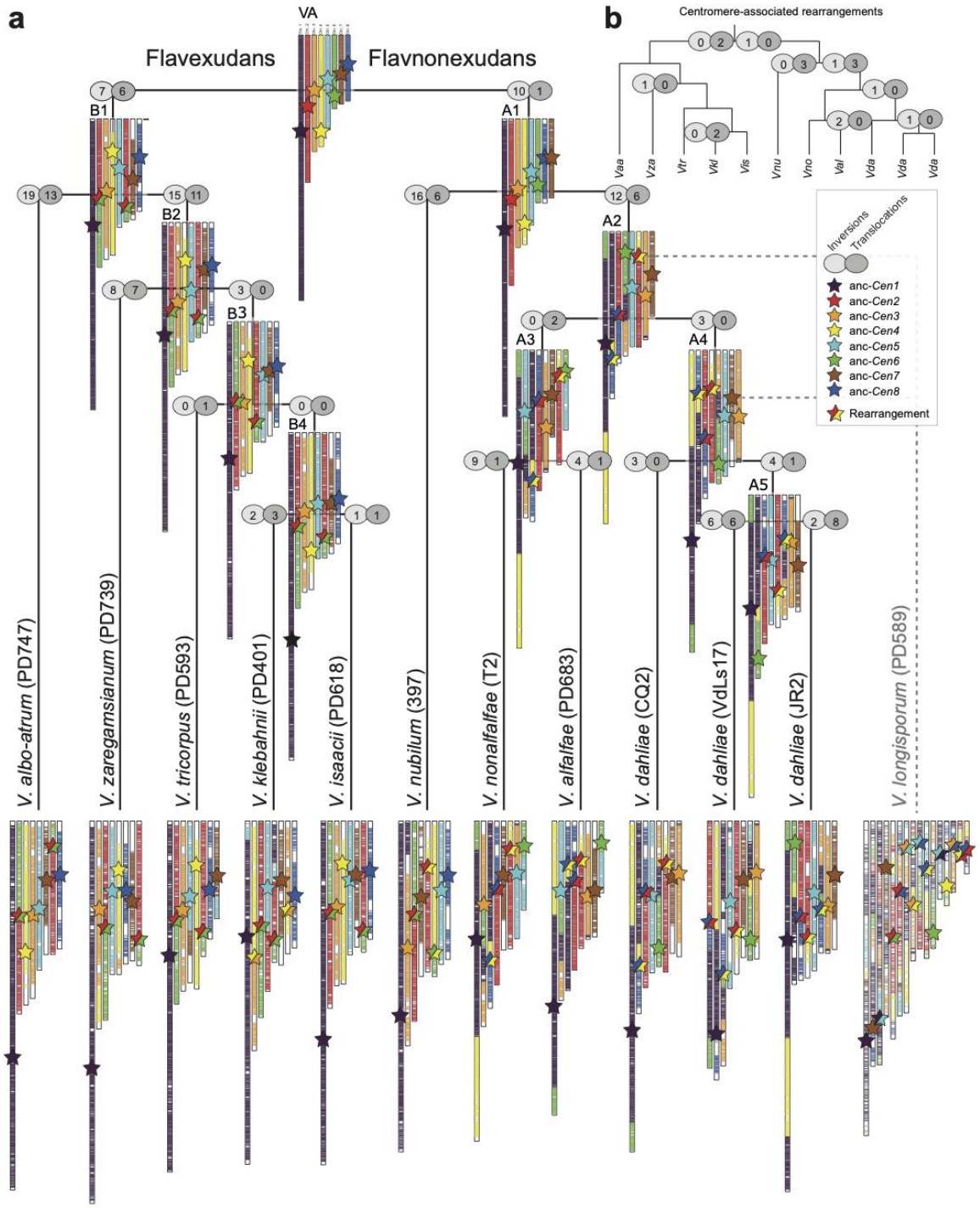


Figure 6 – Centromeres contribute to karyotype evolution in *Verticillium*. (a) Relationship of the ten members of the genus *Verticillium*. The allodiploidization event forming *V. longisporum* is indicated by dashed lines (38, 57). The chromosomal evolution within the haploid members of the genus was reconstructed using AnChro (56). The chromosomal structure of the nine species is shown in relation to the last common ancestor of the genus. The approximate locations of the centromeres are indicated by stars. The number of chromosomal rearrangements (inversions and translocations) are displayed for each branch, and centromeres that co-localize in proximity to chromosomal rearrangements are highlighted by two-colored stars. (b) The number of major

chromosomal rearrangements that occurred at, or in close proximity of, centromeres are shown along the branches depicting the *Verticillium* species phylogeny shown in (a).

DISCUSSION

Centromeric regions are among the most rapidly evolving genomic regions (13-16, 29), yet centromere evolution has only been systematically studied in few fungi (11, 12, 16, 29). Here, we took advantage of the fungal genus *Verticillium* and used a combination of genetic and genomic strategies to identify and characterize centromere organization and evolution. *Verticillium* centromeres are characterized as large regional centromeres that are repeat-rich and embedded in heterochromatin. We furthermore show that centromeres contribute to the karyotype evolution of *Verticillium*. Finally, we demonstrate that *VdLTRE9* is a hallmark of centromeres in some *Verticillium* species, while species that lack *VdLTRE9* display a divergent repeat content.

Centromeres in fungi, plants, and animals co-localize within the nucleus (15, 51-55, 58), a phenomenon that can be exploited for their identification (51, 52). Here, we used Hi-C to first establish chromosome-level genome assemblies and subsequently identify centromeres in every *Verticillium* species, and we demonstrate that centromere locations are in agreement with CenH3-binding. While we obtained chromosome-level genome assemblies for all species, Hi-C scaffolded genome assemblies could still contain partially collapsed repeats and assembly gaps, in particular for short-read assemblies (59). With the exception of *V. nonalfalfae*, we observed only few sequencing gaps and no evidence that would point to collapsed repeats at centromeres, suggesting that the inferred centromeres are of high quality. *Verticillium* centromere sizes differ , which is likely not driven by assembly artefacts, and centromeres in most *Verticillium* species are larger than in *Z. tritici* (27), *C. neoformans*, *M. oryzae*, or *Fusarium graminearum* (13, 16, 29), yet smaller than in *N. crassa* (25). Species of the Flavexudans clade typically encode fewer repeats than species of the clade Flavnonexudans clade (32, 40, 60), and *V. nubilum*, *V. longisporum*, and *V. dahliae* are particularly rich in repeats when compared with other

453 *Verticillium* species (32, 39-41, 60). Thus, increased centromere sizes positively correlate with
454 overall increased repeat contents.

455 Using fragmented genome assemblies, we previously identified chromosomal
456 rearrangements during *Verticillium* evolution (8, 35, 40) that were thought to have contributed to
457 genetic diversity and adaptation in the absence of sexual recombination (7, 35, 40). Chromosome-
458 level genome assemblies for an entire genus enabled unprecedented analyses of the karyotype
459 evolution over longer evolutionary timescales. Here, we observed extensive chromosomal
460 rearrangements and provide evidence that some rearrangements at centromeres contributed to
461 karyotype evolution, most of which occurred early during the divergence of *Verticillium*.
462 Chromosomal rearrangements at centromeres occur in the fungal yeasts *Candida*, *Cryptococcus*,
463 and *Malassezia* (11, 12, 61), and synteny breakpoints have been identified between mammals and
464 chicken (62), suggesting that centromeres often contribute to karyotype evolution. The emergence
465 of chromosomal rearrangements at centromeres could be facilitated by their repeat-rich nature
466 (11, 12). For example, centromeres in *Malassezia* are enriched with an AT-rich motif that could
467 facilitate replication fork stalling, which leads to double strand DNA breaks (11). Repeats
468 localized outside of centromeres in *V. dahliae* contribute to chromosomal rearrangements (8), and
469 thus it seems plausible that centromeric repeats similarly contribute to chromosomal
470 rearrangements. Chromosomal rearrangements often do not only lead to changes in chromosome
471 organization but also in chromosome number (11, 12). While we observed chromosomal
472 rearrangements, all extant and ancestral genomes contained eight chromosomes, suggesting that
473 eight chromosomes are a stable configuration for all *Verticillium* species.

474 Centromere position and function are thought to be driven by the protein complement
475 (e.g. CenH3 localization) and by heterochromatin formation rather than by specific DNA
476 sequences (13, 15, 63). In *V. dahliae*, we observed the co-occurrence of CenH3 with H3K9me3
477 and DNA methylation. This suggests that DNA methylation, as previously reported in *N. crassa*

478 and in *C. neoformans* (16, 25), is also a feature of centromeric DNA in *V. dahliae*. Co-
479 localization of CenH3 with H3K9me2/3 and DNA methylation has been reported for *N. crassa*
480 (25) and *C. neoformans* (16). In contrast, H3K9me3 and H3K27me3 are absent from centromeres
481 in *Z. tritici* (27). H3K4me2 borders most centromeres in *Z. tritici* (27), and is associated with
482 centromeres in *S. pombe* and some animals and plants (64-67). H3K4me2 has not been observed
483 at centromeres in most fungi, including *V. dahliae*, and in the oomycete *P. sojae* (30). Changes in
484 heterochromatin in *N. crassa* leads to altered CenH3 positioning (25), suggesting that
485 heterochromatin is similarly required for centromere maintenance and function in *V. dahliae*.
486 Elevated AT-levels in repeat-rich heterochromatic regions can be caused by RIP mutations (15,
487 25, 26, 42). RIP-like mutations have been previously reported in some repeats in *V. dahliae* (36,
488 44), and we observed strong RIP signals at centromeres. Due to its presumably asexual nature (7),
489 the occurrence of RIP in *V. dahliae* is controversial (8, 43, 44). Noteworthy, mutational signatures
490 resembling RIP have recently been observed in *Z. tritici* propagated through mitotic cell
491 divisions, pointing to the existence of a mitotic version of a RIP-like process (42). Thus, we
492 conclude that RIP was an active process in *V. dahliae* at some point in evolution, or that RIP-like
493 processes outside of the sexual cycle occur in *V. dahliae.*

494 Centromeres are often enriched for a variety of different retrotransposons and other
495 repetitive elements (15, 16, 25, 29, 30). We similarly observed that centromeres in all *Verticillium*
496 species are repeat-rich. Repeats and their remnants identified at centromeres typically also occur
497 outside of centromeres, as observed in *M. oryzae* (29) and *N. crassa* (25). Strikingly, we observed
498 that a single repetitive element, *VdLTRE9*, is strongly associated with centromeres in some
499 *Verticillium* species, which to our knowledge, has only been observed in the fungus *Cryptococcus*
500 where centromeres contain six retrotransposons (*Tcn1-6*) that nearly exclusively occur at
501 centromeres (16). Similarly, centromeres of the oomycete plant pathogen *Phytophthora sojae*
502 contain multiple types of repeats, but they are enriched for a single element called CoLT (*Copia*-

503 Like Transposon) (30). The strong associations of specific repeats to centromeres could directly
504 or indirectly link these elements to centromere function. Functional centromeres as observed here
505 are also heterochromatic and contain CenH3. AT-rich repetitive elements can direct
506 heterochromatin formation via DNA methylation and H3K9me3 deposition in *N. crassa* (45, 68),
507 a phenomenon that can also occurs at repeats outside of centromeres (45). Heterochromatin
508 occurs at centromeres but also at repeat-rich regions outside of centromeres in *V. dahliae*, thus the
509 repeat-rich nature of centromeres is likely not sufficient to direct CenH3 deposition. In *S. pombe*
510 heterochromatin formation is directed by short interfering RNAs (siRNA) derived from flanking
511 repetitive elements via RNAi (69, 70), and RNAi and heterochromatin mediate CenH3
512 localization at centromeres (71, 72). RNAi is also important for centromere maintenance and
513 evolution in *Cryptococcus*, as RNAi deficient species have smaller centromeres than RNAi
514 proficient ones (16). Interestingly, centromere-specific elements (*Tcn1-6*) in RNAi proficient
515 species are typically full-length elements while only remnants can be found in RNAi deficient
516 species, which could be caused by recombination between elements (16). In *Verticillium*,
517 centromere size differences correlate with increase of repeat content and the recruitment of
518 *VdLTRE9*, which is highly fragmented and likely non-active. Furthermore, even though key
519 components of the RNAi machinery exist in at least some *Verticillium* species (73), we know
520 very little about its biological functions. Similarly, to *C. neoformans*, we observed no
521 transcriptional activity of *VdLTRE9* or any other repeat at centromeres, but it is unclear if this
522 silencing is mediated by RNAi, is a consequence of their heterochromatic nature, is due to their
523 fragmentation, or a combination of these. Ultimately, unravelling how specific elements
524 contribute to centromere identify necessitates future experiments. *VdLTRE9* occurs only in some
525 *Verticillium* species and has likely been recruited to centromeres subsequent to the divergence of
526 *V. nubilum*. Conversely, these observations raise further questions on the roles of repeats and
527 mechanisms of centromeric identity in species without *VdLTRE9*. Repeats are important drivers
528 of *Verticillium* genome evolution and function (8, 36), and here we highlight their contributions

529 to centromere diversity within the fungal genus *Verticillium*. Our analyses provide the framework
530 for future research into the diversity or convergence of mechanisms establishing centromere
531 identity and functioning in fungi.

MATERIAL & METHODS

Construction of *Verticillium dahliae* transformants expressing FLAG-tagged CenH3

CenH3 and H3 homologs were identified in the predicted proteomes of *V. dahliae* strain JR2 (32) and selected other fungi through a BLAST sequence similarity search (blastp v2.9.0+; default settings, e-value cutoff 1e-20) (74, 75) using the *N. crassa* CenH3 (Q7RXR3) and H3 (P07041) sequences as queries. Missing homologs of CenH3 or H3 were identified using manual BLAST (tblastn v2.9.0+; default settings) (74, 75) and exonerate (v2.2.0; default settings) (76) searches against the genome sequences. Protein sequences of selected CenH3 and H3 proteins were aligned using mafft (v7.271; default settings, LINSi) (77). A phylogenetic tree was inferred with maximum-likelihood methods implemented in IQ-tree (v1.6.11) (78) and robustness was assessed by 1,000 rapid bootstrap replicates.

To construct the N-terminally FLAG-tagged CenH3 strain of *V. dahliae*, a recombinant DNA fragment was constructed into the binary vector PRF-HU2 (79) or PRF-GU2 for homologous recombination. The CenH3 locus, from *V. dahliae* strain JR2, was amplified as 3 fragments with overlapping sequences (**Table S1f**). The 5' most fragment containing the promoter was amplified using primers A + B, the ORF with primers C+D, the Hyg promoter and ORF with primers E+F, and the 3' end of the CenH3 locus with primers G+H. The four fragments were combined by overlap PCR using primers A + H and cloned into a *Psp*OMI and *Sp*hI linearized vector using Gibson Assembly. The vector construction was confirmed by Sanger sequencing. Vectors were transformed to *Verticillium* with *Agrobacterium*-mediated transformation (80). Correct homologous recombination and replacement at the *CenH3* locus was verified by PCR amplification using primer I+J (**Fig. S1b**, **Table S1f**). Correct translation of the recombinant protein was assessed using Western analyses with anti-FLAG antibody (**Fig. S1c**). Briefly, proteins were extracted from 5-day old cultures grown in 100 ml Potato Dextrose Broth at 22°C with continuous shaking at 120 rpm. Mycelium was collected by straining over a double

557 layer of miracloth and subsequently snap-frozen in liquid nitrogen and ground with a mortar and
558 pestle using liquid nitrogen. Approximately 0.3 g of ground mycelium was resuspended in 600
559 µL protein extraction buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% glycerol,
560 0.02% NP-40, 2 mM Phenylmethanesulfonyl fluoride (PMSF), 100 µM Leupeptin, 1 µg/mL
561 Pepstatin), briefly vortexed, incubated on ice for 15 min and centrifuged at 4°C at 8,000 g for 3
562 min. The supernatant was collected by transferring 20 µL to a new tube to serve as the input
563 control and the remaining ~500 µL was transferred to a fresh microcentrifuge tube with 15 µL of
564 Anti-FLAG M2 affinity gel (catalog number A2220, Sigma-Aldrich, St. Louis, Missouri, United
565 States) and incubated while rotating at 4°C for 1 h. Samples were centrifuged at 5,000 g, 4°C for
566 3 min, after which the supernatant was discarded. Samples were washed with 500 µL of lysis
567 buffer, and the centrifugation and washing were repeated three times. Protein was eluted from the
568 resin by adding 15 µL of lysis buffer, 20 µL of 2x Laemmli loading buffer (4% SDS, 20%
569 glycerol, 0.004% bromophenol blue, 125 mM Tris HCL pH 6.8) and boiled at 95°C for 3 min.
570 Protein samples were separated on a 12% polyacrylamide gel, and subsequently transferred to
571 PVDF membranes, blocked in 5% BSA, washed twice in TBST, and incubated with 1:3500 anti-
572 FLAG antibody (monoclonal anti-FLAG M2; Merck KGaA, Darmstadt, Germany).
573
574 **Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq)**
575 For each *V. dahliae* genotype, one million spores were added to 100 ml Potato Dextrose Broth
576 and incubated for 7 days at 22°C with continuous shaking at 120 rpm. Mycelium was collected by
577 straining over a double layer of miracloth and subsequently snap-frozen in liquid nitrogen and
578 ground with a mortar and pestle using liquid nitrogen. All ground material (0.5-1 gram per
579 sample) was resuspended in 4 mL ChIP Lysis buffer (50 mM HEPES-KOH pH7.5, 140 mM
580 NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% NaDOC) and dounced 40 times in a 10 cm^3 glass
581 tube with tightly fitting pestle on 800 power with a RZR50 homogenizer (Heidolph, Schwabach,
582 Germany), followed by five rounds of 20 seconds sonication on ice with 40 seconds of resting in

583 between rounds with a Soniprep 150 (MSE, London, UK). Samples were redistributed to 2 mL
584 tubes and pelleted for 2 min at maximum speed in a tabletop centrifuge. Supernatants were
585 pooled per sample in a 15 mL tube together with 25 µL α-FLAG M2 magnetic beads (Sigma-
586 Aldrich, St. Louis, Missouri, United States), incubated overnight at 4°C and continuous rotation.
587 Beads were captured on a magnetic stand and washed with wash buffer (50 mM Tris HCl pH 8, 1
588 mM EDTA, 1% Triton X-100, 100 mM NaCL), high-salt wash buffer (50 mM Tris HCl pH 8, 1
589 mM EDTA, 1% Triton X-100, 350 mM NaCL), LiCl wash buffer (10 mM Tris HCl pH8, 1 mM
590 EDTA, 0.5% Triton X-100, 250 mM LiCl), and TE buffer (10 mM Tris HCl pH 8, 1mM EDTA).
591 Chromatin was eluted twice from beads by addition of 100 µL pre-heated TES buffer (100 mM
592 Tris HCl pH 8, 1% SDS, 10 mM EDTA, 50 mM NaCl) and 10 minutes incubation at 65°C. 10
593 mg/mL 2 µL Proteinase K was added and incubated at 65°C for 5 hours, followed chloroform
594 extraction. DNA was precipitated by addition of 2 volumes 100% ethanol, 1/10th volume 3 M
595 NaOAc pH 5.2 and 1/200th volume 20mg/mL glycogen, and overnight incubation at -20°C.

596 Sequencing libraries were prepared using the TruSeq ChIP Library Preparation Kit
597 (Illumina, city, country) according to the manufacterer's instructions, but without gel purification
598 and with use of the Velocity DNA Polymerase (BioLine, Luckenwalde, Germany) for 12 cycles
599 of amplification for the FLAG-CenH3. H3K4me2 ChIP was performed as described previously
600 (36), using an α-H3K4me2 antibody (#39913, ActiveMotif; Carlsbad, California, United States).
601 Single-end (125 bp) sequencing was performed on the Illumina HiSeq2500 platform at KeyGene
602 N.V. (Wageningen, the Netherlands).
603

604 **Chromatin confirmation capturing followed by high-throughput sequencing (Hi-C)**

605 We determined the inter- and intra-chromosomal contact frequencies using Hi-C in *V. dahliae*
606 strains CQ2, JR2, and VdLs17, as well as in *V. albo-atrum* strain PD747, *V. alfalfae* strain
607 PD683, *V. isaacii* strain PD618, *V. klebahnii* strain PD401, *V. longisporum* strain PD589, *V.*

608 *nonalfalfae* strain T2, *V. nubilum* strain 397, *V. tricorpus* strain PD593, and *V. zaregamsianum*
609 strain PD739. For each strain, one million spores were added to 400 mL Potato Dextrose Broth
610 and incubated for 6 days at 22°C with continuous shaking at 120 rpm. Mycelium was collected by
611 straining over double layer miracloth and 300 mg (fresh weight) was used as input for generating
612 Hi-C sequencing libraries with the Proximo Hi-C kit (Microbe) (Phase Genomics, Seattle, WA,
613 USA), according to manufacturer's instructions. Hi-C sequencing libraries of V. *dahliae* strains
614 CQ2, JR2. and VdLs17 were paired end (2x125 bp) sequenced on the Illumina HiSeq2500
615 platform at KeyGene N.V. (Wageningen, the Netherlands). Hi-C sequencing libraries of the other
616 *Verticillium* species were paired-end (2x150 bp) sequenced on the NextSeq500 platform at USEQ
617 (Utrecht, the Netherlands).
618
619 ***In vitro* transcriptome profiling using RNA-seq**
620 RNA sequencing of *V. albo-atrum* strain PD747, *V. isaacii* strain PD618, *V. klebahnii* strain
621 PD401, *V. longisporum* strain PD589, *V. nonalfalfae* strain T2, *V. nubilum* strain 397, *V.*
622 *tricorpus* strain PD593, and *V. zaregamsianum* strain PD739 as described previously (36). Single-
623 end (50 bp) sequencing was performed on the BGISeq500 platform at BGI (BGI Hong Kong).
624
625 **Analyses of high-throughput sequencing data**
626 High-throughput sequencing libraries (**Table S1a**) have been analyzed as follows: Illumina reads
627 were quality-filtered and trimmed using trimmomatic (version 0.36) (81). Sequencing reads were
628 trimmed and filtered by removing Illumina TruSeq sequencing adapters (settings seed
629 mismatches 2, palindrome clip threshold 30, and simple clip threshold 10), removal of low-
630 quality leading or trailing bases below quality 5 and 10, respectively, and 4-base sliding window
631 trimming and cutting when average quality per base dropped below 15. Additionally, filtered and
632 trimmed reads < 90 nt were removed from further analyses. Filtered and trimmed reads were
633 mapped to the corresponding genome assembly with Bowtie2 (default settings) (82), and mapping

files were converted to bam-format using samtools (v 1.8) (83). Genomic coverage was determined using deepTools (v3.4.1; bamCoverage) (84) by extending sequencing reads to 147 bp followed by RPGC normalization with a bin-size of 1,000 bp and smoothening of 3,000 bp. To assess between sample variability, we used deepTools (v3.4.1, plotPCA) (84) to generate principle component analyses. Furthermore, we employed deepTools (v3.4.1, multiBigwigSummary) (84) to summarize genomic coverages of over genes, repetitive elements, and genomic windows (5 kb windows with 500 bp slide). Genomic regions enriched for FLAG-CenH3 were identified using MACS2 (v2.1.1) (broad peak option; broad cutoff 0.0025) (85).

To determine DNA (cytosine) methylation, we utilized sequencing data of bisulfite treated genomic DNA previously generated for *V. dahliae* strain JR2 (36). Sequencing reads were mapped to the *V. dahliae* strain JR2 genome assembly as previously described (36). Subsequently, the number of reads supporting cytosine methylation in CG-context were extracted, and weighted CG-methylation levels were calculated over genes, repetitive elements, and genomic windows (5 kb window size with 500 bp slide) (86); weighted CG-methylation was defined as the sum of reads supporting cytosine methylations divided by the sum of all reads occurring at all CG sites in the respective regions. Sites with less than four reads were not considered.

To improve the genome assemblies of the *Verticillium* species, we mapped Hi-C sequencing reads to genome assemblies of *V. dahliae* strain CQ2, *V. albo-atrum* strain PD747, *V. alfalfae* strain PD683, *V. isaacii* strain PD618, *V. klebahnii* strain PD401, *V. longisporum* strain PD589, *V. nonalfalfae* strain T2, *V. nubilum* strain 397, *V. tricorpus* strain PD593, and *V. zaregamsianum* strain PD739 using Juicer (v1.6) with early stage setting (87). The contact matrices generated by juicer were used by the 3D de novo assembly (3D-DNA) pipeline (88) (v180922) with a contig size threshold of 1000bp to eliminate mis-joints in the previous assemblies and to generate improved assemblies. The genome assemblies were manually

659 improved using Juicebox Assembly Tools (JBAT) (v1.11.08) (89) and improved genome
660 assemblies were generated using the 3D-DNA post-review asm pipeline (88). Centromere
661 locations were determined using a 1 kb-resolution contact matrix in JBAT, by identifying a region
662 per chromosome that displays strong inter-chromosomal interactions, yet weak intra-
663 chromosomal interactions (see Figure S12, S13).

664 To assess potential repeat collapses during genome assemblies at centromeric regions, we
665 mapped previously generated short-read data *V. dahliae* strain JR2 and VdLs17, *V. albo-atrum*
666 strain PD747, *V. alfalfae* strain PD683, *V. isaacii* strain PD618, *V. klebahnii* strain PD401, *V.*
667 *longisporum* strain PD589, *Verticillium nonalfalfae* strain T2, *V. tricorpus* strain PD593, and *V.*
668 *zaregamsianum* strain PD739 (39, 40, 90, 91) to the genome assemblies using BWA (v0.7.17;
669 mem) (83). We first used bedtools (v2.29.2) (92) to identify few genomic regions with > 500x
670 coverage. We then applied deepTools (v3.4.1, computeGCBias) (84) to compute GC biases of
671 read depth across the genome, excluding the identified high coverage regions, and used
672 deepTools (v3.4.1, correctGCBias) (84) to correct GC biases, which addresses known biases in
673 sequencing library preparation to ensure even read coverage throughout the genome irrespective
674 of their base composition (93). We used deepTools (v3.4.1, bamCoverage, bins 50 bp, CPM
675 normalization) (84) to obtain the read coverage throughout the genome, excluding regions
676 containing sequence assembly gaps (Ns). Assuming that collapsed repeats would lead to a local
677 increase in read depth, we used the ratio of the average read coverage at the centromeres and
678 outside of the centromere at each chromosome to correct the inferred centromere sizes. To further
679 validate the genome assembly of regions identified as centromeres of *V. dahliae* strain JR2, the
680 genome assembly was compared to the previously generated optical map (35) using MapSolver (v
681 3.2; OpGen, Gaithersburg, MD).

682 The transcriptional activity for genes and repetitive elements in *V. dahliae* strain JR2 was
683 assessed *in vitro* (in Potato Dextrose Broth) using previously generated deep transcriptome

684 datasets (36). To this end, single-end sequencing reads of three biological replicates were mapped
685 to the *V. dahliae* strain JR2 genome assembly (32) using STAR (v2.4.2a; max. intron size 1 kb
686 and outFilterMismatchNmax to 5) (94). The resulting mapped reads were summarized per
687 genomic feature (gene or repeat) using summarizeOverlaps (95), converted to counts per million
688 (cpm) mapped reads, and averaged over the three biological replicates.
689
690 **Sequence analyses of *Verticillium* genome assemblies, centromeres, repeat and gene content**
691 Repetitive elements in the genomes of *V. dahliae* strains JR2, VdLs17 and CQ2 (32, 33) were
692 identified as previously described (36). Briefly, repetitive elements were identified in each
693 genome independently using a combination of LTRharvest (96) and LTRdigest (97) followed by
694 identification of RepeatModeler. Identified repeats in the different *V. dahliae* strains were
695 clustered into a non-redundant library that contained consensus sequences for each repeat family.
696 The repeat library was manually curated and annotated using PASTEC (98) or by sequence
697 similarity to previously identified and characterized repeat families (32, 44). Genome-wide
698 occurrences of repeat families were determined using RepeatMasker (v 4.0.9; sensitive option and
699 cutoff 250), and the output was postprocessed using 'One code to find then all' (99). We only
700 considered matches to the repeat consensus library, and thereby excluded simple repeats and low-
701 complexity regions.

702 *De novo* gene and repeat annotation for the Hi-C-improved *Verticillium* genome
703 assemblies, and for *V. dahliae* strains JR2 and VdLs17 as a comparison was performed using the
704 funannotate pipeline (100). Briefly, repetitive elements were first *de novo* identified using
705 RepeatModeler and masked for gene prediction using RepeatMasker. Subsequently, gene
706 prediction parameters were estimated using *in vitro* RNA-seq data (see above for details;
707 exception: *V. alfalfae* for which no RNA-seq data was available, *V. nonalfalfae* for which
708 publicly available RNA-seq data was used (90), and *V. dahliae* strain JR2 for which in addition to

the *in vitro* RNA-seq data generated in this study, also previously generated *in vitro* (xylem sap and half-MS; (36)) as well as long-read nanopore cDNA data (101) was used). Based on the gene prediction parameters, gene prediction was performed with funannotate using a combination of *ab initio* gene predictors, consensus predictions were obtained using Evidencemodeler (v1.1.1) (102), and gene predictions were adjusted using information from the RNA-seq data. Repeat annotation for each genome assembly was based on the *de novo* repeat family consensus sequences obtained with funannotate. Genome-wide occurrences of these repeat families as well as previously defined repeat families for *V. dahliae* (see above) were determined using RepeatMasker (v 4.0.9; sensitive option and cutoff 250), and the output was postprocessed using 'One code to find then all' (99). *De novo* repeat families overlapping with centromeres in the different species were clustered using BLASTClust (v2.2.26; parameter '-S 60 -L 0.55 -b F -p F'), and subsequently visualized using Cytoscape (v.3.8.0) (103). Next to RepeatMasker, genome-wide occurrences of the previously determined *VdLTRE9* (32, 36) were identified by BLAST searches (blastn v2.9.0+; e-value cutoff 1e-5, no soft-masking and dust, fixed database size 10e6) (74, 75), and similarity between VdLTRE9 consensus sequences and the *de novo* predicted repeat families was established using BLAST (blastn, e-value cutoff 1e-5, query coverage > 50%, no soft-masking and dust, fixed database size 10e6).

Repeat and gene density (*V. dahliae* strain JR2 and VdLs17 based on previous gene annotation (101)), GC-content, and composite RIP index were calculated along the genome sequence using sliding windows (5 kb window with 500 bp slide). The composite RIP index (CRI) was calculated according to Lewis et al. (45). CRI was determined by subtracting the RIP substrate from the RIP product index, which are defined by dinucleotide frequencies as follows: RIP product index = TpA / ApT and the RIP substrate index = (CpA + TpG)/(ApC + GpT). Overlaps between different genomic features (for example repetitive elements over centromeric

733 regions) was assessed using bedtools (v2.29.2) (92). Genome-wide data was visualized using R
734 (104) with the packages ggplot2 (105), karyplotR (106), or Gviz (107), as well as EasyFig (108).

735 Whole-genome alignments between *V. dahliae* strains JR2, VdLs17, and CQ2 were
736 performed using NUCmer, which is part of the MUMmer package (v 3.1; --maxmatch) (109). To
737 remove short matches, we only considered alignments longer than 10 kb. Ancestral genome
738 configurations were reconstructed using AnChro (56). We first determined the synteny
739 relationships between all possible pairs of haploid *Verticillium* genomes and two outgroup
740 genomes (*Plectosphaerella cucumerina* and *Sodiomyces alkalinus*) using SynChro with synteny
741 block stringency (delta parameter) ranging from 2-5 (110). We then obtained all ancestors by
742 calculating all possible pairs of genomes (G1 and G2) and outgroups (G3,..,G_n) and by varying
743 the delta' (G1 and G2 comparisons) and delta'' (G1/G3..G1/G_n and G2/G3..G2/G_n comparisons)
744 parameters for AnChro. We additionally reconstructed all ancestors starting from the extent
745 genomes in a sequential approach with multiple successive cycles of SynChro and AnChro (delta
746 parameters varied between 2-5). For each ancestor, we chose the optimal reconstructed by the
747 delta parameter combination (delta' and delta'') that minimizes the number of reconstructed
748 chromosomes and rearrangements and at the same time maximizes the number of genes, both
749 guided by the most commonly observed number of chromosomes and genes in all
750 rearrangements. We obtained the number of large-scale rearrangements between reconstructed
751 ancestral genomes and the extent *Verticillium* genomes using ReChro with a delta parameter of 1
752 (56). The relationship between chromosomes of the reconstructed ancestors and the extent species
753 in relationship to the common ancestor is generated with SynChro with a delta parameter of 1
754 (110). A species phylogeny that uses synteny relationships computed by SynChro (see above) as
755 informative character between the *Verticillium* genomes and the outgroup genomes was
756 reconstructed using PhyChro (111).
757

Data availability

ChIP-seq and Hi-C data were submitted to the Short Read Archive (SRA) under the accession PRJNA641329 (**Table S1a**).

ACKNOWLEDGMENTS

Work in the laboratories of M.F.S and B.P.H.J.T. is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands Organization of Scientific Research (NWO). Furthermore, B.P.H.J.T. would like to acknowledge the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2048/1 – Project ID: 390686111. This work was supported in part by a European Molecular Biology Organization postdoctoral fellowship (EMBO, ALTF 969-2013) and Human Frontier Science Program Postdoctoral Fellowship (HFSP, LT000627/2014-L) to D.E.C. A portion of the work was also carried out in the laboratory of D.E.C. under USDA-NIFA-PBI grant 2018-67013-28492. We thank Utrecht Sequencing Facility for providing sequencing service and data. Utrecht Sequencing Facility is subsidized by the University Medical Center Utrecht, Hubrecht Institute, Utrecht University, and The Netherlands X-omics Initiative (NWO project 184.034.019).

774 **SUPPLEMENTARY MATERIAL**

775 **Figure S1** – (**a**) Phylogenetic analyses of the canonical H3 and the centromeric-specific CenH3 in
776 *Verticillium dahliae* (strain JR2) and other fungal genomes. (**b-c**) Transformation of the coding
777 sequence of N-terminally FLAG-tagged CenH3 directed by its native promoter at the *CenH3*
778 locus in *Verticillium dahliae* strain JR2. (**b**) Correct homologous recombination and replacement
779 at the *CenH3* locus was verified by PCR amplification was assessed using PCR and (**c**) Correct
780 translation of the recombinant protein was assessed using Western Blot analyses with anti-FLAG
781 antibody. (**d**) Sequencing read coverage (RPGC normalization in 1 kb bins with 3 kb
782 smoothening) from ChIP-seq experiments using FLAG-tag antibodies on two independent
783 transformants of *Verticillium dahliae* strain JR2 that express FLAG-tagged CenH3 and the wild-
784 type strain are mapped to the eight chromosomes of *V. dahliae* strain JR2 (32). Gene (red) and
785 repeat (blue) density are shown below each chromosome. (**e**) Principal component analysis of the
786 four FLAG-tag ChIP-seq samples (two wild-type and two CenH3-FLGA). (**f**) Comparison of the
787 centromeric regions with the identified centromeres highlighted as blue block in the genome
788 assembly of *Verticillium dahliae* strain JR2 with a previously generated optical map (35). Vertical
789 lines display corresponding (*in silico*) restriction sites and their alignment.

790 **Figure S2** – Schematic overview of the eight chromosomes of *Verticillium dahliae* strain JR2
791 displaying different heterochromatin-associated chromatin modifications (mC, H3K9me3, and
792 H3K27me3) in relation to the centromeres. The different lanes display the CenH3-FLAG ChIP-
793 seq read coverage (RPGC normalization in 1 kb bins with 3 kb smoothening), the repeat-density,
794 the GC-content, the CRI as well as the weighted cytosine methylation (all summarized in 5 kb
795 windows with 500 bp slide), and the normalized H3K9me3 and H3K27me3 ChIP-seq read
796 coverage (RPGC normalization in 1 kb bins with 3 kb smoothening).

797

798 **Figure S3** – (**a**) Boxplot displaying the composite RIP index (CRI) of C to T in CA recorded in
799 genomic windows (5 kb, 500 bp slide), per gene, per annotated repeat, and per window
800 overlapping with the CenH3-enriched centromeres. Statistical differences for the indicated
801 comparisons were calculated using the one-sided non-parametric Mann-Whitney test; p-values <
802 0.001: ***. (**b**) Summary of H3K4me2 (green), H3K9me3 (red), and H3K27me3 (orange)
803 normalized ChIP-seq read coverage (RPGC normalization in 1 kb bins and 3 kb smoothening) in
804 genomic bins (2.5%) across the chromosomal arms of the eight chromosomes of *Verticillium*
805 *dahliae* strain JR2 (divided into 2.5% bins) and the centromeric regions (divided into 10% bins).
806 The dots indicate the average ChIP-seq coverage and the whiskers indicate ± 1.5 times the
807 interquartile range. (**c-e**) Boxplots displaying the (**c**) weighted methylation levels (CG context),
808 (**d**) the composite RIP index, and (**e**) the expression in PDB growth medium (counts per million)
809 for repetitive elements belonging to ten repeat families identified in the eight centromeres in
810 *Verticillium dahliae* JR2.

811 **Figure S4** (**a-c**) Whole-genome alignments between the eight chromosomes of (**a**) *Verticillium*
812 *dahliae* strains JR2 and VdLs17 (32), (**b**) *V. dahliae* strains CQ2 and JR2 (32, 33), and (**c**) *V.*
813 *dahliae* strains CQ2 and VdLs17 (32, 33). (**d-e**) Schematic overview of the genome assemblies of
814 *Verticillium dahliae* strains (**d**) VdLs17 and (**e**) CQ2. The individual lanes show the GC content,
815 the gene (red) and repeat (blue) density (all summarized in 5 kb windows with 500 bp slide), and
816 the location of the centromere associated *VdLTRE9*. (**f**) Synteny analyses of the eight
817 chromosomes of *V. dahliae* strains JR2 and CQ2. Schematic overview of the eight chromosomes
818 of *V. dahliae* strain JR2 (left) and the corresponding syntenic regions in *V. dahliae* strains CQ2
819 (right). Centromeres are indicated by stars, and syntenic centromeres of *V. dahliae* strain CQ2 are
820 colored according to *Cen1-8* of *V. dahliae* strain JR2.

821 **Figure S5**– Hi-C contact matrix showing the interaction frequencies between genomic regions in
822 (**a**) *V. nonalfalfae* (T2), (**b**) *V. alfalfae* (PD683), (**c**) the allodiploid *V. longisporum* (PD589), (**d**)

V. nubilum (397), (**e**) *V. albo-atrum* (PD747), (**f**) *V. zaregamsianum* (PD739), (**g**) *V. tricorpus* (PD593), (**h**) *V. klebhanii* (PD401), and (**i**) *V. isaacii* (PD618). Regions of high inter-chromosomal interaction frequencies are indicative of centromeres and are highlighted by arrow heads, and the blue line indicated boundaries between the pseudo-chromosomes.

Figure S6 – (**a-b**) Comparison of normalized read coverage and corrected centromere lengths for *Verticillium* species for which short-read data is available. (**a**) Counts per million mapped reads (CPM) normalized read coverage was calculated for GC-biased corrected short-read libraries in 50 bp genomic windows, excluding regions containing assembly gaps (Ns). Genomic windows are summarized in boxplots (outliers not shown) by genomic location, centromeric regions (*Cen*, blue) and non-centromeric regions (non-*Cen*, grey). (**b**) Centromeric lengths inferred by Hi-C data were 'corrected' based on the ratio of normalized read depth between centromeres and non-centromeric regions per chromosomes. Differences for each species compared to the overall mean were computed using unpaired T-tests; p-values < 0.0001: ****, p-values < 0.001: ***, p-values < 0.01: **, p-values < 0.05: *. (**c**) The number of BLASTn matches of the *VdLTRE9* consensus element to the genomes of the *Verticillium* species separated by their genomic location, centromeric regions (*Cen*, blue) and non-centromeric regions (non-*Cen*, grey). The overall number of base pairs (bp) covered by the BLASTn matches in each genome sequence is indicated. The asterisk denotes the high number of *VdLTRE9* matches to unassigned, non-*Cen* regions in the genome assembly of *Verticillium nonalfalfae* (T2). (**d**) The number of repetitive element matches identified by RepeatMasker for each *Verticillium* species based on species/strain-specific repeat libraries generated by RepeatModeler separated by their genomic location, centromeric regions (*Cen*, blue) and non-centromeric regions (non-*Cen*, grey). (**e**) GC-content of the *Verticillium* genomes in 50 bp windows and separated by their genomic location, centromeric regions (*Cen*, blue) and non-centromeric regions (non-*Cen*, grey). (∫) The repeat

847 content of centromeric regions in percent covered sequences in the different *Verticillium* species.
848 Each data point summarized in the boxplot is the repeat content per centromere.

849 **Figure S7** – Schematic overview of the centromeric regions (250 kb) in (**a**) *Verticillium dahliae*
850 strain JR2, in (**b**) species belonging to clade Flavnonexudans, and in (**c**) species belonging clade
851 Flavexudans. The centromeres are indicated by dark grey bars. The predicted genes (black) and
852 repeats (blue) are shown below each centromere, and location of *VdLTRE9* (partial) matches
853 (light green) are shown above each centromere. Repeats that share sequence similarity (BLASTn)
854 to the *VdLTRE9* consensus sequence are shown above each centromere (dark green).

855 **Figure S8** – Sequence comparisons of *de novo* repeat families identified with RepeatModeler and
856 RepeatMasker in the genome assemblies of the different *Verticillium* species. Individual repeat
857 family consensus sequences were clustered using BLASTClust. (**a**) Relationships between
858 different repeat family consensus sequences are displayed as connected graphs. The sub-graph
859 with the consensus sequences with similarity to *VdLTRE9* is highlighted in yellow. (**b**) The
860 presence/absence matrix indicates the occurrences of different repeat families in the analyzed
861 *Verticillium* species (black present, white absent). The cluster containing consensus sequences
862 with similarity to *VdLTRE9* is highlighted.

863 **Figure S9** – Reconstruction of ancestral genomes within the genus *Verticillium* with AnChro
864 (56). The number of (**a**) chromosomes and (**b**) genes predicted by all potential ancestral
865 reconstructions using different combinations of genomes and stringency parameters. The
866 phylogenetic tree in (**a**) depicts the relationships between *Verticillium* species and the
867 abbreviations used for the ancestors. The inlays display boxplots to summarize the number of (**a**)
868 chromosomes and (**b**) genes per ancestral reconstruction. (**c**) The number of chromosomes and
869 genes of the chosen 'optimal' reconstruction for each of the internal ancestors. (**d**) The number of
870 genes per chromosome for each of the reconstructed ancestor and the extant *Verticillium* species.
871 The star highlights the reconstruction for the B1 ancestor that had ten chromosomes, but with two

872 chromosomes with six and two genes. (**e**) Reconstruction of the *Verticillium* species phylogeny
873 based on synteny relationship using PhyChro (111).

874

875 **Table S1 –** (a) Overview of the different *Verticillium* sequencing libraries used in this study. (b)
876 Position of the individual centromeric regions inferred by Hi-C inter-chromosomal interaction
877 frequencies and the overlap (in kb) with CenH3-enriched regions and the centromere associated
878 *VdLTRE9* in *Verticillium dahliae* JR2, VdLs17, and CQ2. (c) Overview of the different
879 *Verticillium* genomes assembled using Hi-C interactions. (d) Position, length, and number of
880 assembly gaps (Ns) of the individual centromeric regions inferred by Hi-C inter-chromosomal
881 interaction in V*erticillium nonalfalfae* (T2), *Verticillium alfalfae* (PD683), the allodiploid
882 *Verticillium longisporum* (PD589), *Verticillium nubilum* (397),*Verticillium albo-atrum* (PD747),
883 *Verticillium zaregamsianum* (PD739), *Verticillium tricorpus* (PD593), *Verticillium klebhanii*
884 (PD401), and *Verticillium isaacii* (PD618). (e) The number of *de novo* repeat consensus
885 sequences identified within and outside of centromeric regions in the *Verticillium* species. Only
886 consensus elements with > 5 matches in centromeric regions are displayed. Note that the
887 consensus names between species/strains are not comparable. (f) The primers used for cloning the
888 CenH3 FLAG tag in *Verticillium dahliae* strain JR2

889

REFERENCES

1. Roy B, Sanyal K. 2011. Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. *Eukaryotic cell* 10:1384-1395.

2. Foley EA, Kapoor TM. 2013. Microtubule attachment and spindle assembly checkpoint signalling at the kinetochore. *Nature reviews Molecular cell biology* 14:25-37.

3. Burrack LS, Berman J. 2012. Flexibility of centromere and kinetochore structures. *Trends in Genetics* 28:204-212.

4. Janssen A, van der Burg M, Szuhai K, Kops GJPL, Medema RH. 2011. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science (New York, NY)* 333:1895-1898.

5. Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, Brito IL, Hiraoka Y, Niwa O, Amon A. 2011. Aneuploidy drives genomic instability in yeast. *Science (New York, NY)* 333:1026-1030.

6. Barra V, Fachinetti D. 2018. The dark side of centromeres: Types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nature Communications* 9:4340.

7. Seidl MF, Thomma BPHJ. 2014. Sex or no sex: Evolutionary adaptation occurs regardless. *BioEssays* 36:335-345.

8. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research* 26:1091-1100.

9. Rancati G, Pavelka N, Fleharty B, Noll A, Trimble R, Walton K, Perera A, Staehling-Hampton K, Seidel CW, Li R. 2008. Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* 135:879-893.

10. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R. 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468:321-325.

11. Sankaranarayanan SR, Ianiri G, Coelho MA, Reza MH, Thimmappa BC, Ganguly P, Vadnala RN, Sun S, Siddharthan R, Tellgren-Roth C, Dawson TLJ, Heitman J, Sanyal K. 2020. Loss of centromere function drives karyotype evolution in closely related *Malassezia* species. *Elife* 9.

12. Ola M, O'Brien CE, Coughlan AY, Ma Q, Donovan PD, Wolfe KH, Butler G. 2020. Polymorphic centromere locations in the pathogenic yeast *Candida parapsilosis*. *bioRxiv* doi:10.1101/2020.04.09.034512:2020.04.09.034512.

13. Yadav V, Sreekumar L, Guin K, Sanyal K. 2018. Five pillars of centromeric chromatin in fungal pathogens. *PLoS pathogens* 14:e1007150.

14. Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* 293:1098-102.

15. Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. 2012. Centromeres of filamentous fungi. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 20:635-656.

16. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, Bakkeren G, Cuomo CA, Heitman J, Sanyal K. 2018. RNAi is a critical determinant of centromere evolution in closely related fungi. *Proceedings of the National Academy of Sciences of the United States of America* 115:3108-3113.

17. Fitzgerald-Hayes M, Clarke L, Carbon J. 1982. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell* 29:235-244.

18. Furuyama S, Biggins S. 2007. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America* 104:14706-14711.

19. Krassovsky K, Henikoff JG, Henikoff S. 2012. Tripartite organization of centromeric chromatin in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America* 109:243-248.

20. Cliften PF, Fulton RS, Wilson RK, Johnston M. 2006. After the duplication: gene loss and adaptation in *Saccharomyces* genomes. *Genetics* 172:863-872.

945 21. Baum M, Sanyal K, Mishra PK, Thaler N, Carbon J. 2006. Formation of functional
946 centromeric chromatin is specified epigenetically in *Candida albicans*. *Proceedings of the*
947 *National Academy of Sciences of the United States of America* 103:14877-14882.

948 22. Padmanabhan S, Thakur J, Siddharthan R, Sanyal K. 2008. Rapid evolution of Cse4p-rich
949 centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida*
950 *dubliniensis*. *Proceedings of the National Academy of Sciences of the United States of America*
951 105:19797-19802.

952 23. Sanyal K, Baum M, Carbon J. 2004. Centromeric DNA sequences in the pathogenic yeast
953 *Candida albicans* are all different and unique. *Proceedings of the National Academy of Sciences*
954 *of the United States of America* 101:11374-11379.

955 24. Cambareri EB, Aisner R, Carbon J. 1998. Structure of the chromosome VII centromere
956 region in Neurospora crassa: degenerate transposons and simple repeats. *Molecular and cellular*
957 *biology* 18:5465-5477.

958 25. Smith KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M. 2011. Heterochromatin
959 is required for normal distribution of *Neurospora crassa* CenH3. *Molecular and cellular biology*
960 31:2528-2542.

961 26. Selker EU. 2002. Repeat-induced gene silencing in fungi. *Advances in genetics* 46:439-
962 450.

963 27. Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M,
964 Stukenbrock EH. 2015. Histone modifications rather than the novel regional centromeres of
965 *Zymoseptoria tritici* distinguish core and accessory chromosomes. *Epigenetics & Chromatin* 8:41.

966 28. Thomma BPHJ, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JAL, Faino L.
967 2016. Mind the gap; Seven reasons to close fragmented genome assemblies. *Fungal Genetics and*
968 *Biology* 90:24-30.

969 29. Yadav V, Yang F, Reza MH, Liu S, Valent B, Sanyal K, Naqvi NI. 2019. Cellular
970 dynamics and genomic identity of centromeres in cereal blast fungus. *mBio* 10.

971 30. Fang Y, Coelho MA, Shu H, Schotanus K, Thimmappa BC, Yadav V, Chen H, Malc EP,
972 Wang J, Mieczkowski PA, Kronmiller B, Tyler BM, Sanyal K, Dong S, Nowrousian M, Heitman

973 J. 2020. Long transposon-rich centromeres in an oomycete reveal divergence of centromere
974 features in Stramenopila-Alveolata-Rhizaria lineages. *PLoS Genetics* 16:e1008646.

975 31. Navarro-Mendoza MI, Perez-Arques C, Panchal S, Nicolas FE, Mondo SJ, Ganguly P,
976 Pangilinan J, Grigoriev IV, Heitman J, Sanyal K, Garre V. 2019. Early diverging fungus *Mucor*
977 *circinelloides* lacks centromeric histone CENP-A and displays a mosaic of point and regional
978 centromeres. *Current Biology* 29:3791-3802 e6.

979 32. Faino L, Seidl MF, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, Thomma
980 BPHJ. 2015. Single-Molecule Real-Time sequencing combined with optical mapping yields
981 completely finished fungal genome. *mBio* 6:e00936-15.

982 33. Depotter JRL, Shi-Kunne X, Missonnier H, Liu T, Faino L, van den Berg GCM, Wood
983 TA, Zhang B, Jacques A, Seidl MF, Thomma BPHJ. 2019. Dynamic virulence-related regions of
984 the plant pathogenic fungus *Verticillium dahliae* display enhanced sequence conservation.
985 *Molecular Ecology* 28:3482-3495.

986 34. de Jonge R, van Esse PH, Maruthachalam K, Bolton MD, Santhanam P, Saber MK,
987 Zhang Z, Usami T, Lievens B, Subbarao KV, Thomma BPHJ. 2012. Tomato immune receptor
988 Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA
989 sequencing. *Proceedings of the National Academy of Sciences of the United States of America*
990 109:5110-5115.

991 35. de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ.
992 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen.
993 *Genome research* 23:1271-1282.

994 36. Cook DE, Kramer M, Seidl MF, Thomma BP. 2020. Chromatin features define adaptive
995 genomic regions in a fungal plant pathogen. *bioRxiv*
996 doi:10.1101/2020.01.27.921486:2020.01.27.921486.

997 37. Inderbitzin P, Bostock RM, Davis RM, Usami T, Platt HW, Subbarao KV. 2011.
998 Phylogenetics and taxonomy of the fungal vascular wilt pathogen Verticillium, with the
999 descriptions of five new species. *PloS one* 6:e28341.

1000 38. Inderbitzin P, Davis RM, Bostock RM, Subbarao KV. 2011. The ascomycete Verticillium
1001 longisporum is a hybrid and a plant pathogen with an expanded host range. *PloS one* 6:e18260.

1002 39. Depotter JRL, Beveren Fv, Rodriguez-Moreno L, van den Berg GCM, Wood TA,
1003 Thomma BPHJ, Seidl MF. 2018. Homogenization of sub-genome secretome gene expression
1004 patterns in the allodiploid fungus *Verticillium longisporum*. *bioRxiv*:341636.

1005 40. Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF. 2018. Evolution
1006 within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene
1007 loss. *Environmental Microbiology* 20:1362-1373.

1008 41. Depotter JRL, Seidl MF, van den Berg GCM, Thomma BPHJ, Wood TA. 2017. A
1009 distinct and genetically diverse lineage of the hybrid fungal pathogen *Verticillium longisporum*
1010 population causes stem striping in British oilseed rape. *Environmental Microbiology* 19:3997-
1011 4009.

1012 42. Moeller M, Habig M, Lorrain C, Feurtey A, Haueisen J, Fagundes WC, Alizadeh A,
1013 Freitag M, Stukenbrock EH. 2020. Recent loss of the Dim2 cytosine DNA methyltransferase
1014 impacts mutation rate and evolution in a fungal plant pathogen. *bioRxiv*
1015 doi:10.1101/2020.03.27.012203:2020.03.27.012203.

1016 43. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z,
1017 Henrissat B, Lee Y-H, Park J, Garcia-Pedrajas MD, Barbara DJ, Anchieta A, de Jonge R,
1018 Santhanam P, Maruthachalam K, Atallah Z, Amyotte SG, Paz Z, Inderbitzin P, Hayes RJ, Heiman
1019 DI, Young S, Zeng Q, Engels R, Galagan J, Cuomo CA, Dobinson KF, Ma L-J. 2011.
1020 Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens.
1021 *PLoS pathogens* 7:e1002137.

1022 44. Amyotte SG, Tan X, Pennerman K, del Mar Jimenez-Gasco M, Klosterman SJ, Ma L-J,
1023 Dobinson KF, Veronese P. 2012. Transposable elements in phytopathogenic Verticillium spp.:
1024 insights into genome evolution and inter- and intra-specific diversification. *BMC genomics*
1025 13:314.

1026 45. Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schubeler D, Selker
1027 EU. 2009. Relics of repeat-induced point mutation direct heterochromatin formation in
1028 *Neurospora crassa*. *Genome Research* 19:427-437.

1029 46. Liu S-Y, Lin J-Q, Wu H-L, Wang C-C, Huang S-J, Luo Y-F, Sun J-H, Zhou J-X, Yan S-
1030 J, He J-G, Wang J, He Z-M. 2012. Bisulfite sequencing reveals that Aspergillus flavus holds a
1031 hollow in DNA methylation. *PloS one* 7:e30349.

1032 47. Seymour M, Ji L, Santos AM, Kamei M, Sasaki T, Basenko EY, Schmitz RJ, Zhang X,
1033 Lewis ZA. 2016. Histone H1 Limits DNA Methylation in *Neurospora crassa*. *G3* 6:1879-1889.

1034 48. Kursel LE, Malik HS. 2016. Centromeres. *Current Biology* 26:R487-R490.

1035 49. Friedman S, Freitag M. 2017. Evolving centromeres and kinetochores. *Advances in*
1036 *Genetics* 98:1-41.

1037 50. Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering transposable element
1038 diversification in de novo annotation approaches. *PloS one* 6:e16526.

1039 51. Galazka JM, Klocko AD, Uesake M, Honda S, Selker EU, Freitag M. 2016. Neurospora
1040 chromosomes are organized by blocs of importin alpha-dependent heterochromatin that are
1041 largely independent of H3K9me3. *Genome research* doi:10.1101/gr.203182.115:gr.203182.115.

1042 52. Winter DJ, Ganley ARD, Young CA, Liachko I, Schardl CL, Dupont P-y, Berry D, Ram
1043 A, Scott B, Cox MP. 2018. Repeat elements organise 3D genome structure and mediate
1044 transcription in the filamentous fungus *Epichloë festucae*. *PLoS Genetics* 14:e1007467.

1045 53. Marie-Nelly H, Marbouty M, Cournac A, Flot J-F, Liti G, Parodi DP, Syan S, Guillén N,
1046 Margeot A, Zimmer C, Koszul R. 2014. High-quality genome (re)assembly using chromosomal
1047 contact data. *Nature Communications* 5:5695.

1048 54. Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P,
1049 Dekker J, Mirny L, Barrowman J, Grewal SIS. 2014. Cohesin-dependent globules and
1050 heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516:432-435.

1051 55. Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert J-P, Noble
1052 WS. 2015. Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic*
1053 *Acids Research* 43:5331-5339.

1054 56. Vakirlis N, Sarilar V, Drillon G, Fleiss A, Agier N, Meyniel JP, Blanpain L, Carbone A,
1055 Devillers H, Dubois K, Gillet-Markowska A, Graziani S, Huu-Vang N, Poirel M, Reisser C,
1056 Schott J, Schacherer J, Lafontaine I, Llorente B, Neuveglise C, Fischer G. 2016. Reconstruction
1057 of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution
1058 in a model yeast genus. *Genome Research* 26:918-32.

1059 57. Depotter JR, Deketelaere S, Inderbitzin P, Tiedemann AV, Hofte M, Subbarao KV,
1060 Wood TA, Thomma BP. 2016. *Verticillium longisporum*, the invisible threat to oilseed rape and
1061 other brassicaceous plant hosts. *Molecular Plant Pathology* 17:1004-16.

1062 58. Muller H, Gil J, Jr., Drinnenberg IA. 2019. The impact of centromeres on spatial genome
1063 architecture. *Trends in Genetics* 35:565-578.

1064 59. Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing:
1065 Computational challenges and solutions. *Nature Reviews Genetics* 13:36-46.

1066 60. Seidl MF, Faino L, Shi-Kunne X, van den Berg GCM, Bolton MD, Thomma BPHJ.
1067 2015. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector
1068 repertoire resembling that of its pathogenic relatives. *Molecular Plant-Microbe Interactions*
1069 28:362-373.

1070 61. Sun S, Yadav V, Billmyre RB, Cuomo CA, Nowrousian M, Wang L, Souciet JL,
1071 Boekhout T, Porcel B, Wincker P, Granek JA, Sanyal K, Heitman J. 2017. Fungal genome and
1072 mating system transitions facilitated by chromosomal translocations involving intercentromeric
1073 recombination. *PLoS Biology* 15:e2002527.

1074 62. International Chicken Genome Sequencing C. 2004. Sequence and comparative analysis
1075 of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.

1076 63. Fukagawa T. 2017. Critical histone post-translational modifications for centromere
1077 function and propagation. *Cell Cycle* 16:1259-1265.

1078 64. Sullivan BA, Karpen GH. 2004. Centromeric chromatin exhibits a histone modification
1079 pattern that is distinct from both euchromatin and heterochromatin. *Nature Structural &*
1080 *Molecular Biology* 11:1076-83.

1081 65. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of
1082 heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833-7.

1083 66. Li X, Wang X, He K, Ma Y, Su N, He H, Stolc V, Tongprasit W, Jin W, Jiang J,
1084 Terzaghi W, Li S, Deng XW. 2008. High-resolution mapping of epigenetic modifications of the
1085 rice genome uncovers interplay between DNA methylation, histone methylation, and gene
1086 expression. *Plant Cell* 20:259-76.

1087 67. Blower MD, Sullivan BA, Karpen GH. 2002. Conserved organization of centromeric
1088 chromatin in flies and humans. *Dev Cell* 2:319-30.

1089 68. Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Knip M, Sack R, Selker EU. 2010. DNA
1090 methylation and normal chromosome behavior in *Neurospora* depend on five components of a
1091 histone methyltransferase complex, DCDC. *PLoS Genetics* 6:e1001196.

1092 69. Bayne EH, White SA, Kagansky A, Bijos DA, Sanchez-Pulido L, Hoe KL, Kim DU,
1093 Park HO, Ponting CP, Rappsilber J, Allshire RC. 2010. Stc1: A critical link between RNAi and
1094 chromatin modification required for heterochromatin integrity. *Cell* 140:666-77.

1095 70. Buhler M, Moazed D. 2007. Transcription and RNAi in heterochromatic gene silencing.
1096 *Nature Structural & Molecular Biology* 14:1041-8.

1097 71. Yang J, Sun S, Zhang S, Gonzalez M, Dong Q, Chi Z, Chen YH, Li F. 2018.
1098 Heterochromatin and RNAi regulate centromeres by protecting CENP-A from ubiquitin-mediated
1099 degradation. *PLoS Genetics* 14:e1007572.

1100 72. Kagansky A, Folco HD, Almeida R, Pidoux AL, Boukaba A, Simmer F, Urano T,
1101 Hamilton GL, Allshire RC. 2009. Synthetic heterochromatin bypasses RNAi and centromeric
1102 repeats to establish functional centromeres. *Science* 324:1716-9.

1103 73. Jesenicnik T, Stajner N, Radisek S, Jakse J. 2019. RNA interference core components
1104 identified and characterised in *Verticillium nonalfalfae*, a vascular wilt pathogenic plant fungi of
1105 hops. *Scientific Reports* 9:8651.

1106 74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
1107 search tool. *Journal of molecular biology* 215:403-410.

1108 75. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.
1109 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

1110 76. Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence
1111 comparison. *BMC Bioinformatics* 6:31.

1112 77. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
1113 improvements in performance and usability. *Molecular biology and evolution* 30:772-780.

1114 78. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and
1115 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular*
1116 *Biology and Evolution* 32:268-74.

1117 79. Frandsen RJ, Andersson JA, Kristensen MB, Giese H. 2008. Efficient four fragment
1118 cloning for the construction of vectors for targeted gene replacement in filamentous fungi. *BMC*
1119 *Mol Biol* 9:70.

1120 80. Santhanam P. 2012. Random insertional mutagenesis in fungal genomes to identify
1121 virulence factors. *Methods Mol Biol* 835:509-17.

1122 81. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
1123 sequence data. *Bioinformatics (Oxford, England)* 30:2114-2120.

1124 82. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature*
1125 *methods* 9:357-359.

1126 83. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
1127 transform. *Bioinformatics (Oxford, England)* 25:1754-1760.

1128 84. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F,
1129 Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis.
1130 *Nucleic Acids Res* 44:W160-5.

1131 85. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers
1132 RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome*
1133 *Biology* 9:R137.

1134 86. Schultz MD, Schmitz RJ, Ecker JR. 2012. 'Leveling' the playing field for analyses of
1135 single-base resolution DNA methylomes. *Trends in Genetics* 28:583-585.

1136 87. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016.
1137 Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell*
1138 *Systems* 3:95-8.

1139 88. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS,
1140 Machol I, Lander ES, Aiden AP, Aiden EL. 2017. *De novo* assembly of the *Aedes aegypti*
1141 genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92-95.

1142 89. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M,
1143 Glenn St Hilaire B, Yao W, Stamenova E, Hoeger M, Nyquist SK, Korchina V, Pletch K,
1144 Flanagan JP, Tomaszewicz A, McAloose D, Pérez Estrada C, Novak BJ, Omer AD, Aiden EL.
1145 2018. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes
1146 with chromosome-length scaffolds for under $1000. *bioRxiv* doi:10.1101/254797:254797.

1147 90. Jakse J, Jelen V, Radisek S, de Jonge R, Mandelc S, Majer A, Curk T, Zupan B, Thomma
1148 B, Javornik B. 2018. Genome sequence of a lethal strain of xylem-Invading *Verticillium*
1149 *nonalfalfae*. *Genome Announcements* 6.

1150 91. de Jonge R, Peter van Esse H, Maruthachalam K, Bolton MD, Santhanam P, Saber MK,
1151 Zhang Z, Usami T, Lievens B, Subbarao KV, Thomma BPHJ. 2012. Tomato immune receptor
1152 Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA
1153 sequencing. *Proceedings of the National Academy of Sciences of the United States of America*
1154 109:5110-5115.

1155 92. Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing
1156 genomic features. *Bioinformatics* 26:841-2.

1157 93. Benjamini Y, Speed TP. 2012. Summarizing and correcting the GC content bias in high-
1158 throughput sequencing. *Nucleic Acids Research* 40:e72.

1159 94. Dobin A, Gingeras TR. 2015. Mapping RNA-seq reads with STAR. *Current Protocols in*
1160 *Bioinformatics* 51:11 14 1-11 14 19.

1161 95. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT,
1162 Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Computational*
1163 *Biology* 9:e1003118.

1164 96. Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software
1165 for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.

1166 97. Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and
1167 classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res* 37:7002-13.

1168 98. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, Quesneville H.
1169 2014. PASTEC: an automatic transposable element classification tool. *PloS one* 9:e91929.

1170 99. Bailly-Bechet M, Haudry A, Lerat E. 2014. "One code to find them all": a perl tool to
1171 conveniently parse RepeatMasker output files. *Mobile DNA* 5:13.

1172 100. Palmer JM, JE S. 2016. Funannotate: Eukaryotic genome annotation pipeline.
1173 http://funannotate.readthedocs.io. Accessed

1174 101. Cook DE, Valle-Inclan JE, Pajoro A, Rovenich H, Thomma B, Faino L. 2019. Long-
1175 Read Annotation: Automated Eukaryotic Genome Annotation Based on Long-Read cDNA
1176 Sequencing. *Plant Physiol* 179:38-54.

1177 102. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman
1178 JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the
1179 Program to Assemble Spliced Alignments. *Genome Biology* 9:R7.

1180 103. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski
1181 B, Ideker T. 2003. Cytoscape: A software environment for integrated models of biomolecular
1182 interaction networks. *Genome Research* 13:2498-504.

1183 104. Team RC. 2013. A Language and Environment for Statistical Computing}.

1184 105. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis}.

1185 106. Gel B, Serra E. 2017. karyoploteR: an R/Bioconductor package to plot customizable
1186 genomes displaying arbitrary data. *Bioinformatics* 33:3088-3090.

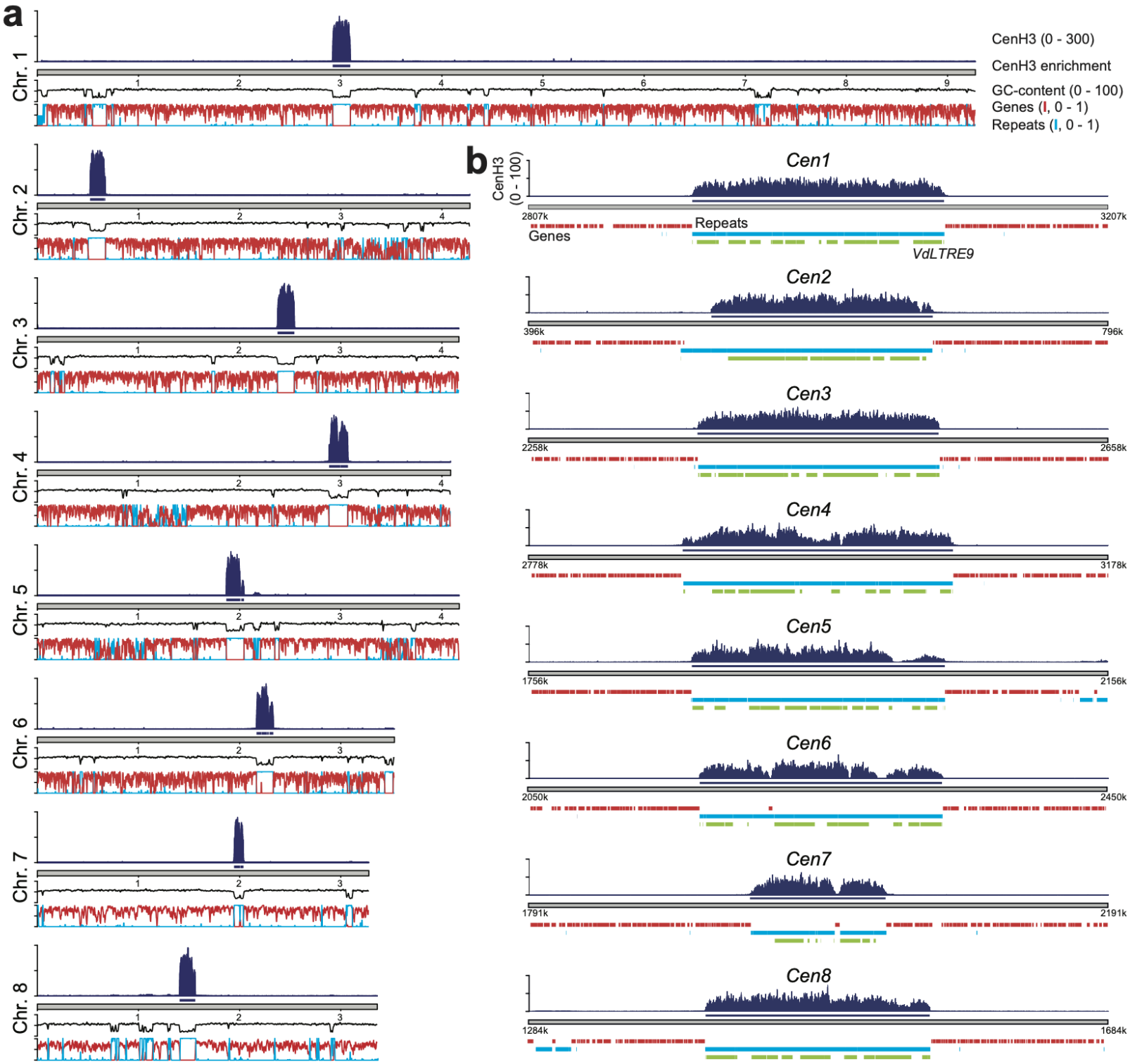
1187 107. Hahne F, Ivanek R. 2016. Visualizing genomic data using Gviz and Bioconductor.
1188 *Methods in molecular biology (Clifton, NJ)* 1418:335-351.

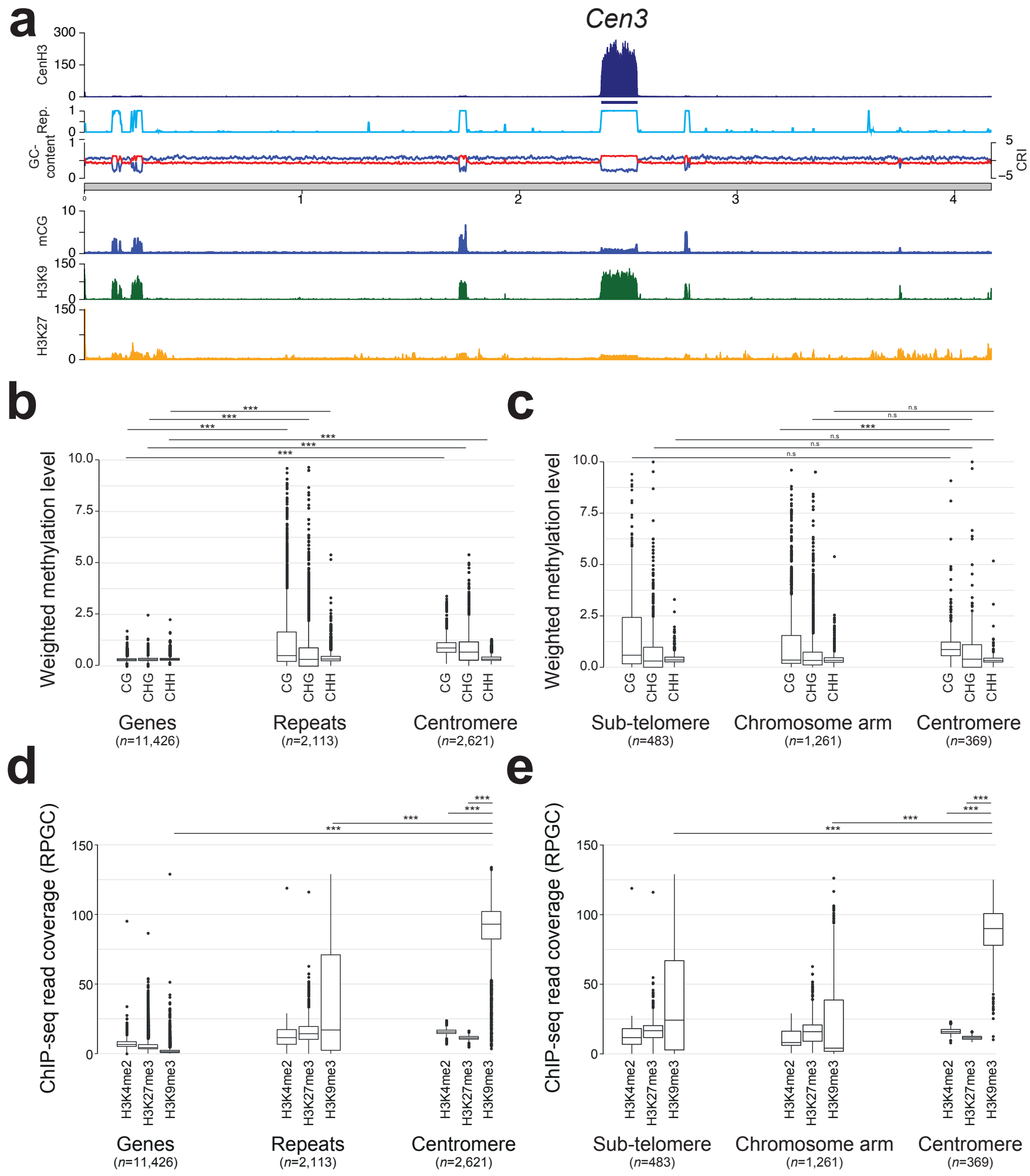
1189 108. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: A genome comparison visualizer.
1190 *Bioinformatics* 27:1009-10.

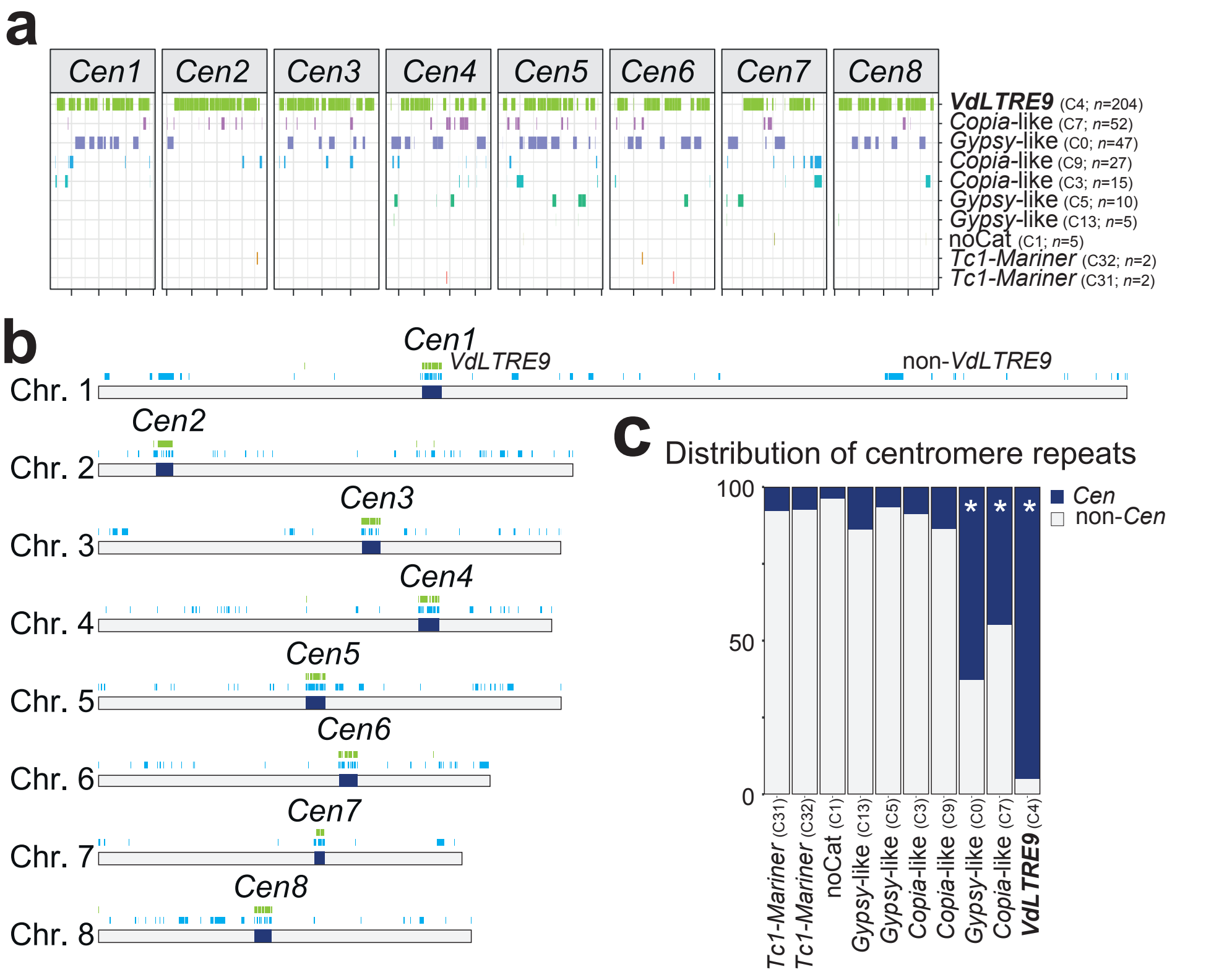
1191 109. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL.
1192 2004. Versatile and open software for comparing large genomes. *Genome Biology* 5:R12.

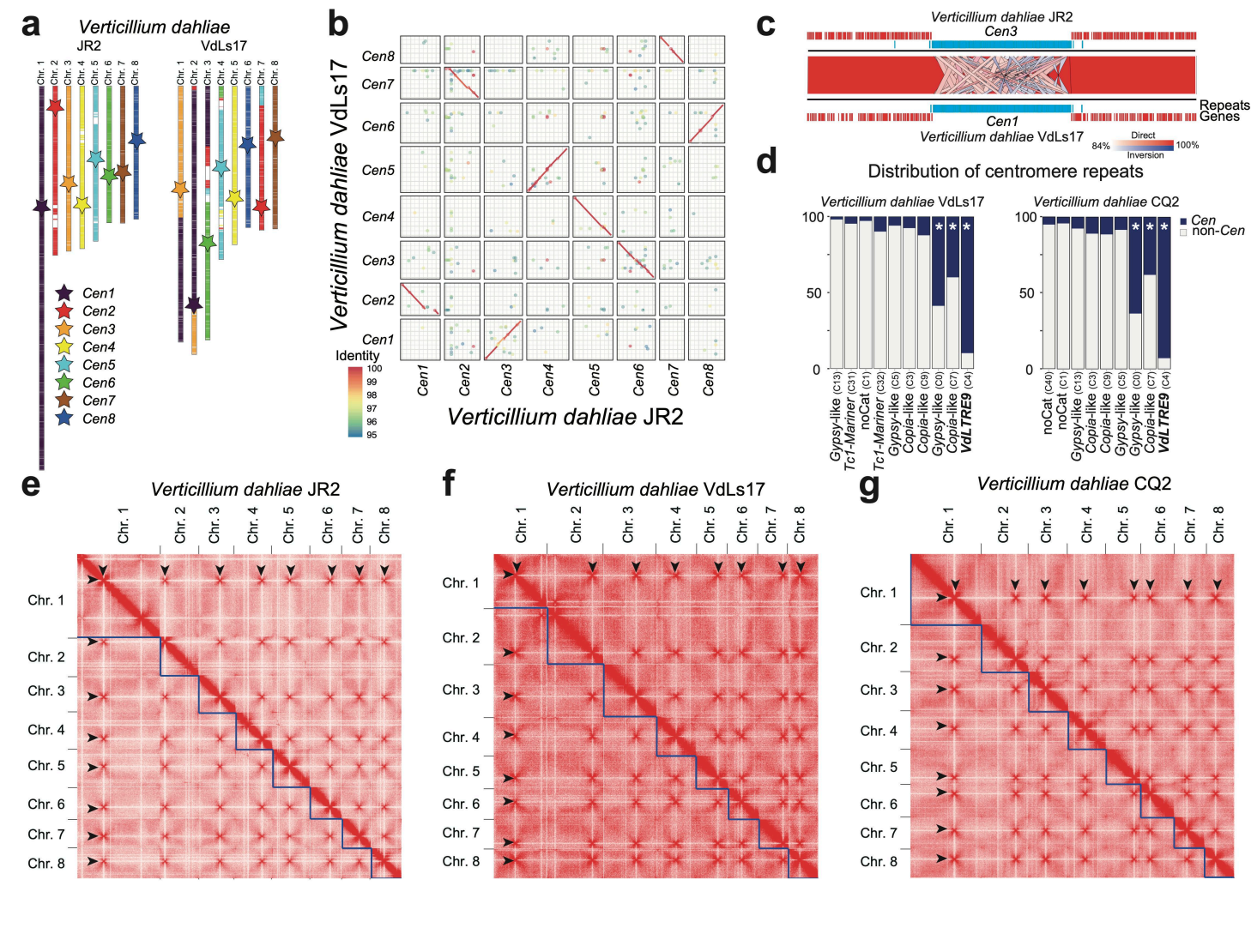
1193 110. Drillon G, Carbone A, Fischer G. 2014. SynChro: A fast and easy tool to reconstruct and
1194 visualize synteny blocks along eukaryotic chromosomes. *PLoS One* 9:e92621.

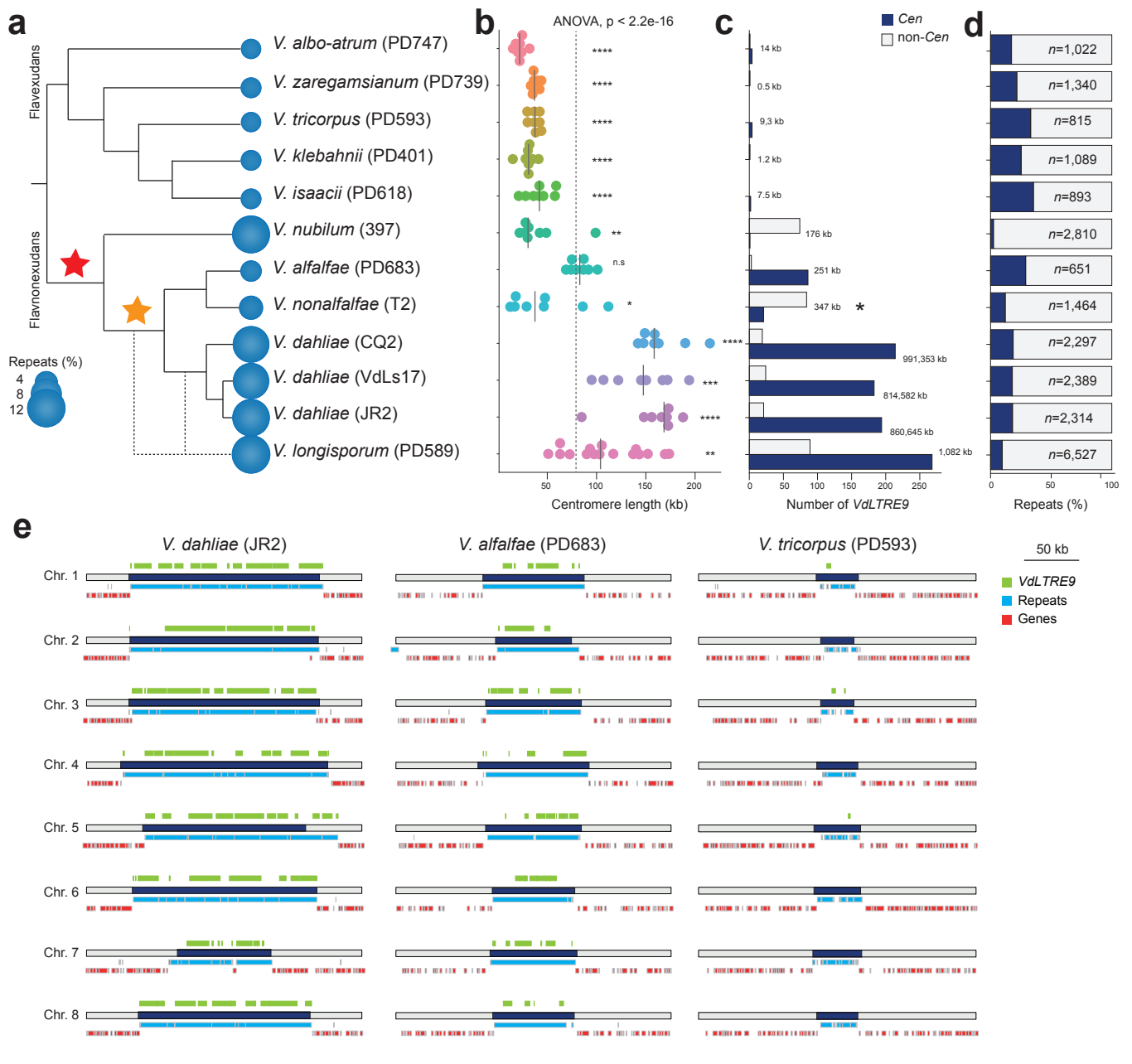
111. Drillon G, Champeimont R, Oteri F, Fischer G, Carbone A. 2020. Phylogenetic reconstruction based on synteny block and gene adjacencies. *Molecular Biology and Evolution* doi:10.1093/molbev/msaa114.




a
Cen3
CenH3
Rep.
GC-content
CRI
mCG
H3K9
H3K27
b
Weighted methylation level
CG
CHG
CHH
Genes
(n=11,426)
Repeats
(n=2,113)
Centromere
(n=2,621)
c
Sub-telomere
(n=483)
Chromosome arm
(n=1,261)
Centromere
(n=369)
d
ChIP-seq read coverage (RPGC)
H3K4me2
H3K27me3
H3K9me3
e


a
Cen1
Cen2
Cen3
Cen4
Cen5
Cen6
Cen7
Cen8
VdLTRE9 (C4; n=204)
Copia-like (C7; n=52)
Gypsy-like (C0; n=47)
Copia-like (C9; n=27)
Copia-like (C3; n=15)
Gypsy-like (C5; n=10)
Gypsy-like (C13; n=5)
noCat (C1; n=5)
Tc1-Mariner (C32; n=2)
Tc1-Mariner (C31; n=2)
b
Cen1
VdLTRE9
non-VdLTRE9
Chr. 1
Cen2
Chr. 2
Cen3
Chr. 3
Cen4
Chr. 4
Cen5
Chr. 5
Cen6
Chr. 6
Cen7
Chr. 7
Cen8
Chr. 8
c
Distribution of centromere repeats
100
50
0
Cen
non-Cen
*
*
*
Tc1-Mariner (C31)
Tc1-Mariner (C32)
noCat (C1)
Gypsy-like (C13)
Gypsy-like (C5)
Copia-like (C3)
Copia-like (C9)
Gypsy-like (C0)
Copia-like (C7)
VdLTRE9 (C4)


a
Verticillium dahliae
JR2
VdLs17
Chr. 1
Chr. 2
Chr. 3
Chr. 4
Chr. 5
Chr. 6
Chr. 7
Chr. 8
Cen1
Cen2
Cen3
Cen4
Cen5
Cen6
Cen7
Cen8
b
Verticillium dahliae VdLs17
Verticillium dahliae JR2
Identity
100
99
98
97
96
95
c
Verticillium dahliae JR2
Cen3
Repeats
Genes
Cen1
Verticillium dahliae VdLs17
84%
Direct
Inversion
100%
d
Distribution of centromere repeats
Verticillium dahliae VdLs17
Verticillium dahliae CQ2
Cen
non-Cen
100
50
0
Gypsy-like (C13)
Tc1-Mariner (C31)
noCat (C1)
Tc1-Mariner (C32)
Gypsy-like (C5)
Copia-like (C3)
Copia-like (C9)
Gypsy-like (C0)
Copia-like (C7)
VdLTRE9 (C4)
noCat (C40)
noCat (C1)
Gypsy-like (C13)
Copia-like (C3)
Copia-like (C9)
Gypsy-like (C5)
Gypsy-like (C0)
Copia-like (C7)
VdLTRE9 (C4)
e
Verticillium dahliae JR2
f
Verticillium dahliae VdLs17
g
Verticillium dahliae CQ2


a
Flavexudans
Flavnonexudans
V. albo-atrum (PD747)
V. zaregamsianum (PD739)
V. tricorpus (PD593)
V. klebahnii (PD401)
V. isaacii (PD618)
V. nubilum (397)
V. alfalfae (PD683)
V. nonalfalfae (T2)
V. dahliae (CQ2)
V. dahliae (VdLs17)
V. dahliae (JR2)
V. longisporum (PD589)
Repeats (%)
4
8
12
b
ANOVA, p < 2.2e-16
****
****
****
****
****
**
n.s
*
****
***
****
**
50
100
150
200
Centromere length (kb)
c
Cen
non-Cen
14 kb
0.5 kb
9,3 kb
1.2 kb
7.5 kb
176 kb
251 kb
347 kb
*
991,353 kb
814,582 kb
860,645 kb
1,082 kb
0
50
100
150
200
250
Number of VdLTRE9
d
n=1,022
n=1,340
n=815
n=1,089
n=893
n=2,810
n=651
n=1,464
n=2,297
n=2,389
n=2,314
n=6,527
0
50
100
Repeats (%)
e
V. dahliae (JR2)
V. alfalfae (PD683)
V. tricorpus (PD593)
50 kb
VdLTRE9
Repeats
Genes
Chr. 1
Chr. 2
Chr. 3
Chr. 4
Chr. 5
Chr. 6
Chr. 7
Chr. 8

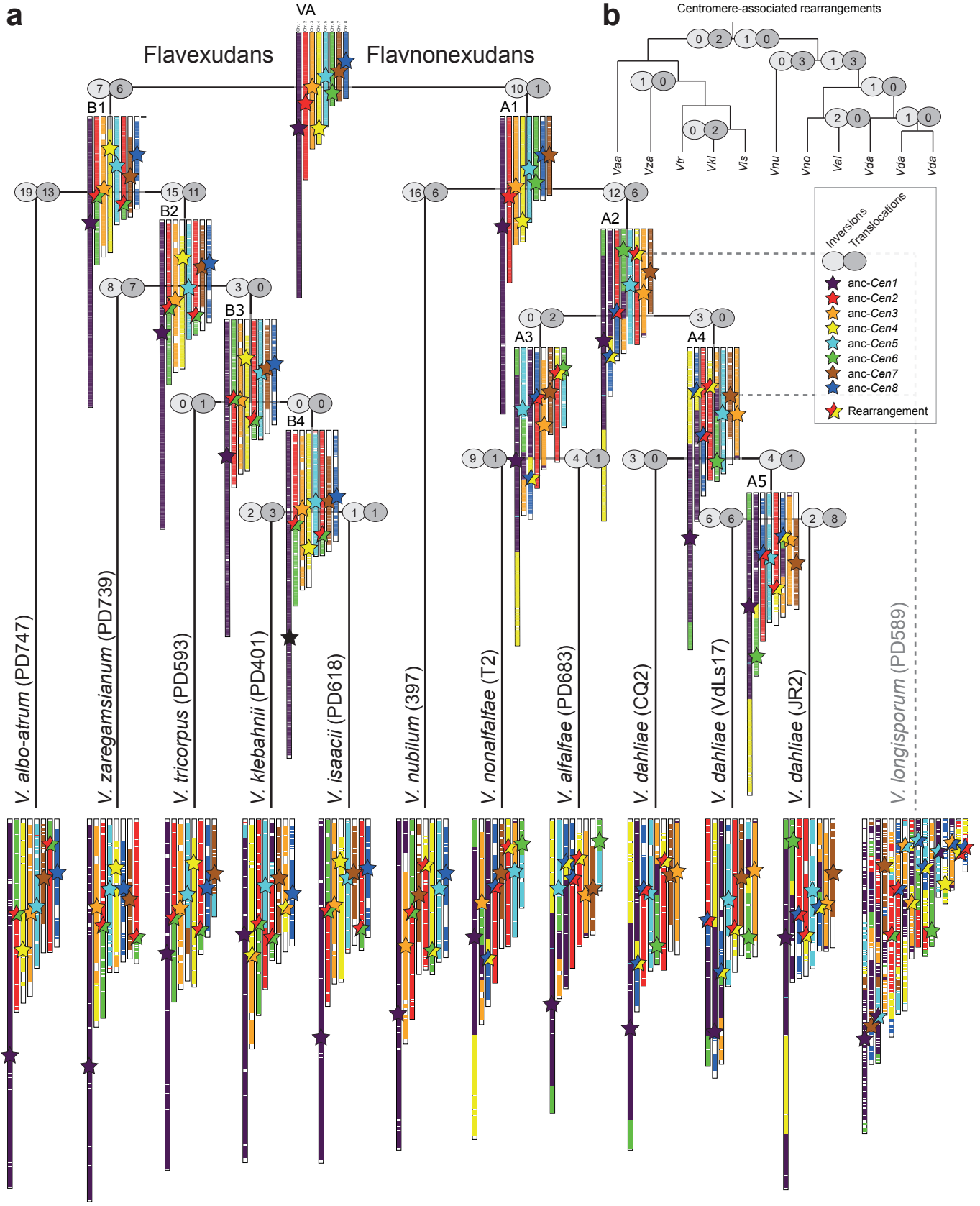

a
VA
Flavexudans
Flavnonexudans
B1
B2
B3
B4
A1
A2
A3
A4
A5
V. albo-atrum (PD747)
V. zaregamsianum (PD739)
V. tricorpus (PD593)
V. klebahnii (PD401)
V. isaacii (PD618)
V. nubilum (397)
V. nonalfalfae (T2)
V. alfalfae (PD683)
V. dahliae (CQ2)
V. dahliae (VdLs17)
V. dahliae (JR2)
V. longisporum (PD589)
b
Centromere-associated rearrangements
Vaa
Vza
Vtr
Vkl
Vis
Vnu
Vno
Val
Vda
Vda
Vda
Inversions
Translocations
anc-Cen1
anc-Cen2
anc-Cen3
anc-Cen4
anc-Cen5
anc-Cen6
anc-Cen7
anc-Cen8
Rearrangement

Table 1: Genome characteristics of the centromeres of *Verticillium dahliae* strain JR2.

		CenH3		AT-rich	AT-content (%)[3]		Repetitive elements	
Chr.	**Locus**	**CenH3 position (bp)[1]**	**CenH3 length (bp)**	**Position (kb)[2]**	**Chr.**	**Cen.**	**# Repeats (%)[4]**	**# VdLTRE9 (%)[4]**
1	*CEN1*	2,920,143-3,094,179	174,037	2,919-3,094	45.7	77.1	50 (99.8)	27 (70.4)
2	*CEN2*	520,698-672,281	151,584	516-672	46.3	77.8	43 (99.7)	26 (83.0)
3	*CEN3*	2,374,294-2,541,026	166,733	2,375-2,542	45.8	77.3	47 (99.8)	31 (80.5)
4	*CEN4*	2,884,316-3,071,412	187,097	2,885-3,072	46.2	75.4	54 (99.5)	24 (53.8)
5	*CEN5*	1,868,317-2,043,260	174,944	1,868-2,044	46.7	73.9	58 (99.5)	25 (63.1)
6	*CEN6*	2,166,972-2,333,060	166,089	2,167-2,334	46.4	75.2	48 (100)	31 (62.6)
7	*CEN7*	1,944,367-2,038,091	93,725	1,945-2,038	44.7	76.5	32 (95.8)	14 (47.8)
8	*CEN8*	1,406,398-1,561,664	155,267	1,406-1,562	47.7	77.0	37 (100)	26 (73.9)

[1]: position of CenH3-enriched domains; enriched domains within 10 kb have been merged
[2]: position of AT-rich domains; AT-rich domains with 20 kb have been merged
[3]: average AT-content of 1 kb windows of the entire chromosome and the AT-rich domain
[4]: percentage of centromeric region covered