AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE
Molecular Biology and Physiology

Repetitive Elements Contribute to the Diversity and Evolution of Centromeres in the Fungal Genus *Verticillium*

Michael F. Seidl,[a,b] H. Martin Kramer,[b] David E. Cook,[b,c] Gabriel L. Fiorin,[b] Grardy C. M. van den Berg,[b] Luigi Faino,[b,d] Bart P. H. J. Thomma[b,e]

[a]Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, the Netherlands
[b]Laboratory of Phytopathology, Wageningen University, Wageningen, the Netherlands
[c]Plant Pathology, Kansas State University, Manhattan, Kansas, USA
[d]Environmental Biology Department, Sapienza Università di Roma, Rome, Italy
[e]University of Cologne, Institute for Plant Sciences, Cluster of Excellence on Plant Sciences (CEPLAS), Cologne, Germany

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 10 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 *Verticillium dahliae*. Using Hi-C, we similarly identified repeat-rich centromeres in the other *Verticillium* species. Strikingly, a single degenerated long terminal repeat (LTR) retrotransposon is strongly associated with centromeric regions in some but not all *Verticillium* species. Extensive chromosomal rearrangements occurred during *Verticillium* evolution, of which some could be linked to centromeres, suggesting that centromeres contributed to chromosomal evolution. 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 10 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.

KEYWORDS centromere, chromosome evolution, heterochromatin, *Verticillium*

Centromeres are crucial for reliable chromosome segregation during mitosis and meiosis. During this process, centromeres direct the assembly of the kinetochore, a multiprotein complex that facilitates attachment of spindle microtubules to chroma-

Citation Seidl MF, Kramer HM, Cook DE, Fiorin GL, van den Berg GCM, Faino L, Thomma BPHJ. 2020. Repetitive elements contribute to the diversity and evolution of centromeres in the fungal genus *Verticillium*. mBio 11:e01714-20. https://doi.org/10.1128/mBio.01714-20.

Editor Joseph Heitman, Duke University

Address correspondence to Michael F. Seidl, m.f.seidl@uu.nl, or Bart P. H. J. Thomma, bart.thomma@wur.nl.

Received 26 June 2020
Accepted 5 August 2020
Published 8 September 2020

tids (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 (nt) 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- to 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 to 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 RNA interference (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 dinucleotides (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 occur 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, are 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 enable studies of 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). The genome of *V. dahliae* is characterized by lineage-specific (LS) regions (7, 8, 33–35) that are hypervariable between *V. dahliae* strains and that contain genes with crucial roles in virulence and host adaptation (7, 8, 33, 35). LS regions evolved by extensive chromosomal rearrangements such as translocations, inversions, duplications, or deletions, that are mediated by erroneous double-strand repair pathways, often involving repetitive elements (8). Repetitive elements within the LS regions display a distinct chromatin state compared with other repetitive regions (36). The *Verticillium* genus consists of 10 species that are all soilborne and presumed asexual but have different lifestyles (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), and regions with characteristics similar to LS regions have been identified in other *Verticillium* species as well (33). Centromeres have been thought to facilitate chromosomal rearrangements and contribute to karyotype evolution (11, 12, 41), and thus deeper knowledge of centromeres might help in understanding mechanisms that drive chromosomal rearrangements in *Verticillium* genome evolution. Facilitated by the availability of *V. dahliae* high-quality genome assemblies and of all other *Verticillium* species (32, 33, 40, 42), we here sought to identify and study the constitution and evolution of centromeres in the *Verticillium* genus and to elucidate their impact on chromosome evolution.

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 (see Fig. S1a in the supplemental material) 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 the locus behind the native *CenH3* promoter (Fig. S1b and 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 and e), while mapping of the sequencing reads derived from the wild-type (WT) strain did not reveal any CenH3-enriched region (Fig. S1d and e). The CenH3-enriched regions, designated *Cen1* to *Cen8*, 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. S1e). Thus, we conclude that CenH3 binding defines large regional centromeres in *V. dahliae* strain JR2.

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

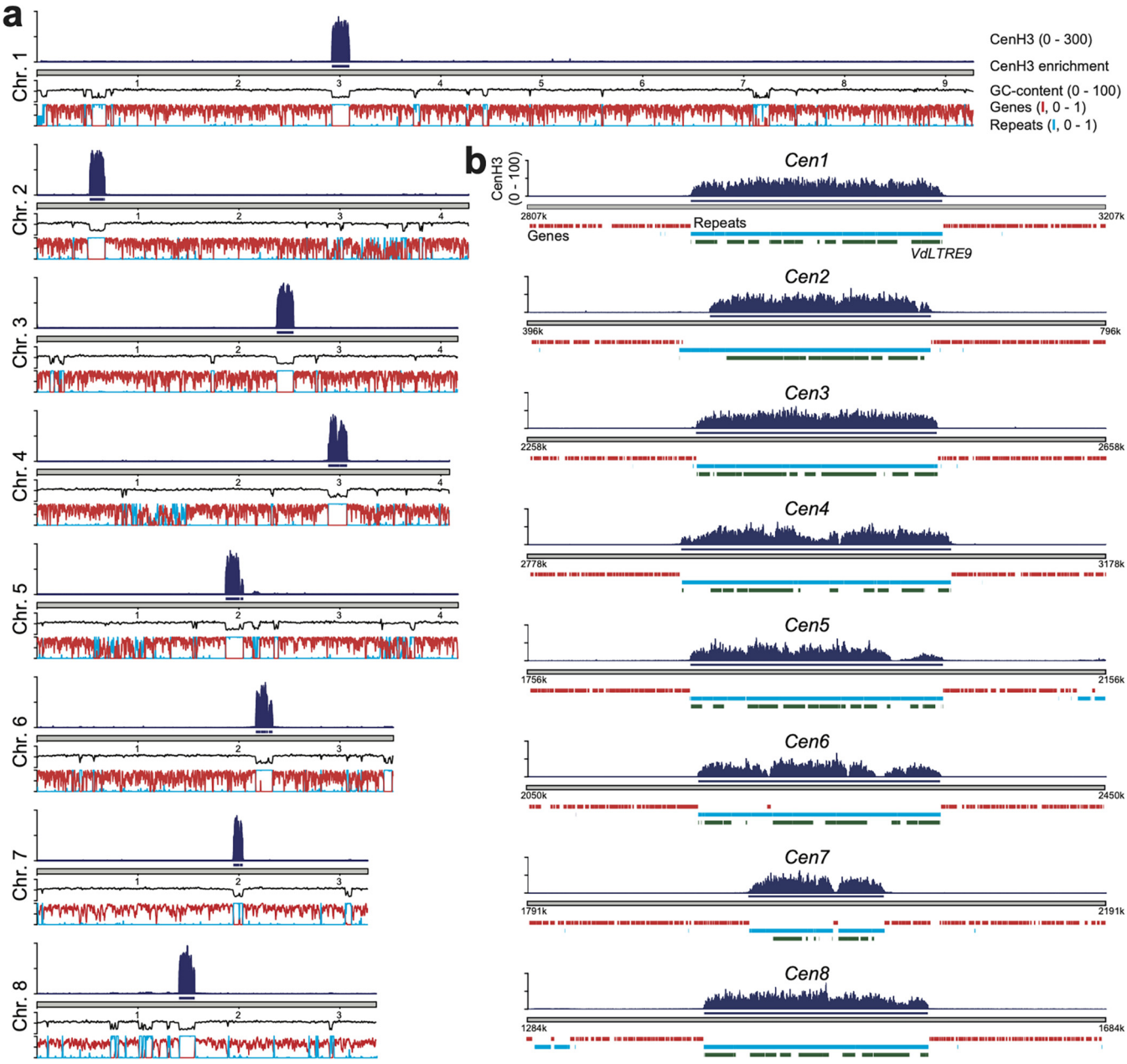

FIG 1 CenH3 binding defines centromeres in *Verticillium dahliae* strain JR2. (a) Schematic overview of the chromosomes of *V. dahliae* strain JR2 showing the normalized CenH3 ChIP-seq read coverage (RPGC normalization in 1-kb bins with 3-kb smoothening), CenH3 enriched regions, GC content, gene density (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* to *-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.

with these characteristics (Fig. 1a; Table 1). However, only one of these overlaps the centromeric regions defined by CenH3 binding (Fig. 1).

Elevated AT levels in repeat-rich regions are caused by RIP mutations in some filamentous fungi (15, 25, 26, 43). Due to its presumably asexual nature (7), the occurrence of RIP in *V. dahliae* is controversial (8, 44, 45), although signatures of RIP have previously been reported in a subset of repeat-rich regions (36). We assessed the occurrence of RIP signatures in centromeres using the composite RIP index (CRI) (46), which considers C-to-T mutations in the CpA context. Intriguingly, genomic regions located at centromeres display significantly higher CRI values than other genomic

TABLE 1 Genome characteristics of the centromeres of *Verticillium dahliae* strain JR2[e]

		CenH3			AT content (%)[c]		Repetitive elements	
Chr.	Locus	Position (bp)[a]	Length (bp)	AT-rich, position (kb)[b]	Chr.	Cen.	No. of repeats (%)[d]	No. of *VdLTRE9* (%)[d]
1	*CEN1*	2920143–3094179	174,037	2919–3094	45.7	77.1	50 (99.8)	27 (70.4)
2	*CEN2*	520698–672281	151,584	516–672	46.3	77.8	43 (99.7)	26 (83.0)
3	*CEN3*	2374294–2541026	166,733	2375–2542	45.8	77.3	47 (99.8)	31 (80.5)
4	*CEN4*	2884316–3071412	187,097	2885–3072	46.2	75.4	54 (99.5)	24 (53.8)
5	*CEN5*	1868317–2043260	174,944	1868–2044	46.7	73.9	58 (99.5)	25 (63.1)
6	*CEN6*	2166972–2333060	166,089	2167–2334	46.4	75.2	48 (100)	31 (62.6)
7	*CEN7*	1944367–2038091	93,725	1945–2038	44.7	76.5	32 (95.8)	14 (47.8)
8	*CEN8*	1406398–1561664	155,267	1406–1562	47.7	77.0	37 (100)	26 (73.9)

[a]Position of CenH3-enriched domains; enriched domains within 10 kb have been merged.
[b]Position of AT-rich domains; AT-rich domains within 20 kb have been merged.
[c]Average AT content of 1-kb windows of the entire chromosome and the AT-rich domain.
[d]Percentage of centromeric region covered.
[e]Abbreviations: Chr., chromosome; Cen., centromere.

regions (e.g., genes or repetitive elements) (Fig. 2a; Fig. S2 and S3a), and thus, RIP signatures at repetitive elements located at centromeres likely contribute to the high AT levels.

In most filamentous fungi and oomycetes, AT- and repeat-rich centromeres are embedded in heterochromatin that is characterized by methylated DNA and by particular histone modifications (H3K9me3 and H3K27me3) (13, 15, 16, 25, 30, 46). We recently determined chromatin states in the genome of *V. dahliae* strain JR2 and revealed that repetitive sequences outside the LS regions display characteristics of heterochromatin (36). To define centromeric chromatin states, we used previously generated bisulfite sequencing data to monitor DNA methylation (mC) and ChIP-seq data to determine the distribution of the heterochromatic marks H3K9me3 and H3K27me3 (36). To also determine the distribution of euchromatin, we performed ChIP-seq with an antibody against the euchromatic mark dimethylation of lysine 4 of histone H3 (H3K4me2). We observed overall low genome-wide DNA methylation levels (36) (Fig. 2a; Fig. S2), similar to the previously reported levels for *Aspergillus flavus* (47) and lower than for *N. crassa* (48). Nevertheless, repetitive elements and centromeres show significantly higher DNA methylation levels in all contexts compared with genes (Fig. 2b). Methylation (in CG context) at repetitive elements at centromeres is significantly higher than at repeats located along the chromosomal arm, but not at subtelomeric regions (Fig. 2c), and more methylation at centromeres correlates with increased CRI (Fig. 2a; Fig. S2 and S3a). DNA methylation colocalizes with H3K9me3 at repeat-rich regions (36) (Fig. 2a; Fig. S2). H3K9me3 occurs predominantly at repetitive elements localized at subtelomeres and centromeres (Fig. 2d and e; Fig. S2 and S3b). In comparison, H3K4me2 and H3K27me3 are largely absent from centromeres (Fig. 2d and e; Fig. S3b). Collectively, these observations indicate that centromeres of *V. dahliae* display typical characteristics of constitutive heterochromatin.

A single repeat associates with centromeres of *Verticillium dahliae* strain JR2. Centromere identity and function are 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, 49, 50). Unsurprisingly, CenH3-bound centromeres are repeat rich in *V. dahliae* (Fig. 1). A detailed analysis of the eight centromeres revealed a nearly complete (>96%) composition of repetitive elements belonging to only 10 different repeat subfamilies (Fig. 1b, Fig. 3a, and 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 nontranscribed (Fig. S3c to e), and thus are likely inactive. Interestingly, a single LTR retrotransposon subfamily, previously designated *VdLTRE9* (8, 32), covers on average ~70% of the DNA sequences at the eight centromeres, ranging from 48% in *Cen7* to 83% in *Cen2* (Fig. 3a; Table 1). We scanned the genome for the localization of the 10 repeat subfamilies (Fig. 3). Intriguingly, although

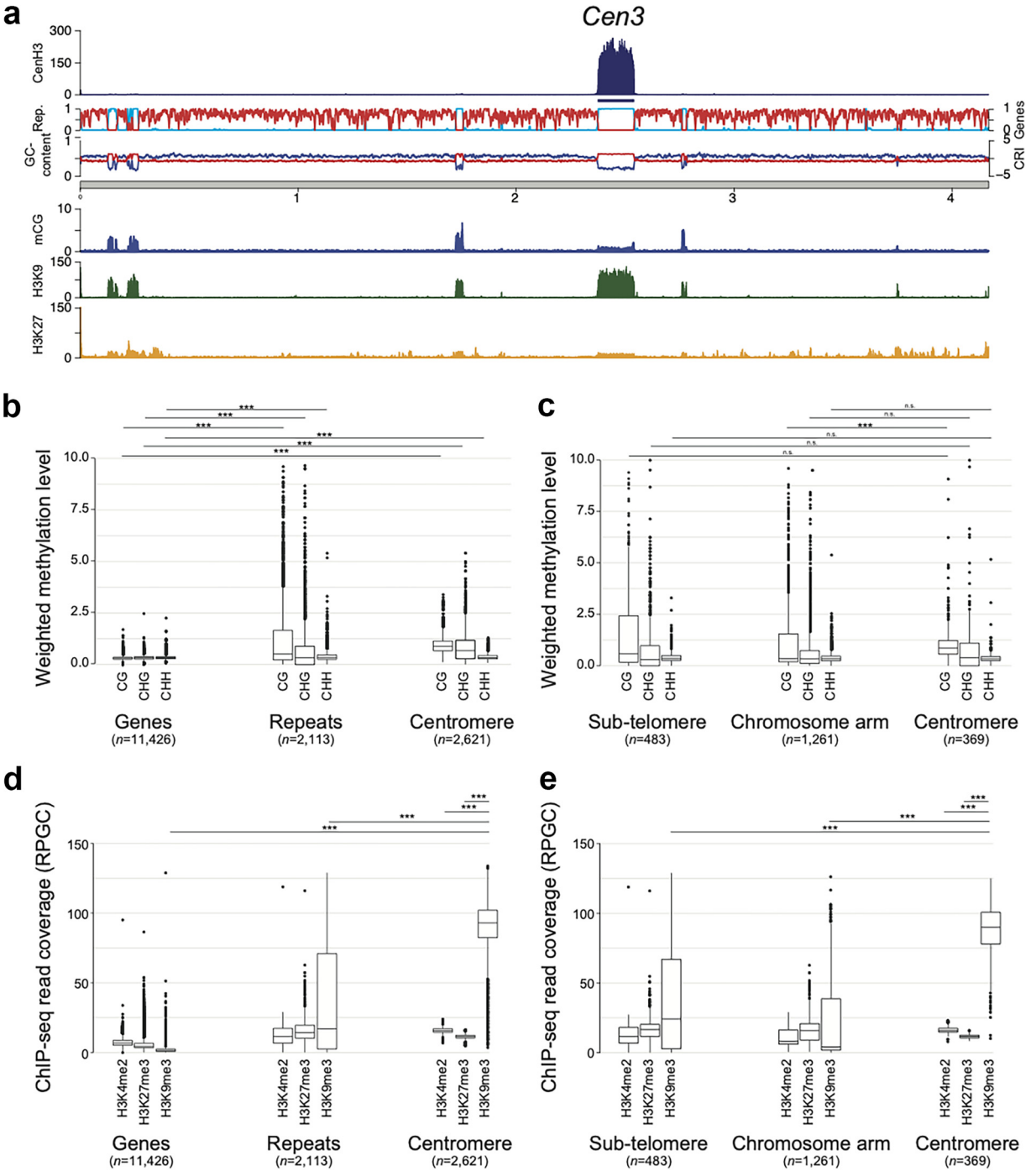

FIG 2 Centromeres in *Verticillium dahliae* strain JR2 are embedded in heterochromatin. (a) Schematic overview of chromosome 3 of *V. dahliae* strain JR2, exemplifying the distribution of heterochromatin-associated chromatin modifications (mC, H3K9me3, and H3K27me3) in relation to the centromeres. The different lanes display the FLAG-CenH3 ChIP-seq read coverage (RPGC normalization in 1-kb bins with 3-kb smoothening), the FLAG-CenH3 enriched regions, the repeat and gene density (light blue and red, respectively), the GC content (blue), the CRI (red) 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 Fig. S2. (b) Boxplots 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 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: subtelomeres (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 panel a for H3K4me2, H3K27m3, and H3K9me3) is summarized over genes, repetitive elements, or 5-kb windows (500-bp slide) overlapping the centromeric regions. (e) ChIP-seq read coverage (RPGC normalized; see panel a for H3K4me2, H3K27m3, and H3K9me3) is summarized over repetitive elements that have been split based on their genomic location: subtelomeres (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 nonparametric Mann-Whitney test; *P* values <0.001, ***; n.s., not significant.

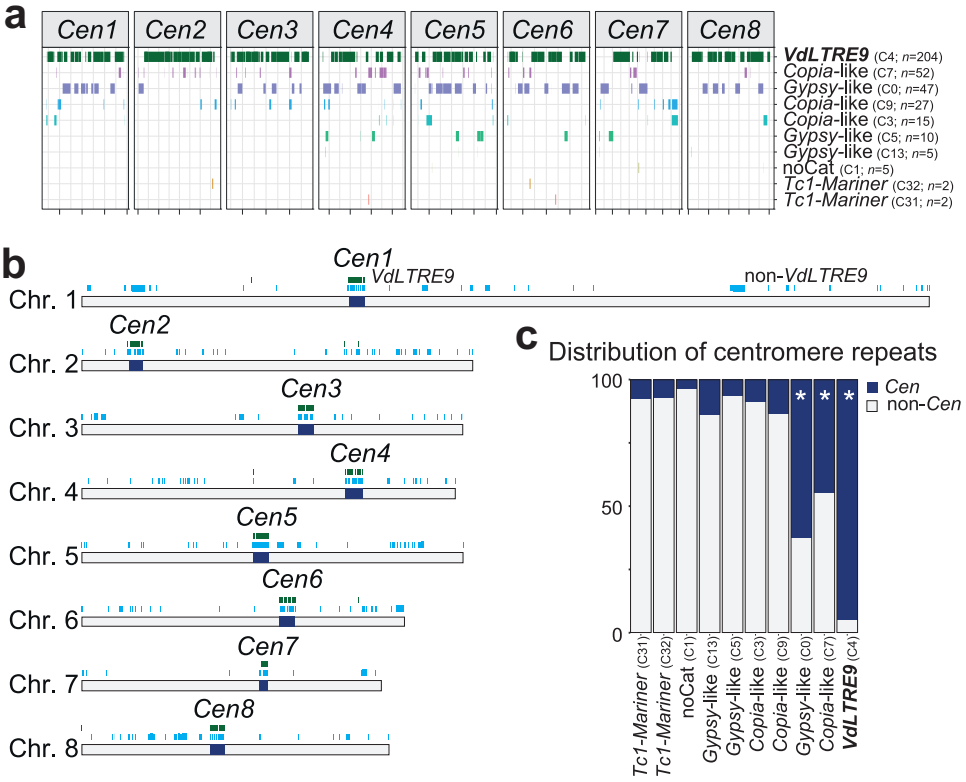

FIG 3 A single repeat family associates with centromeres in *Verticillium dahliae* strain JR2. (a) The presence of different repeat subfamilies is shown across the eight centromeres (*Cen1* to *-8*), and the number of occurrences for each subfamily within the centromeres is indicated. The individual centromeres in the diagram are shown in equal scale. (b) Genome-wide distribution of the 10 repeat subfamilies occurring within the eight centromeres (*Cen1* to *-8*; dark blue); the location of *VdLTRE9* is shown in green, and the location of elements belonging to the other nine repeat subfamilies (from panel a) is shown in light blue. (c) The distribution of different repeat subfamilies in centromeres (*Cen*; dark blue) and across the genome (non-*Cen*; light gray). The enrichment of specific subfamilies at centromeres was assessed using a one-sided Fisher exact test. Significant enrichment (multiple-testing corrected *P* value < 0.01) is denoted with an asterisk.

it is one of the most abundant repeats in the genome with 215 complete or partial matches, *VdLTRE9* is associated with 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 (Fig. 3b and c). The remaining 11 *VdLTRE9* copies (5%) occur outside the CenH3-rich centromeres, yet five out of 11 copies are localized within 50 kb of the centromeric regions (Fig. 3b and c). The nine other repeat subfamilies have additional matches that are located outside the centromeres (Fig. 1a and Fig. 3b and 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 subfamilies occur at the centromeres (Fig. 3c). Repeats at centromeres are often fragmented, and most copies, with the exception of the Tc1/mariner-like elements, are similarly fragmented when located outside the centromeres (Fig. S3f), indicating extensive degeneration of repetitive elements in *V. dahliae*. 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 transposable element classifications using PASTEC (51) indicate 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

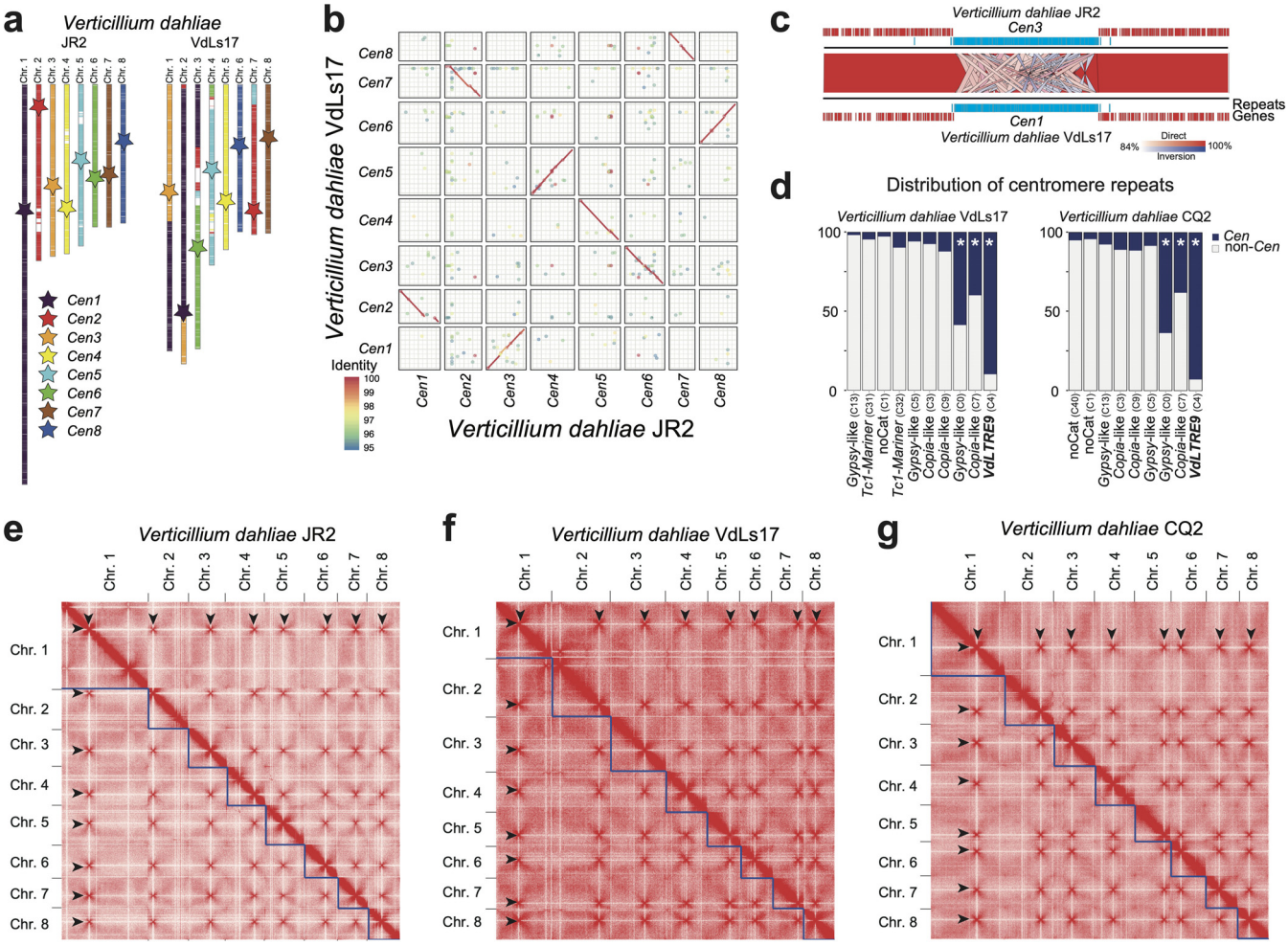

FIG 4 Hi-C contact maps identify *VdLTRE9* as a hallmark of centromeres in *Verticillium dahliae*. (a) Synteny analyses of the eight chromosomes of *V. dahliae* strains JR2 and VdLs17. Schematic overview of the eight chromosomes of *V. dahliae* strain JR2 (left) and the corresponding syntenic regions in *V. dahliae* strain VdLs17 (right). Approximate locations of centromeres are indicated by stars, and syntenic centromeres of *V. dahliae* strain VdLs17 are colored according to *Cen1* to *-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* strain 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 gray) for *V. dahliae* strains VdLs17 and CQ2. The enrichment of specific subfamilies at centromeres was assessed using a one-sided Fisher exact test. Significant enrichment (multiple-testing corrected *P* value < 0.01) is denoted with an asterisk. (e to 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 interchromosomal interaction frequencies are indicative of centromeres and are highlighted by arrowheads. Interaction frequencies are summarized in 50-kb bins along the genome.

caused by RIP (Fig. S3d), indicating that *VdLTRE9*, similar to other repeats in *V. dahliae*, has significantly degenerated.

***VdLTRE9* as hallmark of *Verticillium dahliae* centromeres.** To examine if *VdLTRE9* similarly occurs at centromeres in other *V. dahliae* strains, we made use of the complete genome assembly of *V. dahliae* strain VdLs17 (8, 32, 35). The evolution of *V. dahliae* is characterized by chromosomal rearrangements (8, 35) (Fig. 4a; Fig. S4a to c). Nevertheless, synteny analyses between *V. dahliae* strains JR2 and VdLs17 revealed large regions of colinearity between chromosomes and identified significant sequence and synteny conservation between the centromeres and their flanking regions (Fig. 4b and c; Fig. S4a), suggesting that centromeric sequences and their locations are conserved. We queried the genome of *V. dahliae* strain VdLs17 for the presence of *VdLTRE9* and identified a single region on each chromosome, collectively containing 186 of the 207 (90%) complete or partial matches of *VdLTRE9* in the genome (Fig. 4d) (one-sided

Fisher's exact test; multiple-testing corrected *P* value 3e−146). These *VdLTRE9*-rich regions are ~150 kb in size, AT rich, gene poor, and repeat rich and share similarity to the previously identified CenH3-bound and *VdLTRE9*-enriched regions of *V. dahliae* strain JR2 (Fig. 4b and c; Fig. S4d), suggesting that these regions similarly represent the centromeres of *V. dahliae* strain VdLs17.

Centromeres of *N. crassa* and some other fungi colocalize within the nucleus (15, 52–56). This colocalization can be experimentally determined using chromosome conformation capture (Hi-C), which can identify centromeres by their increased interchromosomal contacts (56). To confirm that Hi-C can be used to identify centromeres in *V. dahliae*, we first applied Hi-C to *V. dahliae* strain JR2. As anticipated, we observed seven strong interchromosomal contacts for each of the eight chromosomes (Fig. 4e). Importantly, the interacting regions overlap the CenH3-bound regions that we identified as centromeres (Table S1b), demonstrating that centromeres in *V. dahliae* strain JR2 colocalize within the nucleus and supporting that Hi-C reliably identifies centromeres (52, 53). We then applied Hi-C to *V. dahliae* strain VdLs17 and similarly identified regions with strong interchromosomal contacts, one for each of the chromosomes (Fig. 4f). These regions overlap 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 compared with *V. dahliae* strain JR2 or VdLs17, namely, strain CQ2, which displays ~1.05% 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 intrachromosomal contacts to assign the contigs into eight pseudochromosomes, leaving ~148-kb unplaced scaffolds (Fig. 4g, Fig. S4e, and Table S1c). We subsequently identified a single region with seven strong interchromosomal contacts for each pseudochromosome that is significantly enriched for *VdLTRE9* (one-sided Fisher's exact test; multiple-testing corrected *P* value 3.4e−166) (Fig. 4d and g, Fig. S4e, and Table S1b). Synteny analyses between *V. dahliae* strains JR2 and CQ2 revealed that the eight *VdLTRE9*-rich regions and their flanking chromosomal regions are colinear, suggesting that centromere locations are conserved between different *V. dahliae* strains (Fig. 4; Fig. S4a to c and f). With an average size of 165 kb, the centromeres of *V. dahliae* strain CQ2 are similar in size to 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 cooccurrence 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.

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 10 *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 intrachromosomal interaction signals, we assigned the vast majority of the previously assembled contigs into eight pseudochromosomes for each of the haploid *Verticillium* species and 16 pseudochromosomes for the diploid *V. longisporum*, leaving between 0.5 kb and 2,022 kb unassigned (Fig. S5; Table S1c). For most genome assemblies, the pseudochromosomes 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 interchromosomal Hi-C interaction signals, we identified a single region with high interchromosomal contacts for each of the pseudo-

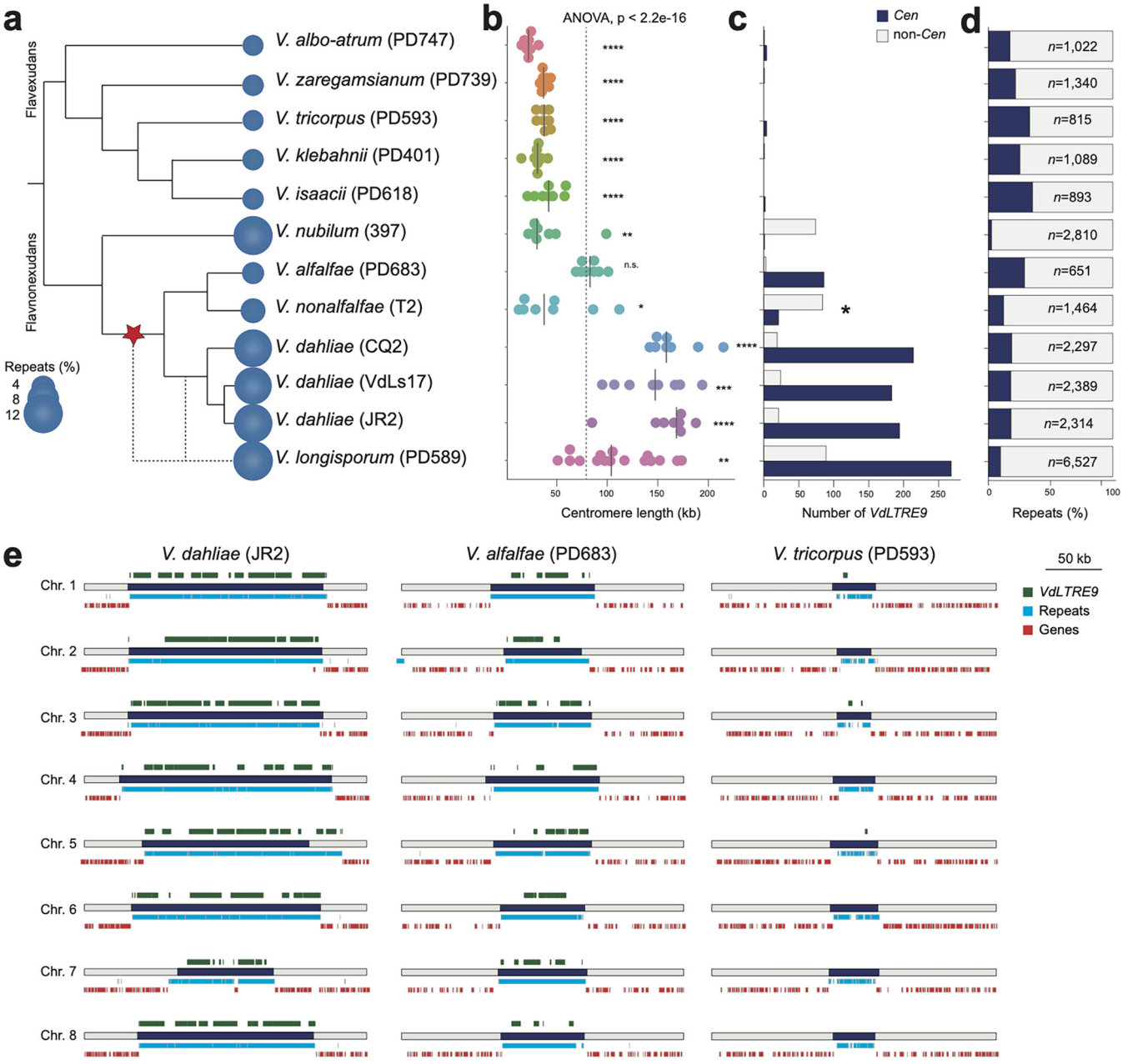

FIG 5 Evolution of centromeres in the genus *Verticillium*. (a) Relationship of the 10 members of the genus *Verticillium*. The predicted repeat content for each of the genomes is indicated (see Table S1c for details). The red star indicates the recruitment of *VdLTRE9* into centromeres. (b) Comparison 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 gray). The asterisk indicates the high number of *VdLTRE9* elements in unassigned contigs for *Verticillium nonalfalfae* strain T2 (see the text for details). (d) Proportion of predicted repeat content localized at centromeres (*Cen*; dark blue) and across the genome (non-*Cen*; light gray). (e) Schematic overview of the eight centromeric regions (250 kb) in *Verticillium dahliae* strain JR2, and *Verticillium alfalfae* strain PD683 and *Verticillium tricorpus* strain 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 locations of (partial) *VdLTRE9* matches (light green) are shown above each centromere. Global statistical differences for the centromere sizes were calculated using one-way analysis of variance (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, *; n.s., not significant.

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; Fig. S6a and b). Centromeres within the Flavexudans clade are similarly sized and significantly smaller than the genus-wide average. In 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 and e; Fig. S6c and d). Similar to *V. dahliae*, the vast majority of matches are fragmented, suggesting that *VdLTRE9* has been significantly degenerated in these species as well. Only very few partial or no matches of *VdLTRE9* consensus could be identified in the genomes of the Flavexudans species (Fig. 5c and e; Fig. S6 and S7; Table S1e). Collectively, these findings suggest that *VdLTRE9* is specific to Flavnonexudans species, yet we cannot exclude the alternative scenario in which *VdLTRE9* was present at the last common ancestor of *Verticillium* and has been lost in all Flavexudans species. Regardless of the origin, *VdLTRE9* has likely been recruited to the centromeres of Flavnonexudans species only after the divergence of *V. nubilum* (Fig. 5a; Fig. S6 and S7).

Since *VdLTRE9* occurs only in a 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. 5d and e; Fig. S6a and 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.

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 pseudochromosomes of the haploid *Verticillium* species to reconstruct ancestral chromosomal configurations using AnChro (Fig. 6a) (57). We reconstructed all potential ancestors that predominantly had eight chromosomes and ~8,000 genes (Fig. S9a and b), yet the number of ancestral chromosomes and genes varied when approaching the last common ancestor (Fig. S9a and 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 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, similarly 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, including other complex rearrangements) (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 and 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-associated translocations (between the ancestral chromosomes 2 and 6) led to the formation of two rearranged chromosomes. In total, we observed that five out of the eight ancestral centromeres were associated with a chromosomal rearrangement at one point during

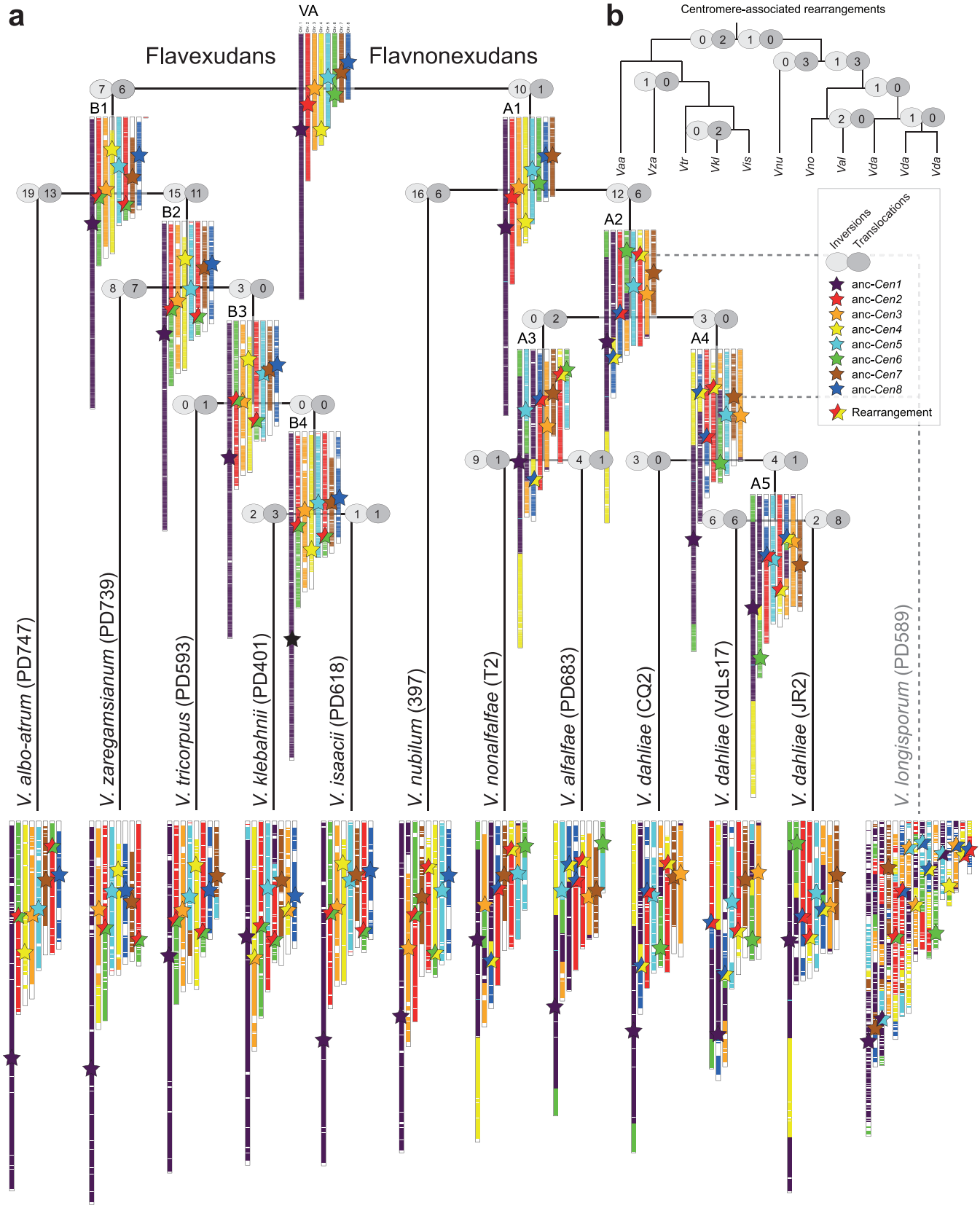

FIG 6 Centromeres contribute to karyotype evolution in *Verticillium*. (a) Relationship of the 10 members of the genus *Verticillium*. The allodiploidization event forming *V. longisporum* is indicated by dashed lines (38, 114). The chromosomal evolution within the haploid members of the genus was reconstructed using AnChro (57). 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; see text) is displayed for each

(Continued on next page)

evolution (Fig. 6a). Nevertheless, comparisons of protein-coding genes that flank centromeres show that these are syntenic in most extant species. Similarly, none of the recent chromosomal rearrangements observed between *V. dahliae* strains is clearly associated with centromeres (Fig. 4a and b and Fig. 6a), even though *CEN2* of *V. dahliae* strain VdLs17 is located near (20 to 25 genes up-/downstream) a chromosomal rearrangement (Fig. 4a). Thus, while chromosomal rearrangements involving centromeres occurred during evolution, they do not account for the majority of the karyotype variation between extant *Verticillium* species.

DISCUSSION

Centromeric regions are among the most rapidly evolving genomic regions (13–16, 29), yet centromere evolution has been systematically studied in only a 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 colocalize within the nucleus (15, 52–56, 58), a phenomenon that can be exploited for their identification (52, 53). 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 a 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 artifacts, and centromeres in most *Verticillium* species are larger than in *Z. tritici* (27), *C. neoformans*, *Magnaporthe 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 compared with other *Verticillium* species (32, 39, 40, 42, 60). Thus, increased centromere sizes positively correlate with overall increased repeat contents.

Using fragmented genome assemblies, we previously identified chromosomal rearrangements during *Verticillium* evolution, which contributed to the formation of hypervariable LS regions containing genes with important roles in pathogen virulence (8, 35, 40). Thus, we proposed that chromosomal rearrangements in *Verticillium* contributed to genetic diversity and adaptation in the absence of sexual recombination (7, 35, 40). Chromosome-level genome assemblies for an entire genus enabled unprecedented analyses of the karyotype evolution over longer evolutionary timescales. Here, we observed extensive chromosomal rearrangements and provide evidence that some rearrangements at centromeres contributed to karyotype evolution, most of which occurred early during the divergence of *Verticillium*. Chromosomal rearrangements at centromeres occur in the yeasts *Candida*, *Cryptococcus*, and *Malassezia* (11, 12, 41), and synteny breakpoints have been identified between mammals and chicken (61), suggesting that centromeres often contribute to karyotype evolution. The emergence of chromosomal rearrangements at centromeres could be facilitated by their repeat-rich nature (11, 12). For example, centromeres in *Malassezia* are enriched with an AT-rich

FIG 6 Legend (Continued)
branch, and centromeres that colocalize 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 to, centromeres is shown along the branches depicting the *Verticillium* species phylogeny shown in panel a.

motif that could facilitate replication fork stalling, which leads to double-strand DNA breaks (11). Repeats localized outside centromeres in *V. dahliae* contribute to chromosomal rearrangements (8), and thus, it seems plausible that centromeric repeats similarly contribute to chromosomal rearrangements. It is tempting to speculate that the additional larger AT- and repeat-rich regions outside the centromeres (e.g., on chromosome 1, 7, or 8 of *V. dahliae* strain JR2) might have been involved in chromosomal rearrangements. However, based on our ancestral chromosome reconstruction, these regions, and even the entire chromosome (e.g., chromosome 8), are conserved and do not colocalize with any of the predicted large-scale translocations, even though smaller rearrangements might have occurred that have remained undetected. Chromosomal rearrangements often do not lead to changes only in chromosome organization but also in chromosome number (11, 12). While we observed chromosomal rearrangements, all extant and ancestral genomes contained eight chromosomes, suggesting that eight chromosomes are a stable configuration for all *Verticillium* species.

Centromere position and function are thought to be driven by the protein complement (e.g., CenH3 localization) and by heterochromatin formation rather than by specific DNA sequences (13, 15, 62). In *V. dahliae*, we observed the cooccurrence of CenH3 with H3K9me3 and DNA methylation. This suggests that DNA methylation, as previously reported in *N. crassa* and in *C. neoformans* (16, 25), is also a feature of centromeric DNA in *V. dahliae*. Colocalization of CenH3 with H3K9me2/3 and DNA methylation has been reported for *N. crassa* (25) and *C. neoformans* (16). In contrast, H3K9me3 and H3K27me3 are absent from centromeres in *Z. tritici* (27). H3K4me2 borders most centromeres in *Z. tritici* (27) and is associated with centromeres in *Schizosaccharomyces pombe* and some animals and plants (63–66). H3K4me2 has not been observed at centromeres in most fungi, including *V. dahliae*, and in the oomycete plant pathogen *Phytophthora sojae* (30). Changes in heterochromatin in *N. crassa* lead to altered CenH3 positioning (25), suggesting that heterochromatin is similarly required for centromere maintenance and function in *V. dahliae*. Elevated AT levels in repeat-rich heterochromatic regions can be caused by RIP mutations (15, 25, 26, 43). RIP-like mutations have been previously reported in some repeats in *V. dahliae* (36, 45), and we observed strong RIP signals at centromeres. Due to its presumably asexual nature (7), the occurrence of RIP in *V. dahliae* is controversial (8, 44, 45). Noteworthy, mutational signatures resembling RIP have recently been observed in *Z. tritici* propagated through mitotic cell divisions, pointing to the existence of a mitotic version of a RIP-like process (43). Thus, we conclude that RIP was an active process in *V. dahliae* at some point in evolution, or that RIP-like processes outside the sexual cycle occur in *V. dahliae*. Furthermore, a mechanistic link between AT-rich RIP mutated DNA, H3K9me3 deposition, and DNA methylation has been established in *N. crassa* (67), suggesting that these processes are also connected in *V. dahliae*.

Centromeres are often enriched for a variety of different retrotransposons and other repetitive elements (15, 16, 25, 29, 30, 68–70). We similarly observed that centromeres in all *Verticillium* species are repeat rich. Repeats and their remnants identified at centromeres typically also occur outside centromeres, as observed in *M. oryzae* (29) and *N. crassa* (25), for instance. Strikingly, we observed that a single degenerated LTR retrotransposon, *VdLTRE9*, is strongly associated with centromeres in some *Verticillium* species, while it is absent from LS regions in *V. dahliae*. The association of specific retrotransposons with centromeres has also been observed in the yeasts *Ogataea polymorpha* (69), *Debaryomyces hansenii* (68), and *Scheffersomyces stipitis* (70), where a retrotransposon related to *Ty5* is enriched at centromeres. Similarly, centromeres in *Cryptococcus* contain six retrotransposons (*Tcn1* to *-6*) that occur nearly exclusively at centromeres (16). Centromeres of *P. sojae* contain multiple types of repeats, but they are enriched for a single element called CoLT (*Copia*-like transposon) (30). The strong associations of specific repeats with centromeres could directly or indirectly link these elements to centromere function. Functional centromeres as observed here are also heterochromatic and contain CenH3. AT-rich repetitive elements can direct heterochro-

matin formation via DNA methylation and H3K9me3 deposition in *N. crassa* (46, 67), a phenomenon that can also occur at repeats outside centromeres (46). Heterochromatin occurs at centromeres but also at repeat-rich regions outside centromeres in *V. dahliae*; thus, the repeat-rich nature of centromeres is likely not sufficient to direct CenH3 deposition. In *S. pombe* heterochromatin formation is directed by short interfering RNAs (siRNAs) derived from flanking repetitive elements via RNAi (71, 72), and RNAi and heterochromatin mediate CenH3 localization at centromeres (73, 74). RNAi is also important for centromere maintenance and evolution in *Cryptococcus*, as RNAi-deficient species have smaller centromeres than RNAi-proficient ones (16). Interestingly, centromere-specific elements (*Tcn1* to *-6*) in RNAi-proficient species are typically full-length elements while only remnants can be found in RNAi-deficient species, which could be caused by recombination between elements (16). Furthermore, the genome size of RNAi-deficient species is smaller than that of RNAi-proficient ones, and centromere size reduction is at least partially responsible for genome size differences (16). In *Verticillium*, centromere size differences correlate with an increase in repeat content and the recruitment of *VdLTRE9*, which is highly fragmented and likely nonactive. Genome size differences exist in haploid *Verticillium* (33 Mb to 36 Mb; see Table S1c in the supplemental material), yet these do not seem to correlate with centromere sizes. Even though key components of the RNAi machinery exist in all *Verticillium* species (75) (Table S1f), we know only little about their biological functions. Similarly to *C. neoformans*, we observed no transcriptional activity of *VdLTRE9* or any other repeat at centromeres, but it is unclear if this silencing is mediated by RNAi, is a consequence of their heterochromatic nature, is due to their fragmentation, or is a combination of these. Ultimately, unraveling how specific elements contribute to centromere identity necessitates future experiments. *VdLTRE9* occurs only in some *Verticillium* species and has likely been recruited to centromeres subsequent to the divergence of *V. nubilum*. Conversely, these observations raise further questions on the roles of repeats and mechanisms of centromeric identity in species without *VdLTRE9*. Repeats drive the formation of chromosomal rearrangements, which are crucial for the formation and maintenance of LS regions, and thus are important drivers of *Verticillium* genome evolution and function (8, 36). Here, we highlight their contributions to centromere diversity within the fungal genus *Verticillium* and demonstrate that also centromeres contributed to chromosomal evolution. Our analyses provide the framework for future research into the diversity or convergence of mechanisms establishing centromere identity and functioning, and to elucidate roles of centromeres in generating genomic diversity in fungi.

MATERIALS AND 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) (76, 77) 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) (76, 77) and exonerate (v2.2.0; default settings) (78) searches against the genome sequences. Protein sequences of selected CenH3 and H3 proteins were aligned using mafft (v7.271; default settings, LINSi) (79), and poorly aligned regions in the alignment were removed using trimAl (v1.2; default settings) (80). A phylogenetic tree was inferred with maximum-likelihood methods implemented in IQ-tree (v1.6.11) (81), 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 (82) or PRF-GU2 for homologous recombination. The CenH3 locus, from *V. dahliae* strain JR2, was amplified as 3 fragments with overlapping sequences (see Table S1g in the supplemental material). The 5′-most fragment containing the promoter was amplified using primers A+B, the open reading frame (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 PspOMI and SphI linearized vector using Gibson assembly. The vector construction was confirmed by Sanger sequencing. Vectors were transformed to *Verticillium* with *Agrobacterium*-mediated transformation (83). Correct homologous recombination and replacement at the *CenH3* locus were verified by PCR amplification using primers I+J (Fig. S1b and Table S1g). 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 layer of Miracloth and subsequently snap-frozen in liquid nitrogen and ground with a mortar and pestle using liquid nitrogen. Approximately 0.3 g of ground mycelium was resuspended in 600 μl protein extraction buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% glycerol, 0.02% NP-40, 2 mM phenylmethanesulfonyl fluoride [PMSF], 100 μM leupeptin, 1 μg/ml pepstatin), briefly vortexed, incubated on ice for 15 min, and centrifuged at 4°C at 8,000 $\times$ *g* for 3 min. The supernatant was collected by transferring 20 μl to a new tube to serve as the input control, and the remaining ~500 μl was transferred to a fresh microcentrifuge tube with 15 μl of anti-FLAG M2 affinity gel (catalog number A2220; Sigma-Aldrich, St. Louis, MO, USA) and incubated while rotating at 4°C for 1 h. Samples were centrifuged at 5,000 $\times$ *g*, 4°C, for 3 min, after which the supernatant was discarded. Samples were washed with 500 μl of lysis buffer, and the centrifugation and washing were repeated three times. Protein was eluted from the resin by adding 15 μl of lysis buffer and 20 μl of 2$\times$ Laemmli loading buffer (4% SDS, 20% glycerol, 0.004% bromophenol blue, 125 mM Tris HCl, pH 6.8) and boiled at 95°C for 3 min. Protein samples were separated on a 12% polyacrylamide gel and subsequently transferred to polyvinylidene difluoride (PVDF) membranes, blocked in 5% bovine serum albumin (BSA), washed twice in Tris-buffered saline–Tween (TBST), and incubated with 1:3,500 anti-FLAG antibody (monoclonal anti-FLAG M2; Merck KGaA, Darmstadt, Germany).

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). For each *V. dahliae* genotype, one million spores were added to 100 ml potato dextrose broth and incubated for 7 days at 22°C with continuous shaking at 120 rpm. Mycelium was collected by straining over a double layer of Miracloth and subsequently snap-frozen in liquid nitrogen and ground with a mortar and pestle using liquid nitrogen. All ground material (0.5 to 1 g per sample) was resuspended in 4 ml ChIP lysis buffer (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate [NaDOC]) and Dounce homogenized 40 times in a 10-cm^3 glass tube with a tightly fitting pestle on 800 power with an RZR50 homogenizer (Heidolph, Schwabach, Germany), followed by five rounds of 20-s sonication on ice with 40 s of resting in between rounds with a Soniprep 150 (MSE, London, United Kingdom). Samples were redistributed to 2-ml tubes and pelleted for 2 min at maximum speed in a tabletop centrifuge. Supernatants were pooled per sample in a 15-ml tube together with 25 μl anti-FLAG M2 magnetic beads (Sigma-Aldrich, St. Louis, MO, USA) and incubated overnight at 4°C and with continuous rotation. Beads were captured on a magnetic stand and washed with wash buffer (50 mM Tris HCl, pH 8, 1 mM EDTA, 1% Triton X-100, 100 mM NaCl), high-salt wash buffer (50 mM Tris HCl, pH 8, 1 mM EDTA, 1% Triton X-100, 350 mM NaCl), LiCl wash buffer (10 mM Tris HCl, pH 8, 1 mM EDTA, 0.5% Triton X-100, 250 mM LiCl), and TE buffer (10 mM Tris HCl, pH 8, 1 mM EDTA). Chromatin was eluted twice from beads by addition of 100 μl preheated TES buffer (100 mM Tris HCl, pH 8, 1% SDS, 10 mM EDTA, 50 mM NaCl) and 10 min incubation at 65°C. Proteinase K (10 mg/ml, 2 μl) was added and incubated at 65°C for 5 h, followed by chloroform extraction. DNA was precipitated by addition of 2 volumes 100% ethanol, 1/10 volume 3 M NaOAc, pH 5.2, and 1/200 volume 20 mg/ml glycogen, and overnight incubation at −20°C.

Sequencing libraries were prepared using the TruSeq ChIP library preparation kit (Illumina, San Diego, CA) according to the manufacturer's instructions, but without gel purification and with use of the Velocity DNA polymerase (BioLine, Luckenwalde, Germany) for 12 cycles of amplification for the FLAG-CenH3. H3K4me2 ChIP was performed as described previously (36), using an anti-H3K4me2 antibody (catalog no. 39913; ActiveMotif, Carlsbad, CA, USA). Single-end (125-bp) sequencing was performed on the Illumina HiSeq2500 platform at KeyGene N.V. (Wageningen, the Netherlands).

Chromatin confirmation capturing followed by high-throughput sequencing (Hi-C). We determined the inter- and intrachromosomal contact frequencies using Hi-C in *V. dahliae* strains CQ2, JR2, and VdLs17, as well as in *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. For each strain, one million spores were added to 400 ml Potato dextrose broth and incubated for 6 days at 22°C with continuous shaking at 120 rpm. Mycelium was collected by straining over double-layer Miracloth, and 300 mg (fresh weight) was used as input for generating Hi-C sequencing libraries with the Proximo Hi-C kit (Microbe) (Phase Genomics, Seattle, WA, USA), according to manufacturer's instructions. Briefly, samples were first cross-linked for 15 min at room temperature. Cross-linked mycelium was treated with fungal cell lysis solution (10 mM beta-mercaptoethanol, 15 mg/ml Glucanex, dissolved in phosphate-buffered saline at pH 7.4) for 1 h at 30°C, followed by snap-freezing in liquid nitrogen and grinding with a plastic pestle to obtain a powder. The resulting material was further lysed using the lysis buffers provided with the Hi-C kit, and chromatin was collected by centrifugation. Next, chromatin was fragmented at 37°C for 1 h and proximity ligation was performed at room temperature for 4 h. Reverse cross-linking was performed overnight at 65°C. The resulting soluble DNA was purified and bound to streptavidin beads. Library preparation was then performed, followed by on-bead library amplification by PCR (72°C for 5 min; 98°C for 30 s; 15 cycles of 98°C for 10 s, 62°C for 20 s, and 72°C for 50 s). Libraries were cleaned up and eluted from the beads. Final yields were determined by quantification using a Qubit 2.0 fluorometer (Invitrogen). Hi-C sequencing libraries of *V. dahliae* strains CQ2, JR2, and VdLs17 were paired-end (2 $\times$ 125 bp) sequenced on the Illumina HiSeq2500 platform at KeyGene N.V. (Wageningen, the Netherlands). Hi-C sequencing libraries of the other *Verticillium* species were paired-end (2 $\times$ 150 bp) sequenced on the NextSeq500 platform at USEQ (Utrecht, the Netherlands).

***In vitro* transcriptome profiling using RNA-seq.** RNA sequencing (RNA-seq) of *V. albo-atrum* strain PD747, *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 was performed

as described previously (36). Single-end (50-bp) sequencing was performed on the BGISeq500 platform at BGI (BGI Hong Kong).

Analyses of high-throughput sequencing data. High-throughput sequencing libraries (Table S1a) have been analyzed as follows. Illumina reads were quality-filtered and trimmed using Trimmomatic (version 0.36) (84). Sequencing reads were trimmed and filtered by removing Illumina TruSeq sequencing adapters (settings seed mismatches 2, palindrome clip threshold 30, and simple clip threshold 10), removal of low-quality leading or trailing bases below quality 5 and 10, respectively, and 4-base sliding window trimming and cutting when average quality per base dropped below 15. Additionally, filtered and trimmed reads of $<$90 nt were removed from further analyses. Filtered and trimmed reads were mapped to the corresponding genome assembly with Bowtie2 (default settings) (85), and mapping files were converted to bam-format using SAMtools (v 1.8) (115). Genomic coverage was determined using deepTools (v3.4.1; bamCoverage) (87) 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) (87) to generate principal-component analyses. Furthermore, we employed deepTools (v3.4.1, multiBigwigSummary) (87) to summarize genomic coverages of values 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) (88).

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 was extracted, and weighted CG-methylation levels were calculated over genes, repetitive elements, and genomic windows (5-kb window size with 500-bp slide) (89); 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 (90). The contact matrices generated by Juicer were used by the three-dimensional (3D) *de novo* assembly (3D-DNA) pipeline (91) (v180922) with a contig size threshold of 1,000 bp to eliminate misjoints in the previous assemblies and to generate improved assemblies. The genome assemblies were manually improved using Juicebox Assembly Tools (JBAT) (v1.11.08) (92), and improved genome assemblies were generated using the 3D-DNA postreview asm pipeline (91). Centromere locations were determined using a 1-kb-resolution contact matrix in JBAT, by identifying a region per chromosome that displays strong interchromosomal interactions, yet weak intrachromosomal interactions (see Fig. S5).

To assess potential repeat collapses during genome assemblies at centromeric regions, we mapped previously generated short-read data for *V. dahliae* strains JR2 and VdLs17, *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. tricorpus* strain PD593, and *V. zaregamsianum* strain PD739 (34, 39, 40, 93) to the genome assemblies using BWA (v0.7.17; mem) (86). We first used bedtools (v2.29.2) (94) to identify genomic regions with $>$500$\times$ coverage. We then applied deepTools (v3.4.1, computeGCBias) (87) to compute GC biases of read depth across the genome, excluding the identified high-coverage regions, and used deepTools (v3.4.1, correctGCBias) (87) to correct GC biases, which addresses known biases in sequencing library preparation to ensure even read coverage throughout the genome irrespective of their base composition (95). We used deepTools (v3.4.1, bamCoverage, bins 50 bp, counts per million [CPM] normalization) (87) to obtain the read coverage throughout the genome, excluding regions containing sequence assembly gaps (N's). Assuming that collapsed repeats would lead to a local increase in read depth, we used the ratio of the average read coverage at the centromeres and outside the centromere at each chromosome to correct the inferred centromere sizes. To further validate the genome assembly of regions identified as centromeres of *V. dahliae* strain JR2, the genome assembly was compared to the previously generated optical map (35) using MapSolver (v 3.2; OpGen, Gaithersburg, MD).

The transcriptional activity for genes and repetitive elements in *V. dahliae* strain JR2 was assessed *in vitro* (in potato dextrose broth) using previously generated deep transcriptome data sets (36). To this end, single-end sequencing reads of three biological replicates were mapped to the *V. dahliae* strain JR2 genome assembly (32) using STAR (v2.4.2a; maximum intron size 1 kb and outFilterMismatchNmax to 5) (96). The resulting mapped reads were summarized per genomic feature (gene or repeat) using summarizeOverlaps (97), converted to counts per million (CPM) mapped reads, and averaged over the three biological replicates.

Sequence analyses of *Verticillium* genome assemblies, centromeres, and repeat and gene content. Repetitive elements in the genomes of *V. dahliae* strains JR2, VdLs17, and CQ2 (32, 33) were identified as previously described (36). Briefly, repetitive elements were identified in each genome independently using a combination of LTRharvest (98) and LTRdigest (99) followed by identification of RepeatModeler. Identified repeats in the different *V. dahliae* strains were clustered into a nonredundant library that contained consensus sequences for each repeat family. The repeat library was, if possible, manually curated and annotated using PASTEC (100) or by sequence similarity to previously identified and characterized repeat families (32, 45). Genome-wide occurrences of repeat families were determined using RepeatMasker (v 4.0.9; sensitive option and cutoff 250), and the output was postprocessed using 'One code to find them all' (101). We considered only matches to the repeat consensus library and thereby excluded simple repeats and low-complexity regions.

De novo gene and repeat annotation for the Hi-C-improved *Verticillium* genome assemblies, and for *V. dahliae* strains JR2 and VdLs17 as a comparison, was performed using the funannotate pipeline (102). Briefly, repetitive elements were first *de novo* identified using RepeatModeler and masked for gene prediction using RepeatMasker. Subsequently, gene prediction parameters were estimated using *in vitro* RNA-seq data (see above for details; exceptions: *V. alfalfae*, for which no RNA-seq data were available; *V. nonalfalfae*, for which publicly available RNA-seq data were used [93]; 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* data [xylem sap and half-strength Murashige and Skoog {36}] as well as long-read nanopore cDNA data [103] were 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) (104), 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 them all' (101). *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) (105). 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) (76, 77), 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 [103]), GC content, and composite RIP index (CRI) were calculated along the genome sequence using sliding windows (5-kb window with 500-bp slide). The CRI was calculated according to the method of Lewis et al. (46). 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 regions) were assessed using bedtools (v2.29.2) (94). Genome-wide data were visualized using R (106) with the package ggplot2 (107), karyplotR (108), or Gviz (109), as well as EasyFig (110).

Whole-genome alignments between *V. dahliae* strains JR2, VdLs17, and CQ2 were performed using NUCmer, which is part of the MUMmer package (v 3.1; –maxmatch) (111). To remove short matches, we considered only alignments longer than 10 kb. Ancestral genome configurations were reconstructed using AnChro (56). We first determined the synteny relationships between all possible pairs of haploid *Verticillium* genomes and two outgroup genomes (*Plectosphaerella cucumerina* and *Sodiomyces alkalinus*) using SynChro with synteny block stringency (delta parameter) ranging from 2 to 5 (112). We then obtained all ancestors by calculating all possible pairs of genomes (G1 and G2) and outgroups (G3, . . ., G_n) and by varying the delta′ (G1 and G2 comparisons) and delta″ (G1/G3...G1/G_n and G2/G3...G2/G_n comparisons) parameters for AnChro. We additionally reconstructed all ancestors starting from the extant genomes in a sequential approach with multiple successive cycles of SynChro and AnChro (delta parameters varied between 2 and 5). For each ancestor, we chose the optimal reconstruction by the combination of delta parameters (delta′ and delta″) that minimizes the number of reconstructed chromosomes and rearrangements and at the same time maximizes the number of genes, both guided by the most commonly observed number of chromosomes and genes in all rearrangements. We obtained the number of large-scale rearrangements between reconstructed ancestral genomes and the extant *Verticillium* genomes using ReChro with a delta parameter of 1 (56). The relationship between chromosomes of the reconstructed ancestors and the extant species in relationship to the common ancestor is generated with SynChro with a delta parameter of 1 (112). A species phylogeny that uses synteny relationships computed by SynChro (see above) as informative character between the *Verticillium* genomes and the outgroup genomes was reconstructed using PhyChro (113).

Data availability. ChIP-seq and Hi-C data were submitted to the Short Read Archive (SRA) under the accession no. PRJNA641329 (Table S1a).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, JPG file, 1.1 MB.

FIG S2, JPG file, 0.8 MB.

FIG S3, JPG file, 0.6 MB.

FIG S4, JPG file, 1.7 MB.

FIG S5, JPG file, 1.7 MB.

FIG S6, JPG file, 2.8 MB.

FIG S7, JPG file, 0.6 MB.

FIG S8, JPG file, 0.5 MB.

FIG S9, JPG file, 1.2 MB.

TABLE S1, XLSX file, 0.04 MB.

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. This work was partially supported by USDA's National Institute of Food and Agriculture (award no. 2018-67013-28492) through the Plant Biotic Interactions Program, which is jointly offered by the National Science Foundation (NSF) and the National Institute of Food and Agriculture (NIFA) and by the National Science Foundation (award no. 1936800) through the Models for Uncovering Rules and Unexpected Phenomena in Biological Systems (MODULUS) program of the Division of Molecular and Cellular Biosciences—Systems and Synthetic Biology to D.E.C. 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). We thank Utrecht Sequencing Facility for providing sequencing service and data.

REFERENCES

1. Roy B, Sanyal K. 2011. Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. Eukaryot Cell 10:1384–1395. https://doi.org/10.1128/EC.05165-11.
2. Foley EA, Kapoor TM. 2013. Microtubule attachment and spindle assembly checkpoint signalling at the kinetochore. Nat Rev Mol Cell Biol 14:25–37. https://doi.org/10.1038/nrm3494.
3. Burrack LS, Berman J. 2012. Flexibility of centromere and kinetochore structures. Trends Genet 28:204–212. https://doi.org/10.1016/j.tig.2012.02.003.
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 333:1895–1898. https://doi.org/10.1126/science.1210214.
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 333:1026–1030. https://doi.org/10.1126/science.1206412.
6. Barra V, Fachinetti D. 2018. The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. Nat Commun 9:4340. https://doi.org/10.1038/s41467-018-06545-y.
7. Seidl MF, Thomma BPHJ. 2014. Sex or no sex: evolutionary adaptation occurs regardless. Bioessays 36:335–345. https://doi.org/10.1002/bies.201300155.
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 Res 26:1091–1100. https://doi.org/10.1101/gr.204974.116.
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. https://doi.org/10.1016/j.cell.2008.09.039.
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. https://doi.org/10.1038/nature09529.
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:e53944. https://doi.org/10.7554/eLife.53944.
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 https://doi.org/10.1101/2020.04.09.034512.
13. Yadav V, Sreekumar L, Guin K, Sanyal K. 2018. Five pillars of centromeric chromatin in fungal pathogens. PLoS Pathog 14:e1007150. https://doi.org/10.1371/journal.ppat.1007150.
14. Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102. https://doi.org/10.1126/science.1062939.
15. Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. 2012. Centromeres of filamentous fungi. Chromosome Res 20:635–656. https://doi.org/10.1007/s10577-012-9290-3.
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. Proc Natl Acad Sci U S A 115:3108–3113. https://doi.org/10.1073/pnas.1713725115.
17. Fitzgerald-Hayes M, Clarke L, Carbon J. 1982. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell 29:235–244. https://doi.org/10.1016/0092-8674(82)90108-8.
18. Furuyama S, Biggins S. 2007. Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc Natl Acad Sci U S A 104:14706–14711. https://doi.org/10.1073/pnas.0706985104.
19. Krassovsky K, Henikoff JG, Henikoff S. 2012. Tripartite organization of centromeric chromatin in budding yeast. Proc Natl Acad Sci U S A 109:243–248. https://doi.org/10.1073/pnas.1118898109.
20. Cliften PF, Fulton RS, Wilson RK, Johnston M. 2006. After the duplication: gene loss and adaptation in *Saccharomyces* genomes. Genetics 172:863–872. https://doi.org/10.1534/genetics.105.048900.
21. Baum M, Sanyal K, Mishra PK, Thaler N, Carbon J. 2006. Formation of functional centromeric chromatin is specified epigenetically in *Candida albicans*. Proc Natl Acad Sci U S A 103:14877–14882. https://doi.org/10.1073/pnas.0606958103.
22. Padmanabhan S, Thakur J, Siddharthan R, Sanyal K. 2008. Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. Proc Natl Acad Sci U S A 105:19797–19802. https://doi.org/10.1073/pnas.0809770105.
23. Sanyal K, Baum M, Carbon J. 2004. Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. Proc Natl Acad Sci U S A 101:11374–11379. https://doi.org/10.1073/pnas.0404318101.
24. Cambareri EB, Aisner R, Carbon J. 1998. Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. Mol Cell Biol 18:5465–5477. https://doi.org/10.1128/mcb.18.9.5465.
25. Smith KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M. 2011. Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. Mol Cell Biol 31:2528–2542. https://doi.org/10.1128/MCB.01285-10.

26. Selker EU. 2002. Repeat-induced gene silencing in fungi. Adv Genet 46:439–450. https://doi.org/10.1016/s0065-2660(02)46016-6.
27. Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, Freitag M, Stukenbrock EH. 2015. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin 8:41. https://doi.org/10.1186/s13072-015-0033-5.
28. Thomma BPHJ, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JAL, Faino L. 2016. Mind the gap; seven reasons to close fragmented genome assemblies. Fungal Genet Biol 90:24–30. https://doi.org/10.1016/j.fgb.2015.08.010.
29. Yadav V, Yang F, Reza MH, Liu S, Valent B, Sanyal K, Naqvi NI. 2019. Cellular dynamics and genomic identity of centromeres in cereal blast fungus. mBio 10:e01581-19. https://doi.org/10.1128/mBio.01581-19.
30. Fang Y, Coelho MA, Shu H, Schotanus K, Thimmappa BC, Yadav V, Chen H, Malc EP, Wang J, Mieczkowski PA, Kronmiller B, Tyler BM, Sanyal K, Dong S, Nowrousian M, Heitman J. 2020. Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. PLoS Genet 16:e1008646. https://doi.org/10.1371/journal.pgen.1008646.
31. Navarro-Mendoza MI, Perez-Arques C, Panchal S, Nicolas FE, Mondo SJ, Ganguly P, Pangilinan J, Grigoriev IV, Heitman J, Sanyal K, Garre V. 2019. Early diverging fungus *Mucor circinelloides* lacks centromeric histone CENP-A and displays a mosaic of point and regional centromeres. Curr Biol 29:3791–3802.e6. https://doi.org/10.1016/j.cub.2019.09.024.
32. Faino L, Seidl MF, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, Thomma BPHJ. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. mBio 6:e00936-15. https://doi.org/10.1128/mBio.00936-15.
33. Depotter JRL, Shi-Kunne X, Missonnier H, Liu T, Faino L, van den Berg GCM, Wood TA, Zhang B, Jacques A, Seidl MF, Thomma BPHJ. 2019. Dynamic virulence-related regions of the plant pathogenic fungus *Verticillium dahliae* display enhanced sequence conservation. Mol Ecol 28:3482–3495. https://doi.org/10.1111/mec.15168.
34. de Jonge R, van Esse PH, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang Z, Usami T, Lievens B, Subbarao KV, Thomma BPHJ. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc Natl Acad Sci U S A 109:5110–5115. https://doi.org/10.1073/pnas.1119623109.
35. de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res 23:1271–1282. https://doi.org/10.1101/gr.152660.112.
36. Cook DE, Kramer M, Seidl MF, Thomma BP. 2020. Chromatin features define adaptive genomic regions in a fungal plant pathogen. bioRxiv https://doi.org/10.1101/2020.01.27.921486.
37. Inderbitzin P, Bostock RM, Davis RM, Usami T, Platt HW, Subbarao KV. 2011. Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. PLoS One 6:e28341. https://doi.org/10.1371/journal.pone.0028341.
38. Inderbitzin P, Davis RM, Bostock RM, Subbarao KV. 2011. The ascomycete *Verticillium longisporum* is a hybrid and a plant pathogen with an expanded host range. PLoS One 6:e18260. https://doi.org/10.1371/journal.pone.0018260.
39. Depotter JRL, Beveren F, Rodriguez-Moreno L, van den Berg GCM, Wood TA, Thomma BPHJ, Seidl MF. 2018. Homogenization of subgenome secretome gene expression patterns in the allodiploid fungus *Verticillium longisporum*. bioRxiv 341636. https://www.biorxiv.org/content/10.1101/341636v1.
40. Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF. 2018. Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene loss. Environ Microbiol 20:1362–1373. https://doi.org/10.1111/1462-2920.14037.
41. Sun S, Yadav V, Billmyre RB, Cuomo CA, Nowrousian M, Wang L, Souciet JL, Boekhout T, Porcel B, Wincker P, Granek JA, Sanyal K, Heitman J. 2017. Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. PLoS Biol 15:e2002527. https://doi.org/10.1371/journal.pbio.2002527.
42. Depotter JRL, Seidl MF, van den Berg GCM, Thomma BPHJ, Wood TA. 2017. A distinct and genetically diverse lineage of the hybrid fungal pathogen *Verticillium longisporum* population causes stem striping in British oilseed rape. Environ Microbiol 19:3997–4009. https://doi.org/10.1111/1462-2920.13801.
43. Moeller M, Habig M, Lorrain C, Feurtey A, Haueisen J, Fagundes WC, Alizadeh A, Freitag M, Stukenbrock EH. 2020. Recent loss of the Dim2 cytosine DNA methyltransferase impacts mutation rate and evolution in a fungal plant pathogen. bioRxiv https://doi.org/10.1101/2020.03.27.012203.
44. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z, Henrissat B, Lee Y-H, Park J, Garcia-Pedrajas MD, Barbara DJ, Anchieta A, de Jonge R, Santhanam P, Maruthachalam K, Atallah Z, Amyotte SG, Paz Z, Inderbitzin P, Hayes RJ, Heiman DI, Young S, Zeng Q, Engels R, Galagan J, Cuomo CA, Dobinson KF, Ma L-J. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog 7:e1002137. https://doi.org/10.1371/journal.ppat.1002137.
45. Amyotte SG, Tan X, Pennerman K, del Mar Jimenez-Gasco M, Klosterman SJ, Ma L-J, Dobinson KF, Veronese P. 2012. Transposable elements in phytopathogenic *Verticillium* spp.: insights into genome evolution and inter- and intra-specific diversification. BMC Genomics 13:314. https://doi.org/10.1186/1471-2164-13-314.
46. Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, Schubeler D, Selker EU. 2009. Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. Genome Res 19:427–437. https://doi.org/10.1101/gr.086231.108.
47. 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-J, He J-G, Wang J, He Z-M. 2012. Bisulfite sequencing reveals that *Aspergillus flavus* holds a hollow in DNA methylation. PLoS One 7:e30349. https://doi.org/10.1371/journal.pone.0030349.
48. Seymour M, Ji L, Santos AM, Kamei M, Sasaki T, Basenko EY, Schmitz RJ, Zhang X, Lewis ZA. 2016. Histone H1 limits DNA methylation in *Neurospora crassa*. G3 (Bethesda) 6:1879–1889. https://doi.org/10.1534/g3.116.028324.
49. Kursel LE, Malik HS. 2016. Centromeres. Curr Biol 26:R487–R490. https://doi.org/10.1016/j.cub.2016.05.031.
50. Friedman S, Freitag M. 2017. Evolving centromeres and kinetochores. Adv Genet 98:1–41. https://doi.org/10.1016/bs.adgen.2017.07.001.
51. Flutre T, Duprat E, Feuillet C, Quesneville H. 2011. Considering transposable element diversification in de novo annotation approaches. PLoS One 6:e16526. https://doi.org/10.1371/journal.pone.0016526.
52. Galazka JM, Klocko AD, Uesaka M, Honda S, Selker EU, Freitag M. 2016. *Neurospora* chromosomes are organized by blocs of importin alpha-dependent heterochromatin that are largely independent of H3K9me3. Genome Res 26:1069–1080. https://doi.org/10.1101/gr.203182.115.
53. Winter DJ, Ganley ARD, Young CA, Liachko I, Schardl CL, Dupont P-Y, Berry D, Ram A, Scott B, Cox MP. 2018. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. PLoS Genet 14:e1007467. https://doi.org/10.1371/journal.pgen.1007467.
54. Marie-Nelly H, Marbouty M, Cournac A, Flot J-F, Liti G, Parodi DP, Syan S, Guillén N, Margeot A, Zimmer C, Koszul R. 2014. High-quality genome (re)assembly using chromosomal contact data. Nat Commun 5:5695. https://doi.org/10.1038/ncomms6695.
55. Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P, Dekker J, Mirny L, Barrowman J, Grewal SIS. 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. Nature 516:432–435. https://doi.org/10.1038/nature13833.
56. Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert J-P, Noble WS. 2015. Accurate identification of centromere locations in yeast genomes using Hi-C. Nucleic Acids Res 43:5331–5339. https://doi.org/10.1093/nar/gkv424.
57. Vakirlis N, Sarilar V, Drillon G, Fleiss A, Agier N, Meyniel J-P, Blanpain L, Carbone A, Devillers H, Dubois K, Gillet-Markowska A, Graziani S, Huu-Vang N, Poirel M, Reisser C, Schott J, Schacherer J, Lafontaine I, Llorente B, Neuvéglise C, Fischer G. 2016. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. Genome Res 26:918–932. https://doi.org/10.1101/gr.204420.116.
58. Muller H, Gil J, Jr, Drinnenberg IA. 2019. The impact of centromeres on spatial genome architecture. Trends Genet 35:565–578. https://doi.org/10.1016/j.tig.2019.05.003.
59. Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36–46. https://doi.org/10.1038/nrg3117.
60. Seidl MF, Faino L, Shi-Kunne X, van den Berg GCM, Bolton MD, Thomma BPHJ. 2015. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic

relatives. Mol Plant Microbe Interact 28:362–373. https://doi.org/10.1094/MPMI-06-14-0173-R.
61. International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716. https://doi.org/10.1038/nature03154.
62. Fukagawa T. 2017. Critical histone post-translational modifications for centromere function and propagation. Cell Cycle 16:1259–1265. https://doi.org/10.1080/15384101.2017.1325044.
63. Sullivan BA, Karpen GH. 2004. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol 11:1076–1083. https://doi.org/10.1038/nsmb845.
64. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837. https://doi.org/10.1126/science.1074973.
65. Li X, Wang X, He K, Ma Y, Su N, He H, Stolc V, Tongprasit W, Jin W, Jiang J, Terzaghi W, Li S, Deng XW. 2008. High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. Plant Cell 20:259–276. https://doi.org/10.1105/tpc.107.056879.
66. Blower MD, Sullivan BA, Karpen GH. 2002. Conserved organization of centromeric chromatin in flies and humans. Dev Cell 2:319–330. https://doi.org/10.1016/s1534-5807(02)00135-1.
67. Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Knip M, Sack R, Selker EU. 2010. DNA methylation and normal chromosome behavior in *Neurospora* depend on five components of a histone methyltransferase complex, DCDC. PLoS Genet 6:e1001196. https://doi.org/10.1371/journal.pgen.1001196.
68. Lynch DB, Logue ME, Butler G, Wolfe KH. 2010. Chromosomal G + C content evolution in yeasts: systematic interspecies differences, and GC-poor troughs at centromeres. Genome Biol Evol 2:572–583. https://doi.org/10.1093/gbe/evq042.
69. Hanson SJ, Byrne KP, Wolfe KH. 2014. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system. Proc Natl Acad Sci U S A 111:E4851–E4858. https://doi.org/10.1073/pnas.1416014111.
70. Coughlan AY, Wolfe KH. 2019. The reported point centromeres of *Scheffersomyces stipitis* are retrotransposon long terminal repeats. Yeast 36:275–283. https://doi.org/10.1002/yea.3375.
71. Bayne EH, White SA, Kagansky A, Bijos DA, Sanchez-Pulido L, Hoe KL, Kim DU, Park HO, Ponting CP, Rappsilber J, Allshire RC. 2010. Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. Cell 140:666–677. https://doi.org/10.1016/j.cell.2010.01.038.
72. Buhler M, Moazed D. 2007. Transcription and RNAi in heterochromatic gene silencing. Nat Struct Mol Biol 14:1041–1048. https://doi.org/10.1038/nsmb1315.
73. Yang J, Sun S, Zhang S, Gonzalez M, Dong Q, Chi Z, Chen YH, Li F. 2018. Heterochromatin and RNAi regulate centromeres by protecting CENP-A from ubiquitin-mediated degradation. PLoS Genet 14:e1007572. https://doi.org/10.1371/journal.pgen.1007572.
74. Kagansky A, Folco HD, Almeida R, Pidoux AL, Boukaba A, Simmer F, Urano T, Hamilton GL, Allshire RC. 2009. Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. Science 324:1716–1719. https://doi.org/10.1126/science.1172026.
75. Jesenicnik T, Stajner N, Radisek S, Jakse J. 2019. RNA interference core components identified and characterised in *Verticillium nonalfalfae*, a vascular wilt pathogenic plant fungi of hops. Sci Rep 9:8651. https://doi.org/10.1038/s41598-019-44494-8.
76. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
77. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
78. Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31. https://doi.org/10.1186/1471-2105-6-31.
79. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
80. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
81. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
82. Frandsen RJ, Andersson JA, Kristensen MB, Giese H. 2008. Efficient four fragment cloning for the construction of vectors for targeted gene replacement in filamentous fungi. BMC Mol Biol 9:70. https://doi.org/10.1186/1471-2199-9-70.
83. Santhanam P. 2012. Random insertional mutagenesis in fungal genomes to identify virulence factors. Methods Mol Biol 835:509–517. https://doi.org/10.1007/978-1-61779-501-5_31.
84. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
85. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
86. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
87. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44:W160–W165. https://doi.org/10.1093/nar/gkw257.
88. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137. https://doi.org/10.1186/gb-2008-9-9-r137.
89. Schultz MD, Schmitz RJ, Ecker JR. 2012. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. Trends Genet 28:583–585. https://doi.org/10.1016/j.tig.2012.10.012.
90. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3:95–98. https://doi.org/10.1016/j.cels.2016.07.002.
91. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356:92–95. https://doi.org/10.1126/science.aal3327.
92. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, Glenn St Hilaire B, Yao W, Stamenova E, Hoeger M, Nyquist SK, Korchina V, Pletch K, Flanagan JP, Tomaszewicz A, McAloose D, Pérez Estrada C, Novak BJ, Omer AD, Aiden EL. 2018. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv https://doi.org/10.1101/254797.
93. Jakse J, Jelen V, Radisek S, de Jonge R, Mandelc S, Majer A, Curk T, Zupan B, Thomma B, Javornik B. 2018. Genome sequence of a lethal strain of xylem-Invading *Verticillium nonalfalfae*. Genome Announc 6:e01458-17. https://doi.org/10.1128/genomeA.01458-17.
94. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033.
95. Benjamini Y, Speed TP. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res 40:e72. https://doi.org/10.1093/nar/gks001.
96. Dobin A, Gingeras TR. 2015. Mapping RNA-seq reads with STAR. Curr Protoc Bioinformatics 51:11.14.1–11.14.19. https://doi.org/10.1002/0471250953.bi1114s51.
97. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol 9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
98. Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9:18. https://doi.org/10.1186/1471-2105-9-18.
99. Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res 37:7002–7013. https://doi.org/10.1093/nar/gkp759.
100. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, Quesneville H. 2014. PASTEC: an automatic transposable element clas-

sification tool. PLoS One 9:e91929. https://doi.org/10.1371/journal.pone.0091929.
101. Bailly-Bechet M, Haudry A, Lerat E. 2014. 'One code to find them all': a perl tool to conveniently parse RepeatMasker output files. Mobile DNA 5:13. https://doi.org/10.1186/1759-8753-5-13.
102. Palmer JM, Je S. 2016. Funannotate: eukaryotic genome annotation pipeline. http://funannotate.readthedocs.io.
103. Cook DE, Valle-Inclan JE, Pajoro A, Rovenich H, Thomma B, Faino L. 2019. Long-read annotation: automated eukaryotic genome annotation based on long-read cDNA sequencing. Plant Physiol 179:38–54. https://doi.org/10.1104/pp.18.00848.
104. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol 9:R7. https://doi.org/10.1186/gb-2008-9-1-r7.
105. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.
106. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
107. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer Verlag, New York, NY.
108. Gel B, Serra E. 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 33:3088–3090. https://doi.org/10.1093/bioinformatics/btx346.
109. Hahne F, Ivanek R. 2016. Visualizing genomic data using Gviz and Bioconductor. Methods Mol Biol 1418:335–351. https://doi.org/10.1007/978-1-4939-3578-9_16.
110. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27:1009–1010. https://doi.org/10.1093/bioinformatics/btr039.
111. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
112. Drillon G, Carbone A, Fischer G. 2014. SynChro: a fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. PLoS One 9:e92621. https://doi.org/10.1371/journal.pone.0092621.
113. Drillon G, Champeimont R, Oteri F, Fischer G, Carbone A. 2020. Phylogenetic reconstruction based on synteny block and gene adjacencies. Mol Biol Evol https://doi.org/10.1093/molbev/msaa114.
114. Depotter JR, Deketelaere S, Inderbitzin P, Tiedemann AV, Hofte M, Subbarao KV, Wood TA, Thomma BP. 2016. *Verticillium longisporum*, the invisible threat to oilseed rape and other brassicaceous plant hosts. Mol Plant Pathol 17:1004–1016. https://doi.org/10.1111/mpp.12350.
115. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.