The multi-speed genome of *Fusarium oxysporum* reveals association of histone modifications with sequence divergence and footprints of past horizontal chromosome transfer events.

Short title: Complex genome organization and adaptivity in *Fusarium oxysporum*

Like Fokkens[1], Shermineh Shahi[1], Lanelle R. Connolly[2], Remco Stam[3], Sarah M. Schmidt[1,4], Kristina M. Smith[2,5], Michael Freitag[2], Martijn Rep[1].

[1] Molecular Plant Pathology, University of Amsterdam, The Netherlands

[2] Department of Biochemistry and Biophysics, Oregon State University, United States of America

[3] Chair of Phytopathology, School of Life Sciences Weihenstephan, Technische Universität München, Germany

[4] Current address: Institute for Molecular Physiology, Heinrich-Heine-Universität Düsseldorf, Germany

[4] Current address: Departement of Biology, Oregon State University-Cascades, United States of America

Abstract

Fusarium oxysporum is an economically important pathogen causing wilting or rotting disease symptoms in a large number of crops. It is proposed to have a structured, "two-speed" genome: i.e. regions containing genes involved in pathogenicity cluster with transposons on separate accessory chromosomes. This is hypothesized to enhance evolvability. Given the continuum of adaptation of all the genes encoded in a genome, however, one would expect a more complex genome structure. By comparing the genome of reference strain Fol4287 to those of 58 other *Fusarium oxysporum* strains, we found that some Fol4287 accessory chromosomes are lineage-specific, while others occur in multiple lineages with very high sequence similarity - but only in strains that infect the same host as Fol4287. This indicates that horizontal chromosome transfer has been instrumental in past host-switches. Unexpectedly, we found that the sequence of the three smallest core chromosomes (Chr. 11, 12 and 13) is more divergent than that of the other core chromosomes. Moreover, these chromosomes are enriched in genes involved in metabolism and transport and genes that are differentially regulated during infection. Interestingly, these chromosomes are –like the accessory chromosomes– marked by histone H3 lysine 27 trimethylation (H3K27me3) and depleted in histone H3 lysine 4 dimethylation (H3K4me2). Detailed genomic analyses revealed a complex, "multi-speed genome" structure in *Fusarium oxysporum*. We found a strong association of H3K27me3 with elevated levels of sequence divergence that is independent of the presence of repetitive elements. This provides new leads into how clustering of genes evolving at similar rates could increase evolvability.

Author summary

Fungi that cause disease on plants are an increasingly important threat to food security. New fungal diseases emerge regularly. The agricultural industry makes

large investments to breed crops that are resistant to fungal infections, yet rapid adaptation enables fungal pathogens to overcome this resistance within a few years or decades. It has been proposed that genome 'compartmentalization' of plant pathogenic fungi, in which infection-related genes are clustered with transposable elements (or 'jumping genes') into separate, fast-evolving regions, enhances their adaptivity. Here, we aimed to shed light on the possible interplay between genome organization and adaptation. We measured differences in sequence divergence and dispensability between and within individual chromosomes of the important plant pathogen *Fusarium oxysporum.* Based on these differences we defined four distinct chromosomal categories. We then mapped histone modifications and gene expression levels under different conditions for these four categories. We found a 'division of labor' between chromosomes, where some are 'pathogenicity chromosomes' - specialized towards infection of a specific host, while others are enriched in genes involved in more generic infection-related processes. Moreover, we confirmed that horizontal transfer of pathogenicity chromosomes likely plays an important role in gain of pathogenicity. Finally, we found that a specific histone modification is associated with increased sequence divergence.

Introduction

Genes encode products that are involved in a myriad of biological processes and are subjected to a wide range of selection pressures that may change over time. In pathogenic micro-organisms, fast-evolving genes are often clustered with transposable elements into separate regions, an observation that has led to the "two-speed genome" hypothesis concept [1-3]. A possible advantage for the separate genome regions with different evolutionary rates is that it might improve a genome's ability to accommodate conflicting evolutionary demands, thus allowing an organism to adapt more quickly to new environments [1-4]. Pathogens secrete specialized proteins, called effectors, that facilitate colonization of a host. These effectors are,

however, sometimes recognized by the host immune system and then trigger an immune response. Effectors that trigger an immune response are called avirulence factors. A pathogen can escape this recognition when the avirulence gene is mutated or no longer expressed during infection. There is accumulating evidence that this is more likely to occur when an avirulence gene is located near transposon sequences. The occurrence of multiple almost identical sequences in a genome increases the likelihood of homologous recombination that can result in genome rearrangements or deletions [5], which can contribute to pathogen adaptation [1-3,6-15]. In addition, transposon insertion into an avirulence gene can result in regain of virulence [16-19]. If the avirulence gene is indispensable for infection the only option to escape recognition is to alter the amino acid sequence of the effector [20-22]. Moreover, an effector needs to coevolve with its target in the host (see e.g. [23]). Effector genes and other genes that are located in 'fast' regions have been found to be under positive selection [24,25] and in general have more polymorphisms [26]. However, with the notable exception of repeat-induced point mutation (RIP) in *Leptosphaeria* species [27,28], no mechanism explaining the link between the proximity of transposons and elevated levels of sequence evolution has yet been firmly established.

Core and accessory chromosomes in fungi typically not only differ in terms of dispensability. They also show differences in gene- and repeat density and -in species that undergo a sexual cycle- in recombination frequencies [29,30] (see e.g. [13,31] for review). Recent studies on the chromatin of core and accessory regions showed that accessory regions are marked by histone H3 lysine 27 trimethylation (H3K27me3), sometimes called 'facultative heterochromatin' [31-34]. One hypothesis is that this allows for concerted regulation of genes involved in similar processes, such as infection [35].

As a basis for investigating the underlying mechanisms of genome compartmentalization and differences in evolutionary 'speeds', a detailed mapping of evolutionary 'volatility' across a genome is required and correlations with transposon proximity, histone modifications and gene expression must be established. In this study, we used 58 *Fusarium oxysporum* genome sequences for a detailed comparison of differences in dispensability and levels of sequence divergence between and within chromosomes. We combined this with functional annotation and gene expression data to study the extent to which evolutionary volatility is linked to involvement in infection. To study whether there are differences in chromatin-mediated regulation between and within chromosomes in Fol4287, and whether these differences correlate with differences in dispensability and sequence divergence and gene expression, we determined the distribution of histone marks associated with euchromatin (H3K4me2) and facultative heterochromatin (H3K27me3) in vitro and correlated these to gene annotation categories as well as levels of gene expression *in vitro* and *in planta*.

The *Fusarium oxysporum* species complex (FOSC) comprises strains that can infect over a 100 economically important crops, though most strains are not known to be pathogenic [36,37]. Individual pathogenic strains are mostly specialized towards a single host and are grouped according to host-preference into *formae speciales* (ff.spp.) [36,38-40]. Strains that belong to the same *forma specialis* (f.sp.) have similar effector repertoires [40]. In 2010, Ma et al. reported the genome sequence of the tomato-infecting strain *Fusarium oxysporum* f. sp. *lycopersici* 4287 (Fol4287). Out of 15 chromosomes in Fol4287, only 11 are largely syntenic with chromosomes of *Fusarium verticillioides*, a sister species that is estimated to have diverged about 11 Mya [37,41]. From this comparison, these 11 chromosomes constitute the core genome of Fol4287, whereas 4 chromosomes (3, 6, 14 and 15) and two large,

130 presumable translocated, regions (chromosome 1b and 2b) are absent in *Fusarium*
131 *verticillioides* and thus constitute the accessory genome of Fol4287. These
132 chromosomes also have been reported to be largely absent in the strains *Fom*-5190a,
133 *Foc*-38-1, *Fop*-37622, *Fo5176*, and *Fom001*, infecting *Medicago* spp., chickpea, pea,
134 *Brassica* spp. and melon respectively [42], indicating that they are conditionally
135 dispensable in the FOSC. Remarkably, strains that have lost 'core' chromosome 12
136 showed no reduction in virulence in experimental conditions and reduction of growth
137 on only a few tested carbon sources. This indicates that chromosome 12 is also
138 conditionally dispensable and thus not a core chromosome *sensu strictu* [43].
139
140 Interestingly, host-preference in FOSC is often polyphyletic [44-56]. Several studies
141 revealed putative horizontal transfer events of effector genes between FOSC strains
142 that infect the same host [40,57-59], indicating that horizontal transfer may play an
143 important role in host switches [39,40,56,60,61]. Of the four accessory chromosomes
144 in Fol4287, chromosome 14 is undoubtedly a pathogenicity chromosome: all fourteen
145 SIX (Secreted In Xylem) genes – known effector genes in tomato-infecting *Fusarium*
146 *oxysporum* – are located on chromosome 14 [62] (and our unpublished observations).
147 Loss of chromosome 14 leads to loss of pathogenicity in Fol4287 [43]. Importantly,
148 horizontal transfer experiments have shown that acquisition of chromosome 14 by
149 the non-pathogenic strain Fo47 is sufficient to turn this strain pathogenic on tomato,
150 albeit less virulent than the original donor [41,63]. The first horizontal transfer
151 experiment reported for the FOSC used the tomato-infecting strain Fol007, which
152 belongs to the same clonal lineage as Fol4287 yet has a different karyotype: Fol007
153 has a small chromosome that is not present in Fol4287. Interestingly, when this small
154 chromosome co-transferred with chromosome 14, the recipient strains were much
155 more virulent [41]. Recently, a similar experiment with the pathogenicity chromosome
156 of Forc016, causing root-rot in cucurbits, shows that host-switches by horizontal
157 chromosome transfer are not unique to tomato-infecting strains [64].

158

159 The polyphyletic origins of host-specificity, the fact that strains that infect the same
160 host have similar effector repertoires, the frequent observations of horizontal transfer
161 of effector genes, and the fact that horizontal transfer of a pathogenicity chromosome
162 can render a non-pathogenic strain pathogenic all indicate that horizontal
163 chromosome transfer causes host-switches in FOSC. Our genome dataset includes
164 tomato-infecting strains that belong to different lineages, hence we also used it to
165 detect putative recent horizontal chromosome transfer events.

166

167

168

169

Results

The accessory genome of Fol4287 consists of pathogenicity-related regions and lineage-specific regions.

When compared to *F. verticillioides*, the accessory genome of Fol4287 consists of four chromosomes and two large segments on core chromosomes. We compared the genome of Fol4287 to that of a closely related strain that is pathogenic on melon, Fom001, and found that the core chromosomes are almost 100% identical between these strains, but that some of the accessory chromosomes are specific to Fol4287 (Fig S1). Chromosomal regions 1b and 2b and chromosome 15, however, align with the genome of the Fom001 with similar levels of sequence similarity as the core chromosomes (Fig S1). This suggests that host-preference is largely determined by genes that are located on three accessory chromosomes. We predict that genes on these chromosomes that are important for tomato infection are also present in other strains that are pathogenic on tomato.

To further assess which accessory regions correlate with phylogenetic clades (lineage-specific chromosomes), and which with *forma specialis* (pathogenicity chromosomes), we aligned 58 FOSC genomes to the genome of Fol4287 (Fig S2, Table S1). We found that the four accessory chromosomes are largely clade-specific. Alignments with genomes of strains outside the lineage of Fol4287 (including Fom001) mostly span less than 5 kilobases (kb) and have less than 92% sequence identity. Notable exceptions are chromosome 14 and a recently duplicated region on chromosome 3 and 6. These regions are present in more distantly related tomato-infecting strains with ~100% sequence identity, compared to ~98% identity between core chromosomes (Fig1A, Fig 2, Fig S2). For these regions synteny is also relatively conserved in tomato-infecting isolates, with aligned segments spanning up to 40 kb, whereas for the rest of the accessory genome this is between 5 and 10 kb (Fig S3).

In conclusion, accessory chromosomes of Fol4287 largely correlate with phylogenetic clade, but chromosome 14 and regions on chromosome 3 and 6 correlate with host and are more conserved in sequence than core chromosomes.

We then asked whether these pathogenicity regions on chromosome 3 and 6 correspond to the small chromosome that was co-transferred with chromosome 14 in in the experiment described in [41], resulting in strains that were more virulent than the ones that received only one chromosome. We sequenced one strain that received chromosome 14 (from Fol007) and one – more virulent – strain that received chromosome 14 and the additional small chromosome. We mapped the reads thus obtained to the genome of Fol4287 and found that exactly those regions in Fol4287 that show hallmarks of horizontal transfer in our genome comparison (Fig 1A) correspond to chromosome 14 and the second, smaller 'pathogenicity' chromosome in Fol007 (Fig 1B).

Our results strongly suggest that horizontal transfer of chromosomes has indeed played an important role in the emergence of new pathogenic clonal lines in natural settings. Moreover, not all accessory regions in Fol4287 are alike: only some regions are associated with pathogenicity towards tomato. This suggests a division between accessory regions in Fol4287: chromosome 14 and part of chromosome 3/6 are involved in virulence towards tomato whereas the other accessory regions are not. From here on, we refer to the chromosomal regions associated with pathogenicity (i.e. chromosome 14 and part of chromosome 3 and 6) as 'pathogenicity regions', and to the rest of the accessory genome of Fol4287 as 'lineage-specific regions', as they only occur in strains that belong to the same lineage (vegetative compatibility group) as Fol4287.

The core genome can be divided into regions with low divergence and regions with high divergence

To compare differences in dispensability among core chromosomes, we applied the same approach to the core genome of Fol4287. We found that all core chromosomes are present in all strains in our dataset, with the notable exception of chromosome 12 that is absent in the cucumber-infecting strain Foc037. This strain was included in several bioassays in a previous study, in which its virulence was confirmed [40]. Recently, flow cytometry was used to screen for loss of chromosome 1, 12 and 14 in Fol4287 [43]. In all three reported cases in which chromosome 12 was lost, no differences were observed in terms of virulence. Chromosome 12 is thus dispensable for growth as well as virulence, and should thus no longer be considered a core chromosome *sensu strictu*. In addition to complete loss of chromosome 12 in Foc037, we observed large deletions in core chromosomes in two melon-infecting strains: ~0.5 Mb in a subtelomeric region of chromosome 12 in Fom010, and ~0.5 Mb in a subtelomeric region of chromosome 13 in Fom013 (Fig 2).

In addition to differences in the propensity for loss or large deletions, there are striking differences in the level of sequence conservation within and among core chromosomes. The three smallest core chromosomes of Fol4287 – Chr. 11, 12 and 13 – are clearly more divergent than the other, larger core chromosomes (Fig 2) in terms of sequence similarity as well as synteny (Fig S2, Fig S3). Subtelomeric regions of all chromosomes and a ~1 Mb central region of chromosome 4, associated with a genomic rearrangement when compared to *F. verticillioides* [41], also show elevated levels of sequence divergence and lower synteny levels (Fig S2, Fig S3). Notably, these chromosomes are not enriched in repetitive elements and have a similar gene density as other core chromosomes [41] (Fig S4). From here on, we will refer to chromosomes 11, 12 and 13 as ‘fast-core’ chromosomes.

Genes on fast-core and accessory chromosomes have lower expression levels. We hypothesized that genes in different genome compartments exhibit differences in overall expression levels. To assess this, we queried an RNA-seq dataset that was generated to compare gene expression of Fol4287 grown in liquid culture (*in vitro*) to Fol4287 infecting tomato plants (*in planta*, 9 days post inoculation) [65] (Table S2). On the accessory genome, most genes -64% and 68% of genes on the lineage-specific and pathogenicity chromosomal regions respectively- are not expressed *in vitro*, (defined here as RPKM <= 0.1). On the fast-core and the core chromosomes, this is 46% and 25% respectively. *In planta*, 53% of genes on the accessory are not expressed (lineage-specific 51%, pathogenicity 57%), compared to <33% of genes on the fast-core and <19% of genes on the core chromosomes. Of the fast-core chromosomes, chromosome 12 - that is absent from Foc037 - has most genes that are not expressed in the conditions we tested (53% *in vitro*, 37% *in planta*). Overall, genes on the accessory genome and genes located in fast-core chromosomal regions (including the central region on chromosome 4) have lower expression than genes on the core genome, both *in vitro* and *in planta* (average over three biological replicates, Fig S5).

Of the accessory genome, only chromosome 14 is enriched for genes that are differentially regulated during infection.

We expected that the pathogenicity regions, required for virulence on tomato (Fig 1), would be enriched for genes that are upregulated during infection. When we compared *in vitro* to *in planta* expression levels, we found that chromosome 14 is indeed significantly enriched for genes that are upregulated during infection (P-value < 9.19×10^{-10} after Bonferroni correction, Fig 3, Fig S6). However, the 'pathogenicity-related' regions on chromosome 3 and 6 – although they correspond to the small chromosome in Fol007 that enhances virulence upon horizontal transfer – are not significantly enriched for upregulated genes. Like the other (lineage-specific)

accessory regions, here only 3.1% of the genes were upregulated *in planta*, compared to > 14.7% of genes on chromosome 14. Hence of the pathogenicity regions, only chromosome 14 is enriched in genes that are differentially expressed during infection.

We then asked whether the upregulated genes on chromosome 14 have similar functions as those encoded on chromosome 3 and 6. Of the 61 genes on chromosome 14 that are upregulated during infection, 16 encode small, secreted proteins, 13 of those are SIX (Secreted In Xylem) genes. About a third (19 out of 61) encodes enzymes, of which four are secreted. The remainder of upregulated genes includes three genes encoding for transcription factors, two of which are homologs of *FTF1*, that induces expression of effector genes [65], three putative transposable elements, a transporter, one encoding a membrane protein and one encoding a putative argonaute-like silencing protein (Table S3). The remaining 19 upregulated genes have no known functional domains and contain no secretion signal peptide. The proteins encoded by the 28 upregulated genes located in the pathogenicity regions on chromosome 3 and 6 include ten transposase-like proteins with a domain of unknown function, five enzymes and two small secreted proteins, one of which is a homolog of *SIX8* (Table S4). The three downregulated genes in the pathogenicity regions on chromosome 3 and 6 encode proteins with no known function and without a secretion signal peptide. In conclusion, more than a quarter of the upregulated genes on chromosome 14 encode small, secreted proteins, compared to less than 8 % of the upregulated genes on the pathogenicity regions on chromosome 3 and 6.

Fast-core regions are enriched for genes upregulated during infection and are involved in metabolism, transport and defence.

Interestingly, while on the 'normal' core chromosomes only 4% of the genes is differentially regulated during infection, on the fast-core chromosomes this is > 10%.

Hence not only chromosome 14, but also the fast-core chromosomes are significantly enriched for differentially expressed genes (P-value < 1.17×10^{-31}), both up- (P-value < 6.5×10^{-25}) and downregulated (P-value < 1.67×10^{-06}) (Fig 3). Not just fast-core chromosomes 11, 12 and 13, but also the other regions that showed elevated levels of sequence divergence (Fig 1, Fig 2, Fig S2) such as the subtelomeric regions of core chromosomes and the central region on chromosome 4 have more differentially expressed genes than core regions (Fig S6).

Compared to chromosome 14, the fast-core chromosomes are relatively conserved within the FOSC (Fig 1, Fig 2, Fig S2), suggesting that the upregulated genes on the fast-core chromosomes are involved in common infection-related processes while chromosome 14 contains host-specific genes. Of the 277 upregulated fast-core genes, 93 (~34%) encode proteins that contain a putative signal peptide for secretion. This is similar to the proportion of upregulated genes on chromosome 14 that encode proteins with a predicted secretion signal peptide, but on chromosome 14 these are mostly small effector genes (Table S4). When we compared the functional annotations of upregulated fast-core genes to that of the rest of the genome, we find that upregulated fast-core genes are enriched for genes involved in metabolism, cellular transport, transport facilitation and transport routes and cell rescue, defense and virulence. More specifically, upregulated fast-core genes are enriched for genes involved in polysaccharide and other C-compound and carbohydrate metabolism, sugar, glucoside, polyol and carboxylate catabolism and extracellular polysaccharide degradation, in transport facilities, C-compound and carbohydrate transport and cellular import, and in degradation/modification of foreign (exogenous) polysaccharides (Table S5). The fact that these genes are differentially expressed during infection suggests that genes in fast-core regions are involved in digestion of plant material and import of products of digestion.

Not only fast-core genes that are upregulated during infection, are involved in metabolism, cellular transport, transport facilitation and transport routes and cell rescue, defense and virulence: when we also include fast-core genes that are not differentially regulated during infection, or are downregulated during infection, we find that these are involved in the same three broad functional categories (Fig S7) and in 'Interaction with the environment'. Metabolic genes on the fast-core are typically involved in 'metabolism of carbon compounds and carbohydrates' (49.6% of metabolic genes) and 'secondary metabolism' (38.7% of metabolic genes). Transport genes are mostly involved in 'transport of carbon compounds and carbohydrates' (50.9% of transport genes), 'cellular import' (61.7% of transport genes) – in particular 'non-vesicular cellular import' (46.1% of transport genes) and 'transport facilities' (85.2% of transport genes). Defence genes are mostly genes involved in 'detoxification' (67.8% of defence genes) including 'detoxification involving cytochrome P450' (15.3% of defence genes) and in 'disease, virulence and defence' (32.2% of defence genes). This indicates that the clustering of genes that are involved in metabolism, transport and defence on fast-core chromosomes is not limited to genes that are involved in infection.

Like accessory chromosomes, fast-core chromosomes are enriched with H3K27me3.

Recent studies have revealed differences in chromatin-mediated regulation of core versus accessory chromosomal regions in pathogenic filamentous fungi: under *in vitro* conditions accessory chromosomes or chromosomal regions are enriched in H3K27me3, which is correlated with gene silencing, whereas core chromosomes are enriched in H3K4me2, which is correlated with gene activation [31,32,34]. To test whether the same holds true for *Fusarium oxysporum*, we performed ChIP-seq experiments for these two different histone marks (Table S6). We found that in *F. oxysporum*, similar to what was reported for *F. graminearum*, *F. fujikuroi* and *Z. tritici*

[31,32,34,66], accessory regions are enriched for H3K27me3 and depleted in H3K4me2 (Fig 3, Fig S8). Interestingly, the same holds true for the fast-core chromosomes and sub-telomeric regions, in sharp contrast to the rest of the core genome that is enriched in H3K4me2 and depleted in H3K27me3 (Fig 4, Fig S8, Fig S9, Table S7).

H3K4me2 was found in or close to coding regions (Kendall's Tau <= 0.39, $p < 1 \times 10^{-300}$), and is negatively correlated with the presence of transposons (Kendall's Tau <= -0.33, $p < 1 \times 10^{-300}$). On the core genome, most genes and promoter regions overlap with an H3K4me2-enriched domain, while most transposons overlap with an H3K27me3-enriched domain (Fig S10). H3K4me2-enriched domains on core chromosomes are larger than those on fast-core and accessory chromosomes: they span between 10 and 100 kb on the core genome, compared to between 2.5 and 5 kb on the fast-core and accessory chromosomes (Fig 4, Fig S9, Table S7). More than 49% of the core genome is covered in H3K4me2-enriched domains, while only 2.5%-13% of the non-core genome is part of an H3K4me2-enriched domain.

In contrast, more than 72% of the accessory genome, ~93% of the fast-core genome and <27% of the core genome is part of an H3K27me3-enriched domain. H3K27me3 is negatively correlated with the presence of coding sequences (Kendall's Tau <= -0.25, $P < 3 \times 10^{-180}$) and is enriched near or in transposons (Kendall's Tau >= 0.35, $P < 5{,}9 \times 10^{-283}$). In general, H3K27me3 occurs in larger blocks than H3K4me2, on all chromosomes (Fig 4, Fig S9, Table S7). The average size of an enriched domain ranges between 15 to 60 kb on core chromosomes, where on the fast-core chromosomes domains vary between 57 and 146 kb (domain size averaged per chromosome). Especially on chromosome 12, H3K27me3-enriched domains are large: averaging ~146 kb in one experiment and ~83 kb in the other (Fig 4, Fig S9, Table S7).

H3K27me3 is associated with facultative heterochromatin, hence we expect low expression of genes that reside in a region that is enriched in this histone modification. In contrast, H3K4me2 is associated with euchromatin, hence we expect high expression levels for genes that are located in regions enriched in this histone mark compared to genes that reside in H3K37me3 domains of enrichment. As mentioned above, the fast-core and accessory chromosomes -enriched in H3K27me3- have more genes with low expression *in vitro* and -to a lesser extent- *in planta* than core chromosomes. Genes that are located in regions that are associated with H3K27me3 have lower transcript levels, both in *in vitro* and in *in planta* conditions (P value ~ 0, Welch's t-test on ranked data), while genes that are located in regions associated with H3K4me2 have higher transcript levels than the rest of the genome (P value ~ 0, Welch's t-test on ranked data), consistent with the assumed role of these histone marks in gene silencing and activation in other fungi [31-34,66]. When we compared gene expression *in vitro* to *in planta* conditions, we found that most transcriptional changes involve *up-regulation* of genes: 896 out of 1036 differentially expressed genes are upregulated *in planta* compared to *in vitro* (>86%) (Table S4, Fig S6). Of these upregulated genes, more than 66% is located in an H3K27me3-enriched domain (P-value < 1.6×10^{-41}, hypergeometric test). The differences in histone codes of core, fast-core and accessory genomes may allow for large-scale coordinated expression of genes that are involved in life-style switches, such as infection.

Fast-core chromosomes are not enriched for genes under positive selection.

The differences in functional gene classes enriched in core, fast-core, accessory and pathogenicity chromosomes raises the question to what extent the observed hierarchy in sequence divergence levels (Fig 1, Fig 2) of these four categories of chromosomes can be explained by positive selection. A previous study identified

420 genes under positive selection based on multiple sequence alignments of genes that
421 belong to nine Fusarium species, including *F. oxysporum*, comparing synonymous
422 and non-synonymous substitution rates and testing for the occurrence of site-specific
423 balancing selection. This study revealed that all Fol4287 accessory chromosomes
424 (pathogenicity and lineage-specific regions) are enriched for genes that are under
425 positive selection [24]. The core and fast-core chromosomes contain relatively few
426 genes that are inferred to be under positive selection: 308 (< 2.7% of tested core
427 genes) and 66 (< 2.8% of tested fast-core genes) respectively (Table S8), compared
428 to 192 genes located in accessory regions (> 9.8% of tested accessory genes; 8.4%
429 lineage-specific and 12.5% pathogenicity). Differentially expressed genes are not
430 significantly enriched for genes under positive selection (P-value > 0.5). Hence
431 positive or diversifying selection on genes that are located on the fast-core
432 chromosomes does not explain the high levels of sequence divergence we observed.
433
434 We then hypothesized that genes on these chromosomes are under relaxed negative
435 selection, which could also explain differences in sequence divergence: many genes
436 have low expression levels. Moreover, fast-core chromosome 12 can be lost in both
437 Fol4287 and Foq037 without losing virulence [40,43] (Fig 2). Per gene, we inferred
438 the number of synonymous substitutions per synonymous site (d_S), non-synonymous
439 substitutions per non-synonymous site (d_N), as well as their ratio (d_N/d_S) based on
440 alignment with their bidirectional best BLAST hit in *F. verticillioides*. We found that
441 higher substitution rates in genes located on fast-core chromosomes compared to
442 those located on other core chromosomes (P-value < 2.8e-39, P-value < 8e-307, P-
443 value < 0.0015, for d_S, d_N and d_N/d_S respectively, Welch's t-test on ranked data, Fig
444 5A, Fig S11). Synonymous substitutions should be little influenced by relaxed or
445 positive selection, yet we observed also a significantly higher *synonymous*
446 substitution rate on fast-core chromosomes. We therefore conclude that relaxed
447 selection on dispensable genes and genes with low expression levels alone, does

not explain differences in sequence divergence between fast-core chromosomes and other core chromosomes.

Rate of sequence divergence is elevated in H3K27me3-enriched regions on core chromosomes.

Interestingly, both genes on fast-core chromosomes and genes on core chromosomes that are located close to a telomere also have higher d_S and d_N values (Fig 5A, Table S8, Fig S11). When we compare substitution levels for genes that are located in a region enriched in H3K27me3 on core chromosomes, the distribution of d_N and d_S of these genes is similar to those located on the fast-core chromosomes (Fig 5A, Fig S12). If we view the synonymous substitution rate as an approximation of the molecular clock, these results suggest that this “clock is ticking faster” in regions that are enriched in H3K27me3.

To study differences in rates of sequence evolution in more detail we inferred Single Nucleotide Polymorphisms (SNPs) on two timescales. The cumulative SNP density of the ten tomato-infecting strains that belong to the same VCG as Fol4287 (the bottom tomato-infecting lineage in Figs 1 and 2, see Materials and Methods) approximates the accumulation of mutations on a short timescale. Assuming no horizontal transfers have occurred within this lineage, we can compare accessory chromosomes with core and fast-core chromosomes, as all strains possess Fol4287’s accessory chromosomes, albeit not all complete chromosomes. We found that the accessory chromosomes accumulated more SNPs than the fast-core and core chromosomes, but that the SNPs density of core and fast-core chromosomes is similar (P-value > 0.35, hypergeometric test, Fig 5, Fig S13, Fig S14). We found that on fast-core chromosomes SNP density was higher in H3K27me3-enriched regions (comprising more than 90% of the fast-core chromosomes), compared to regions that are not enriched in H3K27me3, but that this difference was not significant. However,

476 on core chromosomes H3K27me3-enriched regions have a significantly higher
477 density of SNPs (P-value < 1.5 x 10^{-7}). When we included SNPs in less closely
478 related strains (i.e. all strains in the Clade III, the clade depicted in Fig 1), we found
479 that fast-core chromosomes, the central region on chromosome 4 and sub-telomeric
480 core regions have more SNPs than (other) core regions (Fig S13). Hence, on a
481 longer timescale, as observed when comparing genome sequences, we find that
482 regions that are enriched in H3K27me3, have higher sequence divergence (P-value
483 ~ 0, hypergeometric test, for both core and fast-core chromosomes). Due to the fact
484 that the accessory chromosomes are absent from a number of strains in Clade III,
485 and the fact that the pathogenicity chromosomes were obtained through horizontal
486 transfer, we cannot directly compare accessory chromosomes to core and fast-core
487 chromosomes. We conclude that, especially on core chromosomes and/or on longer
488 timescales, H3K27me3-enriched regions accumulated more SNPs compared to
489 regions that are not enriched in H3K27me3.
490
491 We calculated correlations between change in gene expression, sequence
492 divergence and histone modifications on core or fast-core chromosomes, and found
493 two clusters -each corresponding to a different 'speed'- where repeat density and
494 SNP density cluster with H3K27me3, while gene density and gene expression levels
495 cluster with H3K4me2 (Fig 6A and B). Interestingly, H3K27me3 correlated most
496 strongly with synonymous substitution rates and SNP densities in Clade III, and
497 much less with repeat density. For accessory chromosomes this distinction was not
498 so clear; we observed mostly very low correlation coefficients between different
499 properties (Fig 6C). In conclusion, fast-core and core chromosomes harbour similar
500 "speeds", but in different quantities. The evolutionary processes underlying the
501 different "speeds" on core and fast-core chromosomes probably have had less
502 influence in shaping the organization of accessory chromosomes.
503

Discussion

In this study, we found that *F. oxysporum* has a more complex genome organization than previously proposed [41]. While it has been known for several years that *F. oxysporum* has core and accessory chromosomes, our more detailed comparisons revealed distinct subcategories. Core chromosomes can be divided into 'core' (*sensu strictu*) that are indispensable, conserved, enriched in H3K4me2 and not enriched in genes that are differentially expressed during infection and 'fast-core' that are conditionally dispensable, less conserved, enriched in H3K27me3 and enriched in genes that are differentially expressed during infection. Similarly, we suggest to classify the accessory genome, which is enriched in H3K27me3, into lineage-specific chromosomes that are specific to a clonal line and pathogenicity chromosomes that are present in all strains that infect a certain host. Apart from distinct chromosomes, we also observed regions -parts of chromosomes- that could be assigned to a different category as the rest of the chromosome, such as sub-telomeres, or translocated segments on *bona fide* core chromosomes that can be classified as fast-core, i.e., segments of core chromosomes 1 and 2 that are lineage-specific, and segments of chromosome 3 and 6 that are considered pathogenicity regions.

The genome alignments and the analyses of SNP densities along the genome showed that the different subcategories can be ordered into a hierarchy with respect to their rate of sequence divergence: compared to core regions, fast-core regions accumulated more SNPs and alignments of fast-core regions with genome sequences of other FOSC isolates have lower percent identity. Fol4287's accessory regions are absent from most strains and accumulated many SNPs on a timescale, on which we found no significant difference between core and fast-core chromosomes. Similarly, synteny is most conserved in core regions, less in fast-core regions, and even less in pathogenicity regions and in lineage-specific regions.

Interestingly, we find that -to some extent- different levels in this hierarchy correspond to levels in host-interactions: both fast-core chromosomes and chromosome 14 are enriched for differentially expressed genes, but genes on the less volatile fast-core chromosomes are typically involved in metabolism, transport and defence, whereas chromosome 14 contains many effector genes that are hypothesized to interact with host proteins [67]. The fast-core chromosomes are in general enriched in genes involved in metabolism, transport and defence, importantly, including genes that are not differentially regulated during infection. *F. oxysporum* is not an obligate pathogen and we predict that levels in the observed sequence divergence hierarchy also correspond to different levels on other functional 'axes', including interactions with other soil organisms, where for example lineage-specific regions may play a more important role.

Although we were able to associate different levels of sequence divergence with different levels of host interaction, this correspondence is not perfect. While pathogenicity chromosome 14 of Fol4287 is enriched in genes with induced expression during infection, the regions on chromosome 3 and 6 are not, neither do these regions encode any known effectors. Yet these regions are important for virulence, as combined horizontal transfer of this region with chromosome 14 increases the aggressiveness of the recipient strain compared to strains that receive only chromosome 14 [41]. Moreover, our results show that these regions are present in all tomato-infecting strains, although possibly as a single copy. A remaining question is thus whether these regions on chromosome 3 and 6 contains one or several genes that are important for infection and whether the rest of these regions is 'hitchhiking'. In comparison, Fol strains that have lost large segments of chromosome 14, including *SIX6*, *SIX9* and *SIX11*, whose products have been found in the xylem sap of infected tomato plants [68], are not significantly less virulent than strains with a complete chromosome 14 in laboratory conditions [43]. This suggests that segments

559 of chromosome 14 may be hitchhiking with genes that are important for infection and
560 that reside on separate regions of chromosome 14, such as *SIX1*, *SIX3* and *SIX5*
561 [69,70]. On the other hand, the genes that may seem redundant for virulence in
562 conditions tested in bioassays may perform a role in other conditions. For example,
563 *SIX6*, found in multiple FOSC strains as well as in *Fusarium hostae* and
564 *Colletotrichum* species, is required for virulence of *F. oxysporum* f. sp. *radicis-*
565 *cucumerinum* towards cucumber only at relatively high temperatures [64]. Despite
566 the fact that genetic hitchhiking and incomplete functional characterizations obscure
567 relations between function and sequence divergence, we were still able to associate
568 physical - with functional clustering, as observed on fast-core regions on core
569 chromosomes, on fast-core chromosomes and on chromosome 14.
570
571 We not only found that pathogenicity chromosomes are specific to tomato-infecting
572 strains, we also found that horizontal chromosome transfer (HCT) most likely played
573 an important role in distributing the pathogenicity regions among different FOSC
574 strains. Population genetic studies in FOSC revealed that pathogenic strains group
575 with non-pathogenic ones [45,48,49,52-54,71-73], consistent with a scenario in which
576 a non-pathogenic strains receives at least part of a pathogenicity chromosome and
577 becomes pathogenic on a specific host. This is more likely to occur than switching
578 from one host to another because first of all, non-pathogens are more abundant and
579 secondly, pathogens may secrete effectors that enhance infection in one host but are
580 recognized in another. The importance of HCT in the emergence of new disease
581 depends, among other factors, on how long a pathogenicity chromosome can remain
582 intact in a population. The small pathogenicity chromosome in Fol007 is part of
583 chromosome 3 and chromosome 6 in Fol4287. Interestingly, no horizontal transfer of
584 chromosome 3 and 6 has been observed in the lab [63]. Although we found that
585 chromosome 14 and part of chromosomes 3 and 6 are present in all tomato-infecting
586 strains included in our dataset, whether these segments are present as a single

chromosome, multiple chromosomes or regions attached to other chromosomes, remains to be investigated. The presence of transposons impedes genome assembly from short reads, making it difficult to resolve this question with currently available genome sequences. However, we can conclude that the tomato-infecting strains in our dataset probably arose from a non-pathogenic strain after obtaining one or multiple pathogenicity chromosomes. We found no evidence for horizontal transfer of fast-core chromosomes in this comparison. Nevertheless, given the fact that genes on fast-core chromosomes are involved in infection and that we have observed transfer of core chromosomes in high-throughput experiments [63], we predict that horizontal transfer of core and fast-core chromosomes can also occur in natural settings. Co-transfer of fast-core chromosomes with accessory chromosomes may in some cases even enhance pathogenicity of the recipient strain.

HCT, or -an extreme case- hybridization, can have an enormous evolutionary impact. As we discussed above, it allows for fast spreading of fitness-enhancing pathogenicity chromosomes. Moreover, it provides a means to escape “Muller’s ratchet” in mainly asexual species by exchanging at least parts of core chromosomes. The transposons that reside on pathogenicity chromosomes can play a role in fast adaptation to host-resistance, [19,74] probably mostly by mediating rearrangements [5,8,75], thus potentially disrupting modular genomic organizations and creating novel chromosome segments. Moreover, transposons may disperse into core chromosomes [76] where the chance of harmful effects of transposon insertion is larger than on dispensable chromosomes. This interplay between beneficial and detrimental effects of transposons and mobile and/or accessory chromosomes on their host genomes has not yet been systematically investigated, even though it may strongly influence evolutionary outcomes, including the emergence of novel pathogens by host jumping. Population genetic studies and experimental evolution, combined with long read sequencing, may shed light on these important issues. We

615 demonstrated here how HCT events can be detected and visualized using whole
616 genome sequencing.
617
618 We found that none of the core chromosomes of Fol4287 are rich in transposons (Fig
619 S4), but the three smallest core chromosomes, termed 'fast-core', are distinct from
620 the other core chromosomes in terms of sequence divergence, synteny,
621 dispensability, and selected histone modifications. In that respect they are similar to
622 for example accessory regions in *Fusarium graminearum*, that have a higher SNP
623 density, are enriched in H3K27me3, yet are neither depleted in coding sequences
624 nor enriched for repeats [30]. There has been much debate on the influence of
625 chromatin states on *de novo* mutation rates. Constitutive heterochromatin may be
626 less accessible to the DNA repair machinery. Research in cancer cell lines and
627 mammalian germ line cells yielded conflicting results [77]. A recent study in
628 *Schizosaccharomyces pombe* revealed that in this fungus the DNA mismatch repair
629 machinery favours euchromatin. Similar to what we found here (Fig 5, Fig 6), the
630 authors observed a higher mutation rate in (in their case H3K9me2-enriched)
631 heterochromatin [78]. The fact that on core and fast-core chromosomes, H3K27me3
632 strongly associates with higher synonymous substitution rates (Fig 6A, B) and with
633 density of SNPs accumulated on an intermediate timescale, suggests that
634 H3K27me3 is correlated with or actively enhances mutation rates. However, this
635 effect appears insufficient to generate noticeable differences in density of SNPs
636 accumulated on very short timescales. Moreover, we do not find a strong correlation
637 between H3K27me3 and SNP density on accessory chromosomes, which
638 accumulated many more SNPs than fast-core chromosomes with comparable levels
639 of H3K27me3. Accessory chromosomes, particularly pathogenicity chromosomes of
640 strains that infect crops, may experience different selection pressures than core and
641 fast-core chromosomes, which could affect sequence evolution. Additional histone

642 modifications, possibly associated with a high density of repeats, may also further
643 influence mutation rates.
644
645 Genome organization is shaped by a combination of intrinsic molecular mechanisms
646 and natural selection. It is still not completely clear to what extent physical clustering
647 of genes in the same functional category has a selective advantage, such as robust
648 and stable co-regulation, efficient transferability of co-functional genes or increased
649 evolvability due to locally increased mutation rates (including rearrangements), e.g.
650 due to the presence of transposons or due to histone modifications that influence
651 DNA repair efficiency. We here defined two subcategories of the core and the
652 accessory genome and more fine-grained classifications are possible and probably
653 valuable. Our subclassifications reveal a hierarchical modular genome organization -
654 as is found with proteins and protein networks and is associated with evolvability [79-
655 81]. A more detailed classification of genome 'modules', identifying more than two or
656 four speeds, may help to disentangle the different molecular mechanisms, selection
657 on the organism level -fitness- and long-term selection on genome structure -
658 evolvability.
659
660 ## Materials and methods
661 ### Whole genome alignments.
662 We used nucmer from the MUMmer package (version 3.23) to align the genome of
663 Fol4287 to the 58 FOSC genomes [40] in our dataset (Table S1,
664 (https://github.com/LikeFokkens/whole_genome_alignments). We flagged –
665 maxmatch to allow for more than one alignment per region to be able to identify
666 duplicate regions, other than that we used default settings. We used show-coords
667 with default settings (no filtering) to obtain tab-separated files with aligned regions
668 and custom Python scripts to plot the percent identity (Figs 1, 2, and S2) or length

669 (Fig S3) of these aligned regions (https://github.com/LikeFokkens/genome-
670 wide_plots).
671
672 **Illumina sequencing of horizontal chromosome transfer strains.**
673 Two strains, Fo47-1A that obtained one chromosome and Fo47-2A that obtained two
674 chromosomes in the horizontal transfer experiment [41] were stored at -80 °C and
675 revitalized on potato dextrose agar (PDA) plates at 25 °C. An overgrown agar piece
676 was used to inoculate 100 ml NO_3-medium (0.17% yeast nitrogen base, 3% sucrose,
677 100 nM KNO_3). After 5 days at 25 °C shaking at 250 rpm, mycelium was harvested in
678 miracloth (Merck, pore size of 22-27 µM) and dried overnight in the freeze dryer.
679 Genomic DNA was isolated from freeze-dried mycelium as described in [36] and [40]
680 and sequenced on a Illumina Hiseq2000 sequencer at the Beijing Genomics Institute.
681
682 **Read densities of DNA sequencing reads**
683 We removed adapter sequences and low quality reads from reads obtained from
684 Illumina sequencing as described above with fastq-mcf with '-q 20' and a fasta file
685 with all Illumina adapter sequences. We then first mapped these reads to the
686 genome of Fo47 (the acceptor strain in the experiment), extracted the unmapped
687 reads and mapped these unmapped reads to the genome of Fol4287 to identify the
688 regions in Fol4287 that correspond to the chromosomes that were transferred from
689 Fol007 (the donor strain in the experiment). For each strain (Fo47-1A and Fo47-2A),
690 we mapped two libraries, one with insert size 170 bp, and one with insert size 500 bp.
691 We used Bowtie 2 (version 2.2.6) [82], to map reads to the Fo47 genome, applying
692 the following options: '--end-to-end --fr –reorder', for library with insert size 170 we
693 added the options '-I 140 -X 200' and for the library with insert size 500, we added
694 the options '-I 400 -X 600'. We extracted the unmapped reads using samtools view
695 (version 1.3.1), once with options '-Sbf 4 -F264' to extract unmapped reads for which
696 the mate was mapped to a samfile, once with options '-Sbf 8 -F260' to extract

mapped reads for which the mate was unmapped to a samfile, and once with options '-Sbf 12 -F256' to extract unmapped reads for which the mate was also unmapped to a samfile. We sorted the samfiles using 'samtools sort' and merged them into a single using 'samtools merge'. We used bedtools (version v2.24.0) 'bam2fastq' to convert the merged bamfile to fastq files. These fastq files were then mapped to the genome of Fol4287 using Bowtie2 with the same options as were used for mapping to the genome of Fo47. We removed putative PCR duplicates from the resulting bamfiles using Picard tools version 1.134 (http://broadinstitute.github.io/picard, MarkDuplicates) and used a custom Python script to calculate read densities for 10 kb windows and plot these on the genome (Fig 1B) (https://github.com/LikeFokkens/genome-wide_plots). Reads have been uploaded to ENA (accession PRJEB29294).

Repeats

Repeats were identified using RepeatMasker (Repbase Libraries 20140131) with '–species ascomycota' and otherwise default options. Repeats labeled as 'Low_complexity', Simple_repeat', 'rRNA' or 'Satellite/5S' were filtered out. Repeat densities were calculated for 10 kb non-overlapping sliding windows using 'bedtools coverage'.

Functional enrichment analyses

We generated lists of (upregulated) genes located on fast-core chromosomes and in pathogenicity regions. We used the FungiFun tool [83] to predict functional enrichment and depletion (https://elbe.hki-jena.de/fungifun/fungifun.php), we chose 'Fusarium oxysporum f. sp. lycopersici (strain 4287 / CBS 123668 / FGSC 9935 / NRRL34936)' as species, and set adjusted (~ after FDR correction) P-value < 0.001 as significance level (both under Advanced options). Accessory chromosomes typically contain genes with unknown function and less than 14% of genes in

pathogenicity regions are annotated in FungiFun (Table S5C). Therefore we relied on the annotation published in [62] for chromosome 14. In addition we used HmmerWeb version 2.24.1 to predict Pfam domains (hmmscan, default settings) and the presence of secretion signal peptides. For the upregulated genes on the fast-core chromosomes we predicted secretion signal peptides via the SignalP 4.1 server, using default settings.

Gene expression analyses

To compare *in vitro* and *in planta* gene expression levels, we used a dataset that has been described previously, hence we refer to [65] for an extensive description of methods. We used hypergeometric tests as implemented in scipy with Bonferroni correction to determine whether chromosomal regions are enriched for differentially expressed genes.

ChIP-seq experiments and data analyses.

Fusarium oxysporum f. sp. *lycopersici* 4287 (FGSC9935) was stored as a monoconidial culture at -80 °C and revitalized on potato dextrose agar (PDA) plates at 25 °C. An overgrown agar piece was used to inoculate 100 ml NO3-medium (0.17% yeast nitrogen base, 3% sucrose, 100 nM KNO_3). After 3-5 days at 25 °C shaking at 250 rpm, spores were harvested by filtering through miracloth (Merck, pore size of 22-27 µM). Spores were centrifuged for 10 minutes at 2000 rpm, resuspended in sterile MilliQ water and counted [40]. We inoculated 100 ml of fresh NO3-medium with 10^7 spores and cultured for 2 days at 25 °C and shaking at 250 rpm.

The ChIP-seq protocol [84] was similar to those used for *S. pombe* [85] and *Neurospora crassa* [86,87]. Here we only refer to important changes made for *F. oxysporum* strains. To the 2-day old cultures 20% formaldehyde (final concentration

753 ~0.8%) was added directly to the flask and incubated for 15 min at room temperature
754 while shaking at 100 rpm. After formaldehyde quenching [84], 100-200 mg of the cell
755 pellets were ground in liquid nitrogen. Too much cell mass will decrease ChIP
756 efficiency. Digestion with micrococcal nuclease to release mostly mono- or
757 dinucleosomes was as described [84] but should be optimized for different strains.
758 ChIP-seq libraries were generated as described [84], with Illumina TruSeq adapters,
759 libraries were size-selected between 200-400 bp, amplified with standard Illumina
760 TruSeq PCR primers. Libraries were sequenced on the OSU CGRB Illumina
761 HiSeq2500 machine.
762
763 Adapter sequences were removed from ChIP-seq reads and quality scores converted
764 to Sanger format as described in [32]. Reads were aligned to the genome of Fol4287
765 using 'bwa aln' [88]. Duplicate reads were removed with Picard tools version 1.134
766 (http://broadinstitute.github.io/picard, MarkDuplicates) (Table S2). We used RSEG
767 [89] to predict domains of enrichment using default settings and providing a bedfile
768 with gaps in the assembly (N's in the DNA sequence) as 'deadzones'. We obtained
769 two biological replicates for each experiment, one with relatively low and one with
770 relatively high sequencing depth (Table S2). We estimate the reproducibility of our
771 data by calculating the correlation of average read densities in 10 kb non-overlapping
772 sliding windows. We obtained moderate correlation coefficients, ranging from 0,59 to
773 0,66 (Kendall's Tau, Bonferroni-corrected P-value < 0.001) for H3K27me3 and
774 H3K4me2 respectively. Reads have been uploaded to GEO (accession GSE121283).
775
776 **Synonymous and nonsynonymous substitutions**
777 For Fol4287 genes based on alignments and trees with homologs in eight different
778 *Fusarium* species, others [24] calculated omega values (d_N/d_S ratios) and P-values
779 indicating whether a model including site-specific selection fits the data better than a
780 neutral model, for two different assumed distributions of d_N/d_S ratios: one in which

781 omega is either 0 or 1 (M1M2 in Fig 6), and one in which omega follows a beta
782 distribution between 0 and 1 (M7M8 in Fig 6), where models with sites under positive
783 selection (M2 and M8 respectively) have an additional category omega > 1. We
784 downloaded these data from their Supplemental Online Material. In addition, we
785 aligned Fol4287 protein sequences with their bidirectional best blast hits in the *F.*
786 *verticillioides* proteome with Clustal Omega, used PAL2NAL to infer codon
787 alignments and codeml to calculate d_N and d_S values based on these. We filtered out
788 all genes with d_N or d_S > 1 (i.e. genes with multiple substitutions).
789
790 **SNP calling**
791 Strains that belong to VCG030 are Fol002, Fol004, Fol007, Fol014, Fol018, Fol026,
792 Fol029, Fol038, Fol073 and Fol074. The 'clade III'- set consists of all strains in
793 VCG030 and Fol016, Fol096, Fol072, Fol075, FolMN25, FoMN14, Fom001, Forc016,
794 Forc024 and Forc031. Fol-CL25 also belongs to this clade, but no raw sequencing
795 reads were publicly available for this strain. Raw sequencing reads were filtered and
796 trimmed using fastq-mcf with '-q 20' and a fasta file with Illumina adapter sequences.
797 Trimmed reads were mapped to the genome sequence of Fol4287 using bowtie2
798 with '--end-to-end --fr –reorder' and '-I 100 -X 300' for strains with an estimated
799 fragment length of 200 (all strains in VCG030 except Fol007) and '-I 400 -X 600' for
800 other strains in clade III ((based on an estimated fragment length of 500). Bam files
801 were processed individually for each sample. The files were sorted and duplicate
802 reads were removed using Picard tools version 1.134
803 (http://broadinstitute.github.io/picard, MarkDuplicates). Variant calling was done with
804 the Genome Analysis ToolKit [90] using Haplotypecaller with '-ploidy 1'. Indels were
805 masked and SNP were sorted and filtered using VariantFilter with 'maskExtension 5',
806 'clusterSize 3', 'clusterWindowSize 10', 'MQRankSum < -17.5', 'QD < 1',
807 'ReadPosRankSum < -8.0', 'FS > 60.0', 'MQ < 30.0', 'SOR > 7', 'MQ0 > 3', 'QUAL <
808 30.0', 'DP > 400', 'DP < 10'.

809

810 **Availability of data and scripts.**

811 To facilitate reproducibility of this work and allow others to perform similar analyses,
812 the Python code and Ipython notebooks used to generate the Figures and to
813 calculate statistics in this manuscript and its supplemental online material are
814 available under GNU public license and can be downloaded from
815 https://github.com/LikeFokkens/FOSC_multi-speed-genome . Reads from ChIP-seq
816 experiments and the DNA sequencing of the two horizontal chromosome transfer
817 strains (Fo47-1A and Fo47-2A, Fig 1B) have been submitted to the GEO (accession
818 GSE121283) and the ENA (accession PRJEB29294) respectively.

819

820 **References**

821 1. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why
822 bigger can be better. Nat Rev Microbiol. 2012;10: 417-430.
823 2. Croll D, McDonald BA. The accessory genome as a cradle for adaptive evolution
824 in pathogens. PLoS Pathog. 2012;8: e1002608.
825 3. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous
826 pathogens: waltz with plants. Curr Opin Genet Dev. 2015;35: 57-65.
827 4. Crombach A, Hogeweg P. Chromosome rearrangements and the evolution of
828 genome structuring and adaptability. Mol Biol Evol. 2007;24: 1130-1139.
829 5. Zhang J, Yu C, Krishnaswamy L, Peterson T. Transposable elements as catalysts
830 for chromosome rearrangements. Methods Mol Biol. 2011;701: 315-326.
831 6. de Jonge R, Bolton MD, Kombrink A, van den Berg GC, Yadeta KA, Thomma BP.
832 Extensive chromosomal reshuffling drives evolution of virulence in an asexual
833 pathogen. Genome Res. 2013;23: 1271-1282.
834 7. Zhao C, Waalwijk C, de Wit PJ, Tang D, van der Lee T. Relocation of genes
835 generates non-conserved chromosomal segments in *Fusarium graminearum* that

836 show distinct and co-regulated gene expression patterns. BMC Genomics. 2014;15:
837 191-2164-15-191.
838 8. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GC, Wittenberg AH, et
839 al. Transposons passively and actively contribute to evolution of the two-speed
840 genome of a fungal pathogen. Genome Res. 2016;26: 1091-1100.
841 9. Plissonneau C, Sturchler A, Croll D. The Evolution of Orphan Regions in Genomes
842 of a Fungal Pathogen of Wheat. MBio. 2016;7: 10.1128/mBio.01231-16.
843 10. Dutheil JY, Mannhaupt G, Schweizer G, M K Sieber C, Munsterkotter M,
844 Guldener U, et al. A Tale of Genome Compartmentalization: The Evolution of
845 Virulence Clusters in Smut Fungi. Genome Biol Evol. 2016;8: 681-704.
846 11. Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al.
847 Host specialization of the blast fungus *Magnaporthe oryzae* is associated with
848 dynamic gain and loss of genes linked to transposable elements. BMC Genomics.
849 2016;17: 370-016-2690-6.
850 12. Hartmann FE, Sanchez-Vallet A, McDonald BA, Croll D. A fungal wheat pathogen
851 evolved host specialization by extensive chromosomal rearrangements. ISME J.
852 2017;11: 1189-1204.
853 13. Mehrabi R, Mirzadi Gohari A, Kema GHJ. Karyotype Variability in Plant-
854 Pathogenic Fungi. Annu Rev Phytopathol. 2017;55: 483-503.
855 14. Seidl MF, Thomma BPHJ. Transposable Elements Direct The Coevolution
856 between Plants and Microbes. Trends Genet. 2017;33: 842-851.
857 15. Moller M, Stukenbrock EH. Evolution and genome architecture in fungal plant
858 pathogens. Nat Rev Microbiol. 2017;15: 771.
859 16. Kang S, Lebrun MH, Farrall L, Valent B. Gain of virulence caused by insertion of
860 a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. Mol Plant Microbe
861 Interact. 2001;14: 671-674.
862 17. Inami K, Yoshioka-Akiyama C, Morita Y, Yamasaki M, Teraoka T, Arie T. A
863 genetic mechanism for emergence of races *in Fusarium oxysporum* f. sp. *lycopersici*:

inactivation of avirulence gene AVR1 by transposon insertion. PLoS One. 2012;7: e44101.

18. Kashiwa T, Suzuki T, Sato A, Akai K, Teraoka T, Komatsu K, et al. A new biotype of *Fusarium oxysporum* f. sp. *lycopersici* race 2 emerged by a transposon-driven mutation of avirulence gene AVR1. FEMS Microbiol Lett. 2016;363: 10.1093/femsle/fnw132. Epub 2016 May 17.

19. Biju VC, Fokkens L, Houterman PM, Rep M, Cornelissen BJ. Multiple Evolutionary Trajectories Have Led to the Emergence of Races in *Fusarium oxysporum* f. sp. *lycopersici*. Appl Environ Microbiol. 2017;83: 10.1128/AEM.02548-16. Print 2017 Feb 15.

20. Rep M, Meijer M, Houterman PM, van der Does HC, Cornelissen BJ. *Fusarium oxysporum* evades I-3-mediated resistance without altering the matching avirulence gene. Mol Plant Microbe Interact. 2005;18: 15-23.

21. Parlange F, Daverdin G, Fudal I, Kuhn ML, Balesdent MH, Blaise F, et al. *Leptosphaeria maculans* avirulence gene AvrLm4-7 confers a dual recognition specificity by the Rlm4 and Rlm7 resistance genes of oilseed rape, and circumvents Rlm4-mediated recognition through a single amino acid change. Mol Microbiol. 2009;71: 851-863.

22. Ghanbarnia K, Fudal I, Larkan NJ, Links MG, Balesdent MH, Profotova B, et al. Rapid identification of the *Leptosphaeria maculans* avirulence gene AvrLm2 using an intraspecific comparative genomics approach. Mol Plant Pathol. 2015;16: 699-709.

23. Dong S, Stam R, Cano LM, Song J, Sklenar J, Yoshida K, et al. Effector specialization in a lineage of the Irish potato famine pathogen. Science. 2014;343: 552-555.

24. Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, et al. Genome-Wide Analysis in Three Fusarium Pathogens Identifies Rapidly Evolving Chromosomes and Genes Associated with Pathogenicity. Genome Biol Evol. 2015;7: 1613-1627.

892 25. Grandaubert J, Dutheil JY, Stukenbrock EH. The genomic rate of adaptation in
893 the fungal wheat pathogen *Zymoseptoria tritici*. bioRxiv. 2017.
894 26. Stergiopoulos I, de Wit PJ. Fungal effector proteins. Annu Rev Phytopathol.
895 2009;47: 233-263.
896 27. Fudal I, Ross S, Brun H, Besnard AL, Ermel M, Kuhn ML, et al. Repeat-induced
897 point mutation (RIP) as an alternative mechanism of evolution toward virulence in
898 *Leptosphaeria maculans*. Mol Plant Microbe Interact. 2009;22: 932-941.
899 28. Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, et al.
900 Effector diversification within compartments of the *Leptosphaeria maculans* genome
901 affected by Repeat-Induced Point mutations. Nat Commun. 2011;2: 202.
902 29. Stukenbrock EH, Dutheil JY. Fine-Scale Recombination Maps of Fungal Plant
903 Pathogens Reveal Dynamic Recombination Landscapes and Intragenic Hotspots.
904 Genetics. 2017.
905 30. Wang Q, Jiang C, Wang C, Chen C, Xu JR, Liu H. Characterization of the Two-
906 Speed Subgenomes of *Fusarium graminearum* Reveals the Fast-Speed Subgenome
907 Specialized for Adaption and Infection. Front Plant Sci. 2017;8: 140.
908 31. Galazka JM, Freitag M. Variability of chromosome structure in pathogenic fungi--
909 of 'ends and odds'. Curr Opin Microbiol. 2014;20: 19-26.
910 32. Connolly LR, Smith KM, Freitag M. The *Fusarium graminearum* histone H3 K27
911 methyltransferase KMT6 regulates development and expression of secondary
912 metabolite gene clusters. PLoS Genet. 2013;9: e1003916.
913 33. Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, et al.
914 Epigenetic control of effector gene expression in the plant pathogenic fungus
915 *Leptosphaeria maculans*. PLoS Genet. 2014;10: e1004227.
916 34. Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, et al.
917 Histone modifications rather than the novel regional centromeres of *Zymoseptoria*
918 *tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin. 2015;8:
919 41-015-0033-5. eCollection 2015.

920 35. Soyer JL, Rouxel T, Fudal I. Chromatin-based control of effector gene expression
921 in plant-associated fungi. Curr Opin Plant Biol. 2015;26: 51-56.
922 36. Michielse CB, Rep M. Pathogen profile update: *Fusarium oxysporum*. Mol Plant
923 Pathol. 2009;10: 311-324.
924 37. Ma LJ, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, et al. Fusarium
925 pathogenomics. Annu Rev Microbiol. 2013;67: 399-416.
926 38. Armstrong GM, Armstrong JK. Formae speciales and races of *Fusarium*
927 *oxysporum* causing wilt diseases. In: Nelson, P.E., Toussoun, T.A., and Cook, R.J.,
928 editor. Fusarium: Diseases, Biology, and Taxonomy. ; 1981. pp. 391-399.
929 39. Lievens B, Rep M, Thomma BPHJ. Recent developments in the molecular
930 discrimination of formae speciales of *Fusarium oxysporum*. Pest Manag Sci.
931 2008;64: 781-788.
932 40. van Dam P, Fokkens L, Schmidt SM, Linmans JH, Kistler HC, Ma LJ, et al.
933 Effector profiles distinguish formae speciales of *Fusarium oxysporum*. Environ
934 Microbiol. 2016;18: 4087-4102.
935 41. Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A,
936 et al. Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium.
937 Nature. 2010;464: 367-373.
938 42. Williams AH, Sharma M, Thatcher LF, Azam S, Hane JK, Sperschneider J, et al.
939 Comparative genomics and prediction of conditionally dispensable sequences in
940 legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of
941 candidate effectors. BMC Genomics. 2016;17: 191-016-2486-8.
942 43. Vlaardingerbroek I, Beerens B, Schmidt SM, Cornelissen BJ, Rep M.
943 Dispensable chromosomes in *Fusarium oxysporum* f. sp. *lycopersici*. Mol Plant
944 Pathol. 2016;17: 1455-1466.
945 44. O'Donnell K, Kistler HC, Cigelnik E, Ploetz RC. Multiple evolutionary origins of
946 the fungus causing Panama disease of banana: concordant evidence from nuclear
947 and mitochondrial gene genealogies. Proc Natl Acad Sci U S A. 1998;95: 2044-2049.

948 45. Alves-Santos FM, Benito EP, Eslava AP, Diaz-Minguez JM. Genetic diversity of
949 *Fusarium oxysporum* strains from common bean fields in Spain. Appl Environ
950 Microbiol. 1999;65: 3335-3340.
951 46. Baayen RP, O'Donnell K, Bonants PJ, Cigelnik E, Kroon LP, Roebroeck EJ, et al.
952 Gene Genealogies and AFLP Analyses in the *Fusarium oxysporum* Complex Identify
953 Monophyletic and Nonmonophyletic Formae Speciales Causing Wilt and Rot Disease.
954 Phytopathology. 2000;90: 891-900.
955 47. Skovgaard K, Nirenberg HI, O'Donnell K, Rosendahl S. Evolution of *Fusarium*
956 *oxysporum* f. sp. *vasinfectum* Races Inferred from Multigene Genealogies.
957 Phytopathology. 2001;91: 1231-1237.
958 48. Abo K, Klein KK, Edel-Hermann V, Gautheron N, Traore D, Steinberg C. High
959 Genetic Diversity Among Strains of *Fusarium oxysporum* f. sp. *vasinfectum* from
960 Cotton in Ivory Coast. Phytopathology. 2005;95: 1391-1396.
961 49. Wong JY, Jeffries P. Diversity of pathogenic Fusarium populations associated
962 with asparagus roots in decline soils in Spain and the UK. Plant Pathol. 2006;55:
963 331-342.
964 50. Mbofung GY, Hong SG, Pryor BM. Phylogeny of *Fusarium oxysporum* f. sp.
965 *lactucae* Inferred from Mitochondrial Small Subunit, Elongation Factor 1-alpha, and
966 Nuclear Ribosomal Intergenic Spacer Sequence Data. Phytopathology. 2007;97: 87-
967 98.
968 51. Lievens B, Claes L, Vakalounakis DJ, Vanachter AC, Thomma BP. A robust
969 identification and detection assay to discriminate the cucumber pathogens *Fusarium*
970 *oxysporum* f. sp. *cucumerinum* and f. sp. *radicis-cucumerinum*. Environ Microbiol.
971 2007;9: 2145-2161.
972 52. Ellis ML, Jimenez DRC, Leandro LF, Munkvold GP. Genotypic and Phenotypic
973 Characterization of Fungi in the *Fusarium oxysporum* Species Complex from
974 Soybean Roots. Phytopathology. 2014;104: 1329-1339.

975 53. Pinaria AG, Laurence MH, Burgess LW, Liew ECY. Phylogeny and origin of
976 *Fusarium oxysporum* f. sp. *vanillae* in Indonesia. Plant Pathol. 2015;64: 1358-1365.
977 54. Nirmaladevi D, Venkataramana M, Srivastava RK, Uppalapati SR, Gupta VK, Yli-
978 Mattila T, et al. Molecular phylogeny, pathogenicity and toxigenicity of *Fusarium*
979 *oxysporum* f. sp. *lycopersici*. Sci Rep. 2016;6: 21367.
980 55. Epstein L, Kaur S, Chang PL, Carrasquilla-Garcia N, Lyu G, Cook DR, et al.
981 Races of the Celery Pathogen *Fusarium oxysporum* f. sp. *apii* Are Polyphyletic.
982 Phytopathology. 2017;107: 463-473.
983 56. Zhang Y, Ma LJ. Deciphering Pathogenicity of *Fusarium oxysporum* From a
984 Phylogenomics Perspective. Adv Genet. 2017;100: 179-209.
985 57. Laurence MH, Summerell BA, Liew ECY. *Fusarium oxysporum* f. sp. *canariensis*:
986 evidence for horizontal gene transfer of putative pathogenicity genes. Plant
987 Pathology. 2015;64: 1068-1075.
988 58. van Dam P, Rep M. The Distribution of Miniature Impala Elements and SIX
989 Genes in the Fusarium Genus is Suggestive of Horizontal Gene Transfer. J Mol Evol.
990 2017;85: 14-25.
991 59. Czislowski E, Fraser-Smith S, Zander M, O'Neill WT, Meldrum RA, Tran-Nguyen
992 LTT, et al. Investigation of the diversity of effector genes in the banana pathogen,
993 *Fusarium oxysporum* f. sp. *cubense*, reveals evidence of horizontal gene transfer.
994 Mol Plant Pathol. 2017.
995 60. Rosewich UL, Kistler HC. Role of Horizontal Gene Transfer in the Evolution of
996 Fungi. Annu Rev Phytopathol. 2000;38: 325-363.
997 61. Mehrabi R, Bahkali AH, Abd-Elsalam KA, Moslem M, Ben M'barek S, Gohari AM,
998 et al. Horizontal gene and chromosome transfer in plant pathogenic fungi affecting
999 host range. FEMS Microbiol Rev. 2011;35: 542-554.
1000 62. Schmidt SM, Houterman PM, Schreiver I, Ma L, Amyotte S, Chellappan B, et al.
1001 MITEs in the promoters of effector genes allow prediction of novel virulence genes in
1002 *Fusarium oxysporum*. BMC Genomics. 2013;14: 119-2164-14-119.

1003 63. Vlaardingerbroek I, Beerens B, Rose L, Fokkens L, Cornelissen BJ, Rep M.
1004 Exchange of core chromosomes and horizontal transfer of lineage-specific
1005 chromosomes *in Fusarium oxysporum*. Environ Microbiol. 2016;18: 3702-3713.
1006 64. van Dam P, Fokkens L, Ayukawa Y, van der Gragt M, Ter Horst A, Brankovics B,
1007 et al. A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of
1008 multiple cucurbit species. Sci Rep. 2017;7: 9042-017-07995-y.
1009 65. van der Does HC, Fokkens L, Yang A, Schmidt SM, Langereis L, Lukasiewicz JM,
1010 et al. Transcription Factors Encoded on Core and Accessory Chromosomes of
1011 *Fusarium oxysporum* Induce Expression of Effector Genes. PLoS Genet. 2016;12:
1012 e1006401.
1013 66. Studt L, Rosler SM, Burkhardt I, Arndt B, Freitag M, Humpf HU, et al. Knock-
1014 down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of
1015 cryptic and otherwise silent secondary metabolite gene clusters in *Fusarium fujikuroi*.
1016 Environ Microbiol. 2016;18: 4037-4054.
1017 67. Takken F, Rep M. The arms race between tomato and *Fusarium oxysporum*. Mol
1018 Plant Pathol. 2010;11: 309-314.
1019 68. Houterman PM, Speijer D, Dekker HL, DE Koster CG, Cornelissen BJ, Rep M.
1020 The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. Mol
1021 Plant Pathol. 2007;8: 215-221.
1022 69. Rep M, van der Does HC, Meijer M, van Wijk R, Houterman PM, Dekker HL, et al.
1023 A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of
1024 xylem vessels is required for I-3-mediated resistance in tomato. Mol Microbiol.
1025 2004;53: 1373-1383.
1026 70. Ma L, Houterman PM, Gawehns F, Cao L, Sillo F, Richter H, et al. The AVR2-
1027 SIX5 gene pair is required to activate I-2-mediated immunity in tomato. New Phytol.
1028 2015;208: 507-518.
1029 71. Kistler HC. Genetic Diversity in the Plant-Pathogenic Fungus *Fusarium*
1030 *oxysporum.* Phytopathology. 1997;87: 474-479.

1031 72. Hannachi I, Poli A, Rezgui S, Prassad RD, Cherif M. Genetic and phenotypic
1032 differences of *Fusarium oxysporum* f. sp. *citri* isolated from sweet orange and
1033 tangerine. Eur J Plant Pathol. 2015;142: 269-280.
1034 73. Imazaki I, Kadota I. Molecular phylogeny and diversity of Fusarium endophytes
1035 isolated from tomato stems. FEMS Microbiol Ecol. 2015;91: fiv098-fiv098.
1036 74. Grandaubert J, Lowe RG, Soyer JL, Schoch CL, Van de Wouw AP, Fudal I, et al.
1037 Transposable element-assisted evolution and adaptation to host plant within the
1038 *Leptosphaeria maculans-Leptosphaeria biglobosa* species complex of fungal
1039 pathogens. BMC Genomics. 2014;15: 891-2164-15-891.
1040 75. Gray YH. It takes two transposons to tango: transposable-element-mediated
1041 chromosomal rearrangements. Trends Genet. 2000;16: 461-468.
1042 76. Vanheule A, Audenaert K, Warris S, van de Geest H, Schijlen E, Hofte M, et al.
1043 Living apart together: crosstalk between the core and supernumerary genomes in a
1044 fungal plant pathogen. BMC Genomics. 2016;17: 670-016-2941-6.
1045 77. Makova KD, Hardison RC. The effects of chromatin organization on variation in
1046 mutation rates in the genome. Nat Rev Genet. 2015;16: 213-223.
1047 78. Sun L, Zhang Y, Zhang Z, Zheng Y, Du L, Zhu B. Preferential Protection of
1048 Genetic Fidelity within Open Chromatin by the Mismatch Repair Machinery. J Biol
1049 Chem. 2016;291: 17692-17705.
1050 79. Wagner GP, Pavlicev M, Cheverud JM. The road to modularity. Nat Rev Genet.
1051 2007;8: 921-931.
1052 80. Rorick MM, Wagner GP. Protein structural modularity and robustness are
1053 associated with evolvability. Genome Biol Evol. 2011;3: 456-475.
1054 81. Ten Tusscher KH, Hogeweg P. Evolution of networks for body plan patterning;
1055 interplay of modularity, robustness and evolvability. PLoS Comput Biol. 2011;7:
1056 e1002208.
1057 82. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat
1058 Methods. 2012;9: 357-359.

1059 83. Priebe S, Kreisel C, Horn F, Guthke R, Linde J. FungiFun2: a comprehensive
1060 online resource for systematic analysis of gene lists from fungal species.
1061 Bioinformatics. 2015;31: 445-446.
1062 84. Soyer JL, Moller M, Schotanus K, Connolly LR, Galazka JM, Freitag M, et al.
1063 Chromatin analyses of *Zymoseptoria tritici*: Methods for chromatin
1064 immunoprecipitation followed by high-throughput sequencing (ChIP-seq). Fungal
1065 Genet Biol. 2015;79: 63-70.
1066 85. Nakayama J, Klar AJ, Grewal SI. A chromodomain protein, Swi6, performs
1067 imprinting functions in fission yeast during mitosis and meiosis. Cell. 2000;101: 307-
1068 317.
1069 86. Tamaru H, Zhang X, McMillen D, Singh PB, Nakayama J, Grewal SI, et al.
1070 Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora*
1071 *crassa*. Nat Genet. 2003;34: 75-79.
1072 87. Honda S, Lewis ZA, Shimada K, Fischle W, Sack R, Selker EU. Heterochromatin
1073 protein 1 forms distinct complexes to direct histone deacetylation and DNA
1074 methylation. Nat Struct Mol Biol. 2012;19: 471-7, S1.
1075 88. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
1076 transform. Bioinformatics. 2009;25: 1754-1760.
1077 89. Song Q, Smith AD. Identifying dispersed epigenomic domains from ChIP-Seq
1078 data. Bioinformatics. 2011;27: 870-871.
1079 90. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al.
1080 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
1081 generation DNA sequencing data. Genome Res. 2010;20: 1297-1303.
1082

Figure captions

Fig 1. Lineage-specific and pathogenicity regions in the Fol4287 accessory genome.

A. Horizontal bars -colored according to the percent identity in the alignment- indicate presence of Fol4287 accessory regions 1b and 2b and accessory chromosomes 3, 6, 14 and 15 in 23 other *Fusarium oxysporum* isolates. Leaf nodes in the phylogenetic tree are colored according to *forma specialis*, isolates causing wilting symptoms are represented with a circle, those that cause root rot are represented with squares and those that are non-pathogenic do not have a shape or color. Only alignments that span more than 1 kb and are more than 90% identical are included. Chromosomal regions 1b (Supercontig 2.27) and 2b (Supercontig 2.31), part of chromosome 3, part of chromosome 6 and chromosome 15 are mosty absent or present with relatively low sequence similarity outside the clonal line of Fol4287. In contrast, chromosome 14, part of chromosome 3 and part of chromosome 6 are present with ~100% sequence identity in all tomato-infecting isolates that we queried. **B**. Two strains that respectively received one and two chromosomes in a horizontal transfer experiment involving Fol007 as chromosome donor and Fo47 as recipient, had been sequenced previously [62]. We mapped sequencing reads obtained from these strains on the genome of Fol4287. The density of mapped reads from the strain that received chromosome 14 is depicted in blue. The density of mapped reads from the strain that received a small chromosome in addition to chromosome 14 and is more virulent, is depicted in red. These read densities reveal that that the pathogenicity regions on chromosome 3 and 6 correspond to a small pathogenicity chromosome in Fol007. Read densities in regions that have duplicated or triplicated in Fol4287 are ½ and 1/3 lower respectively than in non-duplicated Fol4287 regions (indicated with green lines).

Fig 2. Core and fast-core chromosomes in the Fol4287 genome.

Horizontal bars -colored according to the % identity in the alignment- indicate presence of Fol4287 core chromosomes 4,5,7,8,11,12, and 13 in 23 other *Fusarium oxysporum* isolates. Leaf nodes in the phylogenetic tree are colored according to *forma specialis*, isolates causing wilting symptoms are represented with a circle, those that cause root rot are represented with squares and those that are non-pathogenic do not have a shape or color. Only alignments that span more than 1 kb and are more than 90% similar are included. Within the Fol4287 clonal line, all chromosomes are mostly 100% identical, while when compared to other, more distant FOSC isolates, the three smallest core chromosomes, 11, 12 and 13, are more divergent than other core chromosomes (see also Figure S1). We denote these three chromosomes as fast-core chromosomes.

Fig 3. Fraction of up- and downregulated genes per chromosome.

The bars represent the fraction of genes that is upregulated (green) or downregulated (red) *in planta* compared to *in vitro* per chromosome and per chromosome category (boxes with black lines). Significant enrichment (P-value < 0.001) of up- or downregulated genes with respect to the rest of the genome is indicated with asterisk signs. Lineage-specific parts of core (Chr. 1 and 2) or accessory (Chr. 3 and 6) chromosomes that are lineage-specific are denoted with a # on the x-axis.

Fig 4. Core chromosomes are enriched for H3K4me2 and fast-core and accessory chromosomes for H3K27me3.

We mapped ChIP-seq reads to the genome of Fol4287 and identified domains of enrichment (DOE). For each chromosome or chromosomal region (x-axis) we plot how much is covered in a DOE (top panels) and the size-distribution of DOEs (bottom panels, violin plots). Lineage-specific parts of core (Chr. 1 and 2) or accessory (Chr. 3 and 6) chromosomes that are lineage-specific are denoted with a #

on the x-axis. The regions on chromosome 3 and on chromosome 6 that are associated with pathogenicity on tomato are denoted with a ^{P}. Bars with thick lines represent the coverage per chromosome category (core, fast-core, lineage specific or pathogenicity) rather than per chromosome. A. Coverage and size distribution of DOE of H3K4me2. For most core chromosomes more than 50% is enriched for H3K4me3, while for the fast-core and accessory chromosomes, this is less than 20%. The DOE of H3K2me2 are larger on the core chromosomes than in the other regions. B. Coverage and size distribution of DOE of H3K27me3. For most core chromosomes less than 50% is enriched for H3K37me3, while for the fast-core chromosomes more than 80% is covered in H3K27me3 and the lineage-specific and accessory chromosomes are more enriched in H3K27me3 than the core chromosomes. In general, the size of the H3K27me3 DOEs is larger than that of H3K4me2 DOEs (note that y-axis is in log-scale).

Fig 5. H3K27me3-enriched regions have higher levels of sequence divergence, independent of repeat density.

These three examples, one core, one fast-core and one accessory chromosome, demonstrate how presence of H3k27me3 (panel 2) coincides with differential expression (panel 1), elevated levels of sequence divergence (panel 3-5), but not necessarily with differences in either gene- or repeat density. The genes on core chromosome 4 that are differentially expressed during infection (panel 1, blue squares) are located mostly in the sub-telomeric regions and one central region. These regions are enriched in H3K27me3 (panel two, orange) and depleted in H3K4me3 (panel two, green). The fast-core chromosome has many differentially expressed genes, most of which are upregulated during infection. There are no large regions that are either enriched or depleted in differentially expressed genes. The same holds true for H3K27me3, which is relatively equally distributed along the chromosome. More or less the same holds true for chromosome 14, the

pathogenicity chromosome and the only accessory chromosome that is enriched for genes that are upregulated during infection. Here also H3K27me3 is relatively uniformly distributed along the chromosome: unlike core chromosomes, H3K27me3-ated regions are not interrupted by regions with high levels of H3K4me2. When looking at sequence divergence on a very short timescale (panel 3, density of SNPs called in strains that belong to VCG030) there are no distinct differences within and between the core and fast-core chromosomes, while chromosome 14 clearly has a higher SNP density (see Figure S14 for more detail). In contrast, on longer timescales (panel 4, density of SNPs called in strains that belong to clade III, panel 5, d_N (black) and d_S (blue) based on bidirectional best hits with sister species *F. verticillioides*), we can clearly see that regions that are enriched in H3K27me3 on the core chromosome correspond with regions with higher d_N and d_S (see also Figure S14 and Figure S15) and a higher SNP density. Fast-core chromosomes resemble

Sub-telomeric core regions. Note that gene- and repeat density are relatively uniform on the core and fast-core chromosomes (see also Figure S4). The SNP density of accessory chromosomes cannot directly be compared to the core and fast-core chromosomes because they are lost in many strains in Clade III and in *F. verticillioides*. The accessory chromosomes are depleted in genes and enriched in repeats.

Figure 6. Clustering of genome characteristics.

We calculated the following genome characteristics for 6186 10 kb non-overlapping windows across the entire Fol4287 genome: **Genes (ORFs**): the number of bp that are part of a coding region divided by the window size, **Repeats**: the number of bp that are part of a repeat region divided by the window size, **RPKM *in planta***: mean (over 3 replicates) RNA-seq read counts for the *in planta* experiment, **RPKM *in vitro***: mean (over 3 replicates) RNA-seq read counts for the *in vitro* experiment, **Change in GE**: log2 fold change in gene expression *in planta* / *in vitro*, **H3K4me2 (a) and (b)**: average read density for the two replicates of the H3K4me2 ChIP-seq experiment, **H3K27me3 (a) and (b)**: average read density for the two replicates of the H3K27me3 ChIP-seq experiment, **dN**: average number of nonsynonymous substitutions per nonsynonymous site dN, **dS**: the average number of synonymous substitutions per synonymous site dS, **dN/dS**: average ratio between synonymous and nonsynonymous substitutions, **SNPs VCG030**: density of SNPs in VCG030 with respect to Fol4287, **SNPs Clade III**: density of SNPs in clade III with respect to Fol4287. Spearman correlation coefficients (R) are depicted in a heatmap, genome characteristics were clustered with average linkage using 1-R as a distance measure. A. Clustering based on core chromosomes. The clustering shows two regimes: one gene-rich, highly expressed, H3K4me2-ated and one repeat-rich, H3K27me3-ated with high levels of sequence divergence. B. Clustering based on fast-core chromosomes (11, 12, 13). As with core chromosomes, the clustering shows two

regimes: one gene-rich, highly expressed, H3K4me2-ated and one repeat-rich, H3K27me3-ated with high levels of sequence divergence, albeit less pronounced as correlation coefficient tend to be lower. C. Clustering based on accessory chromosomes. Unlike the core and fast-core chromosomes, levels of sequence divergence correlate with gene density and expression level. However, this may be an artefact from the SNP calling process, in which reads that map to multiple locations -e.g. reads that map to transposons- are excluded. Due to lack of data points, dN, dS and dN/dS were excluded.

Supporting information captions

Figure S1. Whole genome alignment of Fol4287 with Fom001.

The whole genome alignment of tomato-infecting strain Fol4287 with melon-infecting strain Fom001 is represented here in a dotplot. Lines are coloured according to percent identity of the aligned segments, only alignments that are more than 90% identical and span more than 1 kb are included. Thin black vertical lines indicate Fol4287 chromosomes, grey, semi-transparent lines indicate scaffolds. Unpositioned scaffolds in Fol4287 are not included. The Fol4287 core chromosomes (1,2,4,5,7,8,9-13) are almost 99-100% identical in Fom001. With the exception of one translocation between chromosome 2 and chromosome 10, synteny is conserved in core chromosomes. In contrast, the accessory chromosomes are mostly absent. We expect that differences in host preference are largely determined by these accessory chromosomes. Chromosomal regions 1b, part of chromosome 15 and especially 2b are absent in sister species *F. verticillioides*, but largely present in Fom001 with high levels of sequence similarity. In contrast, chromsome 3, 6 and 14 are absent in Fom001.

Figure S2. Presence/absence -coloured by % identity- of Fol4287 chromosome sequences in the genomes of 58 other *Fo* strains.

The phylogenetic tree on the left was inferred using maximum likelihood from a concatenated alignment of 1194 core genes as described in van Dam et al. (Env. Microb. 2016). Leaf nodes are colored according to forma specialis; red: tomato, light-green: melon, green: cucumber, grey: cotton, petrol: watermelon, yellow: banana, purple: brassicaceae, brown: pea. Leaf nodes are shaped according to disease symptoms; round: wilting, square: root rot. Forc strains (green squares) are colored green but are pathogenic on cucumber, melon and watermelon. Strains that are not pathogenic on plants (to our best knowledge) have no shape, Fo47 and

FoMN14 are non-pathogenic, FOSC-3a is pathogenic on immunocompromised humans. This plot indicates only presence of sequences and are not informative on synteny (see Figure S3 for plots colored according to alignment length): this plot only shows which segments are present, not whether they also accur on the same configuration/on the same chromosome. In Figure S2A and S2B, only alignments that span at least 1000 bp in the query genome and are at least 90% identical are included.

A. Comparison of core chromosomes, i.e. chromosomes that are largely present with high sequence similarity (>97% identity) in all strains. Part of chromosome 1 and chromosome 2 are considered accessory regions. Levels of sequence similarity correspond to the phylogeny. The level of sequence similarity drops in the subtelomeric regions. Chromosome 4 has a region in the middle with lower levels of sequence similarity, this region corresponds to the breakpoint of a chromosomal rearrangement when compared to *F. verticillioides*.

B. Comparison of fast-core and accessory chromosomes. The three smallest core chromosomes differ markedly from the other 8 core chromosomes depicted in Figure S2A. The percent identity in the alignments is lower, the difference is small but very consistent. Moreover, we observed more large-scale deletions in these chromosomes. The accessory chromosomes are present in only in a small subset of genomes. Some sequences are present with low sequence similarity (~90%, depicted in blue), these could correspond to transposons. Alignments of chromosomal regions 1b and 2b and chromosome 15 with the Fom001 genome have higher sequence identity and synteny than of chromosome 3 and 6 (Figure 1, Figure S1). Moreover, chromosome 3 and 6 both contain a region that has been lost in strains Fol029, Fol018, Fol074 and Fol038. Chromosome 15 and chromosomal region 1b are absent Fol014, Fol029 and Fol018 and part of chromosome 14 has been lost in Fol018. Chromosome 1b, 2b, 3, 6, 15 and many of the unpositioned scaffolds hardly occur outside the clonal lineage of Fol4287. Some of the

unpositioned scaffolds have a similar color pattern as core chromosomes, suggesting they are part of core chromosomes but could not be placed there due to lack of resolution in the optical map that was constructed for the reference assembly of Fol4287. Chromosome 14 and part of chromosome 3 and 6 clearly show hallmarks of large-scale horizontal transfer but there are also indications of smaller scale putative transfer events, e.g. part of chromosome 1b that is present in brassicaceae-infecting strains (indicated with purple circles in the phylogenetic tree). Part of chromosome 15 is also present in other strains, but with normal sequence similarity, suggesting that this region is lost in many strains rather than horizontally transferred. Comparisons to outgroup species are needed to confirm this.

C. Presence-absence of core, fast-core and accessory chromosomes with more lenient cutoffs: all alignments that are more than 80% identical in sequence and span more than 100 basepairs are included. The patterns we observed above are not dependent on specific cut-offs, as we observe then with more lenient cut-offs as well.

Figure S3. Presence/absence -coloured by synteny- of Fol4287 genome sequences in the genomes of 58 other *Fo* isolates.

The phylogenetic tree on the left was inferred using maximum likelihood from a concatenated alignment of 1194 core genes as described in van Dam et al. (Env. Microb. 2016). Leaf nodes are colored according to forma specialis; red: tomato, light-green: melon, green: cucumber, grey: cotton, petrol: watermelon, yellow: banana, purple: brassicaceae, brown: pea. Leaf nodes are shaped according to disease symptoms; round: wilting, square: root rot. Forc isolates (green squares) are colored green but are pathogenic on cucumebr, melon and watermelon. Isolates that are not pathogenic on plants (to our best knowledge) have no shape, Fo47 and FoMN14 are non-pathogenic, FOSC-3a is pathogenic on immunocompromised humansIn Figure S3A and S3B, only alignments that span at least 1000 bp in the query genome and are at least 90% identical are included.

A. Comparison of core chromosomes. Part of chromosome 1 and chromosome 2 are considered accessory regions. Within the clonal line of Fol4287 and with Fom001, alignments can span as long as and longer than 50 Mb (plotted in yellow). As with sequence similarity, synteny levels drop at the subtelomeric regions and in the middle region of chromosome 4 that corresponds to the breakpoint of a chromosomal rearrangement when compared to *F. verticillioides*. Fol4287 self-alignment does not produce only alignments of > 50 Mb in length because the assembly contains gaps. Chromosomal region 2B is relatively fragmented (contains > 2 times the number of gaps as e.g. chromosomal region 1B, or Supercontig_2.25 that is of comparable size and is located on chromosome 3) which is reflected in the low level of synteny in the Fol4287 self-alignment.

B. Comparison of fast-core and accessory chromosomal regions. The three smallest core chromosomes differ from the other 8 core chromosomes depicted in Figure S3A, synteny conservation is less conserved in more distant isolates. Part of chromosome 13 is not syntenic to Fom001, it aligns to a number of relatively small contigs. Notably, this region is lost in Fom013. The accessory chromosomes are present in only in a small subset of chromosomes. While for chromosome 14 sequences are conserved in tomato-infecting isolate, synteny is not, in contrast to the region on chromosome 3 and 6 that is shared between tomato-infecting isolates, where we do observe conserved synteny.

C. Presence-absence of core, fast-core and accessory chromosomes with more lenient cutoffs: all alignments that are more than 80% identical in sequence and span more than 100 basepairs are included. The patterns we observed above are not dependent on specific cut-offs, as we observe then with more lenient cut-offs as well.

Figure S4. Gene and repeat density.

For 10 kb non-overlapping windows, we here plot the fraction of base pairs that is part of a gene (blue) or a predicted transposable element (grey). The accessory

chromosomes and chromosomal regions are rich in repeats compared to the core chromosomes, including the fast-core chromosomes. Chromosomes are scaled according to size.

Figure S5. Fast-core and accessory chromosomes have lower gene expression levels.

The expression level in reads per kb per million mapped reads (RPKM), averaged over three biological replicates. Per chromosome, there are two panels: the top panel shows *in vitro* gene expression in black dots and the bottom panel *in planta* gene expression in blue dots. The x-axis show the position of the gene on the genome, the y-axis the RPKM. Note that the y-axis is log-scaled. We drew a red solid line at y = 5 for reference. On core chromosomes, gene expression levels are mostly above this line, where on fast-core and accessory chromosomes, most gene expression levels are under 5 RPKM.

Figure S6. Fast-core and accessory chromosomes have relatively more differentially expressed genes than core chromosomes.

The x-axis show the position of the gene on the genome, the y-axis the log2fold change in gene expression (*in planta/in vitro*) per chromosome. Zero average read counts have been replaced by 0.1 to avoid infinite ratio's (as in Van der Does et al. PloS Genetics 2016). Differentially expressed genes are highlighted in blue squares. Not only fast-core chromosomes and chromosome 14 but also the central fast-core region on chromosome 4 and the subtelomeric regions have many differentially expressed genes.

Figure S7. Functional categories that are overrepresented on the fast-core chromosomes.

Fraction of annotated genes that are assigned to a FunCat category that is significantly overrepresented on the fast-core chromosome, compared to all genes on the genome (P-value < 0.001 after FDR correction). Categories that are also overrepresented when comparing up-regulated genes to the background of all fast-core genes are in bold. Percentages are defined with respect to the total number of annotated genes.

Figure S8. Read density from ChIP-seq experiments targeted at H3K27me3-ated or H3K4me2-ated regions.

Average read densities in 10.000 bp non-overlapping sliding windows are depicted in green (H3K4me2) and orange (H3K27me3). Read densities are normalized with respect to RPKM to allow for comparisons across experiments. Chromosomes are scaled to the length of their sequence in the assembly. The core chromosomes - except chromosome 11, 12 and 13- are largely enriched for H3K4me2, except at the subtelomeric regions. These small core chromosomes, designated fast-core chromosomes, are enriched in H3K27me3, similar to accessory chromosomes. Also the middle region in chromosome 4, which is also more divergent in terms of sequence similarity and synteny (Figure S2A, Figure S3A) is enriched in H3K27me3 as well. The two histone marks appear to be mutually exclusive.

A. Experiments with high sequencing depth: H3K4me2, experiment id 1358: 19735885 reads, of which 18129305 mapped (91,86 %); H3K27me3, experiment id 1360: 21467536 of which 16243857 mapped (75,76 %), see Table S2.

B. Replicate experiments, with lower sequencing depth: H3K4me2, experiment id 806: 9078423 reads, of which 6699238 mapped (73,79 %); H3K27me3, experiment id 808: 7891973 of which 5519933 mapped (69,94 %), see Table S2.

Figure S9. Core chromosomes are enriched for H3K4me2 and fast-core and accessory chromosomes for H3K27me3.

For each chromosome or chromosomal region (x-axis) we plot how much is covered in a DOE (top panels) and the size-distribution of DOEs (bottom panels, violinplots). Lineage-specific regions are denoted with a # on the x-axis. The regions on chromosome 3 and on chromosome 6 that are associated with pathogenicty on tomato are denoted with a P. Bars with thick lines represent the coverage per chromosome category (core, fast-core, lineage specific or pathogenicity) rather than per chromosome.

A. Coverage and size distribution of DOE inferred based on two H3K4me2 experiments (1358 'H3K4me2 (a)' and 806 'H3K4me2 (b)'. B. Coverage and size distribution of DOE inferred based on two H3K27me3 experiments (1360 'H3K27me3 (a)' and 808 'H3K27me3 (b)). The coverage of DOEs is lower for 806 'H3K4me2 (b)' and 808 'H3K27me3 (b), than for 1358 'H3K4me2 (a)' and 1360 'H3K27me3 (a)', but the distribution are not qualitatively different.

Figure S10. Overlap of domains of enrichment with genes, promoters and repeats.

We used RSEG to identify domains enriched in histone marks H3K4me2 (A) or H3K27me3 (B), based on mapped reads generated in two independent ChIP-seq experiments for each mark (left and right panels, 1358 and 806 for H3K4me2, 1360 and 808 for H3K27me3) (see Materials and Methods for more detail). For each chromosome or chromosomal region, we determine the percentage of genes (top panel, light blue), promoter regions (defined as up to 1000 base pairs upstream of first exon, second panel, dark blue), and repeats (as predicted by RepeatMasker, excluding low-complexity regions and simple repeats, third panel, grey) that overlap (for more than 90%) with a domain of enrichment. To compare whether certain elements (e.g. genes) overlap more than expected with a domain of enrichment, we also plot how much of a chromosome or chromosomal region is part of a domain of enrichment (bottom panels, dark grey on light grey background). Average values per

chromosome category are also included in the histogram, and designated with black boxes.

A. H3K4me2 domains occur mainly on the core chromosomes, and very little on the fast-core and accessory chromosomes. Domains of H3K2me2 are relatively underrepresented in repeat regions.

B. H3K27me3 domains occur mainly on the fast-core and accessory chromosomes, and very little on the core chromosomes. Domains of H3K27me3 are overrepresented in repeat regions on the core chromosomes.

Figure S11. d_N, d_S and d_N/d_S values along the genome.

The number of synonymous substitutions per synonymous site d_S, the number of non- synonymous substitutions per nonsynonymous site d_N (top panels, d_S in blue, d_N, in black traingles) and their ratio (bottom panel, grey dots) of genes are plotted on the genome. Genes with $d_N/d_S > 1$ are highlighted with red squares. On the x-axis are chromosomes, dotted vertical lines denote supercontigs that comprise these chromosomes. Y-axes are in log scale. Subtelomeric regions have higher d_N and d_S values, while their ratio is consistent along the genome. The same holds true for the fast-core chromosomes and the middle region on chromosome 4 that is also enriched in H3K27me3 and depleted in H3K4me2. The accessory chromosomes have very little data points, as values are calculated with respect to bidirectional best blast hits in *F. verticillioides* and these chromosomes do not occur in this species.

Figure S12. Frequency distribution of d_N and d_S of core and fast-core genes.

We compare the distributions of all core (light-blue) or all fast-core (light-red) genes to two subsets of genes (dark-blue and dark-red). One subset consists of genes that are differentially expressed during infection (DEGs, left panels) and one of genes that are in an H3K27me3-enriched domain (H3K27me3, right panels).

A. Kernel density estimation of the frequency distribution of d_N. Fast-core genes have higher d_N values than core genes. Fast-core DEGs have low values d_N compared to all fast-core genes, while for core genes there is no difference (left panel). Core genes that are in an H3K27me3-enriched domain have the same distribution as fast-core genes.

B. Kernel density estimation of the frequency distribution of d_S. Fast-core genes have higher d_S values than core genes. Core DEGs have higher d_S values than other core genes, for the fast-core the distribution of d_S of DEGs does not differ much from that of all fast-core genes. Again, core genes that are in an H3K27me3-enriched domain have the same distribution as fast-core genes.

Figure S13. SNP density per 10 kb sliding window, per chromosome.
We here plot the cumulative SNP density in 10000 bp, non-overlapping sliding windows. Per chromosome there are two panels: the top panel shows the cumulative SNP density for all strains in Clade III (see tree in the legend) in purple, whereas the bottom panel shows shows the cumulative SNP density for all strains in Vegetative Compatibility Group 030 (VCG030, dark-grey).

Figure S14. SNP density per 10 kb sliding window, per chromosome.
We here plot the cumulative SNP density in 10000 bp, non-overlapping sliding windows. Per chromosome there are two panels: the top panel shows the cumulative SNP density for all strains in Vegetative Compatibility Group 030 (VCG030, dark-grey), using a different y-axis scale than in Figure S13 to show that the distribution of SNPs on fast-core chromosomes is not different from that on core chromosomes.

Table S1. Strains used in this study.

Table S2. Gene expression levels and fold change in gene expression per gene

Table S3. Functional analysis of differentially regulated genes on pathogenicity region on chromosome 3 and 6.

Table S4. Functional analysis of differentially regulated genes on chromosome 14.

Table S5A. Functional enrichment of genes that reside on chromosome 11, 12 and 13.

Table S5B. Functional enrichment of up-regulatedgenes that reside on chromosome 11, 12 and 13.

Table S6. Sequencing depth, mapping and putative PCR duplicates in the ChIP-seq experiments.

Table S7A. Domains of enrichment of H3K4me2.

Table S7B. Domains of enrichment of H3K27me3.

Table S8. Statistics of selection per gene.

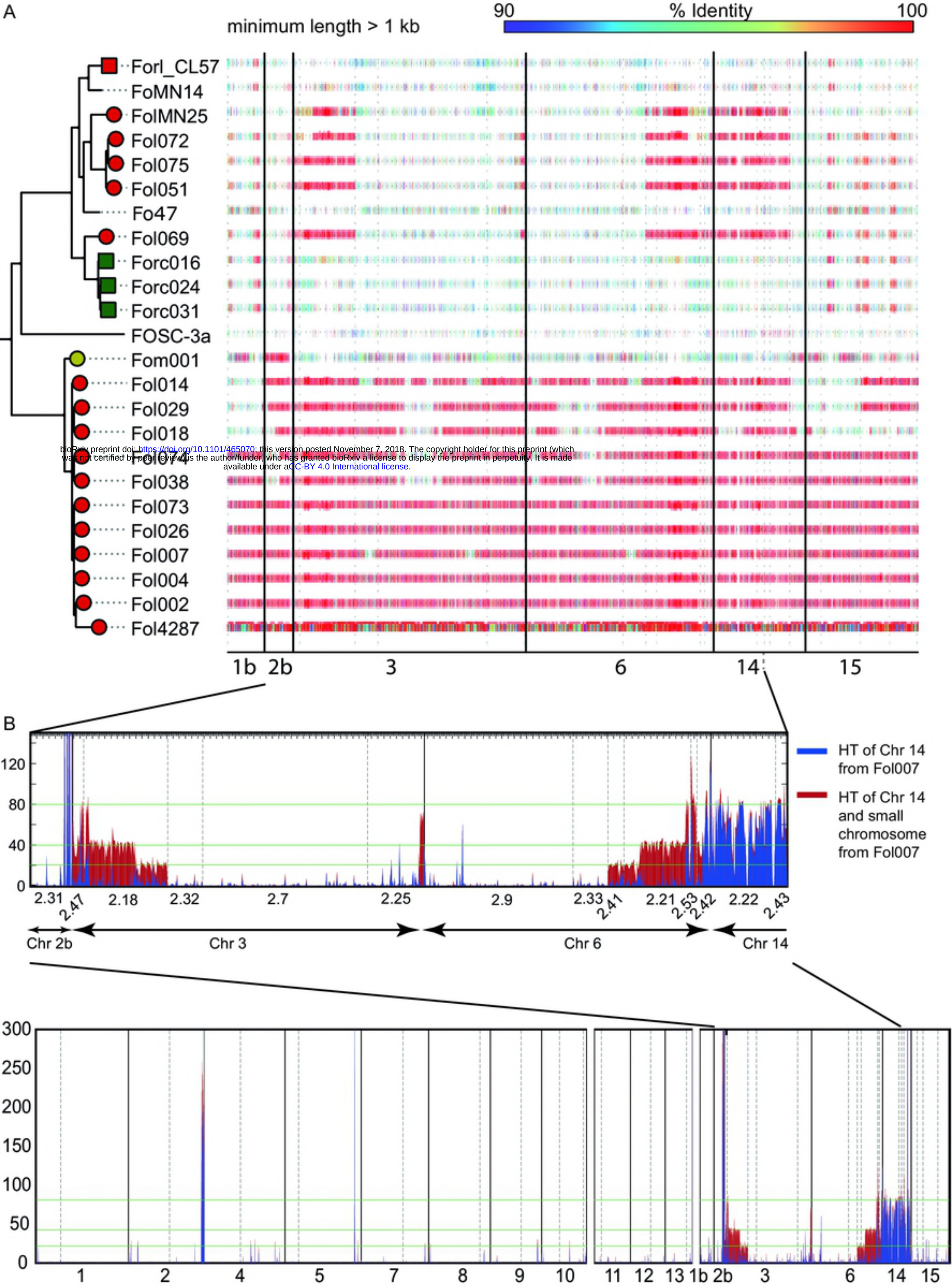


Figure 1

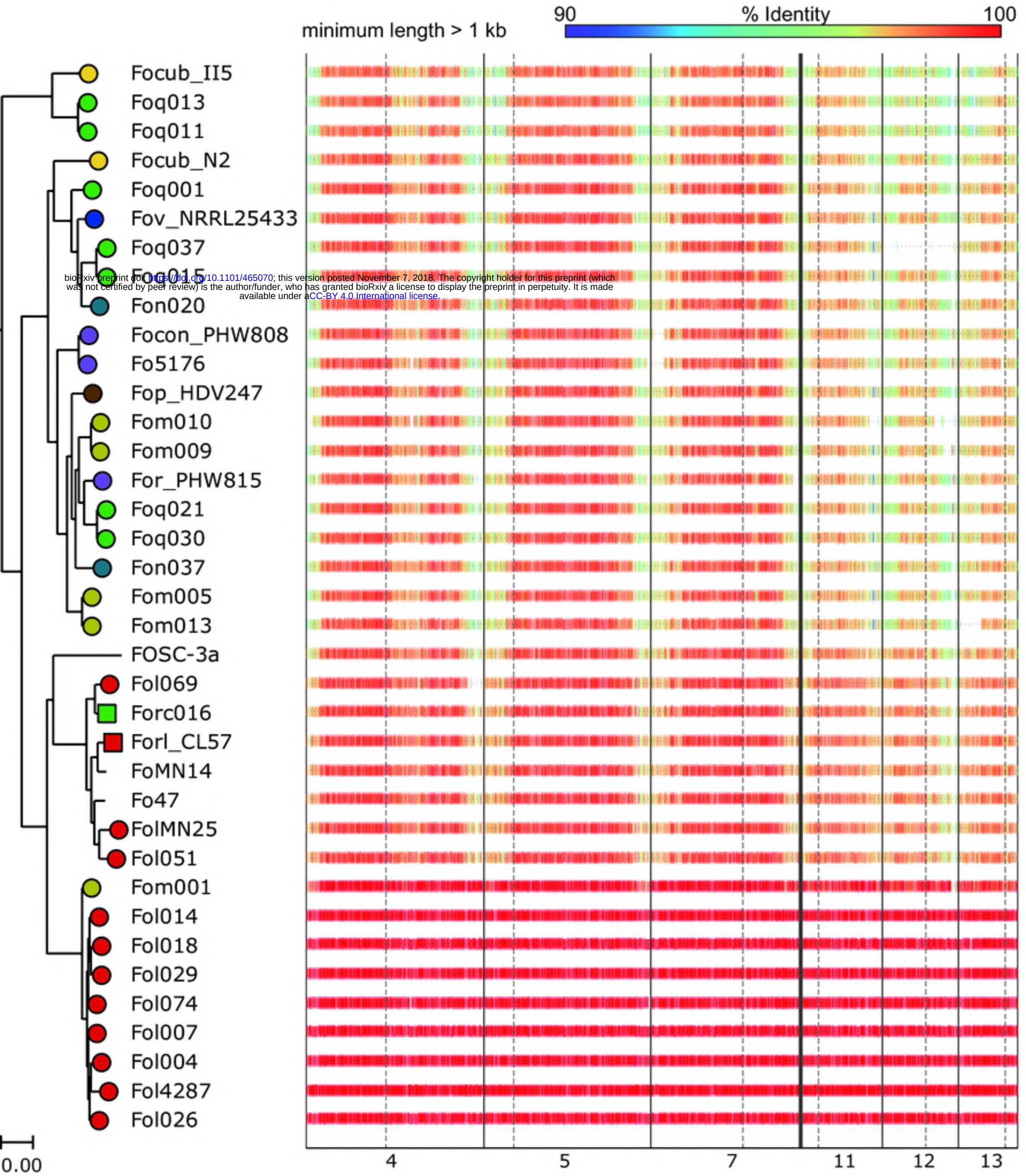


Figure 2

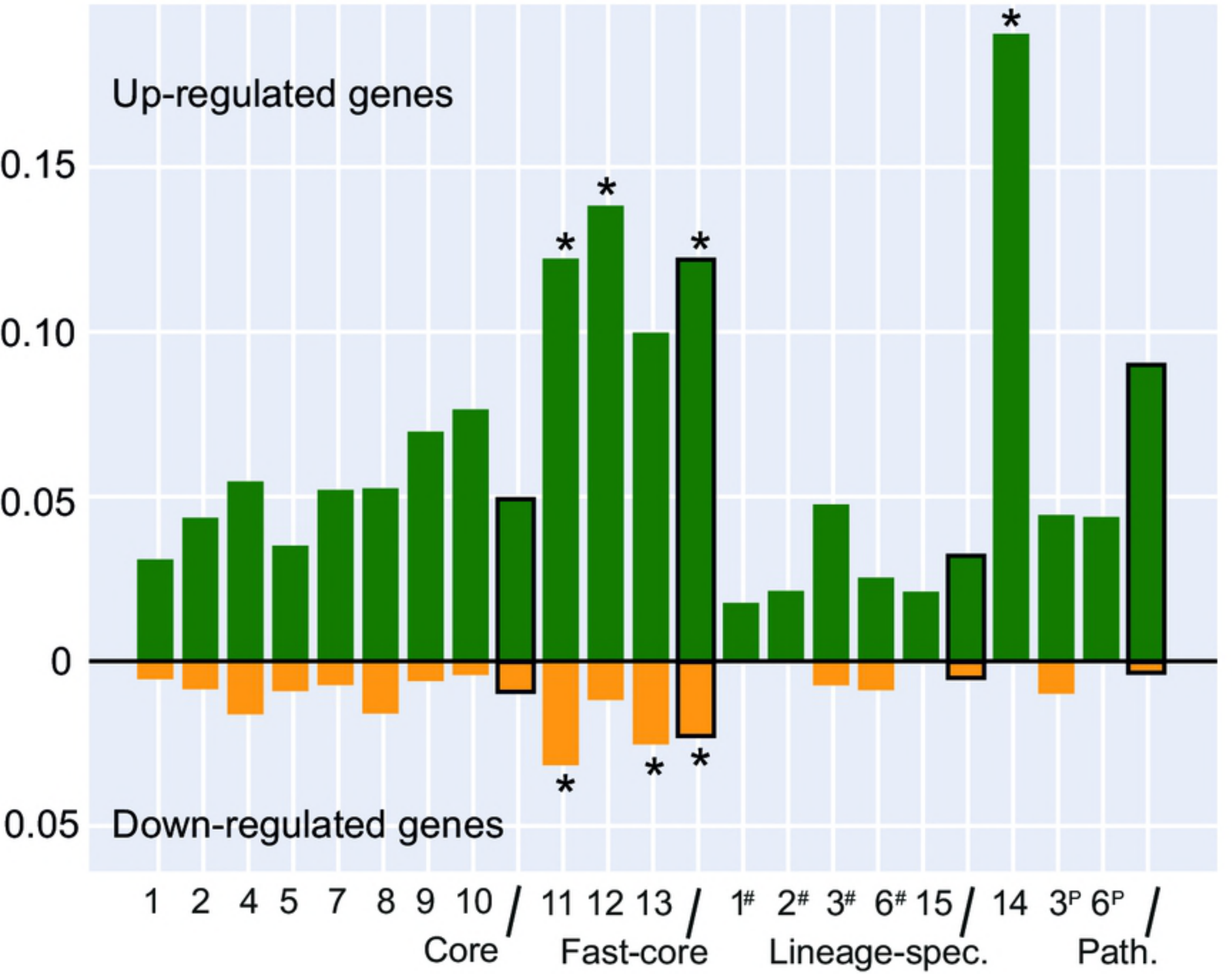

Up-regulated genes
0.15
0.10
0.05
0
0.05
Down-regulated genes
1 2 4 5 7 8 9 10 / 11 12 13 / 1# 2# 3# 6# 15 / 14 3P 6P /
Core
Fast-core
Lineage-spec.
Path.

Figure 3

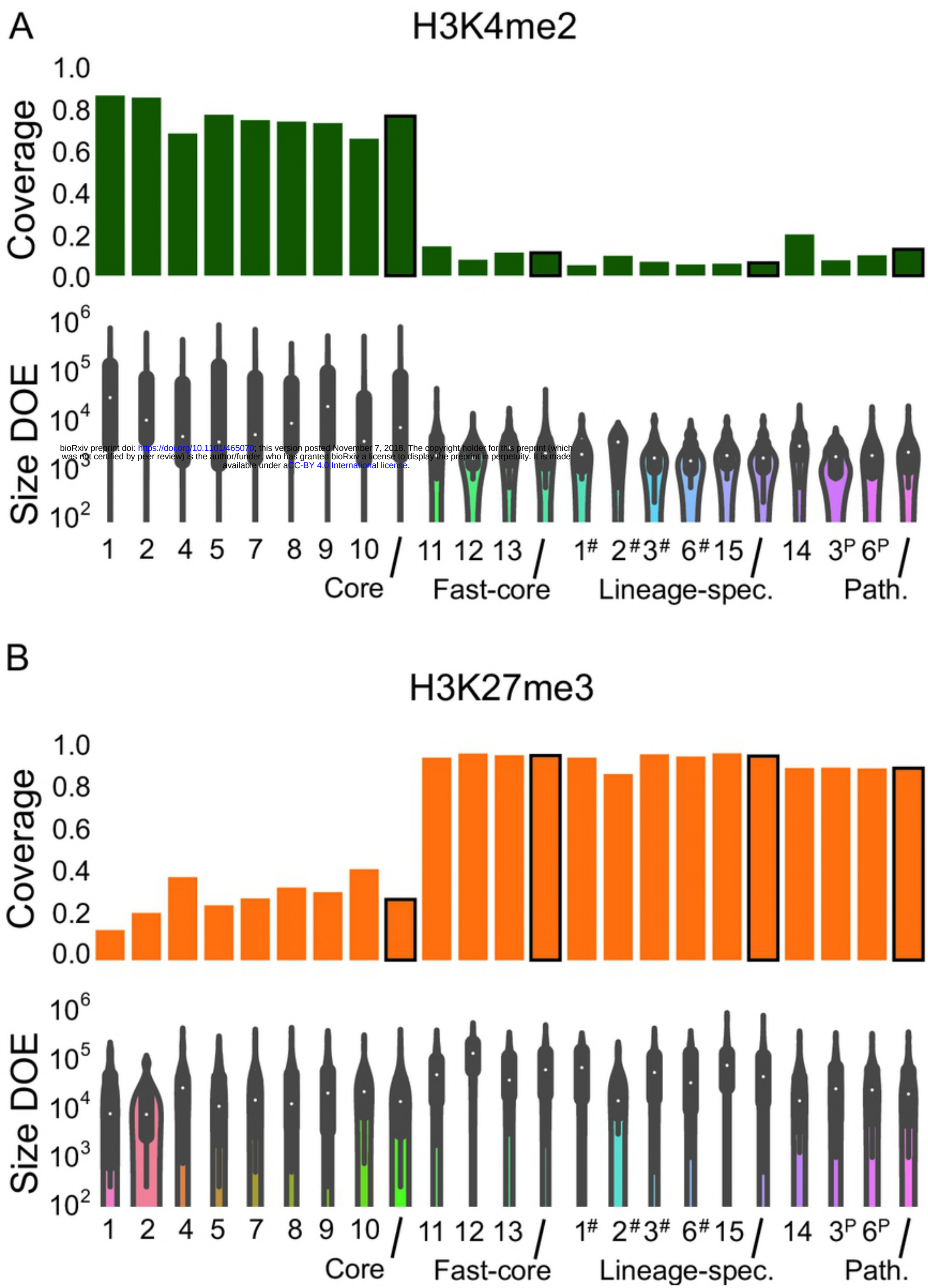
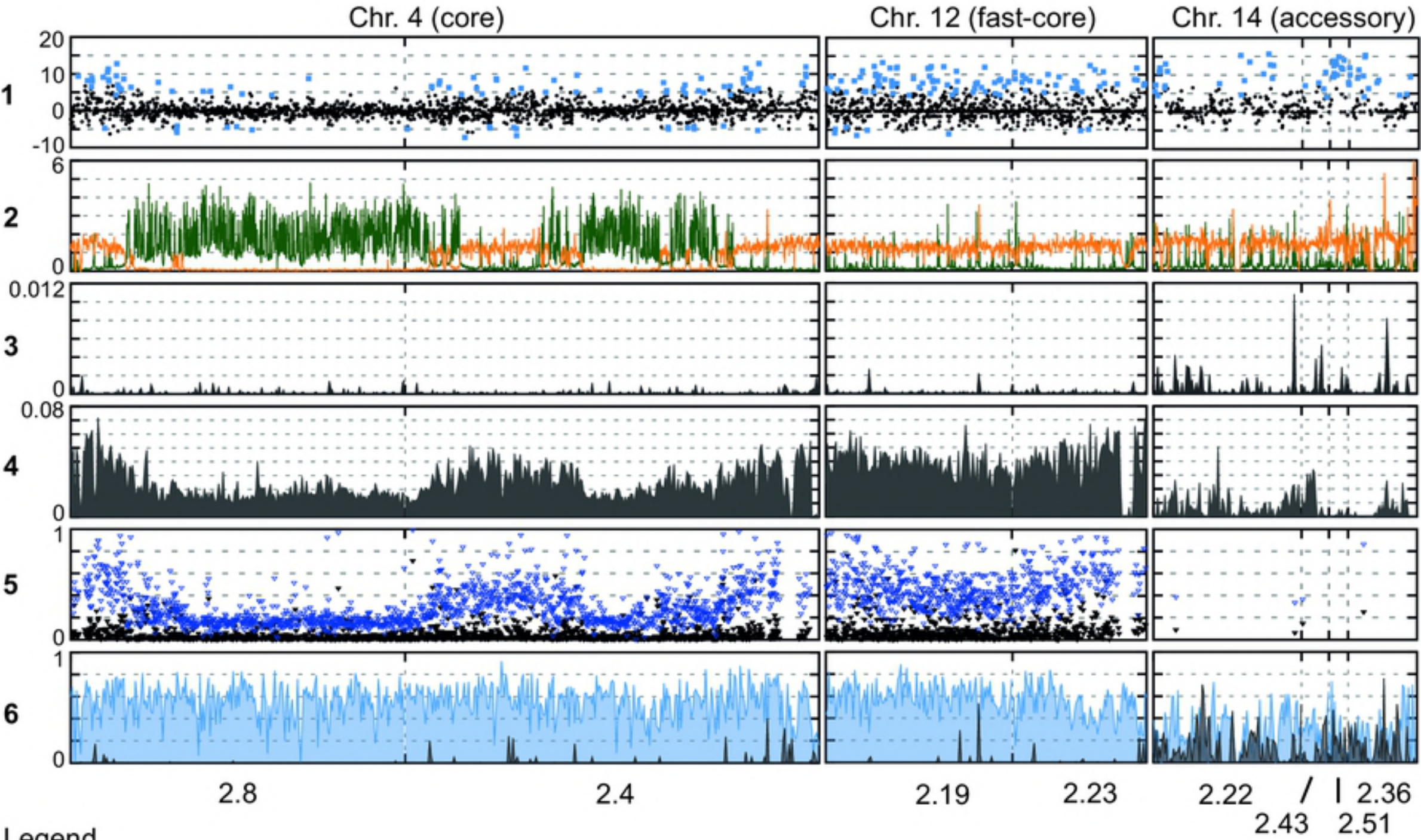


Figure 4



Legend

1 **Log2 fold change in gene expression**
- differentially expressed
- not differentially expressed

2 **Normalized read densities ChIP-seq**
- H3K4me2
- H3K27me3

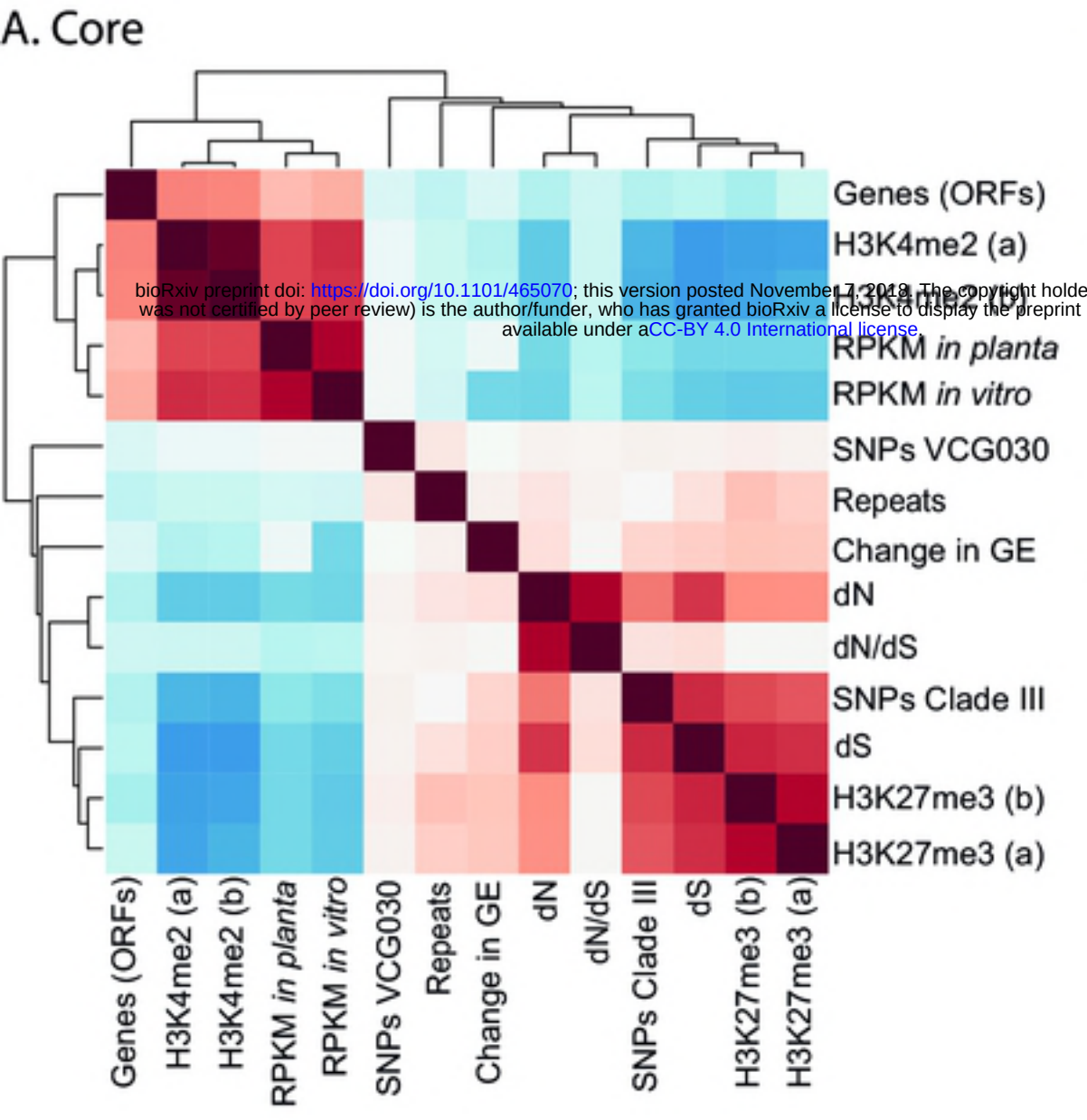
3 **SNPs short timescale**
- number of SNPs in VCG030 / windowsize

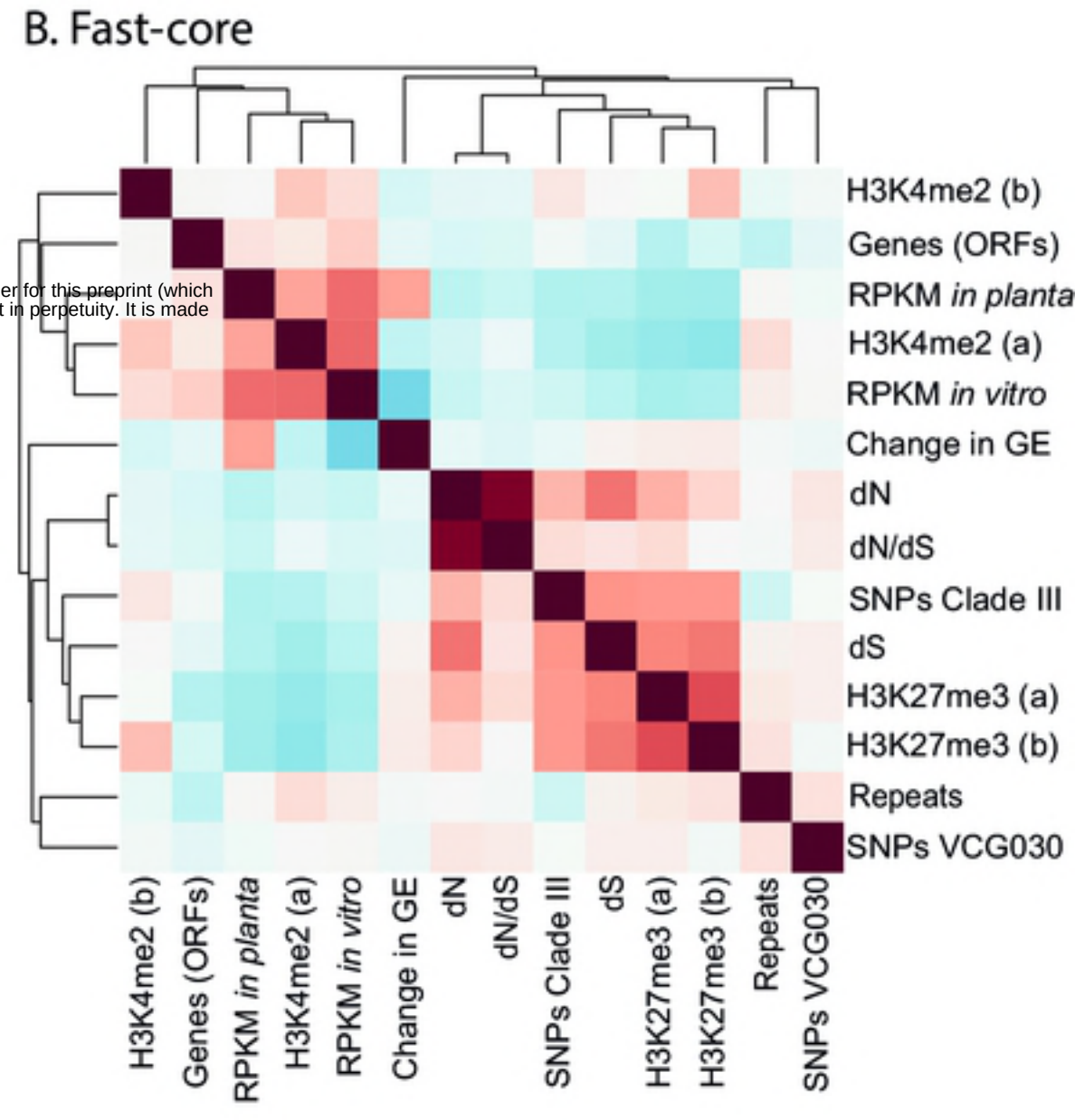
4 **SNPs intermediate timescale**
- number of SNPs in Clade III / windowsize

5 **Sequence divergence coding sequences**
- synonymous substitutions: dS
- nonsynonymous substitutions: dN

6 **Gene and repeat density**
- Gene density
- Repeat density

Figure 5





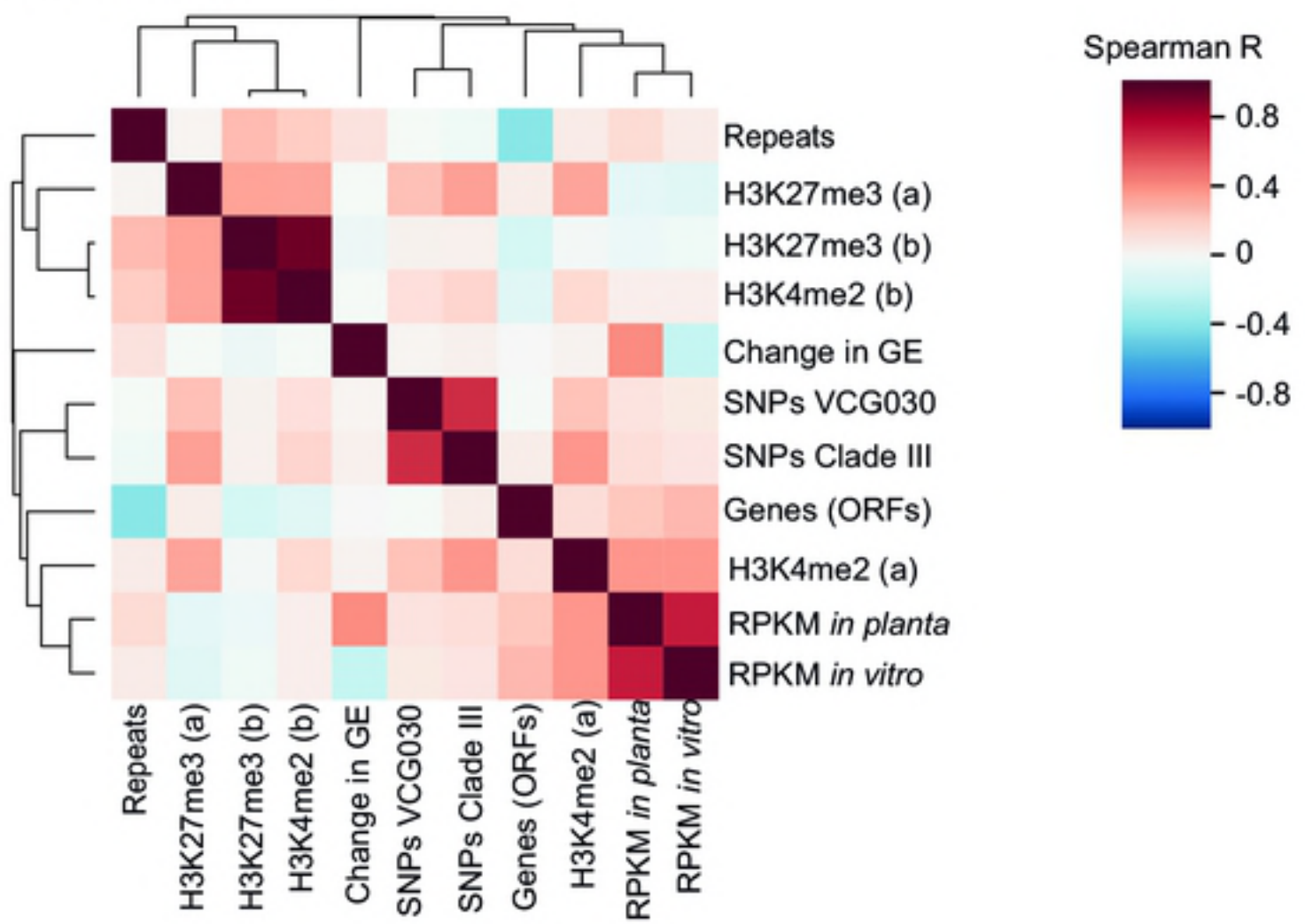


Figure 6