PNAS

Supporting Information for

Frequent horizontal chromosome transfer between asexual fungal insect pathogens

Authors:

Michael Habig [1,2*], Anna V. Grasse[3], Judith Müller[1,2], Eva H. Stukenbrock[1,2], Hanna Leitner[3], Sylvia Cremer[3*]

Author affiliation:

1) Environmental Genomics, Christian-Albrechts University of Kiel, Germany, 2) Max Planck Institute for Evolutionary Biology, Plön, Germany, 3) ISTA (Institute of Science and Technology Austria), Klosterneuburg, Austria

Corresponding authors:

Michael Habig, email: mhabig@bot.uni-kiel.de

Sylvia Cremer, email: sylvia.cremer@ist.ac.at

This PDF file includes:

Other supporting materials for this manuscript include the following:

Content

1. Figures S1 to S9

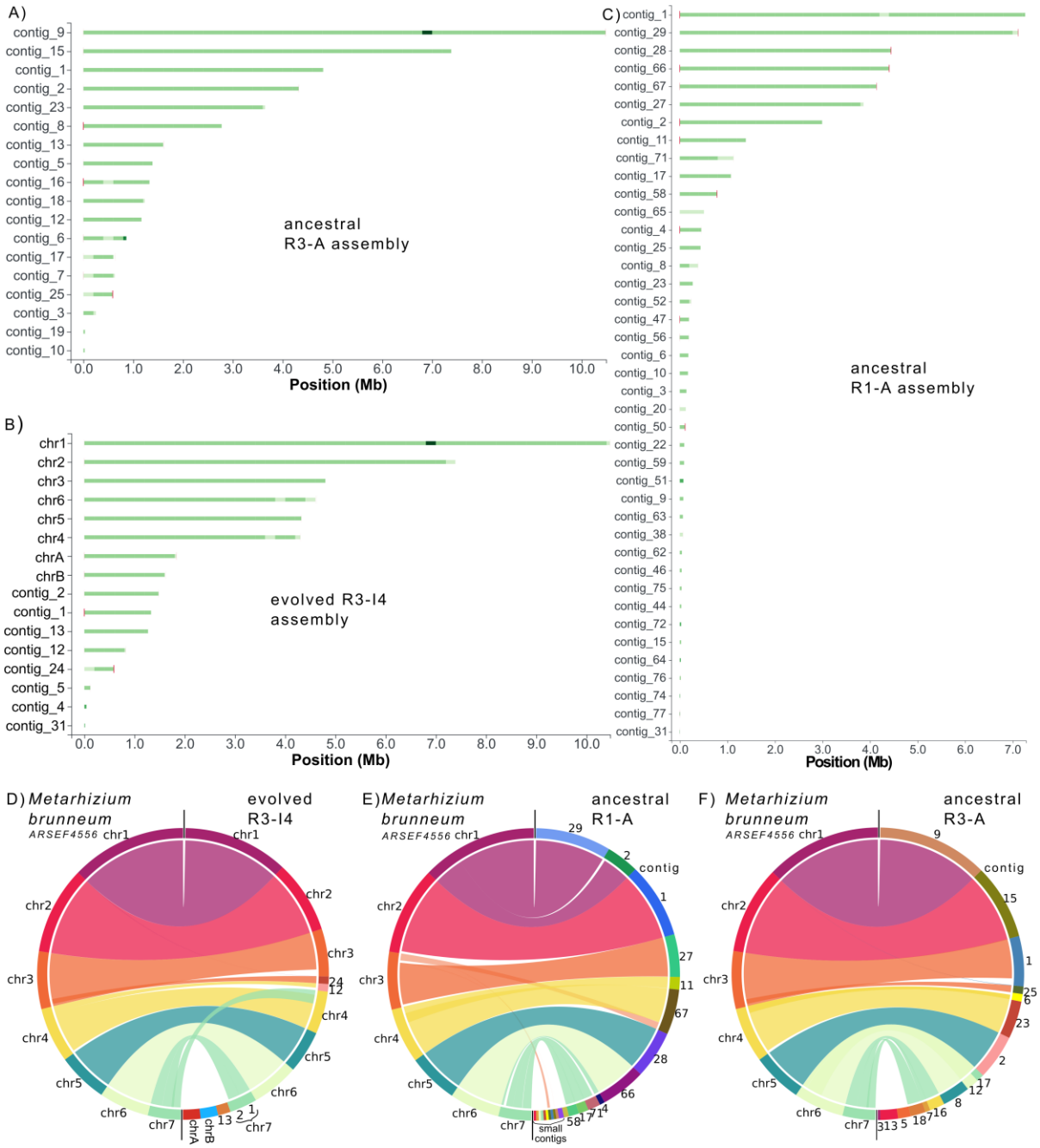
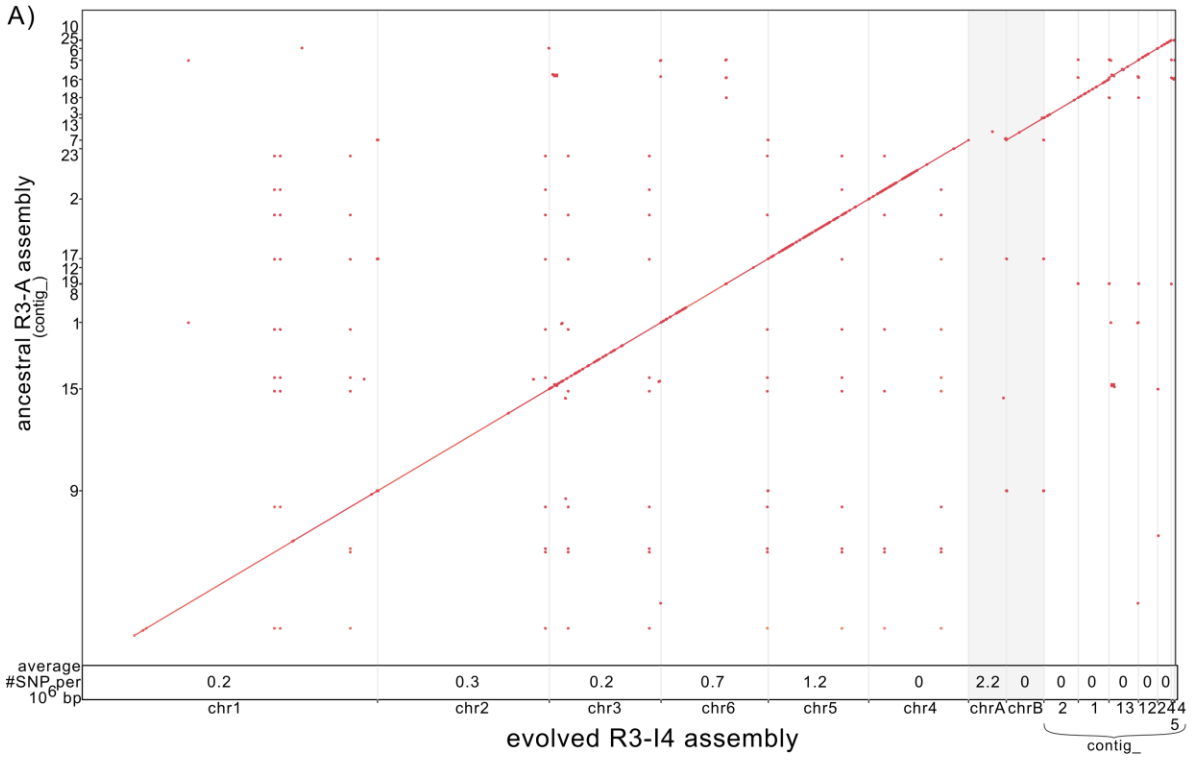


Fig. S1. Nanopore-based assemblies of *Metarhizium robertsii* ancestral R1-A, ancestral R3-A and the evolved R3-I4 strains at near chromosome level. Tapestry reports of the nanopore-based assemblies of A) ancestral R3-A, B) evolved R3-I4 and C) ancestral R1-A strain. Red marks represent the presence of telomere repeats, where the intensity of the red color is proportional to the number of repeats detected. Green intensity is

proportional to coverage. For a detailed description, see also Table S1. D) Synteny between the nanopore-based assemblies of *M. robertsii* strains R3-I4, R1-A and R3-A generated in this study and the *M. brunneum* ARSEF4556 reference assembly (GCA_013426205.1) (36). Note that for R3-I4 chromosomes were labelled based on their synteny to the *M. brunneum* reference assembly.


A)
ancestral R3-A assembly
(contig_)
10
25
6
5
16
18
3
13
7
23
2
17
12
19
8
1
15
9
average
#SNP per
10^6 bp
0.2
0.3
0.2
0.7
1.2
0
2.2
0
0
0
0
0
0
chr1
chr2
chr3
chr6
chr5
chr4
chrA
chrB
2
1
13
12
24
4
5
contig_
evolved R3-I4 assembly

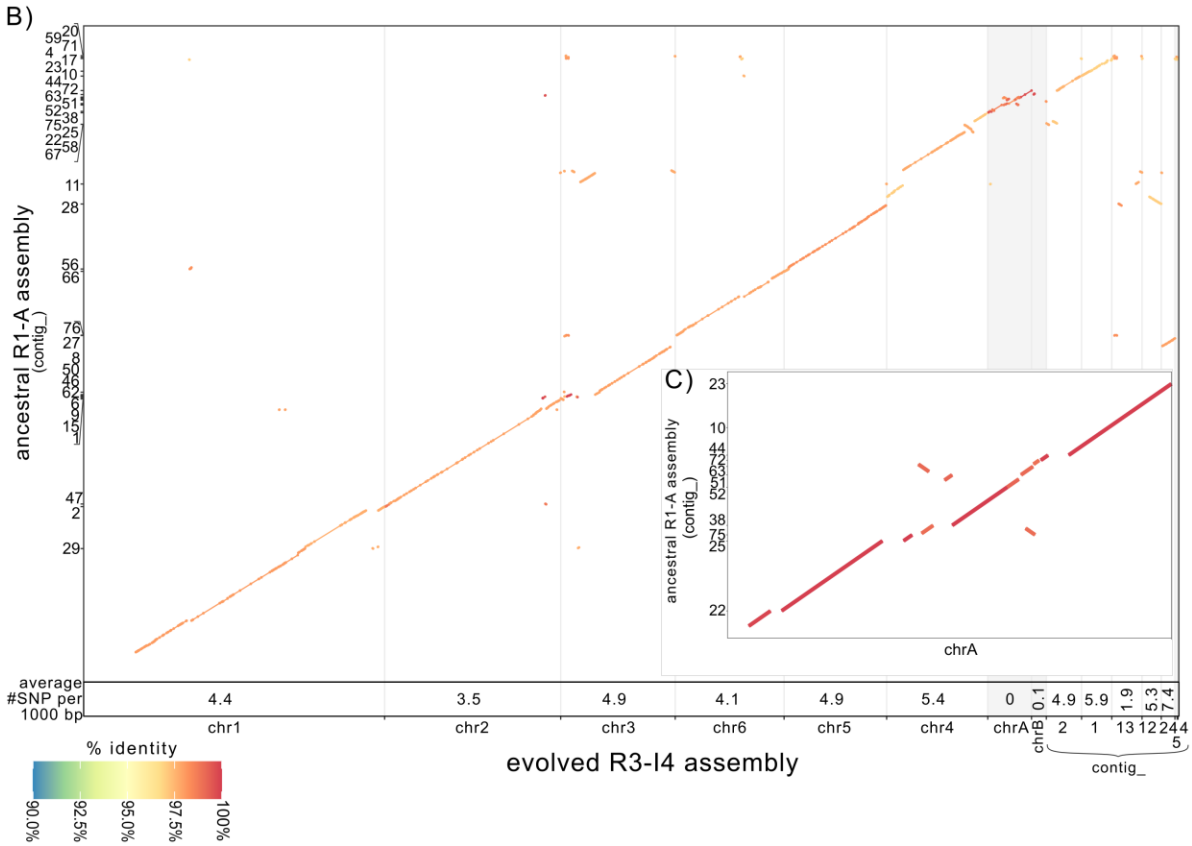

B)
ancestral R1-A assembly
(contig_)
20
59
71
4
17
23
10
44
72
63
51
52
38
75
25
22
58
67
11
28
56
66
76
27
8
50
46
62
47
2
29
C)
23
10
44
52
38
75
25
22
ancestral R1-A assembly
(contig_)
chrA
average
#SNP per
1000 bp
4.4
3.5
4.9
4.1
4.9
5.4
0
0.1
4.9
5.9
1.9
5.3
7.4
chr1
chr2
chr3
chr6
chr5
chr4
chrA
chrB
2
1
13
12
24
4
5
contig_
evolved R3-I4 assembly
% identity
90.0%
92.5%
95.0%
97.5%
100%

Fig. S2. Synteny plot between the nanopore-based assemblies of the A) ancestral R1-A or B) ancestral R3-A with the evolved R3-I4 strain, with synteny for chrA and chrB of the R3-I4 assembly highlighted in grey shade. In C) the alignment of R1-A contigs syntenic with chrA of the evolved R3-I4 is shown. Please note that the apparent structural variation between chrA of the evolved R3-I4 strain and the syntenic contigs of the ancestral R1-A most likely results from the high fragmentation of the R1-A assembly, since cross-mapping of Illumina reads, SNP calling, PFGE, analysis of larger structural variation and sequencing of excised PFGE bands (detailed in Fig. S3) did not find evidence of mutational processes or large-scale reorganization of chrA associated with its transfer from R1 to R3. All contigs of R1-A that are syntenic with chrA of R3-I4 are small (ranging from 30 to 437 kb, as shown in Table S1). The SNP density based on these alignments for each of the contigs is given at the bottom of each graph.

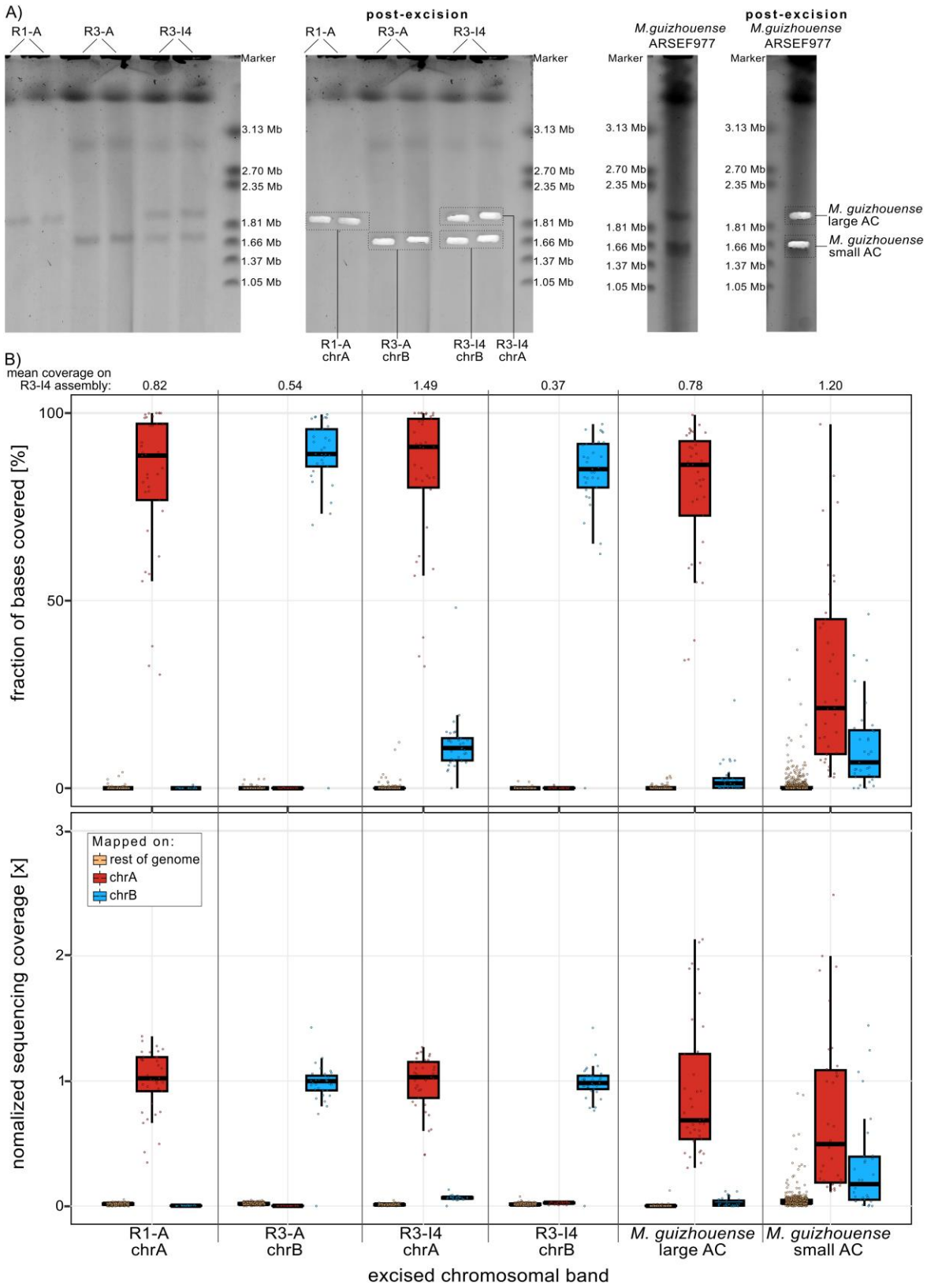

A)
R1-A
R3-A
R3-I4
Marker
3.13 Mb
2.70 Mb
2.35 Mb
1.81 Mb
1.66 Mb
1.37 Mb
1.05 Mb
post-excision
R1-A chrA
R3-A chrB
R3-I4 chrB
R3-I4 chrA
M.guizhouense ARSEF977
M. guizhouense large AC
M. guizhouense small AC
B)
mean coverage on R3-I4 assembly:
0.82
0.54
1.49
0.37
0.78
1.20
fraction of bases covered [%]
nomalized sequencing coverage [x]
Mapped on:
rest of genome
chrA
chrB
R1-A chrA
R3-A chrB
R3-I4 chrA
R3-I4 chrB
M. guizhouense large AC
M. guizhouense small AC
excised chromosomal band

Fig. S3. Verification of inferred chromosomal band identity in PFGE through sequencing of excised bands. A) PFGE-gel of strains *M. robertsii* R1-A, R3-A and R3-I4, as well as *M. guizhouense* ARSEF977 before and after chromosomal bands were excised from the gel. The inferred identity of the chromosomal bands in the PFGE-gel is indicated. To increase the amount of DNA for sequencing, two replicates each for R1-A, R3-A, and R3-I4 were run in the PFGE-gel, and the corresponding chromosomal bands were excised and pooled. B) Results of the Illumina-reads of the DNA from the indicated excised bands mapped on the R3-I4 assembly in 50 kb windows (excluding transposable elements) for chrA (in red), chrB (in blue), and the remaining genome. The top row shows the relative fraction of 50 kb windows covered by at least 5 reads, while the bottom row shows the normalized sequencing coverage (normalized to the sequencing coverage of the contig with the highest average coverage). Both the fraction of bases covered and the normalized sequencing coverage confirm the inferred identity of the chromosomal bands in the case of *M. robertsii* strains R1-A, R3-A and R3-I4. Here, the majority of reads from the excised bands map to the corresponding chromosome and cover large portions. In the case of *M. guizhouense*, two chromosomal bands (large and small accessory chromosome, AC) mainly contain reads that map to chrA, thus confirming the disomy of this chromosome in *M. guizhouense*. Please note that, in agreement with the results of the phylogenetic analysis (Fig. S6), we found some chrB sequences in *M. guizhouense* (small AC). However, their presence is at low coverage (both in terms of the fraction of covered bases and normalized sequencing coverage), and may be due to non-perfect separation during PFGE, rather than their presence in the chromosome represented by the chromosomal band. Moreover, the low amount of DNA has resulted in a low total sequencing coverage (given above the graph), which may have affected the fraction of covered bases, possibly leading to artifacts in the low-coverage 50kb windows. Outliers with normalized sequencing coverage >3 have been excluded from the graph for visual clarity. The supplementary data S1 contains all data.

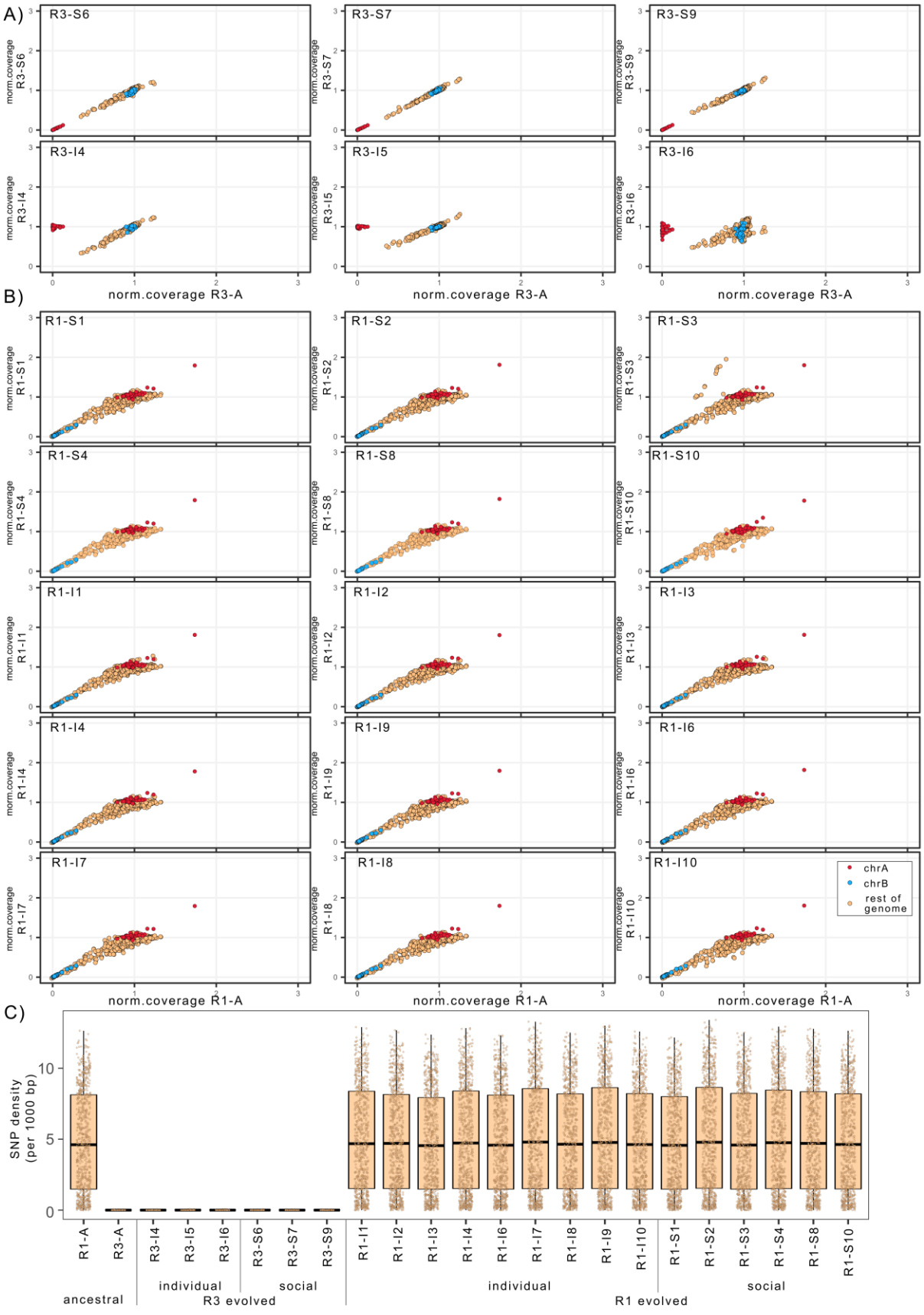

A)
R3-S6
R3-S7
R3-S9
R3-I4
R3-I5
R3-I6
morm.coverage R3-S6
morm.coverage R3-S7
morm.coverage R3-S9
morm.coverage R3-I4
morm.coverage R3-I5
morm.coverage R3-I6
norm.coverage R3-A
B)
R1-S1
R1-S2
R1-S3
R1-S4
R1-S8
R1-S10
R1-I1
R1-I2
R1-I3
R1-I4
R1-I9
R1-I6
R1-I7
R1-I8
R1-I10
morm.coverage R1-S1
morm.coverage R1-S2
morm.coverage R1-S3
morm.coverage R1-S4
morm.coverage R1-S8
morm.coverage R1-S10
morm.coverage R1-I1
morm.coverage R1-I2
morm.coverage R1-I3
morm.coverage R1-I4
morm.coverage R1-I9
morm.coverage R1-I6
morm.coverage R1-I7
morm.coverage R1-I8
morm.coverage R1-I10
chrA
chrB
rest of genome
norm.coverage R1-A
C)
SNP density (per 1000 bp)
R1-A
R3-A
R3-I4
R3-I5
R3-I6
R3-S6
R3-S7
R3-S9
R1-I1
R1-I2
R1-I3
R1-I4
R1-I6
R1-I7
R1-I8
R1-I9
R1-I10
R1-S1
R1-S2
R1-S3
R1-S4
R1-S8
R1-S10
ancestral
individual
social
R3 evolved
individual
social
R1 evolved

Fig. S4. Coverage analysis and SNP/InDel distribution failed to detect horizontal transfer of genetic material in addition to chrA. Illumina sequence coverage analysis of the A) evolved R3 strains compared to the coverage of the ancestral R3-A strain or the B) evolved R1 strains compared to the coverage of the ancestral R1-A strain. With the exception of chrA for the individually-evolved R3 strains, no change in sequence coverage was detected. C) Distribution of SNPs/InDels that are specific to the R1-A (present in the R1-A but absent in the R3-A). No 50 kb windows in the evolved R3 lines showed increased SNP density. Hence no large-scale transfer of genetic material, in addition to chrA, occurred from the ancestral R1-A strain to the evolved R3 strains. In the evolved R1 strains, there was no change in the distribution of SNPs/InDels compared to the ancestral R1-A strain, indicating that no large-scale transfer of genetic material to the evolved R1 strains occurred. Note that the rDNA cluster was excluded from the analysis for visual clarity due to its high coverage.

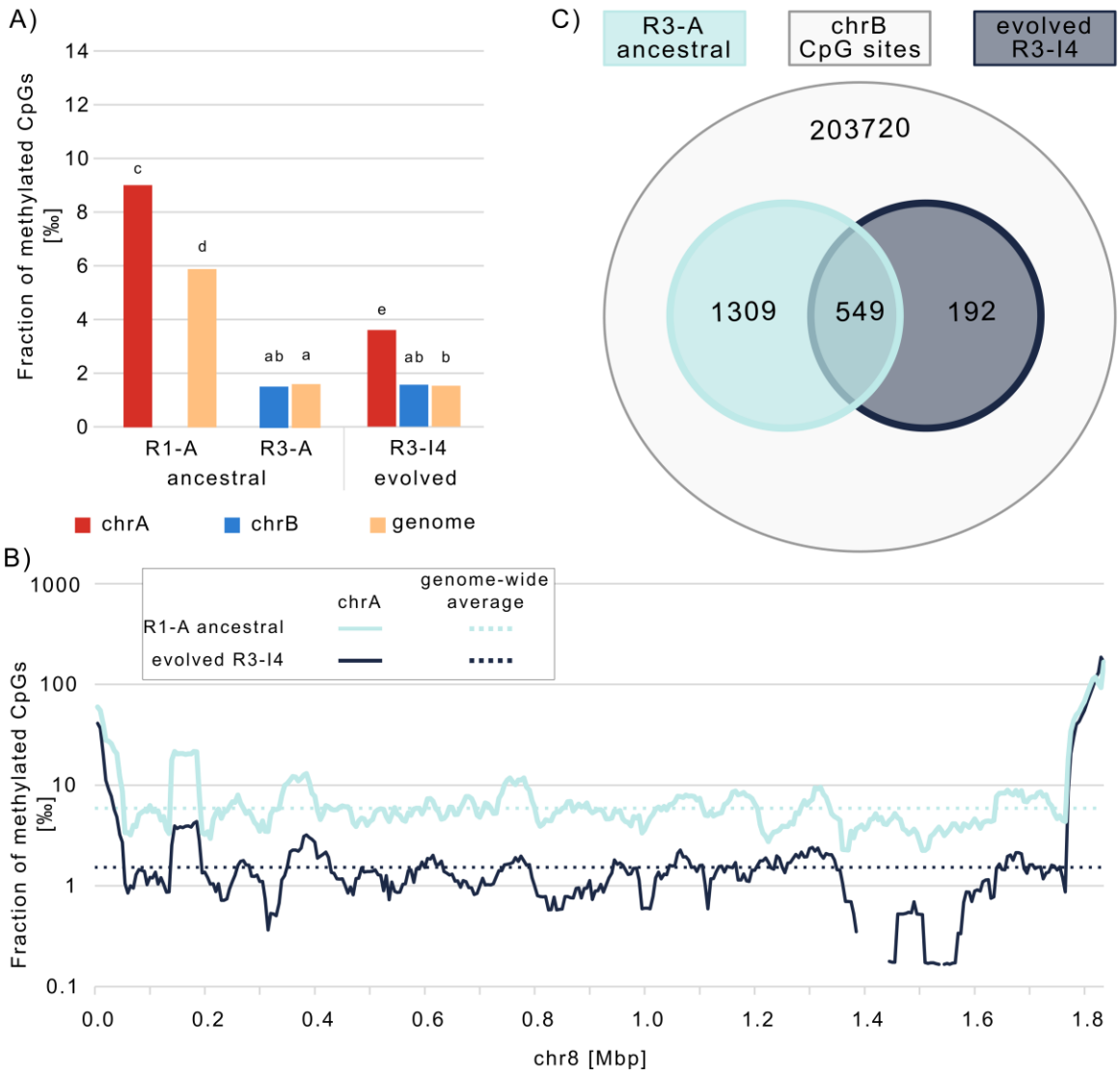


Fig. S5. The methylation pattern differed between ancestral R1-A and R3-A strains. ChrA retained part of the ancestral methylation pattern after horizontal transfer from R1-A to R3-A. A) Fraction of methylated cytosines in CpG contexts for chrA, chrB and the rest of the genome in the ancestral R1-A and R3-A strains and the evolved R3-I4 strain. ChrA showed higher methylation in both the ancestral R1-A strain and the evolved R3-I4 strain than chrB and the rest of the genome (identical letters above groups indicate non-significance at $\alpha<0.05$ determined by Fisher exact test with BH-adjustment for multiple testing). B) Fraction of methylated cytosines in CpG contexts in 50kb windows (sliding: 5 kb) along chrA for the ancestral R1-A (turquoise) and evolved R3-I4 (dark-grey) strains. C) Venn diagram showing the overlap of highly methylated CpG sites (>25%

methylation) of chrA (total number CpG sites 203720) between the ancestral R1-A and the evolved R3-I4 strain.

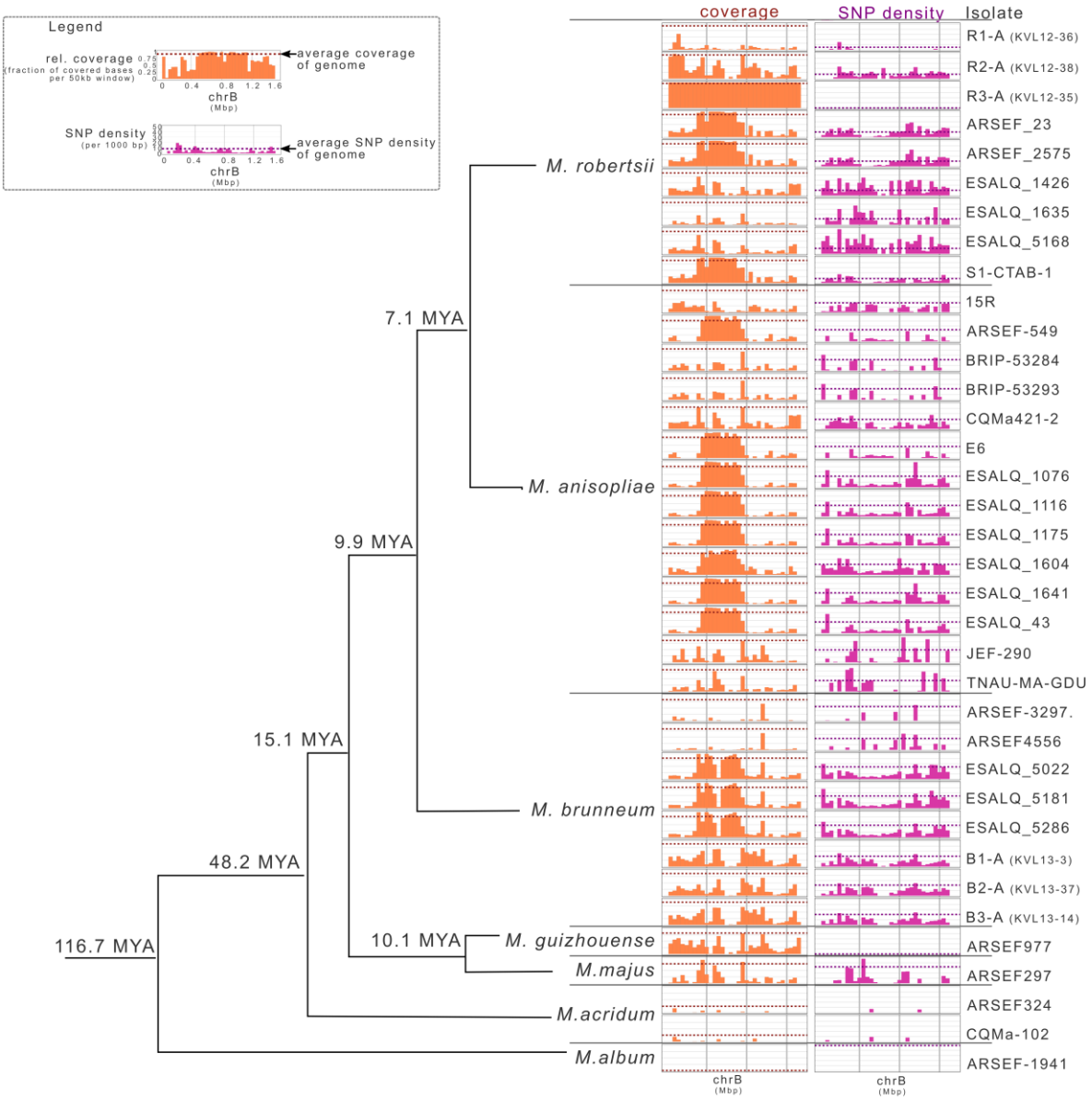


Fig. S6. Presence/absence polymorphism of chrB in published genomes of species of the genus *Metarhizium.* Phylogeny and distribution of relative sequence coverage (fraction of bases covered in 50 kb windows) in orange and SNP density per 1000 bp in 50 kb windows in pink, with the respective genome-wide averages shown as dotted lines. Note: TEs were excluded from the analysis. Phylogeny adapted from published Hu and colleagues, 2014 (41). MYA: Million years ago.

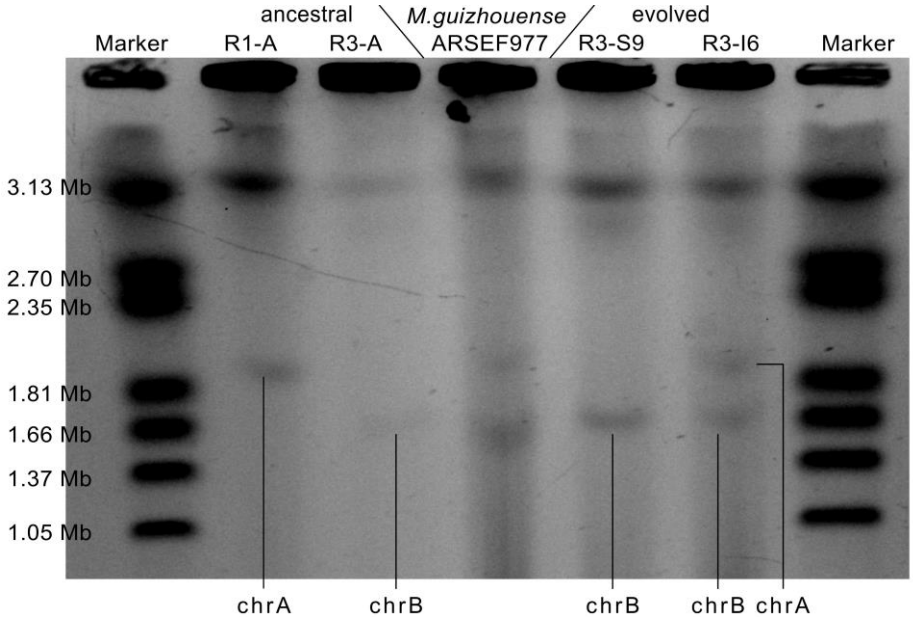


Fig. S7. PFGE-gel of *M. guizhouense* ARSEF977 in comparison to the ancestral *M. robertsii* strains R1-A and R3-A, as well as two evolved R3 strains R3-S9 (lacking chrA) and R3-I6 (including chrA).

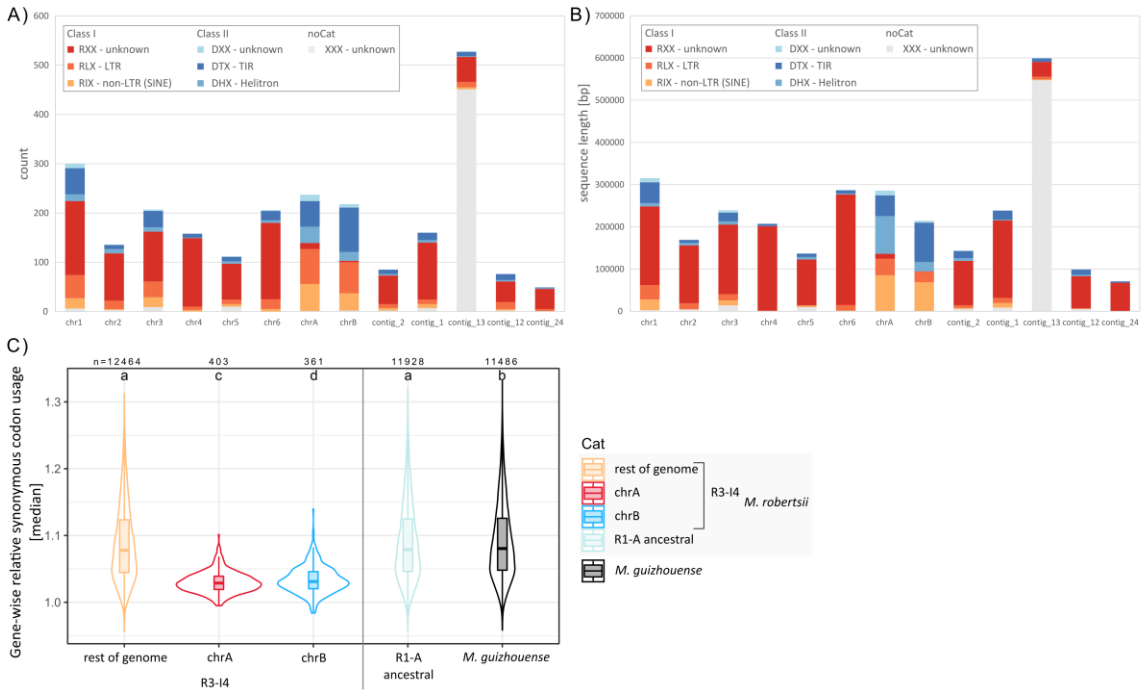


Fig. S8. TE composition and codon usage of the accessory chromosomes chrA and chrB differs from the rest of the genome. A) Number and B) cumulative sequence length of the indicated TE classes and orders of the *M. robertsii* R3-I4 strain. ChrA and chrB have a higher proportion of Class I retrotransposons of LTR and SINE order and virtually lack the unknown order of the Class I retrotransposons that dominates (dark-red) in the other chromosomes. C) Gene-wise relative synonymous codon usage for genes located on chrA and chrB compared to genes located on the rest of the R3-I4 genome and R1-A and the *M. guizhouense* ARSEF977 genomes. Identical letters above the individual plots indicate that the respective groups were not significantly different (pairwise Wilcoxon rank-sum tests with BH correction, α=0.05).

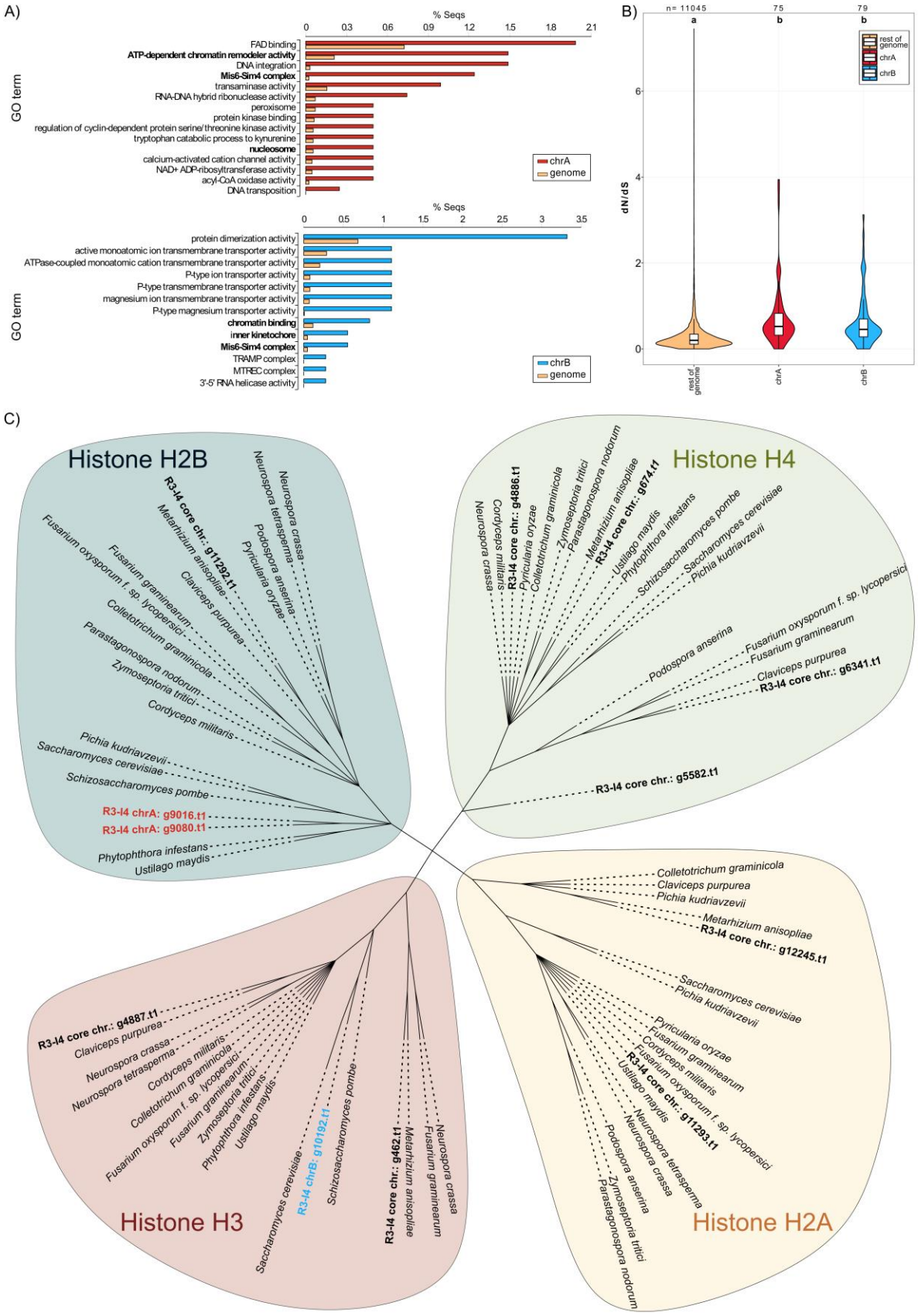

A)
% Seqs
GO term
FAD binding
ATP-dependent chromatin remodeler activity
DNA integration
Mis6-Sim4 complex
transaminase activity
RNA-DNA hybrid ribonuclease activity
peroxisome
protein kinase binding
regulation of cyclin-dependent protein serine/threonine kinase activity
tryptophan catabolic process to kynurenine
nucleosome
calcium-activated cation channel activity
NAD+ ADP-ribosyltransferase activity
acyl-CoA oxidase activity
DNA transposition
chrA
genome
% Seqs
GO term
protein dimerization activity
active monoatomic ion transmembrane transporter activity
ATPase-coupled monoatomic cation transmembrane transporter activity
P-type ion transporter activity
P-type transmembrane transporter activity
magnesium ion transmembrane transporter activity
P-type magnesium transporter activity
chromatin binding
inner kinetochore
Mis6-Sim4 complex
TRAMP complex
MTREC complex
3'-5' RNA helicase activity
chrB
genome
B)
n= 11045
75
79
a
b
b
rest of genome
chrA
chrB
dN/dS
C)
Histone H2B
Neurospora crassa
Neurospora tetrasperma
Podospora anserina
Pyricularia oryzae
R3-l4 core chr.: g11292.t1
Metarhizium anisopliae
Claviceps purpurea
Fusarium graminearum
Fusarium oxysporum f. sp. lycopersici
Colletotrichum graminicola
Parastagonospora nodorum
Zymoseptoria tritici
Cordyceps militaris
Pichia kudriavzevii
Saccharomyces cerevisiae
Schizosaccharomyces pombe
R3-l4 chrA: g9016.t1
R3-l4 chrA: g9080.t1
Phytophthora infestans
Ustilago maydis
Histone H4
Neurospora crassa
Cordyceps militaris
R3-l4 core chr.: g4886.t1
Pyricularia oryzae
Colletotrichum graminicola
Zymoseptoria tritici
Parastagonospora nodorum
Metarhizium anisopliae
R3-l4 core chr.: g674.t1
Ustilago maydis
Phytophthora infestans
Schizosaccharomyces pombe
Saccharomyces cerevisiae
Pichia kudriavzevii
Podospora anserina
Fusarium oxysporum f. sp. lycopersici
Fusarium graminearum
Claviceps purpurea
R3-l4 core chr.: g6341.t1
R3-l4 core chr.: g5582.t1
Histone H3
R3-l4 core chr.: g4887.t1
Claviceps purpurea
Neurospora crassa
Neurospora tetrasperma
Cordyceps militaris
Colletotrichum graminicola
Fusarium oxysporum f. sp. lycopersici
Fusarium graminearum
Zymoseptoria tritici
Phytophthora infestans
Ustilago maydis
Saccharomyces cerevisiae
R3-l4 chrB: g10192.t1
Schizosaccharomyces pombe
R3-l4 core chr.: g462.t1
Metarhizium anisopliae
Fusarium graminearum
Neurospora crassa
Histone H2A
Colletotrichum graminicola
Claviceps purpurea
Pichia kudriavzevii
Metarhizium anisopliae
R3-l4 core chr.: g12245.t1
Saccharomyces cerevisiae
Pichia kudriavzevii
Pyricularia oryzae
Fusarium graminearum
Cordyceps militaris
Fusarium oxysporum f. sp. lycopersici
R3-l4 core chr.: g11293.t1
Ustilago maydis
Neurospora tetrasperma
Neurospora crassa
Podospora anserina
Zymoseptoria tritici
Parastagonospora nodorum

Fig. S9. GO-term enrichment, dN/dS ratio and phylogeny of histones located on the accessory chromosomes chrA and chrB. A) Significantly enriched GO terms (at α=0.05) for genes located on chrA (in red) and chrB (in blue). GO terms associated with chromatin, nucleosome, and/or chromosome segregation are highlighted in bold. B) Box- and Violin-plot of gene-wise dN/dS ratio for all genes located on chrA (in red), chrB (in blue), and the rest of the genome (in yellow) between *M. robertsii* R3-I4 and *M. guizhouense* ARSEF977. Identical letters above the individual plots indicate that the respective groups were not significantly different (pairwise Wilcoxon rank-sum tests with BH correction, α=0.05). C) Unrooted phylogenetic tree of the four core histones (H2A, H2B, H3 and H4) for the putative histone-encoding genes in *M. robertsii* R3-I4 and the correspondingly annotated genes in 16 fungi and one oomycete. Genes located on the accessory chrA are highlighted in red, those on chrB are highlighted in blue, and those located on the core chromosomes of *M. robertsii* are highlighted in bold. The putative histone-encoding genes located on chrA and chrB are not paralogues of the corresponding histone-encoding genes located on the core chromosomes of *M. robertsii* R3-I4.

2. Tables S1 to S7

Table S1: Comparison statistics of Nanopore-based assemblies

Strain	Type	Contig	Chr. (chr1-7: based on synteny to *M. brunneum*)	Length (bp)	GC%	Median ReadDepth	# Start Telomeres	# End Telomeres	Unique Bases	Unique%	Syntenic to chrA in R3-I4
R3-I4	individually-evolved	contig_16	chr1	10465571	51.1	44.3	1	1	10026287	96	
R3-I4	individually-evolved	contig_20	chr2	7387335	51.3	45.4	2	0	7160197	97	
R3-I4	individually-evolved	contig_3	chr3	4807498	50	42.4	1	0	4161460	87	
R3-I4	individually-evolved	contig_7	chr6	4615637	49.7	44.1	1	1	4357934	94	
R3-I4	individually-evolved	contig_14	chr5	4319916	50.9	44.7	1	1	4231323	98	
R3-I4	individually-evolved	contig_6	chr4	4300840	49.9	45.5	0	1	4140497	96	
R3-I4	individually-evolved	contig_18	chrA	1832736	50	42.5	3	3	726620	40	
R3-I4	individually-evolved	contig_19	chrB	1619728	49.8	43.6	3	1	962047	59	
R3-I4	individually-evolved	contig_2	chr7	1477798	48.5	44.8	1	1	1279223	87	
R3-I4	individually-evolved	contig_1	chr7	1328681	46.2	43.7	25	0	1200979	90	
R3-I4	individually-evolved	contig_13	-	1268002	49	42.6	0	0	514021	41	
R3-I4	individually-evolved	contig_12	-	823663	47.3	39.3	1	1	573669	70	
R3-I4	individually-evolved	contig_24	-	587825	48.3	40.4	0	24	521764	89	
R3-I4	individually-evolved	contig_5	-	118431	49	41.3	0	1	35854	30	
R3-I4	individually-evolved	contig_4	-	42546	48	68.2	0	0	1596	4	
R3-I4	individually-evolved	contig_31	-	13566	49	57.9	0	0	4340	32	
R3-A	ancestral	contig_9		10479110	51.1	47.9	1	3	10045194	96	
R3-A	ancestral	contig_15		7382587	51.3	49.1	2	0	7173378	97	
R3-A	ancestral	contig_1		4812484	50	46.7	1	0	4178569	87	
R3-A	ancestral	contig_2		4319998	50.9	46.9	1	2	4238826	98	

R3-A	ancestral	contig_23	3646070	50.7	46.9	0	0	3570947	98	
R3-A	ancestral	contig_8	2774337	51.2	48.7	26	0	2684941	97	
R3-A	ancestral	contig_13	1613873	49.8	43.7	0	2	1169465	72	
R3-A	ancestral	contig_5	1386847	49	43.4	1	0	551284	40	
R3-A	ancestral	contig_16	1325639	46.2	42.6	25	0	1211252	91	
R3-A	ancestral	contig_18	1230992	49	45	0	0	1131971	92	
R3-A	ancestral	contig_12	1165225	49	46.7	1	0	1072860	92	
R3-A	ancestral	contig_6	861996	47.3	45.7	3	0	598882	69	
R3-A	ancestral	contig_17	640873	44.7	36.6	1	1	600832	94	
R3-A	ancestral	contig_7	629488	47.3	38.4	3	0	574862	91	
R3-A	ancestral	contig_25	587802	48.3	41.1	0	26	521944	89	
R3-A	ancestral	contig_3	248960	45.7	32.6	0	0	156642	63	
R3-A	ancestral	contig_19	30752	48.7	40.9	0	0	1	0	
R3-A	ancestral	contig_10	26233	47.2	45	0	0	5487	21	
R1-A	ancestral	contig_1	7267018	51	43.7	25	0	7160995	99	
R1-A	ancestral	contig_29	7120443	50.5	45	0	15	6929065	97	
R1-A	ancestral	contig_28	4444060	50	44.5	0	25	4266863	96	
R1-A	ancestral	contig_66	4406720	49.4	42.5	21	24	4324693	98	
R1-A	ancestral	contig_67	4139298	49.3	43.7	8	14	3973138	96	
R1-A	ancestral	contig_27	3863933	49.9	44.9	0	0	3709959	96	
R1-A	ancestral	contig_2	2994606	51.4	44.5	23	0	2917671	97	
R1-A	ancestral	contig_11	1389962	46.8	43.6	28	0	1261440	91	
R1-A	ancestral	contig_71	1130232	45.7	39.7	0	1	798946	71	
R1-A	ancestral	contig_17	1075475	49.7	42.5	1	1	1002826	93	
R1-A	ancestral	contig_58	784453	48.2	40.8	1	22	751924	96	
R1-A	ancestral	contig_65	508985	49.7	28.5	0	0	33483	7	
R1-A	ancestral	contig_4	451755	48.9	39.6	23	1	430927	95	
R1-A	ancestral	contig_25	436588	50	38.1	0	0	308264	71	yes

R1-A	ancestral	contig_8	383379	36.6	39	0	1	342110	89	
R1-A	ancestral	contig_23	272488	50.5	39.4	2	0	204004	75	yes
R1-A	ancestral	contig_52	238311	49.8	36.3	0	1	183500	77	yes
R1-A	ancestral	contig_47	192533	48.8	37.7	17	1	178869	93	
R1-A	ancestral	contig_56	188225	47.6	39	3	1	178771	95	
R1-A	ancestral	contig_6	180082	47.4	42.3	0	0	124287	69	
R1-A	ancestral	contig_10	172426	50.6	44.9	0	0	126783	74	yes
R1-A	ancestral	contig_3	140387	48.7	33.6	0	0	82551	59	
R1-A	ancestral	contig_20	125788	48.8	30.7	1	0	63815	51	
R1-A	ancestral	contig_50	112477	42	33.7	0	22	79755	71	
R1-A	ancestral	contig_22	94572	48.7	40.8	1	0	57382	61	yes
R1-A	ancestral	contig_59	90690	48.6	44	0	0	60573	67	
R1-A	ancestral	contig_51	76745	48	55.2	0	0	76745	100	yes
R1-A	ancestral	contig_9	74764	49.4	36	0	0	61366	82	
R1-A	ancestral	contig_63	66292	49.6	41.2	0	0	631	1	yes
R1-A	ancestral	contig_38	65268	50	30.8	0	1	1033	2	yes
R1-A	ancestral	contig_62	44691	51.1	41.1	0	0	40351	90	
R1-A	ancestral	contig_46	39964	48.2	41.6	0	0	17692	44	
R1-A	ancestral	contig_75	37454	53.5	46.9	0	1	4130	11	yes
R1-A	ancestral	contig_44	32834	51.9	48	0	0	4392	13	yes
R1-A	ancestral	contig_72	30035	51.2	62.7	0	0	1978	7	yes
R1-A	ancestral	contig_15	29486	49.3	35.2	0	0	21194	72	
R1-A	ancestral	contig_64	22534	49.8	56.2	1	0	2012	9	
R1-A	ancestral	contig_76	21574	48.6	40.5	0	1	0	0	
R1-A	ancestral	contig_74	9854	50.4	55.6	0	1	2098	21	
R1-A	ancestral	contig_77	7940	51.3	57	2	1	67	1	
R1-A	ancestral	contig_31	4621	49.9	70.1	1	0	4621	100	

Table S2: Number of SNPs and small InDels compared to R3-I4 that were not already present in the R3-A ancestral strain

Type	treatment group	Strain	SNPs/Indels in chrA	chrB	rest of genome	Grand Total	SNPs/InDels not present in R3-A in chrA	chrB	rest of genome	Grand Total	but present in R1-Al in chrA	chrB	rest of genome	Grand Total	of these without any evidence of being present in R3-A chrA	chrB	rest of genome	Grand Total
	ancestral	R1-A	0	141	197200	197341	0	141	197193	197334	0	141	197193	197334	0	141	197185	197326
		R3-A	4	1	24	29	0	0	0	0	0	0	0	0	0	0	0	0
R3	individually-evolved	R3-I4	0	1	15	16	0	1	4	5	0	0	1	1	0	0	0	0
		R3-I5	0	0	36	36	0	0	15	15	0	0	7	7	0	0	1*	1*
		R3-I6	0	0	36	36	0	0	18	18	0	0	6	6	0	0	0	0
	socially-evolved	R3-S9	0	2	31	33	0	2	15	17	0	0	4	4	0	0	0	0
		R3-S6	3	1	31	35	1	1	14	16	0	0	3	3	0	0	1*	1*
		R3-S7	4	0	25	29	2	0	11	13	0	0	2	2	0	0	0	0
R1	individually-evolved	R1-I1	0	164	205206	205370	0	164	205198	205362	0	129	189920	190049	0	129	189916	190045
		R1-I2	0	160	202184	202344	0	160	202176	202336	0	132	189115	189247	0	132	189112	189244
		R1-I3	0	158	196293	196451	0	158	196284	196442	0	133	185832	185965	0	133	185826	185959
		R1-I4	0	182	206340	206522	0	182	206332	206514	0	135	190830	190965	0	135	190823	190958
		R1-I6	0	171	199513	199684	0	171	199506	199677	0	131	187689	187820	0	131	187685	187816
		R1-I7	0	202	210943	211145	0	202	210934	211136	0	135	192274	192409	0	135	192269	192404
		R1-I8	1	156	201985	202142	1	156	201978	202135	0	126	188748	188874	0	126	188743	188869
		R1-I9	0	186	211198	211384	0	186	211189	211375	0	136	192238	192374	0	136	192233	192369
		R1-I10	0	165	200522	200687	0	165	200512	200677	0	130	188376	188506	0	130	188368	188498
	socially-evolved	R1-S1	0	155	197095	197250	0	155	197087	197242	0	126	186368	186494	0	126	186365	186491
		R1-S2	0	197	213034	213231	0	197	213027	213224	0	137	192947	193084	0	137	192943	193080
		R1-S3	0	178	201364	201542	0	178	201357	201535	0	133	188583	188716	0	133	188579	188712
		R1-S4	0	181	207754	207935	0	181	207746	207927	0	138	191344	191482	0	138	191337	191475
		R1-S8	0	177	205642	205819	0	177	205635	205812	0	134	190603	190737	0	134	190599	190733

R1-S10	1	165	200027	200193	1	165	200020	200186	0	135	187999	188134	0	135	187998	188133

* InDel at chr5:3,759,062. Wrongly called (based on visual inspection of aligned reads)

Table S3: Number of methylated Cytosines in CpG context

chr/contig of R3-I4 assembly	size [bp]	Number of C in CpG context (on both strands)	R1-A		R3-A		R3-I4	
			Count	Metfrac [x10^{-3}]	Count	Metfrac [x10^{-3}]	Count	Metfrac [x10^{-3}]
chr1	10465571	1240792	7319	5.90	1545	1.25	1411	1.14
chr2	7387335	879786	5570	6.33	1120	1.27	1082	1.23
chr3	4807498	550886	3018	5.48	680	1.23	649	1.18
chr6	4615637	525614	3345	6.36	720	1.37	689	1.31
chr5	4319916	509060	3096	6.08	735	1.44	729	1.43
chr4	4300840	495058	3089	6.24	678	1.37	653	1.32
chrA	1832736	205770	1858	9.03			741	3.60
chrB	1619728	179874	27	0.15	277	1.54	281	1.56
contig_2	1477798	162308	849	5.23	280	1.73	245	1.51
contig_1	1328681	137250	722	5.26	209	1.52	192	1.40
contig_13	1268002	127278	590	4.64	1624	12.76	1421	11.16
contig_12	823663	89124	317	3.56	140	1.57	155	1.74
contig_24	587825	66852	414	6.19	83	1.24	82	1.23
contig_5	118431	12986	22	1.69	39	3.00	31	2.39
contig_4	42546	4394	2	0.46	10	2.28	4	0.91
contig_31	13566	1344	0	0	1	0.74	1	0.74
chrA	1832736	205770	1858	9.03	n.d.	n.d.	741	3.60
chrB	1619728	179874	27	0.15	277	1.54	281	1.56
rest of genome	41557309	4802732	28353	5.90	7864	1.64	7344	1.53

Table S4: Overview of previously published Assemblies and Reads included in this study

Species	Strain	Location	BioSample	Assembly Accession	N75 (assembly)	assembly size [bp]	Sequencing run
M. robertsii	KVL12-36 (C17) = R1	Denmark	SAMN37404239				SRR26068569
M. robertsii	KVL12-35 (E81) = R2	Denmark	SAMN37404240				SRR26068568
M. robertsii	KVL12-38 (F19) = R3	Denmark	SAMN37404241				SRR26068557
M. robertsii	ARSEF 23	USA	SAMN20219665				SRR15182435
M. robertsii	ARSEF 2575	USA	SAMN20219666				SRR15182434
M. robertsii	ESALQ 1426	Brazil	SAMN20243868				SRR15182431
M. robertsii	ESALQ 1635	Brazil	SAMN20243869				SRR15182430
M. robertsii	ESALQ 5168	Brazil	SAMN20219667				SRR15182429
M. robertsii	S1-CTAB-1	China	SAMN21033591				SRR15665281
M. anisopliae	15R	Korea	SAMN26650411				SRR18320042
M. anisopliae	ARSEF-549	Brazil	SAMN03268434	GCA_000814975.1	1007133	38504274	
M. anisopliae	BRIP-53284		SAMN02981522	GCA_000426985.1	99853	38088637	
M. anisopliae	BRIP-53293	Australia	SAMN02981521	GCA_000426965.1	32194	38672492	
M. anisopliae	CQMa421-2	China	SAMN15446573				SRR12224770
M. anisopliae	E6	Brazil	SAMN02840975	GCA_000739145.1	319537	38477109	
M. anisopliae	ESALQ 1076	Brazil	SAMN20243862				SRR15182425
M. anisopliae	ESALQ 1116	Brazil	SAMN20219669				SRR15182427
M. anisopliae	ESALQ 1175	Brazil	SAMN20243863				SRR15182424
M. anisopliae	ESALQ 1604		SAMN20243864				SRR15182423
M. anisopliae	ESALQ 1641	Brazil	SAMN20219670				SRR15182426
M. anisopliae	ESALQ 43	Brazil	SAMN20219668				SRR15182428
M. anisopliae	JEF-290	South Korea	SAMN11321365	GCA_013305495.1	6254943	42848098	
M. anisopliae	TNAU-MA-GDU	India	SAMN26879068	GCA_023212845.1	193594	39437477	
M. brunneum	ARSEF-3297	Mexico	SAMN03268432	GCF_000814965.1	943571	37066166	

M. brunneum	ARSEF 4556	USA	SAMN14166897				SRR11149706
M. brunneum	ESALQ 5022	Brazil	SAMN20243865				SRR15182422
M. brunneum	ESALQ 5181	Brazil	SAMN20243866				SRR15182433
M. brunneum	ESALQ 5286	Brazil	SAMN20243867				SRR15182432
M. brunneum	KVL13-13 (G39) =B1	Denmark	SAMN37404242				SRR26068546
M. brunneum	KVL12-37 (J65) = B2	Denmark	SAMN37404243				SRR26068541
M. brunneum	KVL13-14 (L105) = B3	Denmark	SAMN37404244				SRR26068540
M. guizhouense	ARSEF977	France	SAMN03268433	GCA_000814955.1	317737	43465197	
M. majus	ARSEF297	Samoa	SAMN03268436	GCF_000814945.1	99788	42062993	
M. acridum	ARSEF324	Australia	SAMN18235592	GCA_019434415.1	2337527	44714781	
M. acridum	CQMa-102		SAMN02981259	GCF_000187405.1	54747	39422329	
M. album	ARSEF-1941	Philipines	SAMN03268435	GCF_000804445.1	238756	30449065	

Table S5: Number and phases of SNPs and InDels in *M. guizhouense* ARSEF977 on the *M. robertsii* R3-I4 assembly.

		homozygous SNPs/InDels	heterozygous SNPs/InDels						
chr/contig	Contig size [bp]	1/1	0/1	0|1	1|0	Total SNPs/InDels	Average coverage [per site]	Sites (excluding TEs) [bp]	Average SNP density [per 1 kb]
chr1	10465571	206718	99	293	338	207448	0.873	10002236	20.7
chr2	7387335	150103	62	209	182	150556	0.888	7143658	21.1
chr3	4807498	81897	29	34	77	82037	0.749	4461518	18.4
chr6	4615637	82468	39	34	122	82663	0.833	4284019	19.3
chr5	4319916	87469	45	27	45	87586	0.865	4160338	21.1
chr4	4300840	84871	32	41	106	85050	0.855	4088191	20.8
chrA	1832736	9	6	2790	294	3099	1.000	1265300	2.4
chrB	1619728	2613	15	262	183	3073	0.326	1093624	2.8
contig_2	1477798	22832	5	33	37	22907	0.751	1315475	17.4
contig_1	1328681	15953	6	2	5	15966	0.545	1084083	14.7
contig_13	1268002	4120	3	8	1	4132	0.481	726023	5.7
contig_12	823663	9380	2	23	8	9413	0.497	663248	14.2
contig_24	587825	10841	2			10843	0.754	495643	21.9
contig_5	118431	213	1	12	10	236	nd	100294	2.4
contig_4	42546	13		1	1	15	nd	37130	0.4

Table S6: Comparison of genome annotations.

Strain	Compartment	size [bp]	genes	transcripts	SignalP	EffectorP	Secondary Metabolite cluster (AntiSmash6.0)	CAZyme (GH)	GH18	Chitinases
R3-I4	chrA	1832736	364	403	30	13	0	10	5	3
R3-I4	chrB	1619728	328	361	38	8	1	7	2	1
R3-I4	genome (excluding chrA and chrB)	41557309	11803	12464	1317	370	62	195	26	17
R1-A	genome	42768942	12027	12710	1344	381	73	201	27	18
R3-A	genome	43163266	12254	12962	1365	380	63	200	28	18

Table S7: Overview of sequencing information generated within this study

ID	Experimental ID (as in Stock et al 2023)	Type	Strain	Strain ID	Illumina Sequencing 2*150 bp PE-Reads	Long read sequencing Sequencing Nanopore	PacBio
R1-A	R1	field isolate	C17	KVL12-36 (C17)	22015508	2496817	
R2-A	R2	field isolate	E81	KVL12-35 (E81)	22256812		
R3-A	R3	field isolate	F19	KVL12-38 (F19)	19319614	829861	
B1-A	B1	field isolate	G39	KVL13-13 (G39)	20194796		
B2-A	B2	field isolate	J65	KVL12-37 (J65)	21682732		
B3-A	B3	field isolate	L105	KVL13-14 (L105)	23054504		
R1-I1	R1-I1	individually-evolved	C17		22225480		
R1-I2	R1-I2	individually-evolved	C17		22198284		
R1-I3	R1-I3	individually-evolved	C17		19574520		
R1-I4	R1-I4	individually-evolved	C17		25471732		
R3-I4	R3-I4	individually-evolved	F19		19686382	2619730	
R1-I9	R1-I9	individually-evolved	C17		29332410		
R3-I5	R3-I5	individually-evolved	F19		19893662		
R1-I6	R1-I6	individually-evolved	C17		20092630		
R3-I6	R3-I6	individually-evolved	F19		21851188		
R1-I7	R1-I7	individually-evolved	C17		29468220		
R1-I8	R1-I8	individually-evolved	C17		22521214		
R1-I10	R1-I10	individually-evolved	C17		22189218		
R1-S1	R1-S1	socially-evolved	C17		20603772		
R1-S2	R1-S2	socially-evolved	C17		31655904		
R1-S3	R1-S3	socially-evolved	C17		22553034		
R1-S4	R1-S4	socially-evolved	C17		25893830		
R3-S9	R3-S9	socially-evolved	F19		20853102		
B2-S9	B2-S9	socially-evolved	J65		24756712		

B2-S5	B2-S5	socially-evolved	J65		19393938	
R3-S6	R3-S6	socially-evolved	F19		22975156	
R3-S7	R3-S7	socially-evolved	F19		23254436	
R1-S8	R1-S8	socially-evolved	C17		24098488	
B2-S8	B2-S8	socially-evolved	J65		22818934	
R1-S10	R1-S10	socially-evolved	C17		21397322	
M. guizhouense ARSEF977		field isolate	*M. guizhouense*	ARSEF977		1214214
R1-A chrA		excised PFGE chromosomal band			2863910	
R1-B chrB		excised PFGE chromosomal band			2594854	
R3-I4 chrA		excised PFGE chromosomal band			2139872	
R3-I4 chrB		excised PFGE chromosomal band			2364024	
M. guizhouense ARSEF977 large chromosome		excised PFGE chromosomal band			2475900	
M. guizhouense ARSEF977 small chromosome		excised PFGE chromosomal band			2171724	

3. Supporting text S1. Detailed methods and materials

3.1. Fungal strains and selection experiment outline

This study analyzed 30 fungal lines of the entomopathogenic fungal genera *Metarhizium robertsii* and *brunneum,* six of which had been used as the starting strains of a selection experiment by Stock et al. (1) in ant hosts (ancestral strains), and 24 of which represent the evolved lines of these starting strains at the end of the experiment. The six starting strains (3 *M. robertsii* strains KVL 12-36 (C17), KVL 12-38 (F19), and KVL 12-35 (E81) and 3 *M. brunneum* strains KVL 13-13 (G39), KVL 12-37 (J65), and KVL 13-14 (L105; all obtained from the University of Copenhagen, Denmark (B. Steinwender, J. Eilenberg and N.V. Meyling)) had been collected from an agricultural field in Denmark by Steinwender et al. (2). After being grown as monospore cultivar, the six strains were mixed in equal amounts (total concentration of 1×10^6 spores ml^{-1}) and were used to co-infect workers of the Argentine ant, *Linepithema humile*, over ten host infection cycles in each of two selection treatments (each in ten independent replicate lines), as detailed in Stock et al. (1). In the "individual treatment", the worker ant remained alone after exposure, whilst it was accompanied by two untreated nestmates in the "social treatment". After each infection cycle, the spores growing out of the first eight dying workers per replicate line (ants dying within the first 24 h after exposure, as well as nestmates were not considered) were harvested, mixed and used to infect a new round of hosts, which were then consistently kept either under the individual resp. social treatment conditions depending on the replicate. After passage 5 and 10 of the experiment, it was determined, which strains were still present in the mix (Stock et al., Fig. 1). At the end of the experiment, 16 lines contained only a single strain genotype (identified by microsatellite analyses) whilst four lines were a mix of two persisting spore types from two different starting strains (see Stock et al. (1)). One spore of each type of these evolved lines was expanded to obtain monospore cultivars of all the strains that had been able to persist after the ten host passages. Phenotypic analyses of the lines (16 single-strain lines and the four 2-strain lines, mixed in the proportion of strain presence at the end of the experiment) found that the lines that had adapted to only the individual immune defenses of their single ant hosts showed increased virulence, whilst the lines adapted to the social immunity of the ants kept in groups, showed increased production of spores with a reduced content of the fungal cell membrane compound ergosterol (1).

In this study, we Illumina-sequenced these 24 evolved lines, as well as their six ancestral strains. We follow the terminology used in Stock et al. (1), in that the three *M. robertsii* starting strains are named R1 to R3 (R1: KVL 12-36 (C17), R2: KVL 12-38 (F19), R3: KVL 12-35 (E81)) and the three *M. brunneum* strains B1-B3 (B1: KVL 13-13 (G39), B2: KVL 12-37 (J65), B3: KVL 13-14 (L105)). To distinguish them from the evolved strains at the end of the experiment, we here extend their numbering by an "-A" for "ancestral, i.e. R1-A, for example, representing the ancestral strain of R1. The evolved strains are numbered

following their replicate line number (Fig. 1, with strain identity (given by microsatellite identification performed in Stock et al.), using “I” for the “individual” and “S” for the “social” treatment, i.e., as an example, the strain that was identified as R3 in the individual treatment replicate line 4, is abbreviated as R3-I4.

In addition to the analysis of the ancestral and evolved strains at the end of the experiment, we were here interested in the spore diversity within our evolved lines at different time points of the experiment, that is after passage 1, 3, 5 and 10. To this end, we plated the stored spore suspensions, picked single-clone colonies and determined their strain identity and the presence / absence of chrA, as detailed below. This was done for all evolved lines that showed R3 being present in passage 5, independent of whether it succeeded into passage 10 or not. Therefore, we analyzed replicates I4, I5 and I6 from the individual treatment and S1, S6, S7, S8, S9 from the social treatment.

3.2. Pulsed-field gel electrophoresis (PFGE)

Spores were grown in liquid LB medium (2% sucrose, 1% peptone, 0.3% yeast extract and 0.5% NaCl [w/V]) at 23°C, 200 rpm for 5-10 days, filtered through a sieve, centrifuged (3000 x g, 10 min, RT), the pellet was washed and centrifuged again in 1 X TE buffer. The pellet was resuspended in 500 µl H_2O and 500 µl 2.2% low melting agarose and filled into plug casts. After solidification, ten plugs were incubated in 5 mL lysis buffer (0.45 M EDTA, pH 8.0, 1% SDS, 1.5 mg/mL Proteinase K) for 24 h at 55°C. After 24 h, the lysis buffer was replaced with fresh lysis buffer and the incubation was repeated. After 48 h, the plugs were washed three times with 1x TE for 20 min each and stored in 0.5 M EDTA. PFGE was performed for 72 h in 1x TBE buffer at 14°C, 3 V/cm, 106° and 250-1000 s switching time in a CHEF Dr III system using Hansenula wingei chromosomes (1.05-3.13 Mb) as size marker (Bio-Rad, Hercules, CA, USA). DNA was stained for 30 min in 0.1 µg/mL Ethidiumbromid in H_2O and destained in H_2O for 10 min before documentation.

To identify the accessory chromosomes represented by the individual bands visible in the PFGE gel, these bands were excised. To approximately 100 mg of excised PFGE band, 260 µl H_2O and 40 µl β-agarase buffer (10x) (New England Biolabs) were added and the agarose plug was melted by incubation at 99°C for approximately 15 min until completely melted, followed by 15 min at 80°C. The mixture was brought to 42°C and 4 units of β-agarase (New England Biolabs) added and further incubated at 42°C, 350 rpm for 90 min. 44 µl of 3 M NaOAc was added and the mixture cooled on ice for 15 min before centrifugation for 15 min at 15000 x g. 350 µl was transferred to a new tube and 240 µl of isopropanol was added and centrifuged for 15 min at 15000 x g, 4°C and the pellet was washed with 700 µl of 70% ethanol, dried and dissolved in 20 µl of H_2O and sequenced. Sequencing was performed using the Ultra Low Input DNA library and Illumina sequenced with 150 PE reads on a NextSeq 2000 sequencer at the Max Planck Genome Centre Cologne, Germany.

3.3. Proportion of chrA-containing spores over the course of the experiment

The spore suspensions of passages 1, 3, 5 and 10 from the eight replicate lines that contained the R3 strain at least until passage 5 of the experiment (see Fig. 1 of Stock et al. (1)), were plated on selective medium agar plates, containing 6.5% Sabouraud dextrose agar (Sigma-Aldrich), 1 ml each of Syllit 450 SC (110 mg/ml; Kwizda), Chloramphenicol (100 mg/ml; Sigma-Aldrich) and Streptomycin (100 mg/ml; Sigma-Aldrich) and incubated for one week at 23°C in the dark. Individual clones were selected and transferred to DNeasy 96-well plates (Qiagen) containing 50 µl of nuclease-free water (Sigma-Aldrich). Samples were homogenized in a TissueLyser II (Qiagen) using approx. 100 mg of glass beads (425-600 µm; Sigma-Aldrich) in two steps (2 × 2 min at 30 Hz). Total DNA of the spore-cultivars was extracted using DNeasy 96 Blood & Tissue Kit (Qiagen) according to manufacturer's instructions, with a final elution volume of 50 µl Buffer AE. R3 clones were identified using the same microsatellite loci as in Stock et al. (1) Ma307 and Ma2054. Presence of chrA was determined in all R3 spore clones using primers designed in this study. For primer sequences see Table S8. Each PCR run included a no-template control and two further control reactions (negative control: R3-A genomic DNA which does not contain the chrA; positive control: R1-A genomic DNA, which contains chrA). To verify that the absence of detection of chrA was not due to PCR failure, each PCR that aimed at detecting the presence/absence of chrA was multiplexed with a set of general *M. robertsii* primers that amplify a region on the core chromosomes, and the results of PCR reactions for the absence/presence of chrA were only included in the analysis if the general *M.robertsii* primers did produce the specific PCR product. All reactions were performed using MyTaq HS Red Mix (Bioline), 10 pmol of each primer (Ma307 only 5 pmol per primer) and 4 µl of genomic DNA. PCR amplifications for strain identification were performed as follows: initial denaturation at 95 °C for 1 min, followed by 35 cycles of 30 s at 95 °C, 1 min at 59°C (Ma307) resp. 60 °C (Ma2054) and 1 min at 72°C and a final extension step at 72°C for 7 min. For the detection of chrA in R3 clones we used three sets of primers targeting different regions (left chromosomal arm, center, right chromosomal arm) of the chrA sequence, to make sure that the whole chrA was transferred. To ensure the specificity of the primers to the intended PCR template, we used Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and tested for off-target amplification against the assembled genomes of R1-A, R3-A and R3-I4. Amplifications were performed as follows: initial denaturation at 95 °C for 1 min, followed by 30 cycles of 15 s at 95 °C, 15 s at 60 °C and 10 s at 72°C. In total 2626 individual spore clones (872 of which were already published in Stock et al. (1) and 1754 of which were produced in the course of the current study) were used. 883 of these clones were found to be R3 and were further analyzed for the presence of chrA. A detailed summary of the results can be found in Supplementary Data S1.

Table S8: Primers used within this study

Aim	Primer name	Sequence (5' → 3')	Product size [bp]
M. robertsii confirmation	Mrobertsii_F Mrobertsii_R	GACGAAAAGATTTGCGGCACC TCACGATCCTTGAGACACGC	R1: 122 R3: 116
R3 strain identification	Ma307_F Ma307_R	CATGCTCCGCCTTATTCCTC GGGTGGCGAAGAAGTAGACG	161
	Ma2054_F Ma2054_R	GCCTGATCCAGACTCCCTCAGT GCTTTCGTACCGAGGGCG	230
chrA detection	ChrA_left_F ChrA_left_R	ACTTGCCGCTCGAAAAACAC CCTATTGGACCTTCCGTGCAA	150
	ChrA_center_F ChrA_center_R	AAAGATCGCGGGTGACAACT AACAGGACATCACGCTCTGC	200
	ChrA_right_F ChrA_right_R	ACAGGCACGTGAGCTTCTAC GTTCTAGCAGTGGTTGACGGA	250

3.4. Generating Genome Assemblies

DNA preparation for sequencing

DNA for Nanopore and PacBio sequencing was extracted using the Blood & Cell Culture DNA Midi Kit (Qiagen). Cells were grown in LB media at 23°C, 200 rpm for five to seven days and cells were collected by centrifugation (2000xg, 10min). For Nanopore and Illumina sequencing the fungal material was harvested by vacuum filtration (Filtermax filter top 0.22 µm, 500 mL). The cells were washed twice using ddH2O and afterwards freeze dried using a FreeZone 2.5 Liter Benchtop Freeze Dry System (Labcono) with the following settings: -50°C, 0.04 mbar and 23°C plate-temperature, ~ 18 hours. For PacBio sequencing the material was harvested by centrifugation (10 min, 3000 x g). Cells were ground to a fine powder in liquid nitrogen and 100 mg of the resulting powder was used for DNA isolation according to the manufacturer's instructions. Nanopore sequencing was performed by the Next Generation Sequencing Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria. PacBio sequencing of *M. guizhouense* was performed at the Max Planck Genome Centre Cologne, Germany using Sequel IIe (Pacific Biosciences). Illumina Sequencing was performed at Eurofins Genomics GmbH (Ebersberg). An overview of the sequencing reads generated in this study is given in Table S8.

Nanopore Sequencing and Assembly generation

Sequencing: R9.4.1 Minion Flow cells. Base calling was performed using guppy v5.0.11 using the model dna_r9.4.1_450bps_hac (high accuracy).For the generation of the nanopore-based assemblies we used the pipeline described in (3). In short: Called Nanopore reads were filtered to reads longer than 5000 bases using NanoFilt (v2.3.0) (4). The 150 PE Illumina Reads employed for the correction of the Nanpore reads were trimmed using Trimmomatic V0.39 (5) The Nanopore reads were then corrected by the trimmed 150 PE Illumina reads using FMLRC 2 (0.1.4) (6) followed by further trimming using Canu (v2.1.1) (7). These corrected and trimmed Nanopore reads were assembled by flye (V 2.8.3) including two rounds of polishing (8). The resulting assembly was further polished using the uncorrected Nanopore reads by Racon (V1.4.20) (9) in two rounds and further polished by Medaka (V 1.4.3). Finally, the assembly was polished four times using the 150 PE Illumina reads by Pilon (V 1.24) (10) after mapping with BWA-MEM2 (V 2.2.1) (11) and samtools (1.12) (12).The final assembly was analyzed using Tapestry (V 1.0.0) using the telomeric sequence TTAGGG (13). Due to the lower number of reads >5000 bases for the R1-A reads >3000bp were used for the Tapestry analysis. All contigs that showed a lower average coverage than 30 and larger than 100 were excluded from the assembly (this affected five contigs of the R3-I4 and 22 contigs of the R1-A and three contigs (130 kb) of the R3-A genome assembly. A script for the bioinformatic pipeline is available at michaelH-git/Metarhizium_chromosome_transfer (github.com) (14).

Table S9: Summary statistics of final genome assemblies (generated with Quast (V 5.0.2) (15).

Statistics without reference	R1-A	R3-A	R3-I4
# contigs	41	18	16
# contigs (>= 0 bp)	41	18	16
# contigs (>= 1000 bp)	41	18	16
# contigs (>= 5000 bp)	40	18	16
# contigs (>= 10000 bp)	38	18	16
# contigs (>= 25000 bp)	36	16	15
# contigs (>= 50000 bp)	30	18	14
Largest contig	7267018	10479110	10465571
Total length	42768942	43163266	45009773
Total length (>= 0 bp)	42768942	43163266	45009773
Total length (>= 1000 bp)	42768942	43163266	45009773
Total length (>= 5000 bp)	42764321	43163266	45009773
Total length (>= 10000 bp)	42746527	43163266	45009773
Total length (>= 25000 bp)	42702419	43106281	44996207
Total length (>= 50000 bp)	42487955	43163266	44953661
N50	4406720	4812484	4807498
N75	2994606	2774337	4300840
L50	4	3	3

L75	7	6	6
GC (%)	49.76	50.29	50.24
Mismatches			
# N's	0	0	0
# N's per 100 kbp	0	0	0

3.5. Generation of Annotations

Transposable Element annotation was generated using the REPET pipeline (16, 17) following the pipeline as described in (18). In short, first a *denovo* annotation (TEdenovo) was created which allowed for the generation of the consensus sequences of TEs. These were subsequently used to annotate the TEs in the genome by two rounds of TEannot.

Gene annotation was generated using Braker 2.0 (version 2.1.6) (19) based on protein homology information from the OrthoDB fungal database (Version 10). The Augustus gene prediction generated by Braker 2.0 was used for all subsequent analysis.

Statistics of Proteins: A total of 12027, 12254, and 12131 genes were predicted in the R1-A, R3-A or R3-I4 genome, respectively. The predicted proteins had the following results of BUSCO Analysis 99.7% (R1-A: 3809/3817 BUSCO groups) or 99.8% (R3-I4: 3810/3817 BUSCO groups) or 99.8% (R3-A: 3809/3917 BUSCO groups)

The proteins were functionally annotated using Blastp Swissprot reference database using blastp (V 2.12.0): For proteins encoded by genes located on chrA the number of GO Annotations were created, mapped and merged using Blast2GO (V6.0.3).

CAZymes were annotated using dbCAN2 (20) using HMMER:dbCAN with thresholds: E-Value < $1e^{-15}$, coverage >0.35) and DIAMOND: CAZy (E-Value <$1e^{-102}$) und HMMER:dbCAN-sub (E-Value <$1e^{-15}$, coverage >0,35).

Secreted Proteins were identified using SignalP (v 6.0) (21) and of these putative effectors were predicted using EffectorP (v 3.0) (22). Secondary Metabolite cluster were predicted by AntiSMASH (v 6.0) (23).

A script for the bioinformatic pipeline is available at michaelH-git/Metarhizium_chromosome_transfer (github.com) (14).

3.6. Comparison with existing reads and assemblies

Publicly available whole genome sequencing reads and assemblies of 30 isolates from species of the genus *Metarhizium* were included in the analysis of the distribution of the accessory chromosomes within the genus. For twelve samples these were assemblies and for 18 samples these consisted of WGS reads (See Table S8 – overview of assemblies and reads, (2))

For the synteny analysis between assemblies the following the assemblies were aligned with nucmer (version 4.0.0rc1) and the matches were filtered for those of min 1000bp length with a minimum identity of 90%. Of these bedfiles for the coverage-analysis were generated and the SNPs in covered regions were determined. Alignments were visualized by using dotPlotly (https://github.com/tpoorten/dotPlotly) in R (see example below):

Rscript --vanilla pafCoordsDotPlotly.R -i R3-I4_Mguizhouense.i90.l1000.paf -q 10000 -m 10000 -p 15 -s -o R3-I4_Mguizhouense.i90.l1000 -k 16

For the analysis of the coverage and SNPs of the whole genome sequencing reads deposited for 18 isolates we first deinterleaved the reads using bbmap (version 39.01) and then removed adapter using trimmomatic (V0.39) before mapping and SNP calling by bowtie2 (version 2.4.4) and bcftools mpileup (version= 1.14) and filtering the called SNPs to high quality (Q>50) in callable regions (DP>6). A script for the bioinformatic pipeline is available at michaelH-git/Metarhizium_chromosome_transfer (github.com) (14).

3.7. SNP calling using the Illumina reads of ancestral and evolved strains

Illumina reads were trimmed as described above and mapped onto the three Nanopore-based (R1-A, R3-A, R3-I4) assemblies using bowtie2 (version 2.4.4), samtools (version 1.3.1). The resulting BAM file was reformatted using the Picard functionality (version 2.24.0) and SNPs called using bcftools mpileup (version 1.14). The resulting raw VCF file was further filtered using bcftools (version 1.14). A script for the bioinformatic pipeline is available at michaelH-git/Metarhizium_chromosome_transfer (github.com) (14).

3.8. Calling of structural variants using nanopore reads

Nanopore reads were mapped to the R3-I4 nanopore-based assembly using minimap2 (version 2.26-r1175) and reformatted using samtools (version 1.17) and structural variants were called using Sniffles2 (version 2.2) (24) and called structural variants filtered by having a AF>0.7 within the sample, to only report structural variants that are supported by the majority of reads. A script for the bioinformatic pipeline is available at michaelH-git/Metarhizium_chromosome_transfer (github.com) (14).

3.9. Phasing of SNPs and small InDels for Metarhizium guizhouense ARSEF977

In order to phase the SNPs and small InDels of the *M. guizhouense* ARSEF977 that showed duplicated coverage of the accessory chrA *M. guizhouense* was re-sequenced PacBio HiFi chemistry. The reads were mapped onto the onR3-I4 assembly using minimap2 (version 2.24-r1122) and samtools (version 1.3.1). SNPs and small InDels were phased using WhatsHap (version1.6). Please note that due to the fact that chrA of the R3-I4 assembly showed a disomic sequencing coverage, a ploidy of two was used for the calling of SNPs and filtering was based on quality (>50) and readdepth (DP>10) using

bcftools (version 1.14). A detailed script for the bioinformatic pipeline is available at michaelH-git/Metarhizium_chromosome_transfer (github.com) (14).

Out of the total 3090 heterozygous SNPs on chrA, 2790 were phased as 0|1 and 294 as 1|0. The relative position of all the heterozygous SNPs to each other remained unclear as not all of them were located in one phase block. We used the distribution of heterozygous SNPs within and between phase blocks to estimate the number of SNPs on each copy of chrA.

The distribution of phased SNPs was non-random, indicating that the phase blocks did not contain an equal number of heterozygous SNPs of both phases. This suggests that the number and distribution of SNPs on each copy of ChrA are not similar. Specifically, most phase blocks contained SNPs of only one phase (25 out of 28 phase blocks in total), while only three phase blocks contained SNPs of both phases, meaning that the location of the SNPs is on two different copies. For example, the largest phase block with SNPs of both phases (chrA: 587009-1035720) consisted of 1076 phased SNPs. Of these, 1064 (98.9%) were phased as 0|1 and only 12 (1.1%) were phased as 1|0 (Fig. 4 B). This distribution of SNPs between the two copies of chrA in one phase block would be highly unlikely if both copies of chrA had the same density and distribution of SNPs ($p < 2.2x10-16$, binomial test). As a result, it was concluded that the two copies of chrA differ in SNP density and distribution.

Due to the presence of several phase blocks along chrA, it was not possible to directly assign phases 0|1 or 1|0 to a specific copy of chrA. Therefore, the number of mixed phase blocks (containing SNPs from both phases) and unmixed phase blocks (containing SNPs from only one phase) were used to estimate the number of SNPs on each of the two copies of chrA. We argued that, conservatively, the relative frequency of unmixed phase blocks is representative of the relative frequency of SNPs on one copy, while the relative frequency of mixed phase blocks is representative of the relative frequency of SNPs on the other copy of chrA. Therefore, we assumed that the frequency distribution of mixed and unmixed phase blocks represents the true distribution of SNPs among the two copies of chrA. Out of the 28 phase blocks, 25 (relative frequency: 89.3%) exclusively contained SNPs from the same phase, while three (10.7%) contained SNPs from both phases. Therefore, copy a contains an estimated 2769 SNPs (calculated as 3084 phased SNPs x 25/28 + 6 non-phased SNPs + 9 homozygous SNPs), while copy b contains approximately 345 SNPs (calculated as 3084 phased SNPs x 3/28 + 6 non-phased SNPs + 9 homozygous SNPs).

3.10. Phylogenetic analysis of putative histone encoding genes

Sequences of genes annotated to encode histones in 16 different fungal species and one oomycete were obtained from FungiDB (release 66). These sequences and the 12 sequences from genes encoding putative histone in the genome of *M. robertsii* strain R3-I4 were aligned using MUSCLE and a maximum likelihood phylogeny with uniform

rates and the Tamura-Nei model for substitutions was determined and tested using 500 bootstrap replications as implemented in the MEGA software package (version 11.0.13). The resulting phylogeny was visualized using the iTOL software (https://itol.embl.de/).

3.11. GO enrichment analysis and determination of gene-wise dn/ds ratios.

GO-term enrichment analysis was performed using the blast2go software package (version 6.0.3). The SNPs determined for *M. guizhouense* on the R3-I4 assembly were used to determine the dN/dS ratios. The SNPs in the VCF were used by bcftools (version 1.17) consensus to generate *M. guizhouense* consensus sequences, and the cds were extracted from these using the AGAT (v1.0.0) software package. The corresponding fasta sequences of the cds from the R3-I4 and *M. guizhouense* consensus were aligned pairwise using clustalo (version 1.2.4), and these pairwise alignments for each transcript were then used in the ape package (version 5.7.1) in R (version 4.2.1) to determine the genewise dN/dS ratio. A detailed script for the bioinformatics pipeline is available at michaelH-git/Metarhizium_chromosome_transfer (github.com) (14).

3.12. Sequencing coverage analysis and distribution of SNPs in 50 kb windows

Calculation of median coverage in 50000 bp non-overlapping windows and the fraction of the windows that is covered by at least 5 reads using mosdepth (version 0.3.3) and recovering the number of SNPs in those windows.

Generating 50 kb windows using bedtools (version v2.25.0)

```
bedtools makewindows -b assembly.bed -w 50000 -i srcwinnum >
assembly_window_50000.bed
mosdepth -t 10 -n -T 5 -b assembly_window_50000.bed out out_RG_Dedup.bam
bedtools intersect -c -a assembly_window_50000.bed -b filtered.vcf > SNP_50kb.counts
```

3.13. Genewise relative synonymous codon usage

Genewise relative codon usage was estimated using BioKIT (version 0.1.3) with the following command:

```
biokit gene_wise_relative_synonymous_codon_usage cds.fasta >
cds.genewise.codonusage
```

3.14. Determining the Cytosine methylation within CpG context.

To determine the whether a cytosine in a CpG context was methylated the base calling of the nanopore reads was repeated using guppy (Version 6.4.8+31becc9), minimmap2 (version 2.24-r1122) Samtools (version 1.13) was used to merge the BAM-files and mod with the model dna_r9.4.1_450bps_modbases_5mc_cg_sup.cfg. Modified bases were called using mod_kit (version 0.1.4) with the following command:

```
modkit pileup mod.bam out.bed --cpg --ref R3-I4.fasta
```

3.15. Statistical tests

Statistical tests were performed on the methylation data (Fig. S4 A) using a two-sided Fisher exact test comparing the count data for the two indicated groups for the total number of CpG sites and the number of methylated sites. The p-values were adjusted for multiple testing using the Benjamini-Hochberg (BH) correction. To compare codon usage bias (Fig. S6 C), all groups were first compared using a Kruskal-Wallis rank-sum test and later pairwise comparisons were calculated using the Wilcoxon rank-sum test (for unpaired data). To compare the dN/dS ratios (Fig. S9 B) all groups were first compared using a Kruskal-Wallis rank-sum test and later pairwise comparisons were calculated using the Wilcoxon rank-sum test (for unpaired data) The exact p-values can be found in Supplementary tables S10-S12.

Table S10: Statistical comparison of methylated CpGs in FigS5A

Results of Fisher's exact text (two sided, BH correction for multiple testing)

Comparison		p value (exact)
R3-A_chrB	R3-A_rest	1
R3-A_chrB	R1-A_chrA	<4.62E-15
R3-A_chrB	R1-A_rest	<4.62E-15
R3-A_chrB	R3-I4_chrA	<4.62E-15
R3-A_chrB	R3-I4_chrB	1
R3-A_chrB	R3-I4_rest	1
R3-A_rest	R1-A_chrA	<4.62E-15
R3-A_rest	R1-A_rest	<4.62E-15
R3-A_rest	R3-I4_chrA	<4.62E-15
R3-A_rest	R3-I4_chrB	1
R3-A_rest	R3-I4_rest	0,0001518
R1-A_chrA	R1-A_rest	<4.62E-15
R1-A_chrA	R3-I4_chrA	<4.62E-15
R1-A_chrA	R3-I4_chrB	<4.62E-15
R1-A_chrA	R3-I4_rest	<4.62E-15
R1-A_rest	R3-I4_chrA	<4.62E-15
R1-A_rest	R3-I4_chrB	<4.62E-15
R1-A_rest	R3-I4_rest	<4.62E-15
R3-I4_chrA	R3-I4_chrB	<4.62E-15
R3-I4_chrA	R3-I4_rest	<4.62E-15
R3-I4_chrB	R3-I4_rest	1

Table S11: Statistical comparison of Gene-wise synonymous codon usage in FigS8C

Kruskal-Wallis rank sum test

data: codon_sub$median by codon_sub$Cat
Kruskal-Wallis chi-squared =
1047.5, df = 4, p-value < 2.2e-16

Pairwise comparisons using Wilcoxon rank sum test

	R3-I4 (rest of genome)	chrA	chrB	R1-A
chrA	<2e-16	-	-	-
chrB	<2e-16	0,016	-	-
R1-A	0,083	<2e-16	<2e-16	-
M. guizhouense	6,00E-05	<2e-16	<2e-16	0,021

p-value adjustment method: BH

Table S12: Statistical comparison of dNdS ratios in FigS9B

Kruskal-Wallis rank sum test

data: dnds by cat
Kruskal-Wallis chi-squared = 121.17, df = 2, p-value < 2.2e-16

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: dnds_data$dnds and dnds_data$cat

	rest	chrA
chrA	<2e-16	-
chrB	8,50E-13	0,24

p-value adjustment method: BH

4. Legends for Datasets S1 to S14

Dataset S1: R3-I4: Functional annotation of accessory chrA

Dataset S2: R3-I4: Genome-wide functional annotation

Dataset S3: Source data for Figure 1C. Sequencing coverage on R3-I4 assembly

Dataset S4: Source data for Figure 1E. Sequencing coverage on R3-I4 assembly

Dataset S5: Source data for Figure 2. Clone and ChrA frequencies

Dataset S6: Source data for Figure 3 AB and Fig S6. Coverage and SNP densities

Dataset S7: Source data for Figure 4. Sequencing coverage of *M. guizhouense* on R3-I4 assembly and SNP densities.

Dataset S8: Source data for Figure 5 B. Fraction of transposable elements (TEs) and genes.

Dataset S9: Source data for Figure S3. Fraction of covered bases and normalized sequencing coverage of excised PFGE chromosomal bands.

Dataset S10: Source data for Figure S4. Normalized sequencing coverage on R3-I4 assembly

Dataset S11: Source data for Figure S5B. Methylation of CpG sites.

Dataset S12: Source data for Figure S8AB. Number and Sequence length of Tes

Dataset S13: Source data for Figure S8C. Gene-wise relative synonymous codon usage

Dataset S14: Source data for Figure S9B. dNdsS ratios.

5. Supplemental References

1. M. Stock, *et al.*, Pathogen evasion of social immunity. *Nat Ecol Evol*, 1–11 (2023).
2. B. M. Steinwender, *et al.*, Molecular diversity of the entomopathogenic fungal Metarhizium community within an agroecosystem. *Journal of Invertebrate Pathology* **123**, 6–12 (2014).
3. Z. Saud, A. M. Kortsinoglou, V. N. Kouvelis, T. M. Butt, Telomere length de novo assembly of all 7 chromosomes and mitogenome sequencing of the model entomopathogenic fungus, Metarhizium brunneum, by means of a novel assembly pipeline. *BMC Genomics* **22**, 87 (2021).
4. W. De Coster, S. D'Hert, D. T. Schultz, M. Cruts, C. Van Broeckhoven, NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
5. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
6. J. R. Wang, J. Holt, L. McMillan, C. D. Jones, FMLRC: Hybrid long read error correction using an FM-index. *BMC bioinformatics* **19**, 1–11 (2018).
7. S. Koren, *et al.*, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
8. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540–546 (2019).
9. R. Vaser, I. Sović, N. Nagarajan, M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
10. B. J. Walker, *et al.*, Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963 (2014).
11. M. Vasimuddin, S. Misra, H. Li, S. Aluru, Efficient architecture-aware acceleration of BWA-MEM for multicore systems in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, (IEEE, 2019), pp. 314–324.
12. H. Li, *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
13. J. W. Davey, S. J. Davis, J. C. Mottram, P. D. Ashton, "Tapestry: validate and edit small eukaryotic genome assemblies with long reads" (2020).
14. Habig, M. Metarhizium Chromosome Transfer. GitHub. https://github.com/michaelH-git/Metarhizium_chromosome_transfer. Deposited 27 December 2023.
15. A. Mikheenko, A. Prjibelski, V. Saveliev, D. Antipov, A. Gurevich, Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).
16. T. Flutre, E. Duprat, C. Feuillet, H. Quesneville, Considering transposable element diversification in de novo annotation approaches. *PloS one* **6**, e16526 (2011).

17. K. Simin, *et al.*, pRb inactivation in mammary cells reveals common mechanisms for tumor initiation and progression in divergent epithelia. *PLoS biology* **2**, e22 (2004).
18. C. Lorrain, A. Feurtey, M. Möller, J. Haueisen, E. Stukenbrock, Dynamics of transposable elements in recently diverged fungal pathogens: lineage-specific transposable element content and efficiency of genome defenses. *G3 (Bethesda)* **11**, jkab068 (2021).
19. T. Br\uuna, K. J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* **3**, lqaa108 (2021).
20. H. Zhang, *et al.*, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **46**, W95–W101 (2018).
21. F. Teufel, *et al.*, SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–1025 (2022).
22. J. Sperschneider, P. N. Dodds, EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. *Mol Plant Microbe Interact* **35**, 146–156 (2022).
23. K. Blin, *et al.*, antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* **49**, W29–W35 (2021).
24. F. J. Sedlazeck, *et al.*, Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15**, 461–468 (2018).