Supporting Information. Boritsch *et al.,*

SI Material and Methods

Mycobacterial strains and growth conditions

M. tuberculosis and *M. canettii* strains were grown in Middlebrook 7H9 medium (Becton-Dickinson) supplemented with ADC or on Middlebrook 7H11 medium (Becton-Dickinson) supplemented with OADC at 37°C. *M. smegmatis* strains were grown in Luria-Bertani (LB) broth or on LB agar plates. Hygromycin (50 µg/ml), kanamycin (25 µg/ml), apramycin (50 µg/ml) or rifampicin (2 µg/ml) was used for mycobacterial selection. *M. canettii* and *M. tuberculosis* donor strains were generated by integration of a hygromycin resistance cassette, encoded either on cosmids pYUB412 or F10, which both integrate into the attB site in the *glyV*-tRNA of *M. tuberculosis* and *M. canettii* genomes (2). F10 generates green fluorescent colonies due to the high expression of an EsxA-GFP fusion construct (3, 4).

Mating assay on filter plates

Mating assay was performed as described previously (5). Briefly, bacteria were grown in either LB medium (*M. smegmatis)* or 7H9 supplemented with ADC +0.05% Tween (*M. tuberculosis* and *M. canettii)* until an OD of 1-1.5. 1 ml of OD 1.5 of bacteria suspensions was harvested, washed once and taken up in 7H9 supplemented with ADC. 0.5 ml of two mating pair suspensions were combined with a slight excess of recipient cells, filtered through a 0.22 µm membrane filter (Swinex filter holder, Merck Millipore) and filters placed with the bacterial side up on either trypticase soy agar (TSA) or 7H11 supplemented with OADC agar plates without antibiotics. After incubation at 30˚C or 37˚C for 3-7 days, bacteria were washed off the filters and plated on double-antibiotic selection plates.

For analysis of whether linear or circular DNA is taken up by bacteria, cosmid pEB1a was used either in its circular form or linearized through digestion with *EcoRI* (NEB). This linearized fragment contains an apramycin resistance cassette flanked by > 700 bp of homologous genome sequence of *M. tuberculosis* (genome position 1885.355 kb to 1887.022 kb) on either side. 500 ng of DNA was added to bacteria on filter plates and incubated at 30˚C for 2-4 days. For Mitomycin C assay 1.7 ng, 30 ng or 100 ng were added to bacteria mixes on filter plates and incubated at 37˚C for 3 days. Finally, bacteria were washed off the filters and plated on double-antibiotic selection plates. For stress assays, either 200 µM diethylenetriamine NONOate (Acros Organics) or 4 mM H_20_2 (Sigma-Aldrich) were added to bacterial suspensions and incubated at 37°C for 24 hrs and 2 hrs prior mating, respectively. For low pH conditions, 7H9-grown bacteria were pelleted, taken up in phosphate citrate buffer (prepared from 200 mM sodium phosphate and 100 mM citric acid and adjusted to a pH of 6.1 with HCl) and incubated at 37°C for 24 hrs prior mating.

Resistant colonies were verified to contain the antibiotic resistance cassettes with the following oligo pairs: hygroF (AGAGCACCAACCCCGTACT) and hygroR (TCCGGGAAGACCTCGGAA); kanaF (GCGATAATGTCGGGCAATCA) or kanaR (GAGGCAGTTCCATAGGATGG) or apraF (TCGGTCAGCTTCTCAACCTT) or apraR (TACCTGCCCATCGAGTTCAT). Strain-specific oligo pairs were used as follows : *M.can*-L-F (GCTGCTCATGCAGTTTGGTA) and *M.can*-L-R (CAAGCCATGCTACGTCTGAA); *M.can*-A-F (TGAGTCAGCGAAACAGGTTG) and *M.can*-A-R (CGGTGAGTTCGATTCCTGTT).

Whole genome sequencing of recombinants and Deep Sequencing Analysis
Chromosomal DNA of recombinants was prepared as previously described above (6). After library preparation, recombinant genomes were sequenced using Illumina HiSeq 2000 technology. Genome sequencing yielded on average 14 million reads for the *M. smegmatis* and *M. canettii* recombinants. All reads were submitted to FastQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and, if necessary, trimmed off using Trimmomatic v0.30 (7). Spades 3.1.1 (8) was used for assembly of trimmed reads with the kmer option '-k 21,51,71 and the option '--careful'. The metrics of sequence assemblies are shown in Supplementary Table 5. Contigs were ordered using CONTIGuator 2.7.4 (9) with the *M. smegmatis* strain mc^2155 complete genome (Genbank NC_008596.1, May 2014) as reference for *M. smegmatis* recombinants and with the *M. canettii* CIPT 140070008 genome (Genbank F0203508) as reference for the *M. canettii* recombinants. An overview of the general sequence assembly is shown in Supplementary Table 5. DNA sequence comparison and SNPs/Indels calling were performed with the MUMmer package programs (nucmer, delta-filter, show-snps, show-coords) (10) from the alignment between the ordered contigs to either *M. smegmatis* strains mc^2155 and mc^2874 (MKD8) genomes (Genbank CM001762.1) or to *M. canettii* CIPT 140010059 (Genbank NC_015848.1) and *M. canettii* CIPT 140070008 (Genbank F0203508) genomes. ACT was used for mismatch visualization (11) . A python script was used to parse the positions of mutations in the show-snps output. If two mutations were positioned less than 100 kb from one another, they were considered to be within the same block. The different sizes of each transferred fragments were reexamined manually using ACT.

Generation of spontaneously rifampicin resistant mutants
$1x10^8$ bacteria were plated on 7H11 supplemented with OADC agar plates containing 20 µg/ml rifampicin (10x MIC) and incubated at 37˚C for 3-4 weeks. Resistant colonies were verified by Sanger sequencing to contain mutations in *rpoB* with the oligos rpoB-F (CAGGACGTGGAGGCGATCAC) and rpoB-R (ACGGGTTGACCCGCGCGT).

Intracellular survival of *M. canettii* inside *A. castellanii* and co-infection study
Amoeba cells were grown in amoeba growth medium (AGM; 2% protease bacto peptone, 0.1% yeast extract, 4 mM $MgSO_4$, 0.4 M $CaCl_2$, 0.1% sodium citrate dihydrate, 0.05 mM $Fe_3(PO_4)_2$, 2.5 mM Na_2HPO_4 dibasic, 2.5 mM KH_2PO_4 monobasic, pH 6.5, supplemented with 0.1 M of glucose). Prior infection, amoeba cells were washed and resuspended in minimal medium (AGM without protease bacto peptone, yeast extract and glucose). $7.5x10^4$ amoeba/well were distributed in 96 well plates and infected with an MOI of 10 (10 bacteria per amoeba cell) of bacteria resuspended in amoeba minimal medium. After 4 hours of infection at 37˚C and 5% CO_2, extracellular bacteria were removed by incubation of amoeba with 0.1 $mg{\cdot}ml^{-1}$ amikacin for 1 h and finally amoeba were incubated for 4 days in new minimal medium at 37˚C and 5% CO_2. At various time points amoeba were lysed by centrifugation for 3 min at 14500 rpm and subsequent vortexing for 30 seconds. Bacteria were plated in serial dilutions on 7H11 supplemented with OADC plates to determine intracellular survival. For co-infection study of *M. smegmatis* donor and recipient cells, amoeba were seeded in 12 well plates and infected at a MOI of 50. Amoeba were lysed every second day for 10 days as described above and bacteria were plated on 7H11 supplemented with OADC agar plates containing hygromycin and kanamycin to select for recombinants. For *M. canettii* and *M. tuberculosis* co-infection studies amoeba were seeded in 96 well plates and infected at a MOI of 10 of a 1:1 mix of donor and recipient bacteria cells and infection performed as described above. Prior to microscopy, samples were fixed in 4% PFA at RT overnight. Samples were imaged using an automated inverted fluorescent widefield microscope Nikon Ti with 20X objective driven by Metamorph (*M. canettii* and *M. tuberculosis)* or using a Leica TCS SP5 Confocal with 63.0x1.40 objective (*M. smegmatis).*

Recombination analysis. Two genomic approaches using SplitsTree and ClonalFrameML were taken to measure recombination in the *M. canettii* laboratory recombinants RC1 and RC2 by comparison to a reference collection of 23 previously described tubercle bacilli (12, 13). The accession numbers for the sequence files used are listed in Supplementary Table 6. Raw fastq files for RC1, RC2 and the 23 reference isolates were mapped to the *M. canettii* A reference genome (13) using Snippy (v2.9, github.com/tseemann/snippy.git). The resulting core SNP alignment was used as input to NeighbourNet analysis (SplitsTree, v 4.13.1) (14). For ClonalFrameML, a maximum likelihood phylogeny was first inferred using the same core SNP alignment and FastTree with the GTR model of nucleotide substitution. The resulting tree was rooted using *M. marinum* as an outgroup for subsequent ClonalFrameML analysis with the whole genome alignment file from Snippy (includes invariant sites) as inputs (15). The final ClonalFrameML recombination map was re-rendered using a custom Python script to more readily highlight extant and ancestral recombination events in the population (github.com/kwongj/cfml-maskrc/blob/master/cfml-maskrc.py).

1 Derbyshire KM & Gray TA (2014) Distributive Conjugal Transfer: New Insights into Horizontal Gene Transfer and Genetic Exchange in Mycobacteria. *Microbiol Spectr* 2(1).(pii):04.

2 Pham TT, Jacobs-Sera D, Pedulla ML, Hendrix RW, & Hatfull GF (2007) Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. *Microbiology.* 153(Pt 8):2711-2723.

3 Brodin P*, et al.* (2005) Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of Mycobacterium tuberculosis, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity. *J Biol Chem* 280(40):33953-33959.

4 Bange FC, Collins FM, & Jacobs WR, Jr. (1999) Survival of mice infected with Mycobacterium smegmatis containing large DNA fragments from Mycobacterium tuberculosis. *Tuber Lung Dis* 79(3):171-180.

5 Parsons LM, Jankowski CS, & Derbyshire KM (1998) Conjugal transfer of chromosomal DNA in Mycobacterium smegmatis. *Mol Microbiol* 28(3):571-582.

6 Boritsch EC*, et al.* (2016) *pks5*-recombination-mediated surface remodelling in *Mycobacterium tuberculosis* emergence. *Nature Microbiology* 1:15019

7 Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114-2120. doi: 2110.1093/bioinformatics/btu2170. Epub 2014 Apr 2111.

8 Nurk S*, et al.* (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol.* 20(10): 714-737. doi: 710.1089/cmb.2013.0084.

9 Galardini M, Biondi EG, Bazzicalupo M, & Mengoni A (2011) CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med.* 6:11.(doi):10.1186/1751-0473-1186-1111.

10 Kurtz S*, et al.* (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12. Epub 2004 Jan 2030.

11 Carver TJ*, et al.* (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* 21(16):3422-3423.

12 Comas I*, et al.* (2010) Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nat Genet* 42(6): 498-503.

13 Supply P*, et al.* (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. . *Nat Genet* 45:172-179.

14 Huson DH & Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2):254-267.

15 Didelot X & Wilson DJ (2015) ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol.* 11(2):e1004041. doi: 1004010.1001371/journal.pcbi.1004041. eCollection 1002015 Feb.

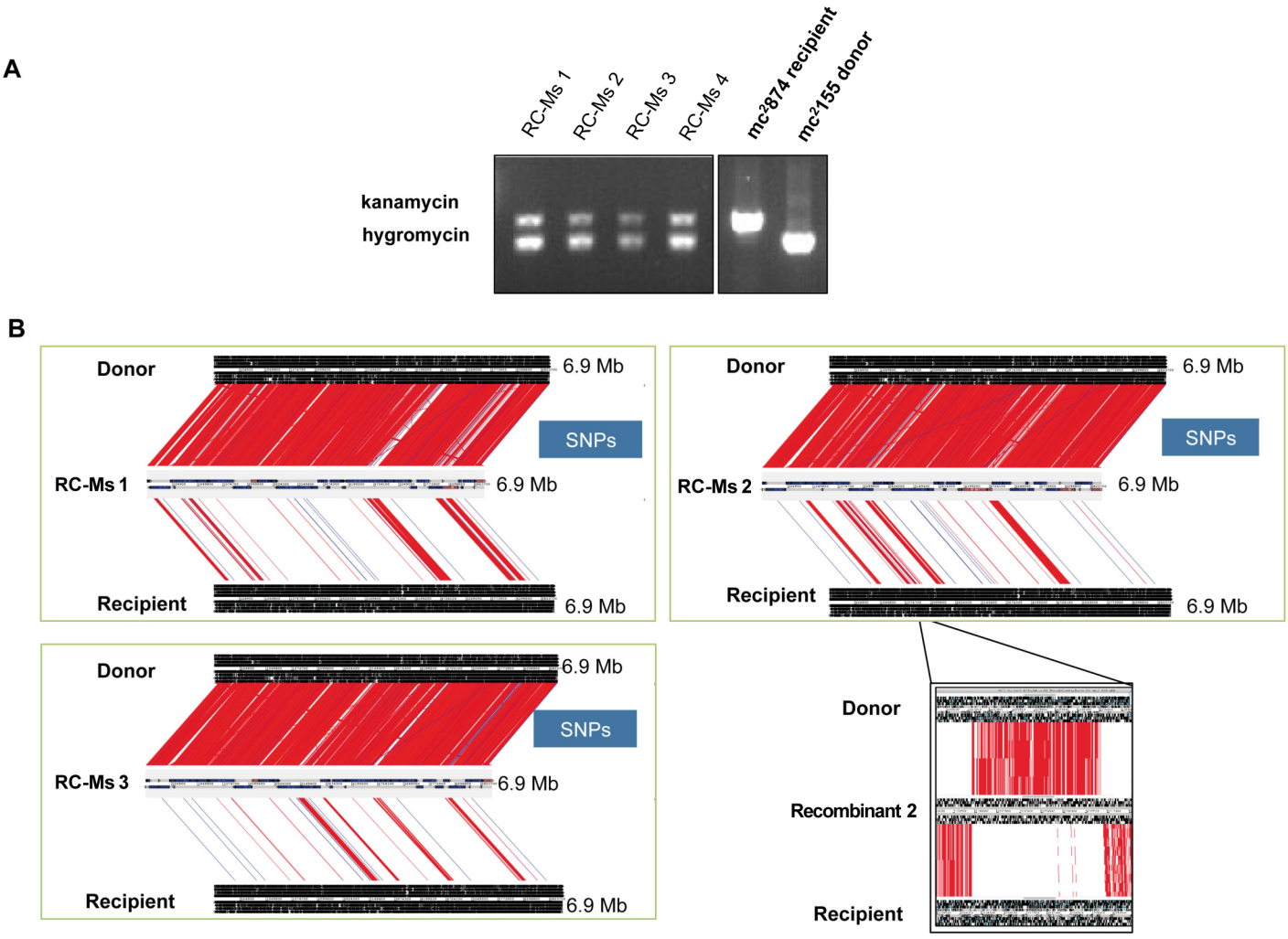


Fig. S1. Distributive conjugal transfer in *Mycobacterium smegmatis*.
(A) Results of PCR analysis of recombinant strains, as well as the recipient *M. smegmatis* mc²874 and the donor *M. smegmatis* mc²155 with oligonucleotides amplifying either the kanamycin or the hygromycin resistance cassettes. (B) ACT visualization of SNPs identified between 3 recombinant (TC) genomes (middle genome of each panel) and either the donor *M. smegmatis* mc²155 genome (top) or the recipient *M. smegmatis* mc²874 genome (bottom). Note that the whole genomes of the strains are displayed (6.9 Mb). The enlargement of one transferred sequence block in the genome of transconjugant 2 is shown (bottom right). SNPs are represented by red lines and Indels by blue lines. RC, recombinant.

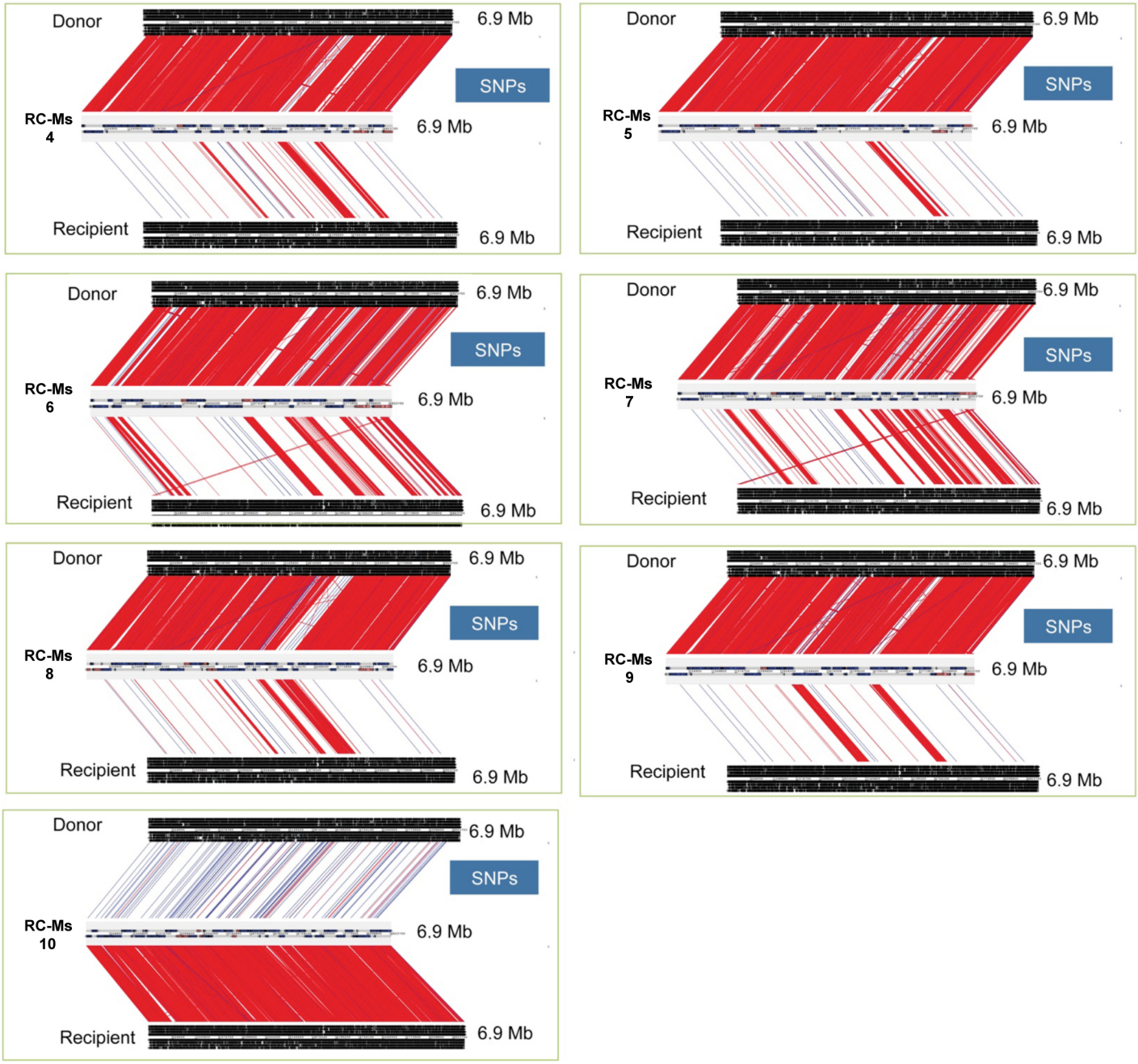


Fig. S2 Distributive conjugal transfer in *M. smegmatis*.
ACT visualization of SNPs identified between the 7 additional recombinant (RC) genomes (middle genome of each panel) and either the donor *M. smegmatis* mc^2155 genome (top) or the recipient *M. smegmatis* mc^2874 genome (bottom). Note that the whole genomes of the strains are shown (6.9 Mb). SNPs are represented by red lines and Indels by blue lines. RC, recombinant.

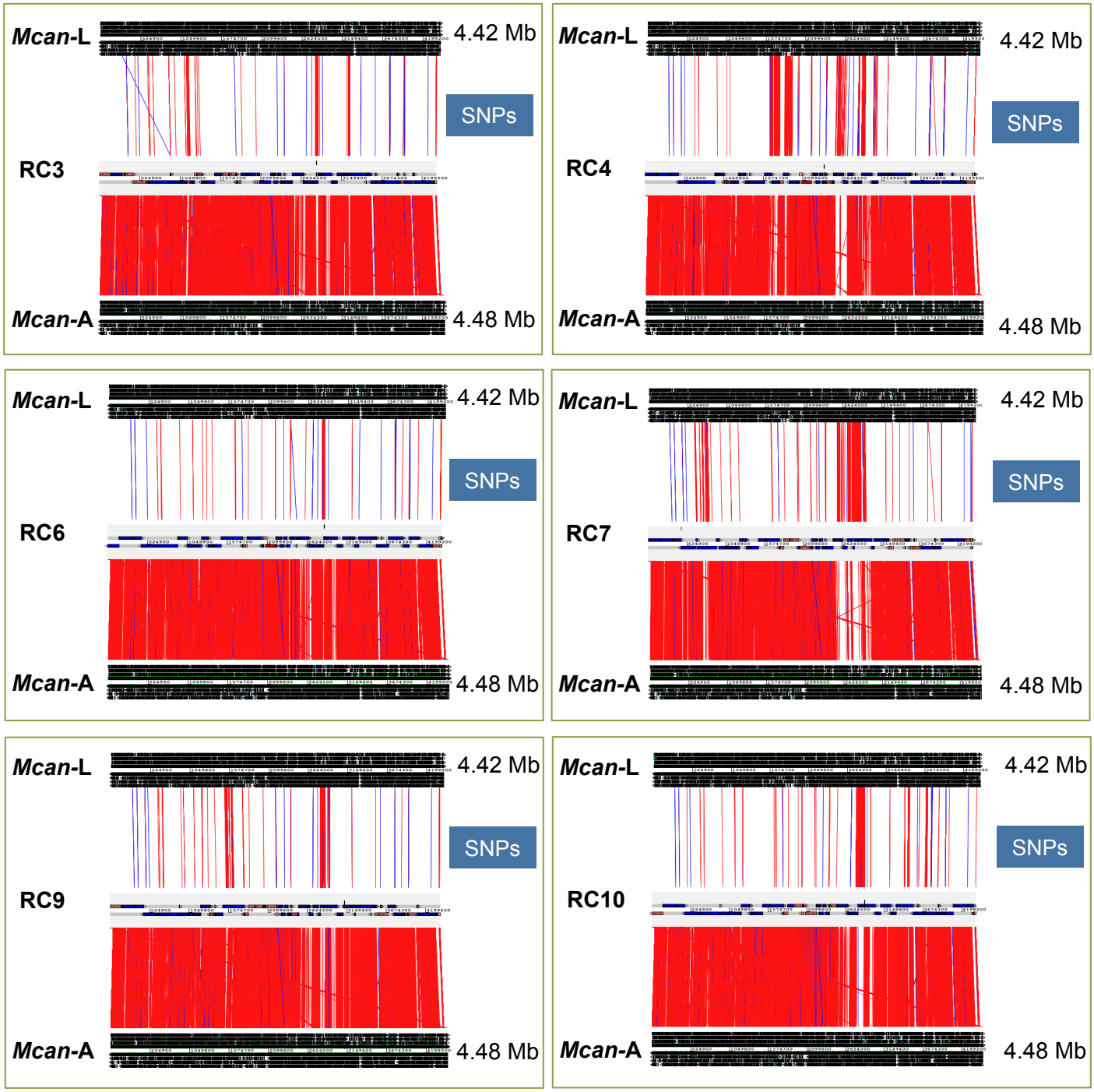


Fig. S3. Horizontal DNA transfer in *Mycobacterium canettii.*
ACT visualization of SNPs identified between the 6 additional recombinant (RC) genomes (middle genome of each panel) and either the recipient *M. canettii* L genome (top) or the donor *M. canettii* A genome (bottom). Note that the whole genomes of the strains are shown (4.42 and 4.48 Mb). SNPs are represented by red lines and Indels by blue lines. *Mcan*, *M. canettii*; RC, recombinant.

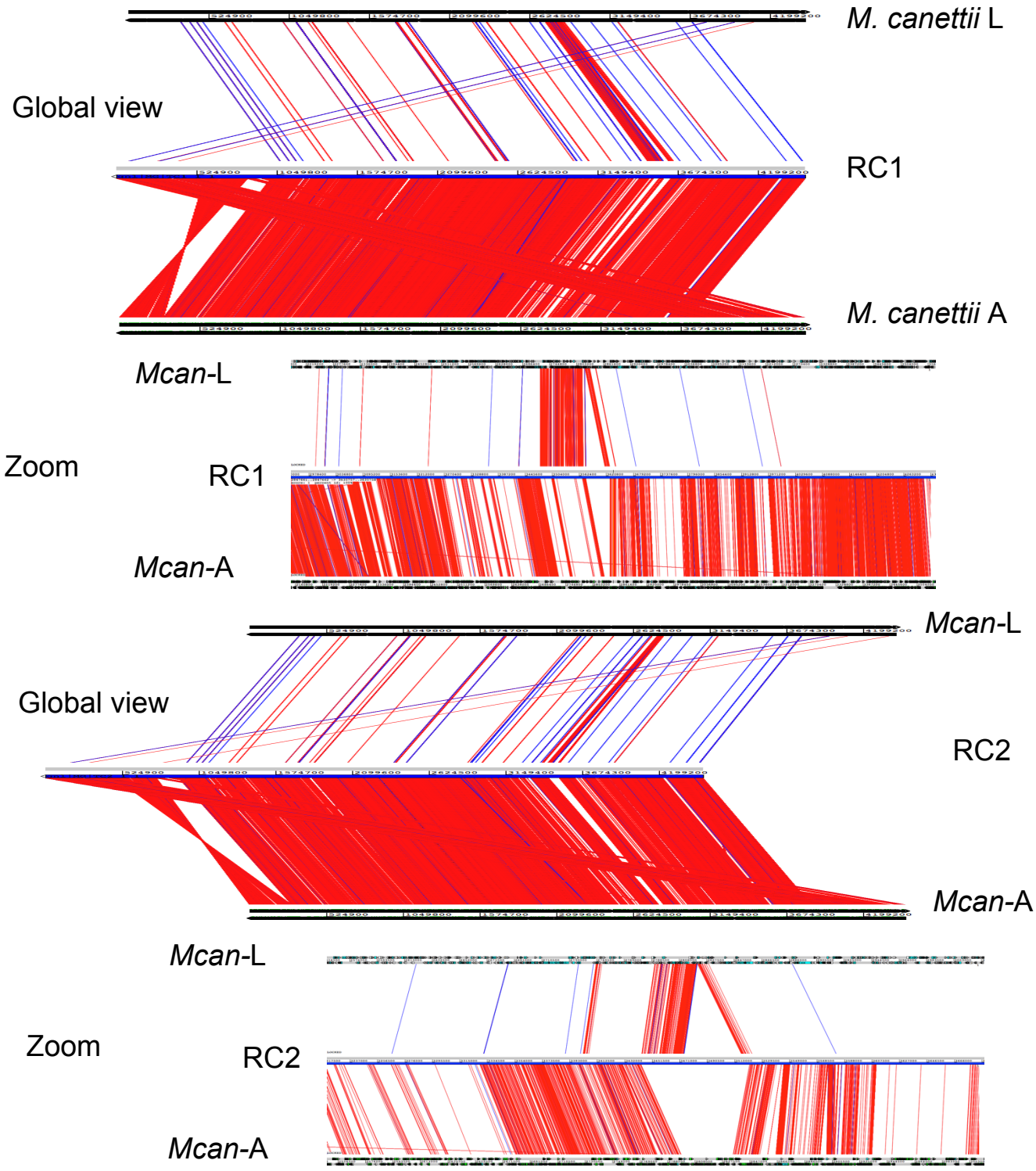
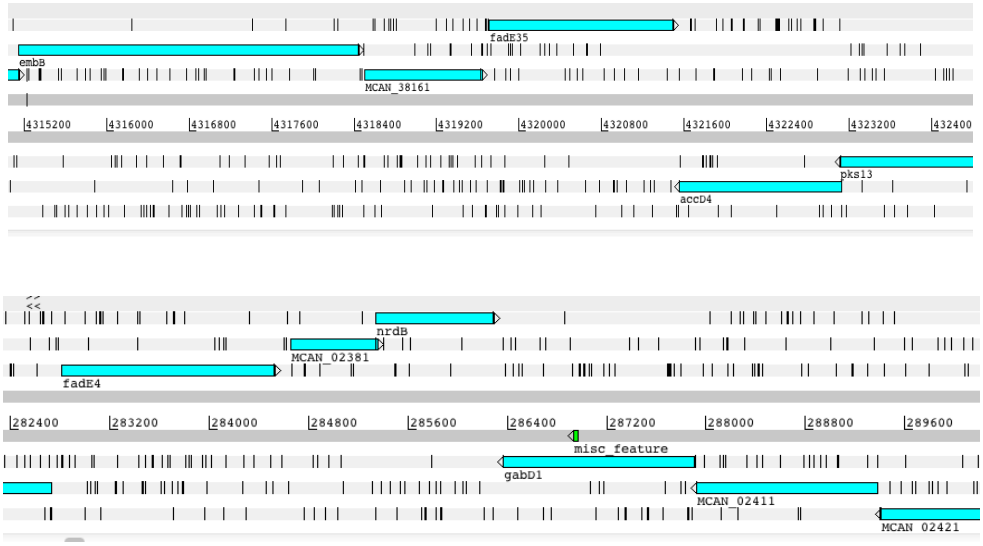
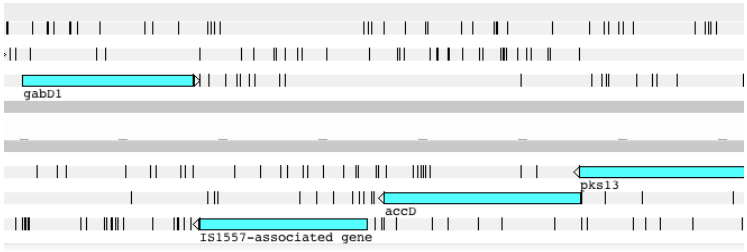


Fig. S4 : ACT visualization of SNPs identified by PacBio sequencing of recombinants (RC1 and RC2) of *M. canettii* strains A (donor) and L (recipient) confirming results obtained by Illumina based sequencing. Note that this analysis using 10 kb long sequence reads also identified a putative assembly error in the original *M. canettii* L (STB-L, CIPT 140070008) GenBank FO203508) genome sequence, where a fragment between two insertion sequences in the region over-spanning the origin of replication seems to be inversed relative to other *M. canettii* strains and strains of the MTBC.



Situation in *M. canettii* strain A (CIPT 140010059) → no IS*1557* inserted between *fadE35* & *accD4* or *nrdB* & *gabD1*

(same for *M. canettii* strains D, J or K)



Situation in *M. canettii* strain L (140070008) as determined by PacBio genome sequencing → IS*1557* between *fadE35* & *nrdB* and IS1557 between *accD4* & *gabD1* → probable inversion of genetic segment between *fadE35* and *gabD1*, spanning the origin of replication.

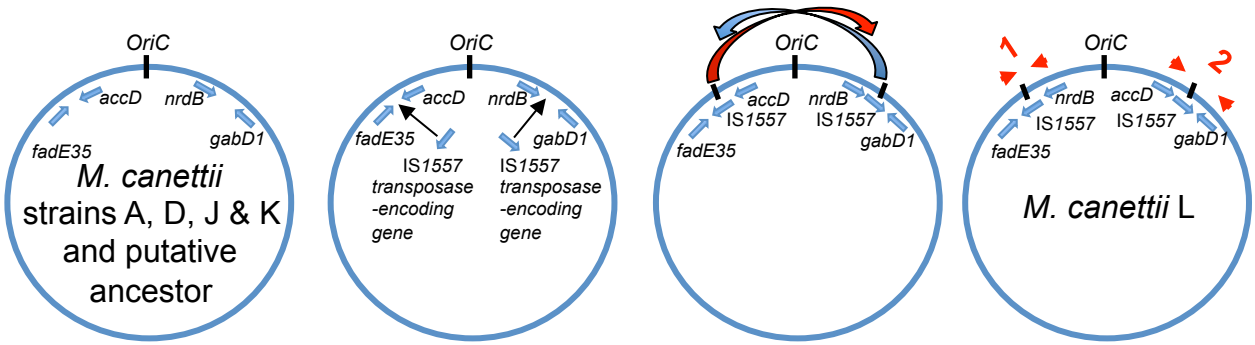


Fig. S5A: Model for occurrence of putative genomic inversion of a 435 kb-sized genomic segment spanning the origin of replication in strain *M. canettii* L, as observed by large sequence read genome sequencing (PacBio) of recombinants RC1 and RC2 (Fig. S4). Results suggest that the inversion is already present in wildtype strain *M. canettii* L (STB-L), unlike previously reported in GenBank entry F0203508. 1, 2, fragments to be amplified as shown in Fig. S5B.

B

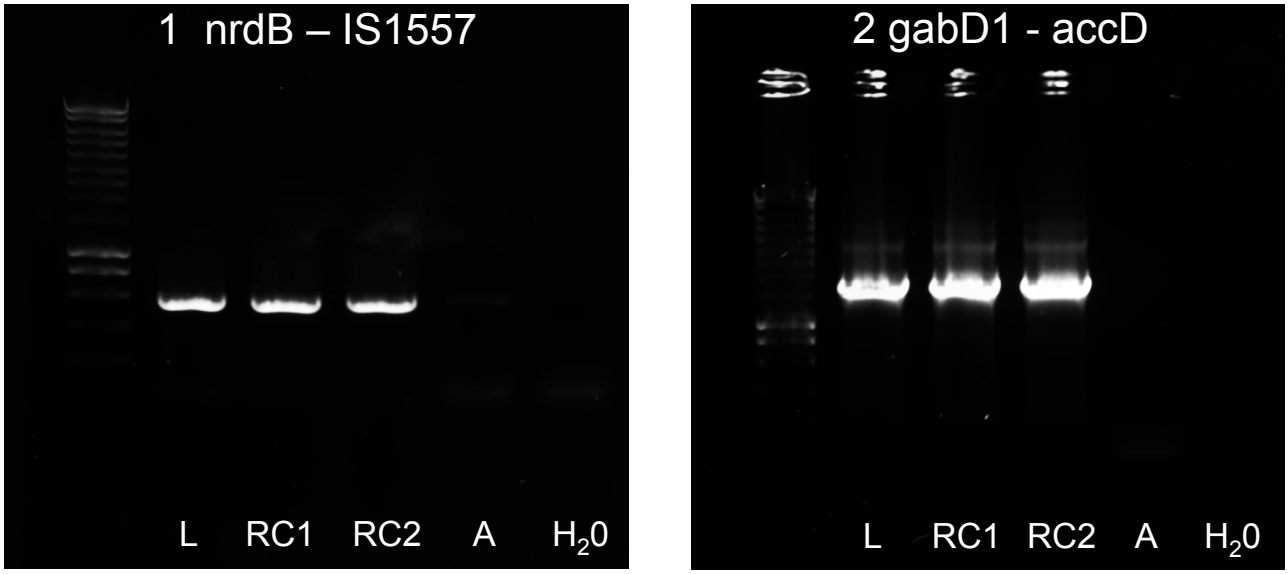


Fig. S5B: PCR results suggest that the inversion is already present in wildtype strain *M. canettii L* (STB-L), unlike previously reported in GenBank entry F0203508. PCR #1, spanning the genomic region from *nrdB* to IS1557, was done using the oligos **L-inv-279000_F** 5' CTACTACGCCTGGCACAAGA 3' and **L-inv-279000_R** 5' GTTGTCGGCTCCTTGTGTTT 3'; PCR #2; spanning the genomic region from *gabD1* to *accD,* was done using the oligos **L-inv-gabD1_F** 5' ACGGTGTCCTATCCCGAAT 3' and **L-inv-accD5_R** 5'TCATCGACCCGCATGAGAC 3'.

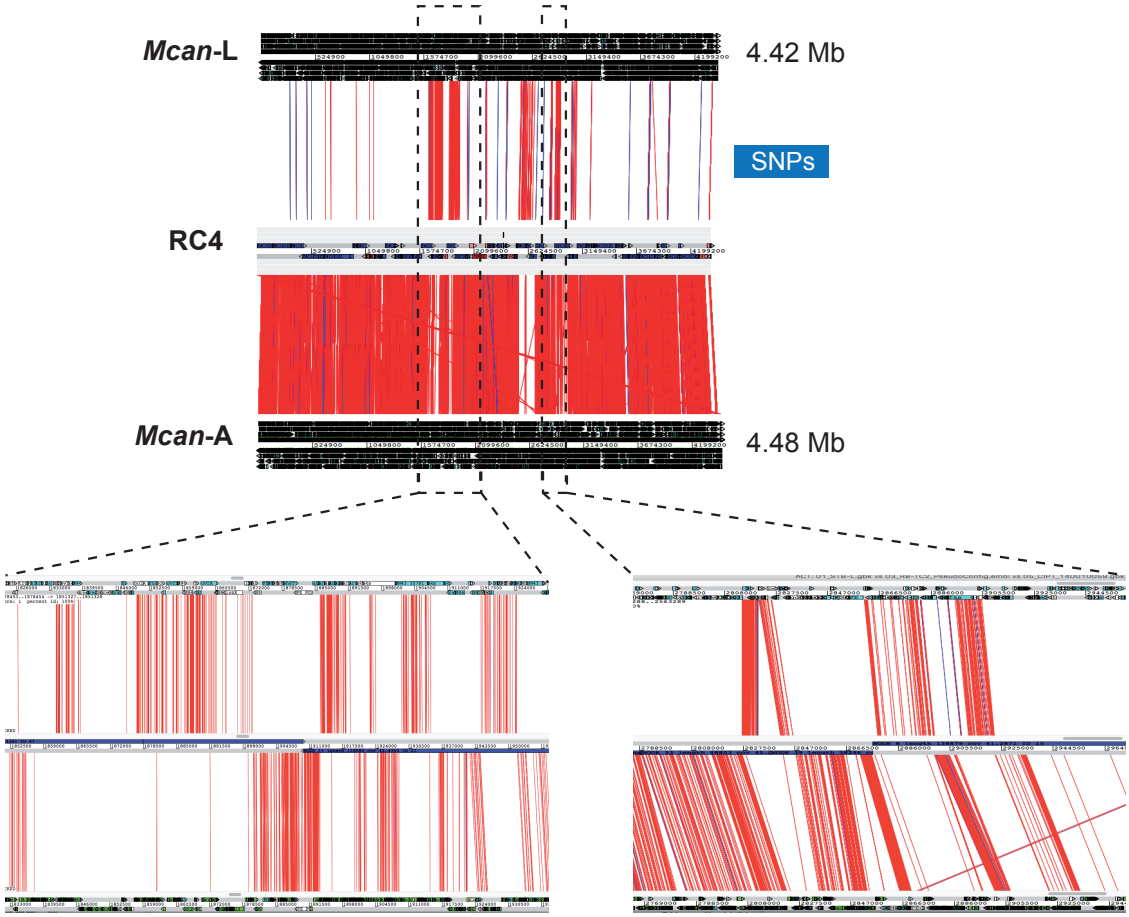


Fig. S6. Microheterogeneity in *M. canettii* recombinant 4.
ACT visualization of SNPs identified between the genome of RC4 (middle) and either the recipient *M. canettii* L genome (top) or the donor *M. canettii* A genome (bottom) including an enlargement of selected donor-derived sequence blocks interspersed by recipient DNA (microheterogeneity) (bottom panels). Note that the whole genomes of the strains are shown (4.42 and 4.48 Mb). SNPs are represented by red lines and Indels by blue lines. RC, recombinant.

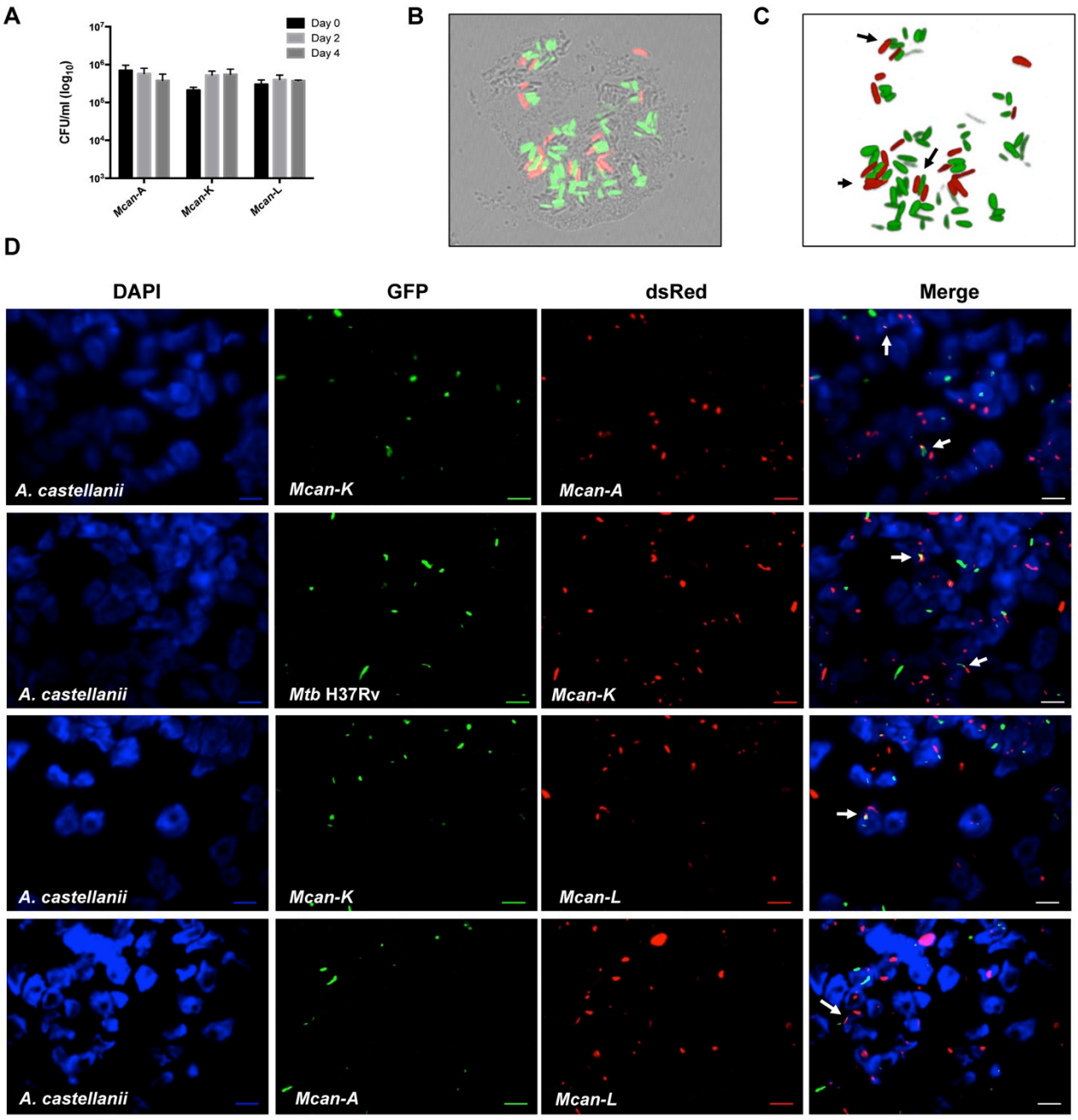


Fig. S7. Co-localization of different mycobacterial strains inside single amoeba cells. (A) Survival of *M. canettii* strains A, K and L inside *A. castellanii* amoeba cells. Amoeba were infected with various strains at an MOI of 10. CFU of intracellular bacteria were determined 3 h and at indicated timepoints post infection. The figure shows CFU numbers of intracellular bacteria. Data are represented as means and standard deviation of three independent experiments. (B-C) Co-infection of *A. castellanii* with *M. smegmatis* mc^2155 (green) and *M. smegmatis* mc^2847 (red). Arrows indicate potential contacts of the different strains. (D) Co-infection of *A. castellanii* with different combinations of *M. canettii* strains or *M. tuberculosis* H37Rv 4 h post infection. White arrows indicate potential inter-strain contacts. Scale bar, 10 µm.

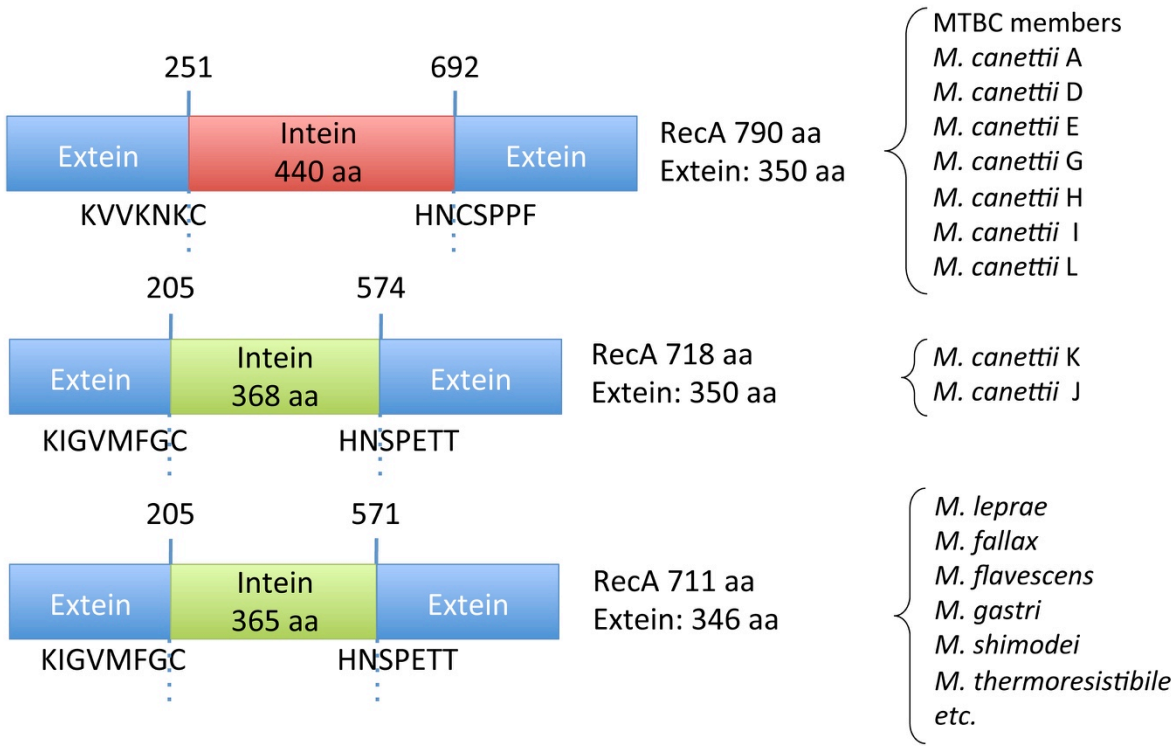


Fig. S8. Sequence comparisons between intein-containing RecA proteins of different *M. tuberculosis* complex and *M. canettii* strains show sequence differences between *M. canettii* strains J and K with other tubercle bacilli. The latter sequences resemble the intein-containing RecA proteins of non-tuberculous mycobacteria.

Fig. S9: CLUSTAL 2.0.12 multiple sequence alignment of N-terminal sequence (350 of the 1094 amino acids) of RecB in different mycobacterial strains

```
M. canettii J |recB|    MDRFNLLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. canettii K |recB|    MDRFNLLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. canettii A |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. canettii I |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. canettii L |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. canettii E |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. canettii H |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. canettii D |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. tub.CDC1551|recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. tub. H37Rv |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. afric. L6  |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
M. bovis      |recB|    MDRFELLGPLPREGTTTVLEASAGTGKTFALAGLVTRYLAETAATLDEML
                        ****:**************************************************
M. smegmatis  |recB|    VKVFDLLGPLPAPNTTTVLEASAGTGKTFALAGLVTRFVAEGVATLDQML


M. canettii J |recB|    LITFNRAASRELRERVRGQIVEAVGALQGGAPPSGALVEHLLRGSDAERA
M. canettii K |recB|    LITFNRVASRELRERVRGQIVEAVGALQGGAPPSGELVEHLLRGSDAERA
M. canettii A |recB|    LITFNRAASRELRERVRGQIVEAVGALQGGAPPSGELVEHLLRGSDAERA
M. canettii I |recB|    LITFNRAASRELRERVRGQIVEAVGALQGGAPPSGELVEHLLRGSDAERA
M. canettii L |recB|    LITFNRAASRELRERVRGQIVEAVGALQGGAPPSGELVEHLLRGSDAERA
M. canettii E |recB|    LITFNRAASRELRERVRGQIVEAVGALQGGAPPSGELVEHLLRGSDAERA
M. canettii H |recB|    LITFNRAASRELRERVRGQIVEAVGALQGGAPPSGELVEHLLRGSDAERA
M. canettii D |recB|    LITFNRAASRELRERVRGQIVEAVGALQGGAPPSGELVEHLLRGSDAERA
M. tub.CDC1551|recB|    LITFNRAASRELRERVRGQIVEAVGALQGDAPPSGELVEHLLRGSDAERA
M. tub. H37Rv |recB|    LITFNRAASRELRERVRGQIVEAVGALQGDAPPSGELVEHLLRGSDAERA
M. afric. L6  |recB|    LITFNRAASRELRERVRGQIVEAVGALQGDAPPSGELVEHLLRGSDAERA
M. bovis      |recB|    LITFNRAASRELRERVRGQIVEALGALQGDAPPSGELVEHLLRGSDAERA
                        ******.****************:*****.***** **************
M. smegmatis  |recB|    LITFGRAASQELRERVRAQIVAALVALDDPSRACNDLEEYLVK---TDQQ


M. canettii J |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. canettii K |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. canettii A |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. canettii I |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. canettii L |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. canettii E |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. canettii H |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. canettii D |recB|    QKRLRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. tub.CDC1551|recB|    QKRSRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. tub. H37Rv |recB|    QKRSRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. afric. L6  |recB|    QKRSRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
M. bovis      |recB|    QKRSRLRDALANFDAATIATTHEFCGSVLKSLGVAGDNAADVELKESLTD
                        *** **********************************************
M. smegmatis  |recB|    ARRRLRDALAGFDAATIATTHQFCQIVLKSLGVAGDSDAGVTLVESLDD


M. canettii J |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVNDPCAQLRPRDPEP
M. canettii K |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVDDPCAQLRPRDPEP
M. canettii A |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVDDPCAQLRPRDPEP
M. canettii I |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVDDPCAQLRPRDPEP
M. canettii L |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVDDPCAQLRPRDPEP
M. canettii E |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVDDPCAQLRPRDPEP
M. canettii H |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVDDPCAQLRPRDPEP
M. canettii D |recB|    LVTEIVDDRYLANFGRQETDPDLTYAEALALALAVVDDPCAQLRPRDPEP
M. tub.CDC1551|recB|    LVTEIVDDRYLANFGRQETDPELTYAEALALALAVVDDPCAQLRPPDPEP
M. tub. H37Rv |recB|    LVTEIVDDRYLANFGRQETDPELTYAEALALALAVVDDPCAQLRPPDPEP
M. afric. L6  |recB|    LVTEIVDDRYLANFGRQETDPELTYAEALALALAVVDDPCAQLRPPDPEP
M. bovis      |recB|    LVTEIVDDRYLANFGRQETDPELTYAEALALALAVVDDPCAQLRPPDPEP
                        **********************:**************:******** ****
M. smegmatis  |recB|    LVSEIVDDLYLAHFGGQKDDPELSYPEALKLARVVVGNPATQLRPRDPDP
```

Fig. S9 (cont.): CLUSTAL 2.0.12 multiple sequence alignment of N-terminal sequence (350 of the 1094 amino acids)of RecB in different mycobacterial strains

```
M. canettii J |recB|              GSKAAVRLRFATEVLEALEHRKGRLRVQGFDDLLIRLASALEAADSPARD
M. canettii K |recB|              GSKAAVRLRFATEVLEELEHRKGRLRVQGFDDLLIRLASALEAADSPARD
M. canettii A |recB|              GSKAAVRLRFATEVLEELEHRKGRLRVQGFNDLLIRLASALEAADSPARD
M. canettii I |recB|              GSKAAVRLRFATEVLEALEHRKGRLRVQGFNDLLIRLASALEAADSPARD
M. canettii L |recB|              GSKAAVRLRFATEVLEALEHRKGRLRVQGFNDLLIRLASALEAADSPARD
M. canettii E |recB|              GSKAAVRLRFATEVLEALEHRKGRLRVQGFNDLLIRLASALEAADSPARD
M. canettii H |recB|              GSKAAVRLRFATEVLEALEHRKGRLRVQGFNDLLIRLASALEAPDSPARD
M. canettii D |recB|              GSKAAVRLRFATEVLEALEHRKGRLRVQGFNDLLIRLASALEAADSPARD
M. tub.CDC1551|recB|              GSKAAVRLRFAAEVLEELERRKGRLRAQGFNDLLIRLATALEAADSPARD
M. tub. H37Rv |recB|              GSKAAVRLRFAAEVLEELERRKGRLRAQGFNDLLIRLATALEAADSPARD
M. afric. L6  |recB|              GSKAAVRLRFAAEVLEELERRKGRLRAQGFNDLLIRLATALEAADSPARD
M. bovis      |recB|              GSKAAVRLRFAAEVLEELERRKGRLRAQGFNDLLIRLATALEAADSPARD
                                  ***********:**** **:******.***:*******:****.******
M. smegmatis  |recB|              DSPAAVRLKFARDVLAELEIRKRRRGVLGYDDLLTRLADALEPEDSPARV

M. canettii J |recB|              RMRERWRIVLVDEFQDTDPIQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. canettii K |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. canettii A |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. canettii I |recB|              RMQERWRIVLVDEFQDTDPMQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. canettii L |recB|              RMQERWRIVLVDEFQDTDPMQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. canettii E |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. canettii H |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. canettii D |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSGHSALILIGDPKQAIYGFRG
M. tub.CDC1551|recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSRHSALILIGDPKQAIYGFRG
M. tub. H37Rv |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSRHSALILIGDPKQAIYGFRG
M. afric. L6  |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSRHSALILIGDPKQAIYGFRG
M. bovis      |recB|              RMRERWRIVLVDEFQDTDPMQWRVLERAFSRHSALILIGDPKQAIYGFRG
                                  **:****************:********** ******************
M. smegmatis  |recB|              RMQQRWPIVMVDEFQDTDPVQWQVIERAFSGRSTLVLIGDPKQAIYAFRG

M. canettii J |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. canettii K |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. canettii A |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. canettii I |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. canettii L |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. canettii E |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. canettii H |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. canettii D |recB|              GDIHTYLTAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. tub.CDC1551|recB|              GDIHTYLKAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. tub. H37Rv |recB|              GDIHTYLKAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. afric. L6  |recB|              GDIHTYLKAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
M. bovis      |recB|              GDIHTYLKAAGTADARYTLGVNWRSDRALVESLQTVLRDATLGHADIVVR
                                  *******.****************************************
M. smegmatis  |recB|              GDIATYLRAAATAGDKQTLGTNWRSDRALVDRLQAVLRGAQLGGPDIVVH
```

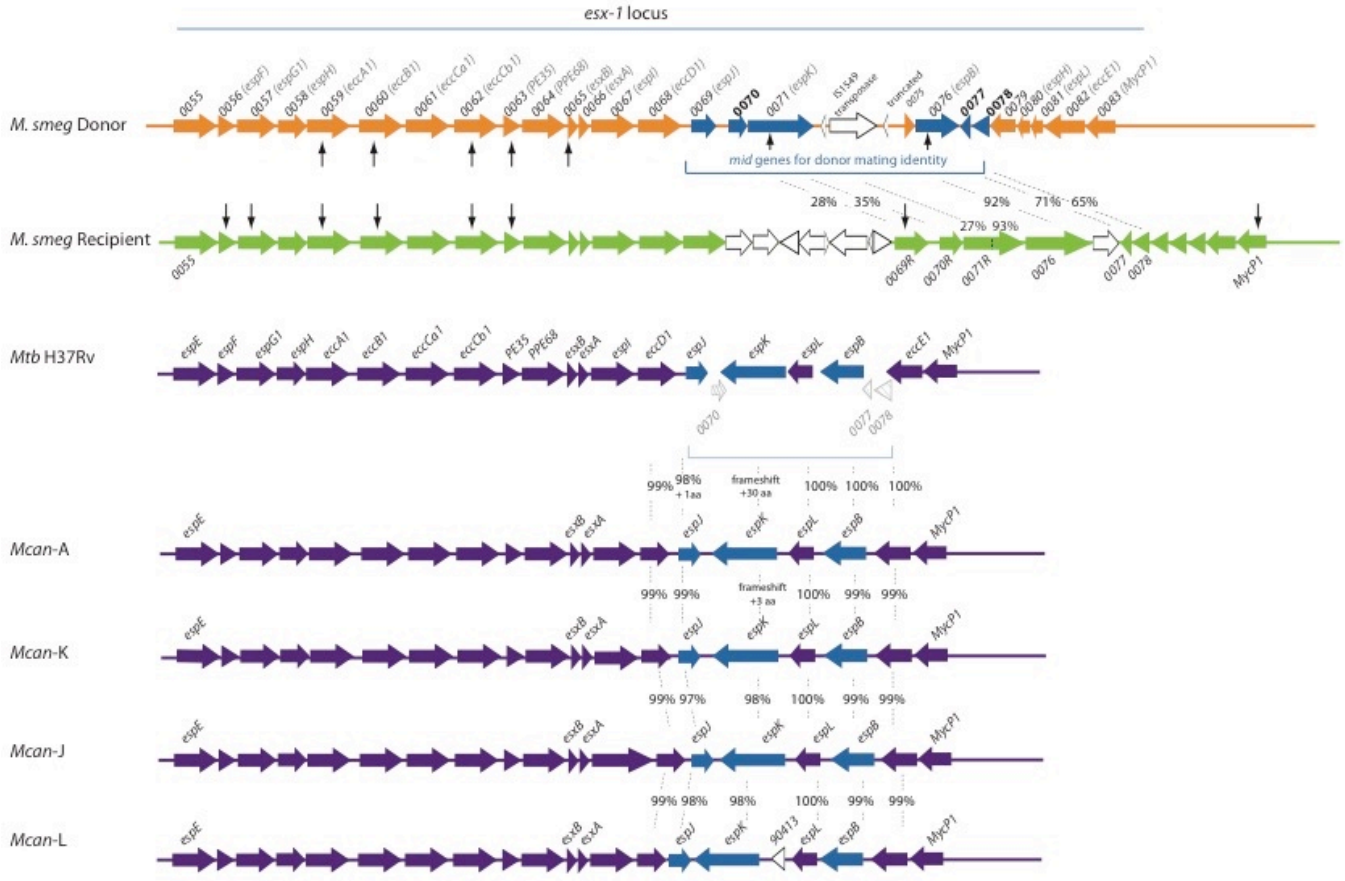


Fig. S10. Comparison of the *esx-1* locus between *M. smegmatis* donor and recipients, as well as *M. tuberculosis* and *M. canettii* strains A, K, J and L.
The *esx-1* locus spans the genes from *espE* (*msmeg_0055;* Rv3864*)* to *MycP1 (msmeg_0083;* Rv3883c) and contains 26 homologous genes in the *M. smegmatis* donor (orange) and the recipient (green), as well as specific genes and repetitive elements (white). The black arrows indicate transposon insertions in distinct donor or recipient genes with impact on DNA transfer. Backcrossing of donor-proficient transconjugants with the recipient revealed 6 genes involved in the mating type identity (*mid* genes, *msmeg_0069-0078;* blue*)* (1). Sequence comparison of the *M. smegmatis esx-1* locus with *M. tuberculosis* H37Rv or *M. canettii* strains A, K, J and L (violet) showed that only 3 *mid* genes are conserved in the tubercle bacilli, *espJ, espK* and *espB* (blue). Gene names are indicated above arrows, gray gene descriptions in brackets after *M. smegmatis* gene numbers represent respective homologues in *M. tuberculosis.* Light gray arrows below *M. tuberculosis esx-1* locus illustrate the 3 *M. smegmatis mid* genes without homologues in *M. tuberculosis*. Percentages on dotted lines represent % amino acid identity between respective strains.

Table S1. Overview of mating results between different *M. canettii* and *M. tuberculosis* strains.

Numbers represent recombinants obtained/ average resistant colonies after mating per 2.10^8 cells of donor strains carrying pYUB412 and recipient strains carrying the pMRF1-dsRed, except for *M. canettii* J (*), which contains an integrative pMV306.Kan plasmid, due to restricted plasmid maintenance by the *eptABCD* gene cluster. *M. canettii* donor strains A, K, L and I additionally carried spontaneous rifampicin resistance mutations. Data represent total numbers of at least two mating attempts per pair. N.D. not determined.

Hygʳ donor strains \ Kmʳ recipient strains	*M. canettii* A	*M. canettii* K	*M. canettii* L	*M. canettii* J*	*Mtb* H37Rv	*Mtb* H37Rv ΔRD1
M. canettii A	-	0/2	0/22	N.D.	0/7	0/0
M. canettii I	0/0	0/0	0/31	N.D.	0/12	0/0
M. canettii J	0/0	0/0	0/23	-	0/5	N.D.
M. canettii K	0/2	-	0/36	N.D.	0/26	0/0
M. canettii L	0/1	0/9	-	N.D.	0/21	0/0
Mtb H37Rv	0/2	0/11	0/35	0/0	-	-
Mtb H37Rv ΔRD1	0/1	0/0	0/53	N.D.	0/1	-

Hygʳ donor strains \ Kmʳ recipient strains	*Mtb* H37Rv	*Mtb* 79112	*Mtb* Tb36
Mtb H37Rv	-	0/0	0/2
Mtb 79112	0/0	-	0/1
Mtb Tb36	0/1	0/2	-

Table S2. Overview of mating results after treatment with different concentrations of Mitocmycin C (MitoC).

The donor *M. canettii Mcan*-A, with either the cosmid pYUB412 or F10 stably integrated into the genome, was mated with *M. canettii Mcan*-L or *M. tuberculosis* H37Rv at either 30˚C or 37˚C with different concentrations of mitomycin C (recombinants obtained/average resistant colonies after mating per 2.10^8 cells). N.D. not determined.

		***M. canettii* L Km^r**	**7H9**
***M. canettii* A Hyg^r**	**30°C**	0/0	N.D.
	37°C	0/5	9
	MitoC (1.7 ng/ml)	0/3	N.D.
	MitoC (30 ng/ml)	0/5	N.D.
	MitoC (100 ng/ml)	0/2	19

Table S3: Size and genomic positions of transferred fragments in the obtained *M. canettii* recombinants RC1-RC10

Re-combinant	Position of transferred DNA on *M. canettii* L genome as reference		Size of transferred fragment (kb)	Number of transferred fragments	Size of total transferred DNA (kb)	Distance between transferred sequence blocks (kb)	Number of transferred segments in regions of genomic heterogeneity (less than 45 kb of recipient DNA between transferred donor DNA blocks)
	Start	End					
RC1	1807261	1810528	3.3			91.4	
	2724616	2749046	24.4			5.1	
	2754144	2795769	41.6			1.2	
	2796993	2797765	0.7	8		1.6	7
	2799324	2812957	13.6		96.6	5.6	
	2813517	2814741	1.2			5.6	
	2820380	2831841	11.5			12.7	
	2844510	2844786	0.3				
RC2	2003946	2004369	0.4			744.7	
	2749031	2749982	1	4	36.7	40.6	2
	2790564	2825806	35.2			1543.4	
	4369241	4369393	0.1				
RC3	1114036	1142495	28.5			1655.3	
	2797765	2829206	31.4	3	86.8	393	
	3222159	3249096	26.9				
RC4	1619840	1657882	38			22.8	2
	1680692	1758710	78			75.6	
	1834317	1871088	36.8			15	
	1886063	1891768	5.7			3.9	5
	1895704	1896779	1			5.1	
	1901894	1907591	5.7			9.9	
	1917492	1924425	6.9	13	327.6	594.1	
	2518532	2573480	54.9			30.2	2
	2603691	2651681	48			162.9	
	2814593	2825806	11.2			35.1	3
	2860912	2882104	21.2			10.1	
	2892197	2904902	12.7			110	
	3014949	3022434	7.5				
RC6	2806238	2838239	32	1	32		
RC7	708400	723549	15.1			22.9	2
	746478	788849	42.4			1761.3	
	2550166	2566512	16.3			3.6	
	2570158	2651681	81.5	7	334.5	41.4	5
	2693084	2728329	35.2			16.1	
	2744387	2862001	117.6			37.8	
	2899795	2926157	26.4				
RC9	443620	443810	0.19			1064	
	1507793	1539394	31.6			22.1	3
	1561502	1564998	3.5			24.2	
	1589157	1607118	18	7	130.2	1169.4	
	2776536	2841816	65.3			34.5	3
	2876320	2879316	3			16.6	
	2895878	2904482	8.6				
RC10	2765324	2861843	96.5			542.8	
	3404661	3408205	3.5			40.6	2
	3448768	3469022	20.2	5	142.7	222.3	
	3691330	3697668	6.3			0.5	2
	3698205	3714420	16.2				

Table S4. Overview of mating results between *M. canettii* strains A and L under various stress conditions.

Numbers represent putative recombinants obtained/ average resistant colonies after mating per 2.10^8 cells of donor strains carrying the F10 cosmid and recipient strains carrying the pMRF1-dsRed plasmid**.** Ctrl; control conditions without stress. NO; nitrogen monoxide.

		***M. canettii* L Km^r**	**7H9**
***M. canettii* A::F10**	**ctrl**	**35**/65	5
	pH 6.1	**3**/6	8
	H_2O_2	**32**/84	2
	NO	**5**/14	3

Table S5. Sequence comparison of conjugation-associated proteins between *M. tuberculosis* and *M. canettii* strains A, K, J and L.
Shown are the % amino acid (AA) identity to the homologues in *M. tuberculosis* (AA *Mcan* strain/ AA *M. tuberculosis).*

M. tuberculosis H37Rv	*M. canetti Mcan*-A	*M. canetti Mcan*-K	*M. canetti Mcan*-J	*M. canetti Mcan*-L	gene function	*M. smeg* mutant phenotype	
espG1	**99%** (282/283)	**99%** (281/283)	**100%**	**99%** (282/283)	ESX-1 secretion-associated protein EspG1	**recipient-defective**	
eccA1	**99%** (571/573/)	**99%** (569/573)	**99%** (569/573)	**99%** (571/573)	ESX-1 type VII secretion system protein	**recipient-defective and donor hyper-conjugative**	
eccB1	**99%** (478/480)	**99%** (478/480)	**99%** (479/480)	**99%** (478/480)	ESX-1 type VII secretion system protein		
eccCb1	**99%** (589/591)	**99%** (589/591)	**100%**	**99%** (589/591)	ESX-1 type VII secretion system protein		
PE35	**98%** (97/99)	**98%** (97/99)	**99%** (98/99)	**98%** (97/99)	PE-family protein with unknown function		
cfp10	**100%**	**100%**	**100%**	**100%**	10 kDa culture filtrate antigen EsxB	**donor hyper-conjugative**	
espJ	**98 %** (275/280; -1 aa)	**97%** (273/280;-1 aa)	**97%** (271/280; -1 aa)	**98 %** (275/280; -1 aa)	putative translation initiation factor IF-2		***mid* genes for donor mating identity**
espK	**30 aa longer**	**33 aa longer**	**33 aa longer**	**33 aa longer**	ESX-1 secretion-associated protein EspK	**recipient-defective and donor hyper-conjugative**	
espB	**100%**	**99%** (457/461)	**99%** (456/461)	**99%** (456/461)	Secreted ESX-1 substrate protein B		
mycP1	**100%**	**99%** (445/446)	**100%**	**99%** (445/446)	Membrane-anchored mycosin MycP1 (serine protease)	**recipient-defective**	
Rv3196	**100%**	**99%** (297/300)	**99%** (299/300)	**100%**	conserved hypothetical		
Rv3193c	**99%** (992/993)	**99%** (988/993)	**99%** (991/993)	**99%** (992/993)	putative membrane protein		
Rv2972c	**100%**	**99%** (236/238)	**99%** (237/238)	**99%** (237/238)	putative extracellular deoxyribonuclease		
Rv2095c (pafC)	**98%** (312/317)	**99%** (313/317)	**99%** (313/317)	**100%**	putative transcription regulator		
Rv0419	**99%** (496/498)	**99% frameshift** 1 aa less	**99%** (497/498)	**99%** (496/498)	Possible lipoprotein peptidase LpqM	**donor transfer-defective**	
lsr2	**100%**	**100%**	**100%**	**100%**	Iron-regulated H-NS-like protein Lsr2		
Rv3659c	**99%** (351/353)	**97%** (344/353)	**99%** (351/353)	**99%** (351/353)	TadA-like, membrane associated ATPase, macromolecular transport	**recipient-defective and donor hyper-conjugative**	

Table S6. Overview of general sequence assembly.
The table indicates metrics from the sequence assemblies. N50 is median contig size value of the assembly.

Strain Name	Number of Paired-End reads	Total assembly (nuc.)	Number of contigs (>= 1000 bp)	N50 (nuc.)	N90 (nuc.)	Largest Contig (nuc.)
***M. smegmatis* recombinants**						
RC-Ms 1	14 096 938	7 090 137	78	315 186	81 440	820 565
RC-Ms 2	15 800 516	7 184 082	92	219 855	61 131	410 410
RC-Ms 3	14 049 131	6 930 940	80	280 426	65 987	516 599
RC-Ms 4	16 372 265	7 242 167	91	212 995	61 324	351 553
RC-Ms 5	12 309 612	7 130 933	81	290 095	84 873	738 139
RC-Ms 6	16 402 478	7 338 952	91	221 919	61 161	529 643
RC-Ms 7	8 191 119	7 201 166	69	309 477	91 358	734 953
RC-Ms 8	13 993 372	7 303 357	105	208 112	53 587	513 516
RC-Ms 9	13 152 956	7 415 911	95	244 336	56 656	530 747
RC-Ms 10	10 788 513	6 871 301	100	150 896	56 250	292 253
***M. canettii* recombinants**						
RC1	14 964 875	4 425 555	81	123 956	39 479	278 822
RC2	15 632 683	4 417 911	84	122 113	39 479	278 875
RC3	3 013 704	4 423 266	93	103 239	32 058	278 608
RC4	5 836 743	4 445 358	99	96 326	32 058	278 608
RC6	5 736 598	4 562 455	118	122 060	29 736	279 500
RC7	4 495 776	4 421 852	92	103 015	34 700	259 578
RC9	4 461 147	4 419 649	93	118 665	34 700	279 519
RC10	7 119 797	4 425 100	85	121 955	34 700	278 608

Table S7. Overview of genome sequence accession numbers.
The table indicates the various database accession numbers of nine *M. canettii* and 16 *Mycobacterium tuberculosis* complex genomes used for comparative, recombination and NeighbourNet analyses.

Isolate Name	Accession Number
L1_1078602	ERR234155
L1_N72	ERR234236
L2_58	ERR234107
L2_GQ73	ERR234136
L3_855B	ERR234111
L3_N04	ERR234232
L4_141702	ERR234159
L4_DY8	ERR234206
L5_1001003	ERR234097
L5_DY20	ERR234202
L6_533604	ERR234148
L6_N0091	ERR234254
L7_BTBH1012	ERR233346
L7_BTBH935	ERR233349
L7_BTBS610	ERR233354
L7_BTBS746	ERR233355
A	HE572590
D	FO203507
E	CAOL01000000
G	CAOM01000000
H	CAON01000000
I	CAOO01000000
J	FO203510
K	FO203509
L	FO203508