The *Enterprise*, a massive transposon carrying *Spok* meiotic drive genes

Vogan, Aaron A., Ament-Velásquez, S. Lorena, Bastiaans, Eric, Wallerman, Ola, Saupe, Sven J., Suh, Alexander, Johannesson, Hanna

Contents

Supplementary Figures

Supplementary Methods

Fungal material

Strains used in this study were obtained from the Wageningen collection and cultivated as in (Vogan et al, 2019). Strain S was used as the standard reference strain with no *Spok* block. As spore killer strains for the fitness experiments (see below) we used the backcrossed strains Psk1xS$_{14}$ and Psk7xS$_{14}$ (Vogan et al, 2019), which should be isogenic to S but with a *Spok* block in Chromosome 3 and 5, respectively. The more recently isolated strains Wa131, Wa137 and Wa139 were sampled during the fall of 2016 around Wageningen (the Netherlands) from dung of rabbit (Wa131 and Wa137, locality Unksepad Oosterbeek) or horse (Wa139, locality Uiterwaarden Wolfswaard). Morphological differences like smaller perithecia and abundant tomentose apricot-coloured mycelium in HPM medium (Vogan et al, 2019), as well as analyses of sequence data, allowed us to assign Wa131 and Wa139 to the species *P. comata*. Previously, only one strain from this species, T_D, was known (Boucher et al, 2017; Vogan et al, 2019), hence these new strains constitute a new report of this species for the Netherlands.

DNA and RNA extraction and sequencing

Following Vogan et al (2019), we grew monokaryotic strains on PASM0.2 plates covered with a layer of cellophane. Genomic DNA for short-read sequencing was extracted from 80 mg–100 mg of fungal tissue with the Fungal/Bacterial Microprep kit (Zymo; `www.zymo.com`). Paired-end libraries were prepared and sequenced using the Illumina HiSeq X (150-bp-long) technology at the SNP and SEQ Technology platform (SciLifeLab, Uppsala, Sweden). For RNA extraction, around 150 mg of harvested mycelium were frozen in liquid nitrogen and stored at −80 ℃. We extracted total RNA from the grounded frozen tissue using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Quality was checked on the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) and the RNA was treated with DNaseI (Thermo Scientific). The sequencing library was prepared with a NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). We purified poly(A)+ transcripts with the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs). A paired-end library was sequenced with Illumina HiSeq 2500 at the SNP and SEQ Technology platform.

For long-read sequencing, we grew the monokaryotic strains in PASM0.2 plates, from where we sliced small agar cubes to inoculate liquid cultures of 200 }ml 3% malt extract solution, which were subsequently incubated in a shaker for 10 d–14 d at 27 ℃ (Vogan et al, 2019). Mycelium aggregates were filtered from the flasks, any remaining agar was removed, and around 1 g was stored at −20 ℃. As described in Sun et al (2017), the tissue was freeze-dried and macerated, followed by DNA extraction using Genomic Tip G-500 columns (Qiagen) and cleaning with the PowerClean DNA Clean-Up kit (MoBio Labs). Additionally, DNA was purified using magnetic beads (Speed-Beads, GE) and eluted for 20 min at 37 ℃ followed by overnight storage at 4 ℃ twice to increase concentration ($\approx$65 ng µl^{-1}). Wa137- was sequenced on an R9.5.1 Flowcell (Oxford Nanopore Technologies) with a modified SQK-RAD004 protocol using 550 ng DNA to 1.5 µl FRA to increase read lengths. Wa139- was prepared using the ligation protocol (SQK-LSK109) and sequenced on an R9.4.1 flowcell. Basecalling was done using Guppy v. 1.6.

Genome assembly

For most strains we used the assemblies produced in Vogan et al (2019). For newly sequenced strains, we produced new assemblies as follows. The adapters from the Illumina HiSeq reads were identified with cutadapt v. 1.13 (Martin, 2011) and removed using Trimmomatic v. 0.36 (Bolger et al, 2014) using the following options: ILLUMINACLIP:adapters.fasta:1:30:9 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20

MINLEN:30. Pairs with both forward and reverse reads after filtering were used for downstream analyses. For the strain Wa131, which only has Illumina data, we used SPAdes v. 3.12.0 (Bankevich et al, 2012) with the k-mers 21,33,55,77 and the –careful option. For the strains Wa137 and Wa139, the MinION reads with a mean Phred quality (QV) above 9 and longer than 1 kb were assembled using minimap2 v. 2.11 and Miniasm v. 0.2 (Li, 2016, 2018). The resulting assembly was polished twice with Racon v. 1.3.1 (Vaser et al, 2017) using all MinION reads (no filtering). Further polishing was done with the filtered Illumina reads in five consecutive rounds of Pilon v. 1.22 (Walker et al, 2014). We used BWA v. 0.7.17 (Li Durbin, 2010) for short-read mapping, with PCR duplicates marked using Picard v. 2.18.11 (`http://broadinstitute.github.io/picard/`), as well as local indel re-alignment using the Genome Analysis Toolkit (GATK) v. 3.7 (Van der Auwera et al, 2013).

We assigned the scaffolds to chromosomes based on alignments to the reference genome of the S strain (Espagne et al, 2008), available at the Joint Genome Institute MycoCosm website (`https://mycocosm.jgi.doe.gov/mycocosm/home`) as "Podan2" (Grigoriev et al, 2014). If a given chromosome was not assembled completely, the corresponding scaffolds were assigned an additional number (e.g., scaffolds mapping to Chromosome 1 were named Chromosome 1.1, Chromosome 1.2, etc.). We discarded small contigs (<100 kb) of rDNA repeats as well as mitochondrial-derived sequences, except for the largest mitochondrial contig. We assessed the quality of the final assemblies by visual inspection of the mapping of both long and short reads using minimap2 and BWA, respectively. Mean depth of coverage was calculated with QualiMap v.2.2 (Okonechnikov et al, 2016). Other assembly statistics were calculated with QUAST v. 4.6.3 (Mikheenko et al, 2016). As in Vogan et al (2019), the assembly of each strain is named based on the source strain and the mating type. For example, PaWa137m is the assembly of a monokaryon from the minus (-) mating type derived from the dikaryotic strain Wa137. However, the reference genomes of *P. anserina* (strain S) and *P. comata* (T_D) are referred to as Podan2 and PODCO (Espagne et al, 2008; Silar et al, 2019) for consistency with available databases.

Genome annotation

A GitHub repository is available with Snakemake v. 5.4.4 (Köster Rahmann, 2012) pipelines at `https://github.com/johannessonlab/SpokBlockPaper` and in the Supplemental Code.

The TEs and other repeats in *P. anserina* were classified previously by Espagne et al (2008) based on the original reference genome of the S strain or "Podan1", and is hereafterx referred to as the "Espagne library". To explore the diversity of TEs in the newly generated *Podospora* genomes, we identified repeats de novo and manually compared them to the Espagne library to identify duplicates and new elements. Specifically, we ran RepeatModeler v. 1.0.8 (`http://www.repeatmasker.org/RepeatModeler/`) on the scaffolds larger than 50 kb of all available long-read assemblies (Snakemake pipeline `PaTEs.smk`). Each resulting RepeatModeler consensus was BLASTN-searched back to the original genome and the best 20 hits with 2-kb flanks were aligned with T-Coffee v. 12.00.7fb08c2 (Notredame et al, 2000) (`TEManualCuration.smk`), and visually inspected for manual curation. The curated consensuses were assigned to the Espagne library (Espagne et al, 2008) equivalents based on similarity (allowing for RIP-induced mutations) or were given a new name when having no homology to anything in the Espagne library. It was discovered that the gypsy element *crapaud* has numerous diverged copies with unique LTRs. We annotated all *crapaud* LTRs that were in multiple copies within *P. anserina* individually to improve repeat masking. We refer to the final repeat library as "PodoTE-1.00" (available at the GitHub repository).

To generate a genome annotation of all assemblies, we ran an updated version of the pipeline in Vogan et al (2019), named `PaAnnotation.smk`. We used MAKER v. 3.01.2 (Holt Yandell, 2011; Campbell et al, 2014) with the previously produced training files used for the ab initio prediction programs GeneMark-ES v. 4.32 (Lomsadze et al, 2005; Ter-Hovhannisyan et al, 2008) and SNAP release 2013-06-16 (Lomsadze et al, 2005), as well as the following dependencies: RepeatMasker v. 4.0.7 (`http://www.repeatmasker.org/`), BLAST suite 2.6.0+ (Camacho et al, 2009), Exonerate v. 2.2.0 (Slater Birney, 2005), and tRNAscan-SE v. 1.3.1 (Lowe Eddy, 1997). As evidence, we used STAR v. 2.6.1b (Dobin et al, 2013) to produce transcript models (maximum intron length set to 1000 bp) of various RNA-seq data sets. Specifically, we mapped the reads of the monokaryotic isolate Wa63- (*P. anserina*) to the assembly PaWa63m (Vogan et al, 2019), of the monokaryotic isolate Wa131- (*P. comata*) to the assembly PcWa139m (this study), and of the dikaryotic Psk7xS_{14} (*P. anserina*) to the assembly PaWa58m (Vogan et al, 2019). We then processed the mapped reads with Cufflinks v. 2.2.1 (Trapnell et al, 2010) to obtain the transcript models. As external evidence, we used CDS from the Podan2 annotation, protein sequences from the T_D strain of *P. comata* (Silar et al, 2019), and a small dataset of manually curated proteins. To aid in manual curation of selected regions (mostly the *Spok* block), we visually inspected the mapping of RNA-seq reads of the different datasets, along with CDS produced with TransDecoder v. 5.5.0 (Haas et al, 2013) on the Cufflinks models, as well as the output of RepeatMasker ran externally from MAKER with the PodoTE-1.00 library. Additionally, we queried predicted gene models into the NCBI databases (NCBI Resource Coordinators 2016) to verify the annotations.

The *Kirc* protein sequence was analysed with HHPred (Zimmermann et al, 2018) and Gremlin (Balakrishnan et al, 2011). The Gremlin-generated alignment of *Kirc* homologs was used to generate region-specific sequence logos with WebLlogo (Crooks et al, 2004). The relationship of *Kirc* to other YRs was confirmed by comparing the sequence to the crystal structure of known YRs (CRE (PDB code 3mgv), XERD (1a0p), and FLP (1flo)) as well as the protein sequence from the transposable element *Crypton-Cn1* using the software Promals3D (Pei et al, 2008). Other annotated genes present in the *Enterprise* of Wa139 (the "crew") were manually curated as above, and named the following way: **Ch**ronically **e**xpressed **k**inase-containing **O**R**F** – *Chekof* (based on the high expression levels observed in the RNA-seq data); **U**nknown **he**licase **r**elated to **a**ccumulation – *Uhera* (as it is associated to the accumulation or duplication of *Spok* genes within the *Spok* blocks); and ***Sc****lerotinia* **o**rtholog **ty**pical of *Enterprise* - *Scoty* (based off best BLAST homology).

To calculate the total repeat content in bp of a given genome or *Spok* block, we used the output of RepeatMasker produced with our repeat library (in GTF format) and collapsed all overlapping features using the script `totalcovergff.py` v. 2.01 available at the GitHub repository. In order to assess how common the TSD motif of the *Spok* block is in the genome, we used Jellyfish v. 2.2.10 (Marçais Kingsford, 2011) to calculate the distribution of k-mers (substring) of length six in the reference genome of *P. anserina* (Podan2) as a representative of the species. Jellyfish was run with a hash of 100 million elements (-s 100M) in the pipeline `PoJellyfish.smk`.

Comparative genomics

We used the NUCmer program from the MUMmer package v. 4.0.0beta2 (Kurtz et al, 2004) using the parameters -b 200 -c 22 –maxmatch to align the *Spok* blocks to each other, and changed to -c 40 for whole-genome assemblies. To achieve higher sensitivity, we used BLASTN from the BLAST suite 2.9.0 (Camacho et al, 2009) to search for the presence of the unclassified repeats *bufo* and *schoutenella*. Both the NUCmer and the BLAST alignments were plotted using Circos v. 0.69.6 (Krzywinski et al, 2009) along with manual curations of coding regions and repetitive elements. The distribution of TE and gene content along chromosomes was calculated in windows of 50 kb with steps of 10 kb using BEDTtools v. 2.29.0 (Quinlan Hall, 2010; Quinlan, 2014) with the utilities makewindows and coverage. The fraction of conservation between blocks compared to the block in Wa137 was calculated by aligning the block sequences (within the TSD) of Wa28 (*Psk-2*), Wa53 (*Psk-1*), Wa58 (*Psk-7*) and Wa139 with NUCmer and the BEDTtools utility genomecov. The Snakemake pipelines used to produce the Circos plots (`CircosBlock.smk` and `CircosAllBlocks.smk`) are also available at `https://github.com/johannessonlab/SpokBlockPaper`.

The dot plots of *bufo* and *schoutedenella* were produced by using the consensuses of these elements in our repeat library as BLAST queries against all genome assemblies. All sites with hits larger than 150 bp and percentage of identity larger than 70% were extracted along with 3000 (*bufo*) or 5000 (*schoutedenella*) extra base pairs on each flank and aligned with the online server of MAFFT v. 7 (`https://mafft.cbrc.jp/alignment/server/`) (Katoh et al, 2019).

To search MycoCosm for other copies of *Enterprise*, the following approach was taken. The protein sequence of *Kirc* was used as a query with BLASTX against all genomes within MycoCosm (as of February 2019). Genomes with multiple high-confidence positive hits were identified and the regions with putative *Kirc* homologs were manually extracted. Priority was given to genomes where the hits were associated with large duplicated regions ($>$50 kb). *Melanconium sp.* NRRL 54901 (produced as part of the 1KFG project; Spatafora (2011)) had the most copies with clear termini. The genomic regions surrounding these *Kirc* homologs were aligned with NUCmer using the parameters -b 200 -c 22 –maxmatch –nosimplify to produce dot plots. From these regions, the least degraded *Enterprise* was extracted (scaffold 11) and compared to the *Psk-9 Spok* block using PROmer (default parameters except –maxmatch), which produces alignments based upon the six-frame translations of both input sequences. We used the filtered proteins annotations available in MycoCosm to mark the position of *Melanconium sp.* associated genes. To determine the copy number of the various genes of interest from the *Spok* block (*Kirc*, *Uhera*, *Scoty* and *Chekof*), each gene was used as a query with BLASTPp against the NCBI RefSeq database (consulted in October 2020). All hits with e-values $<$ 1 were compiled.

Phylogenetic analyses

Maximum Likelihood analyses were performed using IQ-TREE v. 1.6.8 (Kalyaanamoorthy et al, 2017; Nguyen et al, 2014) with extended model selection (-m MFP) and 1000 standard bootstrap pseudoreplicates to estimate branch support. In the case of *bufo* and *schoutedenella*, we used as input the first 345 (*bufo*) or 278 (*schoutedenella*) bp of the MAFFT alignment produced above but excluding the RGGTAG motif.

To estimate the phylogeny of *Kirc*, homologs were identified from GenBank using BLASTP with a truncated version of *Kirc* from Wa53 that has no CHROMO domain. CHROMO domains are highly conserved and are present in many different types of genes, so including this domain in the search results in numerous additional hits that have no putative YR domain, and likely no relation to *Kirc*. Nucleotide sequences from hits with e-values $< 1e \times 10^{-100}$ were compiled along with a homolog from *P. anserina* (Pa_5_10116), two homologs from *Melanconium sp.* NRRL 54901 extracted from MycoCosm (see above), and the full length sequence of *Kirc* in the *Spok* block of Wa53, and aligned with MACSE v. 2.03 (Ranwez et al, 2018). We used TrimAl v. 1.4.1 (Capella-Gutierrez et al, 2009) to trim the resulting protein alignment with the -gappyout option which was then used as input for IQ-TREE.

Fitness assays

The cultures used for the crosses were revived from the –80 ℃ freezer on PASM0.2 (van Diepeningen et al, 2008) at 27 ℃ for several days and then stored at 4 ℃ until use. Strains were grown for 5 d on fresh PASM0.2 plates before inoculating the cross. In the crosses, one strain was grown as mycelia and thereby assigned the female role, while a compatible strain of the other mating type was assigned the male role by fertilizing the mycelia with microconidia. The strain that was assigned the female role was grown in a 35 mm petri dish with 5 ml HPM medium (Vogan et al, 2019) by inoculating a small cube of agar with mycelium ($\approx$2 $\times$ 2 mm). In parallel, the strain that was assigned the male role was grown on a 90 mm petri dish with micro conidiation medium (King, 2013) by inoculating seven plugs of mycelium spread over the plate. After 7 d of growth, microconidia were harvested by adding 5 ml of sterile water to the plate and sweep over the mycelium with a drigalski spatula for 1 min. The female mycelium was then fertilized with 0.5 ml of the microconidial suspension. The suspension was carefully spread out to make sure all mycelium was covered. The fertilized mycelia were then further incubated under standard conditions (27 ℃, 12/12 light/dark cycle) (Vogan et al, 2019). The cultures were monitored daily for signs of spores shot from the asci in order to score the first day of spore-shooting. To reduce the complexity of the experiment, the strains used as female were always of mating type *mat+*.

At 6 d post-fertilization, single spores were collected with a needle to measure germination frequency and growth speed. From each cross, 10 spores from 4-spored asci were picked, and in cases with spore killing, an additional 10 spores from 2-spored asci were picked. The 10 spores were transferred to a single 90 mm petri dish with PASM2 medium (van Diepeningen et al, 2008) with 0.4% ammonium acetate added (to activate the spores) (King, 2013). Spores were spaced out in a predetermined pattern (4 lines of 2, 3, 3, 2 spores). After two days of incubation, the germination was scored and colony diameter was measured in two directions. If there was no growth microscopic inspection was performed to check whether a spore was present in the agar to avoid scoring no germination in case the inoculation failed.

At 12 d post-fertilization, spores were harvested from the lids of each crossed culture and used for estimating total spore yield. Spores were collected by pipetting 750 µl of harvest liquid (1 M NAOH, 0.025% SDS) in the lid. Spores were then scraped off the lid using the pipette tip. The liquid was then collected into a 2 ml Eppendorf tube. Another 750 µl of harvest liquid was used to repeat the process to make sure most of the spores were collected from the lid. The tubes were then heated for 4 h at 85 ℃, then shaken in a Qiagen Tissuelyser for 90 s at 30 HZ. After this, the tubes were stored at 4 ℃ overnight. The cooled tubes were again shaken in a Qiagen Tissuelyser for 90 s at 30 30 HZ. This process prevents the clumping of spores. Total yield was determined by counting the amount of spores in a volume of 5 µl of 50× diluted suspension pipetted on an object glass using a stereomicroscope. Counts were taken five times for each replicate cross. Ten replicate crosses were conducted for each cross. Statistical analyses were conducted in base R v. 3.5.0 to determine significance and power.

References

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**: 11.10.1–33.

Balakrishnan S, Kamisetty H, Carbonell JG, Lee SI, Langmead CJ. 2011. Learning generative models for protein fold families. *Proteins: Structure, Function, and Bioinformatics* **79**: 1061–1078.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Boucher C, Nguyen TS, Silar P. 2017. Species delimitation in the *Podospora anserina*/ *P. pauciseta*/*P. comata* species complex (sordariales). *Cryptogamie, Mycologie* **38**: 485–506.

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

Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinformatics* **48**: 4.11.1–39.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. Weblogo: a sequence logo generator. *Genome research* **14**: 1188–1190.

van Diepeningen AD, Debets AJM, Slakhorst SM, Hoekstra RF. 2008. Mitochondrial pAL2-1 plasmid homologs are senescence factors in podospora anserina independent of intrinsic senescence. *Biotechnol. J.* **3**: 791–802.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, Couloux A, Aury JM, Ségurens B, Poulain J, et al. 2008. The genome sequence of the model ascomycete fungus *Podospora anserina*. *Genome Biol.* **9**: R77.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. 2014. Mycocosm portal: gearing up for 1000 fungal genomes. *Nucleic acids research* **42**: D699–D704.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* **8**: 1494–1512.

Holt C Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**: 587–589.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**: 1160–1166.

King RC. 2013. *Handbook of Genetics: Volume 1 Bacteria, Bacteriophages, and Fungi*. Springer Science & Business Media.

Köster J Rahmann S. 2012. Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics* **28**: 2520–2522.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**: 1639–1645.

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

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**: 2103–2110.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li H Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–595.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **33**: 6494–6506.

Lowe TM Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.

Marçais G Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764–770.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10.

Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. 2016. Icarus: visualizer for de novo assembly evaluation. *Bioinformatics* **32**: 3321–3323.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**: 268–274.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**: 205–217.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**: 292–294.

Pei J, Kim BH, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**: 2295–2300.

Quinlan AR. 2014. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**: 11.12.1–34.

Quinlan AR Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* **35**: 2582–2584.

Silar P, Dauget JM, Gautier V, Grognet P, Chablat M, Hermann-Le Denmat S, Couloux A, Wincker P, Debuchy R. 2019. A gene graveyard in the genome of the fungus *Podospora comata*. *Mol. Genet. Genomics* **294**: 177–190.

Slater GSC Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31.

Spatafora J. 2011. 1000 fungal genomes to be sequenced. *IMA Fungus* **2**: 41–45.

Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. 2017. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat. Commun.* **8**: 1140.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* **18**: 1979–1990.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**: 511–515.

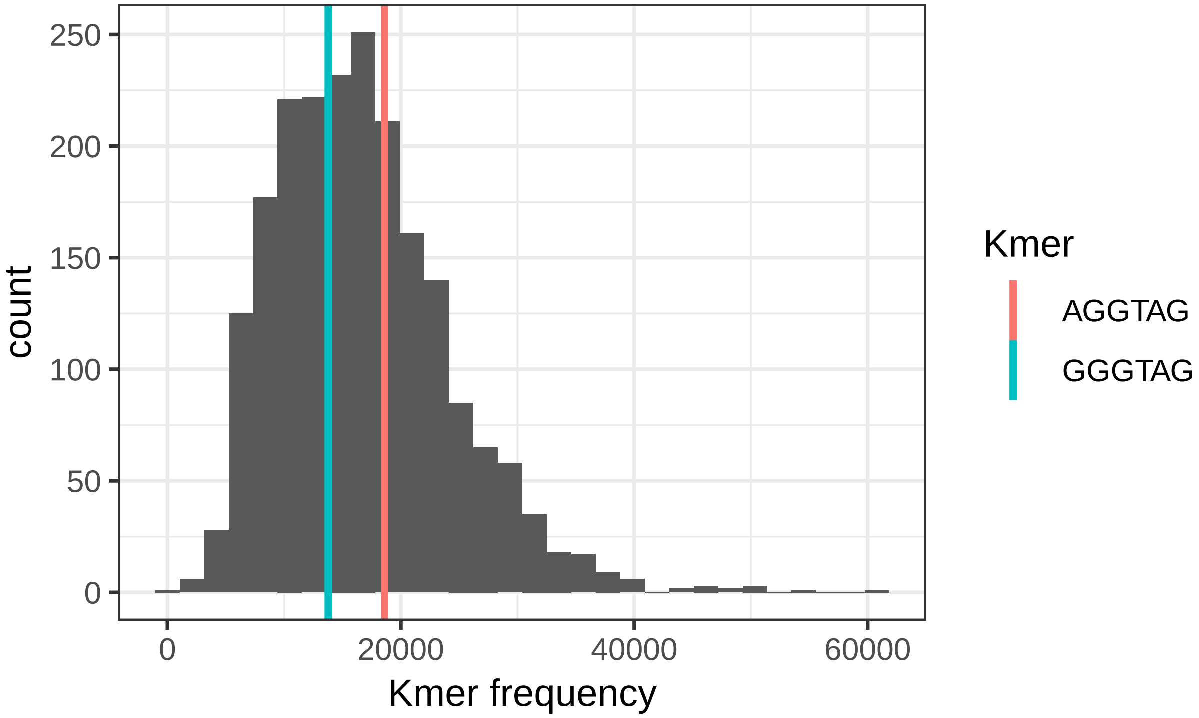
Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**: 737–746.

Vogan AA, Ament-Velásquez SL, Granger-Farbos A, Svedberg J, Bastiaans E, Debets AJ, Coustou V, Yvanne H, Clavé C, Saupe SJ, et al. 2019. Combinations of *Spok* genes create multiple meiotic drivers in *Podospora*. *Elife* **8**: e46454.

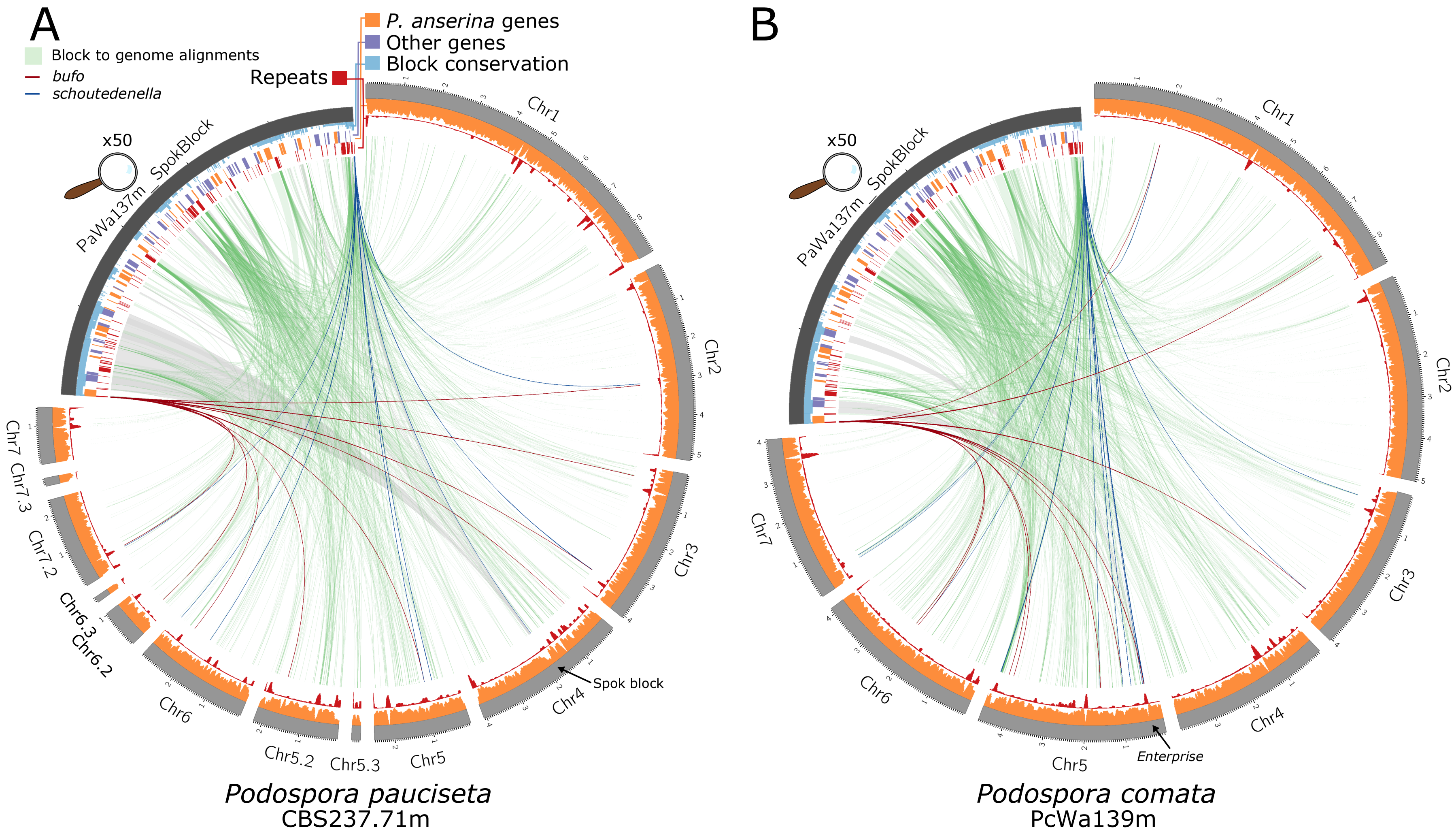
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented mpi bioinformatics toolkit with a new hhpred server at its core. *Journal of molecular biology* **430**: 2237–2243.

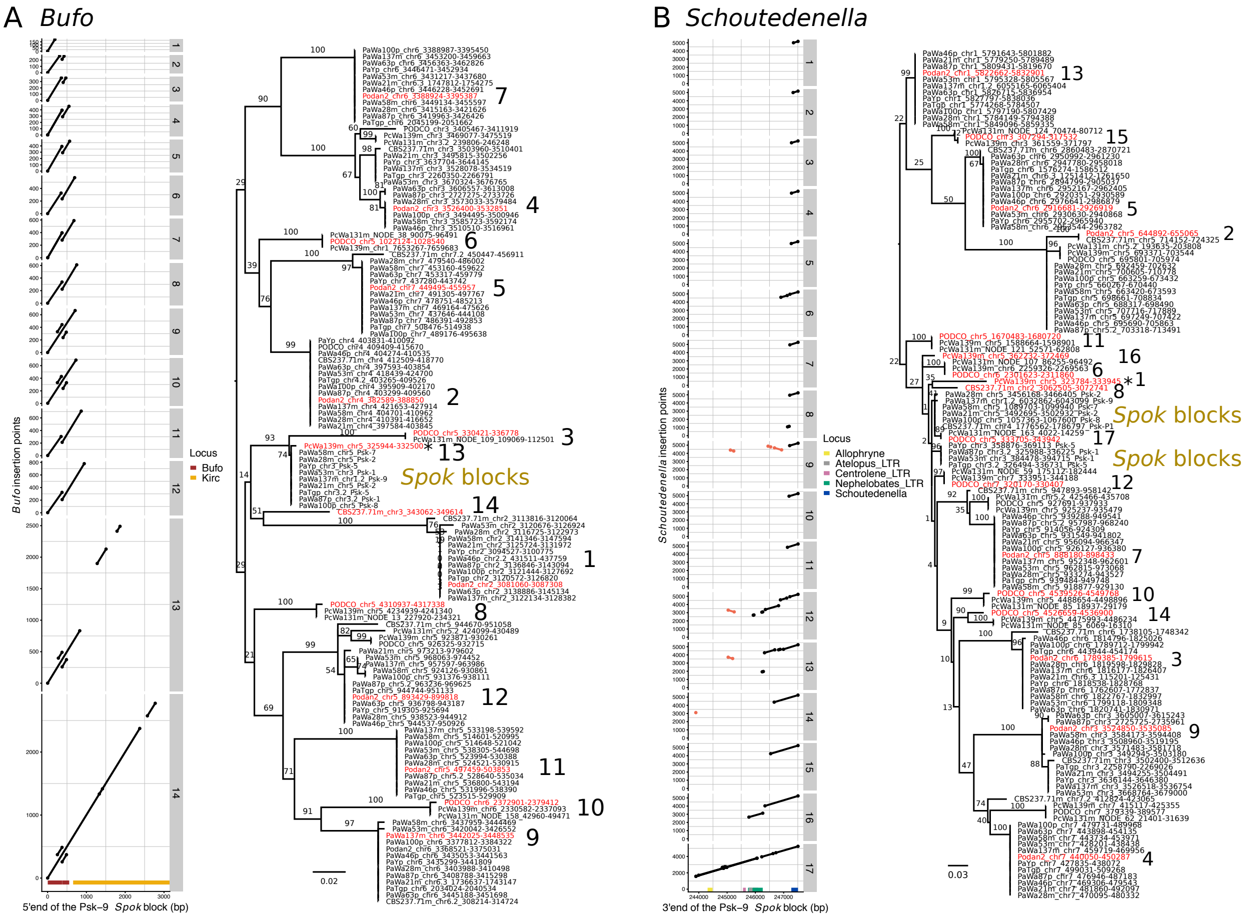
Supplementary Figures



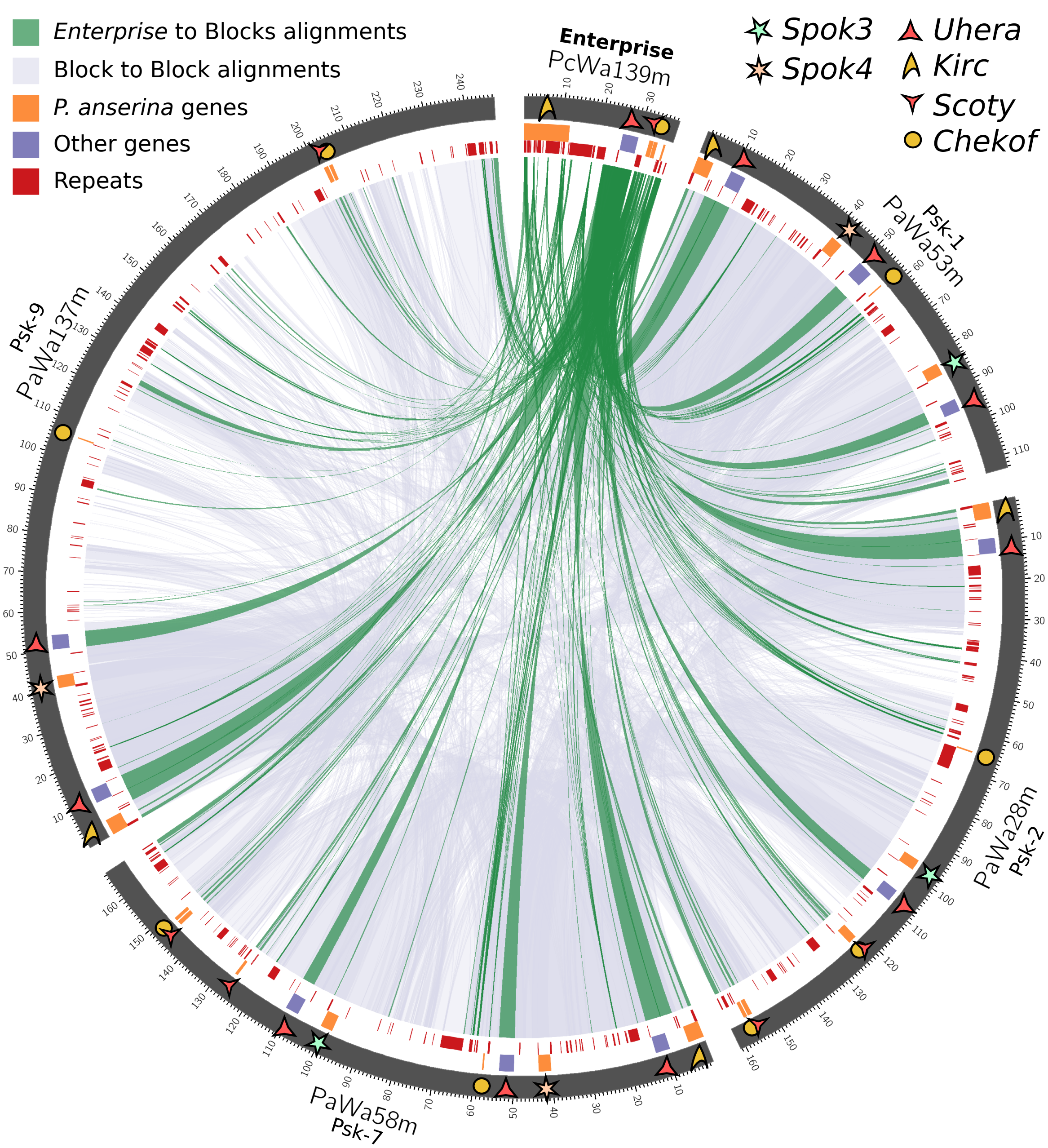
Supplementary Figure S1: A histogram showing the abundance of 6 bp k-mers in the *P. anserina* genome (Podan2). The two potential putative targets of *Kirc* are shown with vertical lines.



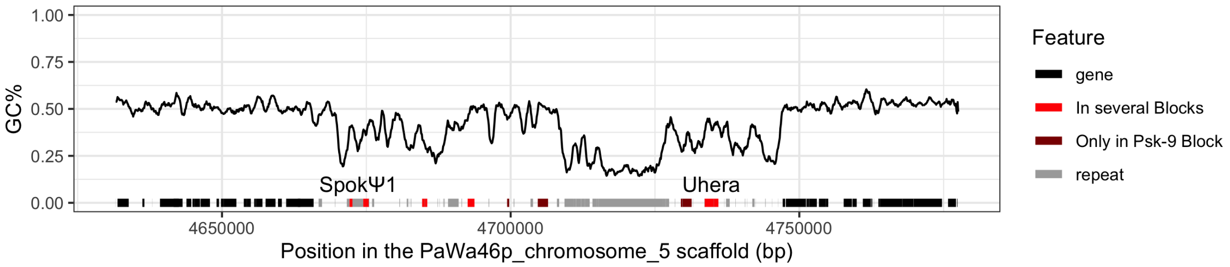
Supplementary Figure S2: Circos plots comparing the largest *Spok* block known (strain Wa137) to the genome of **A** the *P. pauciseta* strain CBS237.71 and **B** the *P. comata* strain Wa139. The comparison revealed large alignment blocks (in gray), corresponding to the known *Spok* block in CBS237.71 (*Psk-P1*) and the *Enterprise* element of Wa139.



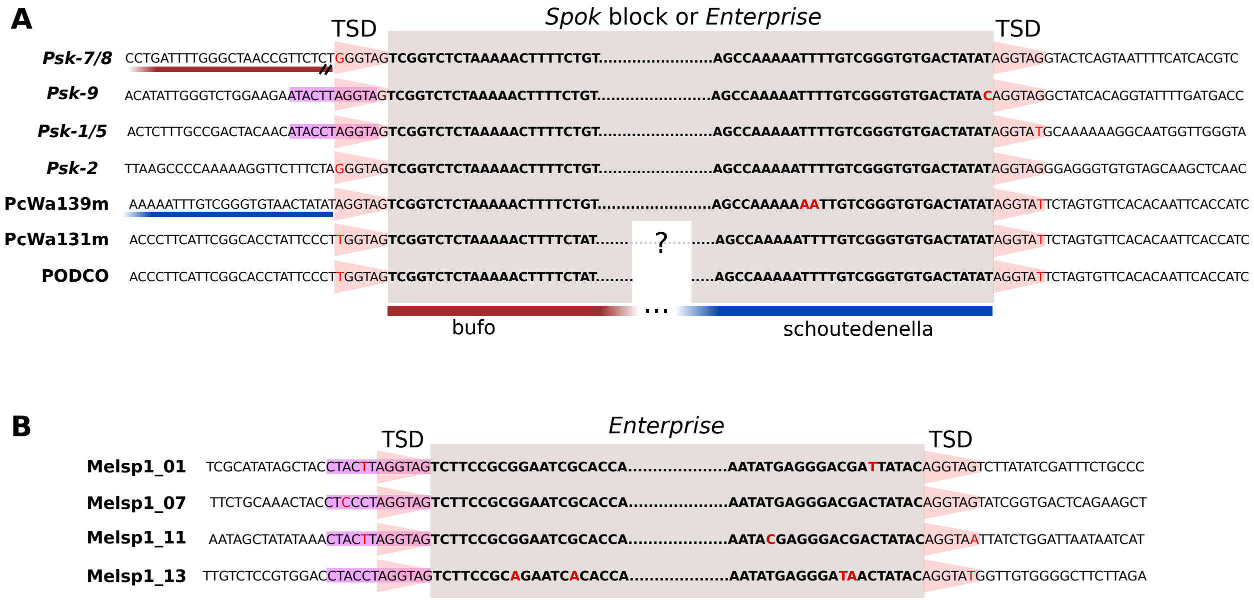
Supplementary Figure S3: The previously unclassified elements *bufo* (**A**) and *schoutedenella* (**B**) constitute the ends of partially deleted *Spok* block or closely related *Enterprise* elements. The dot plots are alignments of all insertion points for *bufo* (14 loci) and *schoutedenella* (17) across the three *Podospora* species against the ends of the Wa137 (*Psk-9*) *Spok* block. The unclassified elements always align well with the *Spok* block edges (starting at the first and last base of the *Spok* block) but their opposite end has decaying extensions of homology towards the interior of the *Spok* block. For some insertions, the homology of *bufo* elements extends into the coding region of *Kirc* (**A**). Loci of interest (all repetitive elements) are marked in colour on the x-axis of the dot plots. Coordinates are with respect to positions in the Wa137 *Spok* block. Black segments represent collinear homology, red segments mark inverse homology. Numbers in facets refer to each insertion point, as reflected by the relationships in the phylogenies of either *bufo* (**A**) and *schoutedenella* (**B**). Sequences used to represent each insertion point are marked in red text. The ends of the *Spok* blocks themselves were also included in the phylogeny (*Psk* sequences). The asterisks mark the sequences of the Wa139 *Enterprise*.



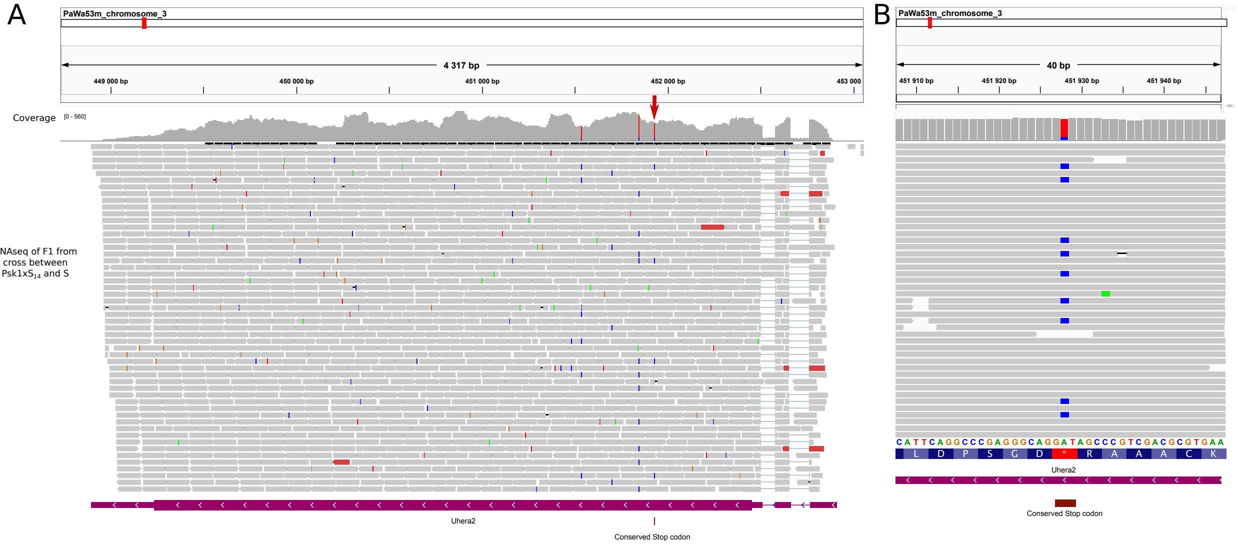
Supplementary Figure S4: A Circos plot comparing the *Enterprise* of five different strains. Four of these represent different *Spok* blocks. Dark green lines connect homologous segments of the Wa139 *Enterprise* to the various *Spok* blocks. Lilac lines show homologous regions among the *Spok* blocks. Genes of interest are marked with symbols as in **Figure 3**.



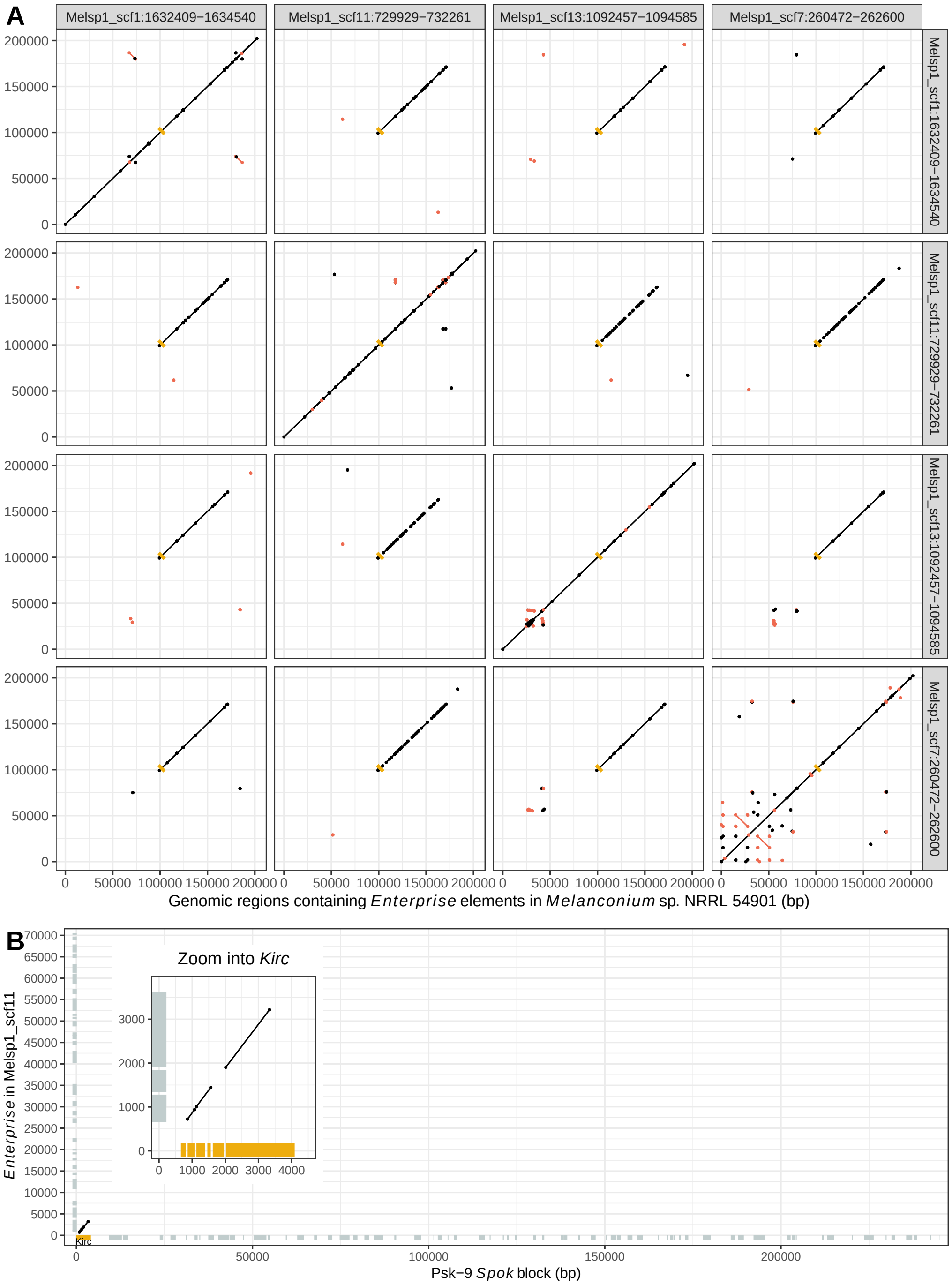
Supplementary Figure S5: The genomic region where the pseudogenized *Spok* gene (*Spok*$\Psi 1$) is located in strain Wa46. Percentage of GC is plotted in windows of 250 bp with steps of 10 bp as an indication of the effect of RIP. Gene and repeat annotations are marked on the x-axis. We inferred that the locus is in fact a degraded *Spok* block based on the presence of multiple genes in one (*Psk-9*) or more *Spok* blocks. *Spok*$\Psi 1$ and *Uhera* are highlighted. Notice that *Spok*$\Psi 1$ is interrupted by a transposable element.


A
Spok block or Enterprise
TSD
TSD
Psk-7/8 CCTGATTTTGGGCTAACCGTTCTCTGGGTAGTCGGTCTCTAAAAACTTTTCTGT.......................AGCCAAAAATTTTGTCGGGTGTGACTATATAGGTAGGTACTCAGTAATTTTCATCACGTC
Psk-9 ACATATTGGGTCTGGAAGAATACTTAGGTAGTCGGTCTCTAAAAACTTTTCTGT.......................AGCCAAAAATTTTGTCGGGTGTGACTATACAGGTAGGCTATCACAGGTATTTTGATGACC
Psk-1/5 ACTCTTTGCCGACTACAACATACCTAGGTAGTCGGTCTCTAAAAACTTTTCTGT.......................AGCCAAAAATTTTGTCGGGTGTGACTATATAGGTATGCAAAAAAGGCAATGGTTGGGTA
Psk-2 TTAAGCCCCAAAAAGGTTCTTTCTAGGGTAGTCGGTCTCTAAAAACTTTTCTGT.......................AGCCAAAAATTTTGTCGGGTGTGACTATATAGGTAGGGAGGGTGTGTAGCAAGCTCAAC
PcWa139m AAAAATTTGTCGGGTGTAACTATATAGGTAGTCGGTCTCTAAAAACTTTTCTGT.......................AGCCAAAAAAATTGTCGGGTGTGACTATATAGGTATTCTAGTGTTCACACAATTCACCATC
PcWa131m ACCCTTCATTCGGCACCTATTCCCTTGGTAGTCGGTCTCTAAAAACTTTTCTAT...... ? .....AGCCAAAAATTTTGTCGGGTGTGACTATATAGGTATTCTAGTGTTCACACAATTCACCATC
PODCO ACCCTTCATTCGGCACCTATTCCCTTGGTAGTCGGTCTCTAAAAACTTTTCTAT....... .....AGCCAAAAATTTTGTCGGGTGTGACTATATAGGTATTCTAGTGTTCACACAATTCACCATC
...
bufo
schoutedenella
B
Enterprise
TSD
TSD
Melsp1_01 TCGCATATAGCTACCTACTTAGGTAGTCTTCCGCGGAATCGCACCA...................AATATGAGGGACGATTATACAGGTAGTCTTATATCGATTTCTGCCC
Melsp1_07 TTCTGCAAACTACCTCCCTAGGTAGTCTTCCGCGGAATCGCACCA...................AATATGAGGGACGACTATACAGGTAGTATCGGTGACTCAGAAGCT
Melsp1_11 AATAGCTATATAAACTACTTAGGTAGTCTTCCGCGGAATCGCACCA...................AATACGAGGGACGACTATACAGGTAATTATCTGGATTAATAATCAT
Melsp1_13 TTGTCTCCGTGGACCTACCTAGGTAGTCTTCCGCAGAATCACACCA...................AATATGAGGGATAACTATACAGGTATGGTTGTGGGGCTTCTTAGA

Supplementary Figure S6: **A**. An alignment of the ends of four versions of the *Spok* block displaying the TSD (red trapezoid) plus the insertion site of *Enterprise* in three *P. comata* strains. The (partial) palindromic motif is marked in magenta. The majority of *Enterprise* is deleted in the strain TD (PODCO) and unassembled in Wa131. Additionally, the Wa139 copy is inserted next to a copy of *schoutedenella* that is absent in both PODCO and Wa131. **B**. Alignment of four *Enterprise* elements within *Melanconium sp.* NRRL 54901 (Melsp1; adjacent numbers correspond to scaffolds in the assembly) revealing the TSD and palindromic motif.



Supplementary Figure S7: Gene expression of the second *Uhera* gene in the *Spok* block of a *Psk-1* strain. **A**. Manual curation of the gene model reveals that a stop codon (red arrow) present in all *Uhera* genes shows indications of RNA A to I editing, as shown in the zoomed in version in **B**. Polymorphic sites in the coverage track are marked as long as they have a minor allele frequency larger than 0.1.



Supplementary Figure S8: The genome of *Melanconium sp.* NRRL 54901 has *Enterprise* elements. **A**. Dot plots representing MUMmer output of alignments between the four regions of *Melanconnium sp.* that are inferred to be copies of an *Enterprise* element, plus flanking sequence. Black segments represent collinear homology, red segments mark inverse homology. The golden bar represents the *Kirc*_Msp gene. Genomic regions and scaffolds are provided. Note that the region from scaffold 11 is the only one fully annotated with gene models and contains the specific *Kirc*_Msp gene analysed here. **B**. Comparison of the *Melanconium sp. Enterprise* element from scaffold 11 to the Wa137 *Spok* block using PROmer. Grey boxes on the axis represent boundaries of annotated genes; *Kirc* is marked in gold. Only the *Kirc* genes show homology to each other. The inset shows the *Kirc* homologs aligning, with boxes now representing exons.