The Enterprise: A massive transposon carrying *Spok* meiotic drive genes

Aaron A. Vogan*[1], S. Lorena Ament-Velásquez[1], Eric Bastiaans[1,2], Ola Wallerman[3], Sven J. Saupe[4], Alexander Suh[1], Hanna Johannesson[1]

[1]Systematic Biology, Department of Organismal Biology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden
[2]Laboratory of Genetics, Wageningen University, Arboretumlaan 4, 6703 BD, Wageningen, The Netherlands
[3]Department of Medical Biochemistry and Microbiology, Comparative genetics and functional genomics; Uppsala University, Husargatan 3, 752 37 Uppsala, Sweden
[4]IBGC, UMR 5095, CNRS Université de Bordeaux, 1 rue Camille Saint Saëns, 33077 Bordeaux, cedex, France
*For Correspondence: Aaron Vogan, aaron.vogan@ebc.uu.se

Abstract

The genomes of eukaryotes are full of parasitic sequences known as transposable elements (TEs). Most TEs studied to date are relatively small (50 -- 12000 bp), but can contribute to very large proportions of genomes. Here we report the discovery of a giant tyrosine-recombinase-mobilized DNA transposon, *Enterprise*, from the model fungus *Podospora anserina*. Previously, we described a large genomic feature called the *Spok* block which is notable due to the presence of meiotic drive genes of the *Spok* gene family). The *Spok* block ranges from 110 kb to 247 kb and can be present in at least four different genomic locations within *P. anserina*, despite what is an otherwise highly conserved genome structure. We have determined that the reason for its varying positions is that the *Spok* block is not only capable of meiotic drive, but is also capable of transposition. More precisely, the *Spok* block represents a unique case where the *Enterprise* has captured the *Spok*s, thereby parasitizing a resident genomic parasite to become a genomic hyperparasite. Furthermore, we demonstrate that *Enterprise* (without the *Spok*s) is found in other fungal lineages, where it can be as large as 70 kb. Lastly, we provide experimental evidence that the *Spok* block is deleterious, with detrimental effects on spore production in strains which carry it. In contrast to the selfish role of the *Enterprise* in *P. anserina*, we speculate that the mobility of the *Enterprise* may also play an adaptive role in many other fungi, through the horizontal transfer of metabolic genes. This union of meiotic drivers and a

35   transposon has created a selfish element of impressive size in *Podospora*, challenging our

36   perception of how TEs influence genome evolution and broadening the horizons in terms of

37   what the upper limit of transposition may be.

38

Introduction

40

41   Transposable elements (TEs) are major agents of change in eukaryotic genomes. Their ability

42   to selfishly parasitise their host replication machinery has large impacts on both genome size

43   and on gene regulation (Chénais et al. 2012). In extreme cases, TEs can contribute up to 85%

44   of genomic content (Schnable et al. 2009) and expansion and reduction of TEs can result in

45   rapid changes in both genome size and architecture (Haas et al. 2009; Talla et al. 2017; Möller

46   and Stukenbrock 2017). Generally, TEs have small sizes (~50 -- 12000 bp) and accomplish

47   these large-scale changes through their sheer number. For example, there are ~1.1 million *Alu*

48   elements in the human genome, which have had a large impact on genome evolution (Bennett

49   et al. 2008; Jurka 2004). The largest known cases among Class I retrotransposons are long

50   terminal repeat (LTR) elements that can be as large as 30 kb, but among Class II DNA

51   transposons, Mavericks/Polintons are known to grow as large as 40 kb through the capture of

52   additional open reading frames (ORFs) (Arkhipova and Yushenova 2019). Recently, a

53   behemoth TE named *Teratorn* was described in teleost fish; it can be up to 182 kb in length,

54   dwarfing all other known TEs. *Teratorn* has achieved this impressive size by fusing a *piggyBac*

55   DNA transposon with a herpesvirus, thereby blurring the line between TEs and viruses (Inoue et

56   al. 2017, 2018). Truly massive transposons may be lurking in the depths of many eukaryotic

57   genomes, but the limitations of short-read genome sequencing technologies and the lack of

58   population-level high-quality assemblies may make them difficult to identify.

59

60   The *Spok* block is a large genomic feature that was first identified thanks to the presence

61   of the **spo**re **k**illing (*Spok*) genes in species from the genus *Podospora* (Vogan, Ament-

62   Velásquez, et al. 2019; Grognet et al. 2014). The *Spoks* are selfish genetic elements that bias

63   their transmission to the next generation in a process known as meiotic drive. Here, drive occurs

64   by inducing the death of spores that do not inherit them, through a single protein that operates

65   as both a toxin and an antidote (Vogan, Ament-Velásquez, et al. 2019; Grognet et al. 2014).

66   The first *Spok* gene described, *Spok1*, was discovered in *P. comata* (Grognet et al. 2014). In *P.*

67   *anserina*, the homologous gene *Spok2* is found at high population frequencies, while two other

68   genes of the *Spok* family, *Spok3* and *Spok4,* are at low to intermediate frequencies (Vogan,

69   Ament-Velásquez, et al. 2019). Unlike *Spok1* and *Spok2*, however, *Spok3* and *Spok4* are

70   always associated with a large genomic region (the *Spok* block). The *Spok* block can be located

71   at different chromosomal locations in different individuals, but is never found more than once in

72   natural strains. The number of *Spok* genes and the location of the *Spok* block (which carries

73   *Spok3*, *Spok4* or both) define the overall meiotic driver behavior of a given genome, which can

74   be classified into the so-called **Podospora s**pore **k**illers or *Psk*s (Vogan, Ament-Velásquez, et al.

75   2019; van der Gaag et al. 2000). The *Spok* block stands out not only because of its size,

76   typically around 150 kb, but also because there is otherwise high genome synteny among

77   strains of *P. anserina* and with the related species *P. comata* and *P. pauciseta* (Vogan, Ament-

78   Velásquez, et al. 2019).

79

80   The fact that the *Spok* block is found at unique genomic positions between otherwise

81   highly similar strains is of prime interest as each novel *Spok* block position creates a unique

82   meiotic drive type (*Psk*) due to the intricacies of meiosis in *Podospora*. In this study, we have

83   determined that the reason the *Spok* block occurs at multiple genomic locations is that the *Spok*

84   block itself is actually a unique version of the *Enterprise*, a novel DNA transposon that is

85   mobilized by a tyrosine recombinase (YR). The *Spok* block represents an *Enterprise* that has

86   captured meiotic drive genes and subsequently grown to a massive size through a gradual

87   accumulation of DNA sequence. We find copies of *Enterprise* without meiotic drive genes in

88   other fungal species which are at least 70 kb, placing *Enterprise* among the largest known TEs,

89   even when not harbouring the *Spok*s. The *Spok* block version of *Enterprise* not only represents

90   the largest TE discovered to date (reaching 247 kb), but is also capable of both transposition

91   and meiotic drive. Furthermore, our data suggest that the *Spok* block is associated with fitness

92   costs in *P. anserina*. This discovery drastically changes what we know about the size limits of

93   transposition, and provides new insight into the role of massive TEs in eukaryotic genome

94   evolution.

Results

Newly isolated *Podospora* strains reveal a novel spore killer type

We isolated and sequenced with Illumina HiSeq technology a strain of *P. anserina* (named Wa137) and two strains of *P. comata* (Wa131 and Wa139) collected in Wageningen, the Netherlands. The strains Wa137 and Wa139 were also sequenced using MinION Oxford Nanopore technology, achieving similar quality to published chromosome-level assemblies (**Supplementary Table 1**). Additionally, we included in our analyses all published long-read genomes of *P. anserina* (10 strains), *P. pauciseta* (one), and *P. comata* (one) (Espagne et al. 2008; Vogan, Ament-Velásquez, et al. 2019; Silar et al. 2019), which are mostly assembled at chromosome level, with respective genome sizes around 36 Mb (**Supplementary Table 1**). Previous work demonstrated that the *Spok* block can be found at three unique chromosomal positions among characterized *P. anserina* strains, defining the killer types *Psk-1/5* (with the *Spok* block located at chromosome 3), *Psk-2* (right arm of chromosome 5) and *Psk-7/8* (left arm of chromosome 5) (Vogan, Ament-Velásquez, et al. 2019) (**Table 1**). The newly isolated strain Wa137 was found to have a *Spok* block with the largest size yet reported (247 kb) and in a novel position (chromosome 1), conferring it a new spore killing phenotype that we named *Psk-9* (see (Vogan, Ament-Velásquez, et al. 2019) for details on nomenclature). The published *P. pauciseta* genome has a *Spok* block in chromosome 4 (*Psk-P1*), but it seems to represent a fragmented version of the *P. anserina* blocks (Vogan, Ament-Velásquez, et al. 2019). Unlike *P. anserina* and *P. pauciseta*, the newly isolated strains of *P. comata* were found to have a single full-copy *Spok* gene (*Spok1*) which is not associated with any *Spok* block-like features, in agreement with the previously published reference genome of *P. comata* (**Table 1**) (Grognet et al. 2014; Silar et al. 2019).

The *Spok* block moves via transposition

A number of mechanisms exist by which a sequence can move within a genome, including reciprocal translocation, ectopic recombination, and transposition (Mieczkowski, Lemoine, and Petes 2006). We can rule out reciprocal translocations in the case of the *Spok* block, as overall chromosome synteny is preserved (Vogan, Ament-Velásquez, et al. 2019). Ectopic recombination is often mediated by TEs, and transposition is, by definition, the process of TE mobility. In order to determine a candidate mode by which the *Spok* block moves, we examined

125   the four unique *Spok* block insertion sites. This analysis revealed that no specific TE was

126   present at the insertion site, but showed that six base pairs (RGGTAG) are always present and

127   are repeated at the end of the *Spok* block (**Figure 1A and B**). This repeated sequence may

128   represent a target site duplication (TSD), which is a hallmark of transposition mechanisms

129   (Wicker et al. 2007). Additionally, the *Psk-1* and *Psk-9 Spok* block insertion sites constitute a

130   partially palindromic sequence ATACYT||AGGTAG (**Figure 1B**), a characteristic of some TE

131   target sites (Linheiro and Bergman 2012). Together, this finding supports an explanation

132   whereby the *Spok* block is mobilized like a TE via transposition.

133

134   The terminal sequences of TEs are often composed of structural features that are

135   intrinsic to the transposition mechanism; LTRs for LTR retrotransposons and terminal inverted

136   repeats (TIRs) for transposase-mobilized DNA transposons, for example (Wicker et al. 2007).

137   These features can thus be used to determine the type of transposition or transposon

138   underlying a given translocation. With this expectation, we examined the ends of the *Spok* block

139   (which are very similar across all instances (Vogan, Ament-Velásquez, et al. 2019)), but found

140   no such structural features. Curiously, the ends of the *Spok* block have in fact been previously

141   identified as two unclassified repetitive elements called *bufo* and *schoutedenella* (Espagne et al.

142   2008) (**Figure 1B**). We suspect that these unrelated elements actually represent the two ends

143   of the *Spok* block (*bufo* representing the first 463 bp and *schoutedenella* representing the last

144   231 bp), and are found on their own in the genome as remains of previous, partially deleted,

145   *Spok* block copies. To test this assumption, we mapped the location of all *bufo* and

146   *schoutedenella* elements present in the *Podospora* genomes. We found that both elements can

147   be identified throughout the chromosomes, with a tendency for *bufo* and *schoutedenella* to

148   locate in TE-rich areas in all species (**Figure 1C and Supplementary Figure 1**). In agreement

149   with the postulate that the genomic copies of *bufo* and *schoutedenella* represent past *Spok*

150   block insertions, at the *Psk-7* insertion site there is ~1200 bp of sequence that is homologous to

151   the beginning of the *Spok* block. This is nearly 700 bp longer than what *bufo* alone represents,

152   and further implies that the *Psk-7 Spok* block was inserted at the same site as a previous, now

153   largely deleted copy. Together, these findings further support the idea that the *Spok* block

154   inserts at the specific target site sequence RGGTAG which is duplicated during transposition

155   (**Figure 1B**).

156

The *Spok* block accumulates sequence from the genome

157

158 As there were no structural features to guide us in our attempt to determine the mechanism
159 through which the *Spok* block translocates, we examined the composition of the *Spok* block to
160 identify candidate genes capable of transposing it. We annotated all genome assemblies using
161 a modified version of the annotation pipeline in (Vogan, Ament-Velásquez, et al. 2019), relying
162 on additional RNA-seq data, as well as an improved, manually curated repeat library (see
163 Methods). The content of the different copies of the *Spok* block is largely overlapping (Vogan,
164 Lorena Ament-Velásquez, et al. 2019), although the Wa137 *Spok* block has a large region of
165 unique sequence resulting in a total size of 247 kb (**Figure 1C and Supplementary Figure 2**).
166 Very few genes were predicted within the *Spok* block by this methodology (as evidenced by the
167 dip in the orange track on the chr 1.2 scaffold; **Figure 1C**), so we manually annotated the *Spok*
168 block of representative strains (Wa53, Wa28, Wa58, and Wa137), which resulted in numerous
169 hypothetical protein-coding genes (e.g., 67 in the Wa137 *Spok block*). The blocks do not
170 generally exhibit unusually high TE load. For example, the *Spok* block in Wa137 is only
171 composed of 10.5% annotated TEs (compared to genome wide estimates of 3--6%,
172 **Supplementary Table S1**), nor do they appear to have a strong signature of repeat-induced
173 point mutation (RIP). RIP is a process that operates in numerous fungi, including *Podospora*
174 (Graïa et al. 2001), by specifically inducing C-to-T mutations of any repeated sequence within
175 the genome (Selker et al. 1987; Cambareri et al. 1989) and results in a drop of GC proportion in
176 repetitive elements. Such a GC-drop is clear in many regions of the *Podospora* genome, but is
177 conspicuously absent within the *Spok* block (Vogan, Ament-Velásquez, et al. 2019). While some
178 of the predicted genes had identifiable homologs within the *P. anserina* reference genome
179 (orange features in **Figure 1C**), many are absent from the reference but have sequence
180 similarity to genes from other fungi (purple features in **Figure 1C**). Of note, many of the
181 predicted genes have potential roles in various metabolic pathways, like metal tolerance or
182 antimicrobial resistance (**Supplementary Table S2**), rather than canonical TE-derived genes,
183 such as transposases or reverse transcriptases. It thus seems likely that the *Spok* block has
184 grown to its large size, at least in part, by accumulating non-repetitive sequence from elsewhere
185 within the genome. The accumulation does not appear to have been a single event, but was
186 more likely many individual events, as the genomic homologs are found on all chromosomes
187 (green links in **Figure 1C**).

The *Enterprise*

188

189 The gene annotation provided us with very few clues as to the agent(s) responsible for the

190 translocation of the *Spok* block, so we turned to another approach. We hypothesize that the

191 *Spok* block may have been a much smaller element in the past, and that such an element may

192 still be transposing in the genomes of *Podospora* species. The closely related species *P.*

193 *comata* possesses both *bufo* and *schoutedenella* repeats, and appears to have large areas of

194 homology to the *Spok* block (**Supplementary Figure 1B**). We discovered that the *P. comata*

195 strain Wa139 contains a 39 kb subtelomeric region on chromosome 5 (henceforth referred to as

196 the *Enterprise*) that is nearly completely composed of TEs and sequence homologous to the

197 *Spok* block (**Supplementary Figure 2**). The ends of the *Enterprise* are consistent with the *Spok*

198 block (both *bufo* and *schoutedenella* are present) and it possesses the RGGTAG target site

199 duplication (**Supplementary Figure 3A**). *Enterprise* is absent at the orthologous location in *P.*

200 *anserina* and *P. pauciseta*, is much smaller in the other *P. comata* strain T_D (5.5 kb), and

201 unassembled in Wa131. The other *P. comata* strains possess *bufo* and *schoutedenella* at this

202 location, suggesting that this specific *Enterprise* insertion is now polymorphic for its state of

203 degradation. However, it is difficult to fully recapitulate the history of the region due to the limited

204 number of strains and the fact that Wa139 also has another *schoutedenella* copy at the insertion

205 site that is absent in the other strains (**Supplementary Figure 3A**). Manual annotation revealed

206 five putative genes within the Wa139 *Enterprise*, four of which are homologous to the genes

207 within the *Spok* block and thus represent good candidates for involvement in the transposition of

208 *Enterprise* and therefore the *Spok* block.

209

210 None of these four genes have a known function and none are *Spok* homologs. We

211 examined their predicted protein domains, searched the *P. anserina* genomes for homologs,

212 and scanned genomic databases for similar sequences in an attempt to discern whether they

213 may be integral to the movement of the *Enterprise* (**Supplementary Table 3**). Only one of the

214 genes appears to be a likely candidate for enacting the transposition of the *Enterprise*. This

215 gene is always present as the first ORF in the *Spok* block, but is degraded and interrupted by

216 multiple TEs in Wa139. This gene possesses two domains which may be characteristic of

217 transposons: a domain of unknown function called DUF3435, and a predicted CHROMO

218 domain (**Figure 1D**). CHROMO domains are implicated in histone binding and may allow TEs to

219 be targeted to specific regions of the genome (Kordis 2005). Although the other domain is of

220 unknown function, the annotation (pfam11917) in the Pfam protein families database (El-Gebali

221   et al. 2019) states that it may be related to phage integrases (pfam00589), which could indicate

222   that DUF3435 represents a tyrosine recombinase (YR) domain. Therefore, we conducted an

223   analysis using Promals3D (Pei, Kim, and Grishin 2008) to compare the protein sequence of the

224   ORF to crystal structures of known YRs (CRE, XERD, and FLP) as well as to the protein

225   sequence from the transposable element *Crypton-Cn1* (a known TE that transposes via a YR,

226   but has no structural features (Kojima and Jurka 2011), as with the *Enterprise*). This analysis

227   revealed that although there is little to no sequence similarity between the protein and known

228   YRs, there is high conservation at the structural level. Importantly, YRs are known to have a

229   catalytic tetrad consisting of an R-H-R-Y motif (Esposito and Scocca 1997), which is present in

230   DUF3435 (**Figure 1D & Supplementary Figure 4**). We therefore name this gene "spore **ki**lling

231   **r**elated **c**rypton" or *Kirc*.

232

233   From the evidence presented here, we propose that the *Spok* block and the homologous

234   region of Wa139 represent copies of a previously unknown group of DNA transposons, which

235   we name *Enterprise*. We define *Enterprise* as being composed of a YR-encoding gene

236   homologous to *Kirc*, and possessing a TSD (**Figure 1E**). *Cryptons* are a type of DNA

237   transposon defined only by the presence of a YR domain (Wicker et al. 2007) and may possess

238   a TSD or not (Kojima and Jurka 2011), thus *Enterprise* can be classified as a novel group of

239   *Crypton.*

240

241   To confirm our hypothesis that *Enterprise* is capable of transposition and selfish replication, we

242   mined fungal genomes available on JGI MycoCosm for homologous proteins of *Kirc* that are

243   present in multiple copies within a single genome. We identified such a case in *Melanconium*

244   *sp.* NRRL 54901. In this genome, a homolog of *Kirc* is found at the beginning of a ~70 kb region

245   that is present in four copies in the genome. Critically, this region is flanked by the same TSD as

246   the *Spok* block and the target site is in its full palindromic context CTACCT||AGGTAG

247   (**Supplementary Figure 5**). The only gene homologous with the *Enterprise* is the *Kirc* homolog,

248   thereby confirming that the minimal feature for transposition is the putative YR-encoding gene,

249   *Kirc,* and classifying this region of *Melanconium sp.* as an *Enterprise.*

Kirc is wide-spread in filamentous fungi

251   Given the fact that *Enterprise* is present in at least one species outside of *Podospora*, we

252   queried GenBank with the sequence of *Kirc* using BLASTp to determine how widespread

253   *Enterprise* is. We recovered a total of 481 protein hits, which were almost exclusively from

254   within the Pezizomycotina, although putative homologs were identified from ten basidiomycete

255   genomes as well. A phylogeny of a representative set of sequences shows that the relationships

256   among the homologs do not follow the expected species phylogeny (**Figure 2**). Numerous

257   species have multiple *Kirc* homologs, some of which are not closely related. Within *P. anserina*,

258   one homolog was recovered, Pa_5_10116 (see Espagne et al. (2008) for *Podospora* gene

259   notation), which appears to be distantly related to *Kirc*, yet highly similar to homologs from

260   *Fusarium*. Furthermore, Pa_5_10116 is pseudogenized and absent in the close relatives of *P.*

261   *anserina*, *P. comata* and *P. pauciseta*, consistent with it being a transposable element. Attempts

262   to describe TSDs in other fungi were largely unsuccessful as conserved flanks could not be

263   identified in most cases, which is not unusual for *Cryptons* in general (Kojima and Jurka 2011).

264   Thus, in the absence of additional evidence from these fungal species, we consider these as

265   *Enterprise*-like elements. Regardless, these results suggest that *Enterprise* is a YR-mobilized

266   group of DNA transposons that is spread throughout fungal genomes.

The *Spok* block can be deleterious

268   In *P. anserina*, wild strains with more than one full copy of the *Spok* block have never been

269   found, although it is easy to generate them in the lab via crosses (Vogan, Ament-Velásquez, et

270   al. 2019). It is possible that the burden imposed by the *Spok* block on the host is quite high,

271   leading to selection to purge most copies in nature. To evaluate this hypothesis, we made use

272   of backcrosses of two of the *Psk* strains (Vogan, Ament-Velásquez, et al. 2019). The

273   backcrossed strains $Psk1xS_{14}$ and $Psk7xS_{14}$ are isogenic with the reference strain S, except that

274   S has no *Spok* block. $Psk1xS_{14}$ has the *Spok* block on chromosome 3 and induces killing in 90%

275   of meioses, and $Psk7xS_{14}$ has the *Spok* block on the left arm of chromosome 5 and induces

276   spore killing in 50% of meioses. We crossed strains either to themselves (no spore killing) or to

277   strain S (spore killing) and evaluated a number of traits related to fitness. Radial growth and

278   time to germination of spores produced by the matings showed no variation between crosses,

279   however significant differences were observed among the amount of spores produced by a

280   cross (**Figure 3**). Specifically, crosses from selfings of $Psk1xS_{14}$ produced significantly fewer

281   spores than from selfings of either S or $Psk7xS_{14}$, despite the fact that no spore killing occurs,

282   indicating that the *Psk-1 Spok* block inhibits spore production. This effect was even greater in

283   crosses between S and $Psk1xS_{14}$, and was most prominent when $Psk1xS_{14}$ was used as the

284   female, suggesting a maternal effect. With both $Psk1xS_{14}$ and $Psk7xS_{14}$ more spores were

285   produced in the killing crosses than expected given the proportion of killing per ascus (Vogan,

286   Ament-Velásquez, et al. 2019). As there were no significant differences in the amount of spores

287   produced by Psk7xS$_{14}$ than by S, this result suggests that the general presence of the *Spok*

288   block itself is not deleterious, but rather that the negative effect is due to the specific content

289   and/or to the genomic location of a given *Spok* block.

Discussion

291   Here we have provided evidence that the *Spok* block is a variant of a newly described TE,

292   *Enterprise*, that moves throughout the genome of *P. anserina* by means of YR-mediated

293   transposition. In addition to this movement, the *Spok* block is also capable of meiotic drive due

294   to the presence of the *Spok* genes. Given these two selfish properties, it can be asked who is

295   parasitizing whom. One possibility is that *Enterprise* has hijacked the *Spoks* in order to increase

296   its rate of transmission, thereby parasitizing a resident genomic parasite and becoming a

297   genomic hyperparasite. Support for this commandeering can be found in the distribution of the

298   *Spok* genes. In both *P. anserina* and *P. comata*, the *Spok* genes that are not found in

299   association with the *Spok* block are at high frequency (*Spok2* is found in 98% of strains isolated

300   in Wageningen (Vogan, Ament-Velásquez, et al. 2019), and so far *Spok1* has been found in all

301   analysed strains of *P.* comata). The *Spok* block is comparatively rare (~18% of strains from

302   Wageningen (Vogan, Ament-Velásquez, et al. 2019)), indicating that it may prevent the *Spoks*

303   from reaching high frequencies. Alternatively, the *Spok* genes may generally benefit from

304   moving throughout the genome. *Spok1* and *Spok2* are found at different locations in the

305   genome and are surrounded by TEs, suggesting they may have moved through other

306   mechanisms like ectopic recombination (Vogan, Ament-Velásquez, et al. 2019). It may thus be

307   advantageous for the *Spoks* to mobilize within TEs, like *Enterprise*, in order to change their

308   genomic position on a regular basis, due to the fact that this relocation will result in a novel

309   spore killing phenotype. Presumably, the population dynamics of meiotic drive ultimately decide

310   the fate of the *Spoks* (Nauta and Hoekstra 1993), but the confederation of the *Spoks* and

311   *Enterprise* as the *Spok* block may fundamentally change how effective selection is at controlling

312   either element.

313

314   Given that the results of the fitness experiments suggest that the *Spok* block can be

315   deleterious, there may be strong selection to purge any copies of *Enterprise*, but this may

316   ultimately be dependent on genomic context as insinuated by the differences between the *Psk-1*

317   and *Psk-7 Spok* blocks. The *Psk-7 Spok* block is larger by nearly 50 kb, some of which includes

318   known retrotransposons. Yet, only *Psk-1* showed a significant decrease in the amount of spores

319   produced. It thus seems probable that the location of the block has a stronger deleterious effect

320   rather than its content or size. The *Psk-1 Spok* block is located close to the centromere on the

321   left arm of chromosome 3. It is possible that the increased amount of killing in *Psk-1* results in

322   its poorer performance, although this cannot explain the observed maternal effect. The rDNA

323   cluster resides on the same chromosome arm as the *Psk-1* block (Espagne et al. 2008). Given

324   that this arm is only ~700 kb and the block itself is 113 kb, the *Spok* block insertion might

325   interfere with recombination, which is necessary for proper segregation of rDNA (Tomson et al.

326   2006) and likely inhibits the ability of the strain to produce viable spores. Moreover, the

327   observation that there is not a large decrease in spore production for *Psk-1* or *Pks-7* strains

328   when involved in spore killing suggests that strains are able to compensate for the lost spores in

329   some way. This results has significant implications to our understanding of spore killing as a

330   meiotic drive, as it shifts the system away from providing the killer genotype a relative fitness

331   advantage to an absolute fitness advantage (Nauta and Hoekstra 1993; Lyttle 1991). Explicitly,

332   in the naïve expectation, a spore killer that is 100% efficient would reduce the total amount of

333   spores produced in a killing cross to half. Thus, it does not produce more total offspring with its

334   genotype than a non-spore killing gene (relative advantage). However, if the strain carrying the

335   spore killer is able to compensate for the loss of spore production, and produce the same or

336   similar number of spores as in a non-killing cross (as observed here), the spore killer observes

337   an absolute advantage, as is the case for other types of meiotic drive, like female meiotic drive.

338   Therefore, spore killers may be more successful at invading and driving through populations

339   than previously thought.

340

341   The *Enterprise* clearly has the ability to move large amounts of genetic material around

342   the genome and TEs are known to be agents of horizontal gene transfer (HGT) (Schaack,

343   Gilbert, and Feschotte 2010; Gilbert and Feschotte 2018). As such, it is plausible that *Enterprise*

344   and related YR-mobilized DNA transposons may jump between species, and thereby transport

345   additional genes. The phylogeny of the *Kirc* homologs is indicative of HGT as it shows closely

346   related *Kirc* homologs distributed amongst unrelated fungi. In at least two cases, fungal TEs

347   have been implicated as the vehicles for gene mobility. In the first case, a hAT element is

348   associated with the HGT of toxin genes among cereal pathogens (McDonald et al. 2019). The

349   second case comes from a recent publication which described a TE named *HEPHAESTUS* in

350   the fungus *Paecilomyces* (Urquhart et al. 2020). It carries multiple genes that provide resistance

351   to at least five different heavy metals and shows evidence of transfer to a distantly related

352   species of *Penicillium*. In both these cases, the HGT appears to have been adaptive. Additional

353   cases of adaptive HGT between fungi have been described multiple times (Wisecaver and

354   Rokas 2015), with examples of transferred regions being as large as 500 kb in the cheese mold

355   *Penicillium roquefortii* (Ropars et al. 2015) and known HGT cases involving *P. anserina* itself

356   (Slot and Rokas 2011). Whether the *Enterprise* can also play a role in adaptive HGT is unknown

357   as of yet, but the potential certainly exists. Future studies should focus on the interplay between

358   adaptive gene transfer and selfish replication to understand how these two possible features of

359   large TEs are impacting genome evolution.

360   # Conclusions

361   The constant 'tug-of-war' between TEs linking themselves to host genes, and the actions of

362   genome defense and selection to purge them is of key importance to the evolution of genome

363   architecture. It is likely that we are witnessing this fight play out to the extreme in *P. anserina*,

364   with the high effectiveness of both RIP and the *Spok* block, making *Podospora* an ideal system

365   in which to continue to study genomic conflict. The discovery that the up to 247 kb large *Spok*

366   block transposes likely through a YR-mediated mechanism moves the upper limit of TE size by

367   nearly 100 kb, and its hyper-selfishness combining TE mobility and meiotic drive adds a new

368   dynamic by which selfish elements exploit their host genome. This not only changes our

369   perception of how TEs influence genome evolution, but also broadens the horizons in terms of

370   what may be possible through genetic manipulations in the laboratory. Understanding the

371   molecular mechanism of YR-mediated transposition should thus be of prime interest in future

372   research.

373

374

375   # Methods

376

377   ## Fungal material

378

379 Strains used in this study were obtained from the Wageningen collection and cultivated as in

380 Vogan, Ament-Velásquez, et al. (2019). Strain S was used as the standard reference strain with

381 no *Spok* block. As spore killer strains for the fitness experiments (see below) we used the

382 backcrossed strains Psk1xS$_{14}$ and Psk7xS$_{14}$ (Vogan, Ament-Velásquez, et al. 2019), which

383 should be isogenic to S but with a *Spok* block in chromosome 3 and 5, respectively. The more

384 recently isolated strains Wa131, Wa137 and Wa139 were sampled during the fall of 2016

385 around Wageningen (the Netherlands) from dung of rabbit (Wa131 and Wa137, locality

386 Unksepad Oosterbeek) or horse (Wa139, locality Uiterwaarden Wolfswaard). Morphological

387 differences like smaller perithecia and abundant tomentose apricot-colored mycelium in HPM

388 medium (Vogan, Ament-Velásquez, et al. 2019), as well as analyses of sequence data, allowed

389 us to assign Wa131 and Wa139 to the species *P. comata*. Previously, only one strain from this

390 species, T$_D$, was known (Boucher, Nguyen, and Silar 2017; Vogan, Ament-Velásquez, et al.

391 2019), hence these new strains constitute a new report of this species for the Netherlands.

392

393 # DNA and RNA extraction and sequencing

394 Following Vogan et al. (Vogan, Ament-Velásquez, et al. 2019), we grew monokaryotic strains on

395 PASM0.2 plates covered with a layer of cellophane. Genomic DNA for short-read sequencing

396 was extracted from 80-100 mg of fungal tissue with the Fungal/Bacterial Microprep kit (Zymo;

397 www.zymo.com). Paired-end libraries were prepared and sequenced using the Illumina HiSeq X

398 (150-bp-long) technology at the SNP and SEQ Technology platform (SciLifeLab, Uppsala,

399 Sweden). For RNA extraction, around 150 mg of harvested mycelium were frozen in liquid

400 nitrogen and stored at -80°C. We extracted total RNA from the grounded frozen tissue using the

401 RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Quality was checked on the agilent 2100

402 Bioanalyzer (Agilent Technologies, USA) and the RNA was treated with DNaseI (Thermo

403 Scientific). The sequencing library was prepared with a NEBNext Ultra Directional RNA Library

404 Prep Kit for Illumina (New England Biolabs). We purified polyA+ transcripts with the NEBNext

405 Poly(A) mRNA Magnetic Isolation Module (New England Biolabs). A paired-end library was

406 sequenced with Illumina HiSeq 2500 at the SNP and SEQ Technology platform.

407

408 For long-read sequencing, we grew the monokaryotic strains in PASM0.2 plates, from

409 where we sliced small agar cubes to inoculate liquid cultures of 200 ml 3% malt extract solution

410 incubated in a shaker for 10-14 days at 27°C (Vogan, Ament-Velásquez, et al. 2019). Mycelium

411   aggregates were filtered from the flasks, any remaining agar was removed, and around 1 g was

412   stored at -20°C. As described in Sun et al. (2017), the tissue was freeze-dried and macerated,

413   followed by DNA extraction using Genomic Tip G-500 columns (Qiagen) and cleaning with the

414   PowerClean DNA Clean-Up kit (MoBio Labs). Additionally, DNA was purified using magnetic

415   beads (Speed-Beads, GE) and eluted for 20 min at 37°C followed by overnight storage at 4°C

416   twice to increase concentration (around 65 ng/µl). An R9.5.1 Flowcell (Oxford Nanopore

417   Technologies) was used for whole-genome sequencing of each strain with the rapid sequencing

418   kit (SQK-RAD004) using the standard protocol. Basecalling was done using Guppy v. 1.6.

Genome assembly

420   For most strains we used the assemblies produced in Vogan & Ament-Velásquez et al. (2019).

421   For newly sequenced strains, we produced new assemblies as follows. The adapters from the

422   Illumina HiSeq reads were identified with cutadapt v. 1.13 (Martin 2011) and removed using

423   Trimmomatic v. 0.36 (Bolger, Lohse, and Usadel 2014) using the following options:

424   ILLUMINACLIP:adapters.fasta:1:30:9 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20

425   MINLEN:30. Pairs with both forward and reverse reads after filtering were used for downstream

426   analyses. For the strain Wa131, which only has Illumina data, we used SPAdes v. 3.12.0

427   (Bankevich et al. 2012) with the k-mers 21,33,55,77 and the *--careful* option. For the strains

428   Wa137 and Wa139, the MinION reads with a mean Phred quality (QV) above 9 and longer than

429   1 kb were assembled using Minimap2 v. 2.11 and Miniasm v. 0.2 (Li 2018, 2016). The resulting

430   assembly was polished twice with Racon v. 1.3.1 (Vaser et al. 2017) using all MinION reads (no

431   filtering). Further polishing was done with the filtered Illumina reads in five consecutive rounds of

432   Pilon v. 1.22 (Walker et al. 2014). We used BWA v. 0.7.17 (Li and Durbin 2010) for short-read

433   mapping, with PCR duplicates marked using Picard v. 2.18.11

434   (http://broadinstitute.github.io/picard/), as well as local indel re-alignment using the Genome

435   Analysis Toolkit (GATK) v. 3.7 (Van der Auwera et al. 2013).

436

437   We assigned the scaffolds to chromosomes based on alignments to the reference

438   genome of the S strain (Espagne et al. 2008), available at the Joint Genome Institute

439   MycoCosm website (https://mycocosm.jgi.doe.gov/mycocosm/home) as "Podan2" (Grigoriev et

440   al. 2014). We discarded small contigs (<100kb) of rDNA repeats as well as mitochondrial-

441   derived sequences, except for the largest mitochondrial contig. We assessed the quality of the

442   final assemblies by visual inspection of the mapping of both long and short reads using

443   Minimap2 and BWA, respectively. Mean depth of coverage was calculated with QualiMap v.2.2

444   (Okonechnikov, Conesa, and García-Alcalde 2016). Other assembly statistics were calculated

445   with QUAST v. 4.6.3 (Mikheenko et al. 2016).

Genome annotation

447   A GitHub repository is available with Snakemake v. 5.4.4 (Köster and Rahmann 2012) pipelines

448   relevant to genome annotation at https://github.com/johannessonlab/SpokBlockPaper.

449

450   The TEs and other repeats in *P. anserina* were classified previously by Espagne et al.

451   (2008) based on the original reference genome of the S strain or "Podan1", and is hereafter

452   referred to as the "Espagne library". To explore the diversity of TEs in the newly generated

453   *Podospora* genomes, we identified repeats de novo and manually compared them to the

454   Espagne library to identify duplicates and new elements. Specifically, we ran RepeatModeler v.

455   1.0.8 (http://www.repeatmasker.org/RepeatModeler/) on the scaffolds larger than 50 kb of all

456   available long-read assemblies (Snakemake pipeline *PaTEs.smk*). Each resulting

457   RepeatModeler consensus was BLASTn-searched back to the original genome and the best 20

458   hits with 2-kb flanks were aligned with T-Coffee v. 12.00.7fb08c2 (Notredame, Higgins, and

459   Heringa 2000) (*TEManualCuration.smk*), and visually inspected for manual curation. The

460   curated consensuses were assigned to the Espagne et al. (2008) equivalents based on

461   similarity (allowing for RIP-induced mutations) or were given a new name when having no

462   homology to anything in the Espagne library. It was discovered that the *gypsy* element *crapaud*

463   has numerous diverged copies with unique LTRs. We annotated all *crapaud* LTRs that were in

464   multiple copies within *P. anserina* individually to improve repeat masking. We refer to the final

465   repeat library as "PodoTE-1.00" (available at the GitHub repository).

466

467   To generate a genome annotation of all assemblies, we ran an updated version of the

468   pipeline in Vogan & Ament-Velásquez et al. (2019), named *PaAnnotation.smk*. We used

469   MAKER v. 3.01.2 (Holt and Yandell 2011; Campbell et al. 2014) with the previously produced

470   training files used for the ab initio prediction programs GeneMark-ES v. 4.32 (Lomsadze et al.

471   2005; Ter-Hovhannisyan et al. 2008) and SNAP release 2013-06-16 (Lomsadze et al. 2005), as

472   well as the following dependencies: RepeatMasker v. 4.0.7 (http://www.repeatmasker.org/),

473   BLAST suite 2.6.0+ (Camacho et al. 2009), Exonerate v. 2.2.0 (Slater and Birney 2005), and

474   tRNAscan-SE v. 1.3.1 (Lowe and Eddy 1997). As evidence, we used STAR v. 2.6.1b (Dobin et

475   al. 2013) to produce transcript models (maximum intron length set to 1000 bp) of various RNA-

476   seq data sets. Specifically, we mapped the reads of the monokaryotic isolate Wa63- (*P.*

477   *anserina*) to the assembly PaWa63m (Vogan, Ament-Velásquez, et al. 2019), of the

478   monokaryotic isolate Wa131- (*P. comata*) to the assembly PcWa139m (this study), and of the

479   dikaryotic Psk7xS$_{14}$ (*P. anserina*) to the assembly PaWa58m (Vogan, Ament-Velásquez, et al.

480   2019). We then processed the mapped reads with Cufflinks v. 2.2.1 (Trapnell et al. 2010) to

481   obtain the transcript models. As external evidence, we used CDS from the Podan2 annotation,

482   protein sequences from the T strain of *P. comata* (Silar et al. 2019), and a small dataset of

483   manually curated proteins. To aid in manual curation of selected regions (mostly the *Spok*

484   block), we visually inspected the mapping of RNA-seq reads of the different datasets, along with

485   CDS produced with TransDecoder v. 5.5.0 (Haas et al. 2013) on the Cufflinks models, as well

486   as the output of RepeatMasker ran externally from MAKER with the PodoTE-1.00 library.

487   Additionally, we queried predicted gene models into the NCBI databases (NCBI Resource

488   Coordinators 2016) to verify the annotations.

Genome alignments

490   We used the NUCmer program from the MUMmer package v. 4.0.0beta2 (Kurtz et al. 2004)

491   using the parameters *-b 200 -c 22 --maxmatch* to align the *Spok* blocks to each other, and

492   *changed to -c 40* for whole-genome assemblies. To achieve higher sensitivity, we used BLASTn

493   from the BLAST suite 2.9.0 (Camacho et al. 2009) to search for the presence of the unclassified

494   repeats *bufo* and *schoutenella*. Both the NUCmer and the BLAST alignments were plotted using

495   Circos v. 0.69.6 (Krzywinski et al. 2009) along with manual curations of coding regions and

496   repetitive elements. The distribution of TE and gene content along chromosomes was

497   calculated in windows of 50 kb with steps of 10 kb using BEDtools v. 2.29.0 (Quinlan and Hall

498   2010; Quinlan 2014), with the utilities *makewindows* and *coverage*. The fraction of conservation

499   between blocks compared to the block in Wa137 was calculated by aligning the block

500   sequences (within the TSD) of Wa28 (*Psk-2*), Wa53 (*Psk-1*), Wa58 (*Psk-7*) and Wa139 with

501   NUCmer and the BEDtools utility *genomecov*. The Snakemake pipelines used to produce the

502   Circos plots (*CircosBlock.smk* and *CircosAllBlocks.smk*) are available at

503   https://github.com/johannessonlab/SpokBlockPaper.

Phylogenetic analyses

504

505 To determine the relationship of *Kirc* to other YRs, the crystal structure of known YRs (CRE

506 (PDB code 3mgv), XERD (1a0p), and FLP (1flo)) as well as the protein sequence from the

507 transposable element *Crypton-Cn1* were analysed using the software Promals3D (Pei, Kim, and

508 Grishin 2008).

509

510 Homologs of *Kirc* were identified from GenBank using BLASTp with the truncated

511 version of *Kirc* from Wa53 that has no CHROMO domain. Nucleotide sequences of hits with e-

512 values < 1e-100 were compiled along with a homolog from *P. anserina* (Pa_5_10116), and two

513 homologs from *Melanconium sp.* NRRL 54901 extracted from MycoCosm (see below), and

514 aligned with MACSE v. 2.03 (Ranwez et al. 2018). We used TrimAl v. 1.4.1 (Capella-Gutierrez,

515 Silla-Martinez, and Gabaldon 2009) to trim the resulting protein alignment with the *-gappyout*

516 option. We then used IQTree v. 1.6.8 (Kalyaanamoorthy et al. 2017; Nguyen et al. 2014) to

517 produce a Maximum Likelihood phylogeny with extended model selection (–m MFP). To

518 estimate the branch support, we produced 1000 standard bootstrap pseudoreplicates.

519

520 To search Mycocosm for other copies of *Enterprise* the following approach was taken.

521 The protein sequence of *Kirc* was used as a query with BLASTX against all genomes within

522 MycoCosm (as of February 2019). Genomes with multiple high confidence positive hits were

523 identified and the regions with putative *Kirc* homologs were manually extracted. Priority was

524 given to genomes where the hits were associated with large duplicated regions (>50 kb).

525 *Melanconium sp.* NRRL 54901 had the most copies with clear termini.

Fitness assays

526

527 The cultures used for the crosses were revived from the -80 freezer on PASM0.2 (van

528 Diepeningen et al. 2008) for several days and then stored at +4°C until use. Strains were grown

529 for 5 days on fresh PASM0.2 plates before inoculating the cross. The strain that was assigned

530 the female role was grown in a 35 mm petri dish with 5 ml HPM medium (Vogan, Ament-

531 Velásquez, et al. 2019) by inoculating a small cube of agar with mycelium (~2 x 2 mm). In

532 parallel, the strain that was assigned the male role was grown on a 90 mm petri dish with micro

533 conidiation medium (Esser 1974; King 2013) by inoculating seven plugs of mycelium spread

534 over the plate. After 7 days of growth, microconidia were harvested by adding 5 ml of sterile

535 water to the plate and sweep over the mycelium with a drigalski spatula for 1 minute. The

536 female mycelium was then fertilized with 0.5 ml of the microconidial suspension. The

537 suspension was carefully spread out to make sure all mycelium was covered. The fertilized

538 mycelia were then further incubated under standard conditions (27ºC, 12/12 light/dark cycle)

539 (Vogan, Ament-Velásquez, et al. 2019). The cultures were monitored daily for signs of spores

540 shot from the asci in order to score the first day of spore-shooting. To reduce the complexity of

541 the experiment, the strains used as female were always of mating type *mat+*.

542 At 6 days post-fertilization, single spores were collected with a needle to measure

543 germination frequency and growth speed. From each cross, 10 spores from 4-spored asci were

544 picked, and in cases with spore killing, an additional 10 spores from 2-spored asci were picked.

545 The 10 spores were transferred to a single 90 mm petri dish with PASM2 medium (van

546 Diepeningen et al. 2008) with 0.4% ammonium acetate added (to activate the spores) (Esser

547 1974; King 2013). Spores were spaced out in a predetermined pattern (4 lines of 2, 3, 3, 2

548 spores). After two days of incubation, the germination was scored and colony diameter was

549 measured in two directions. If there was no growth microscopic inspection was performed to

550 check whether a spore was present in the agar to avoid scoring no germination in case the

551 inoculation failed.

552 At 12 days post-fertilization, spores were harvested from the lids of each crossed culture

553 and used for estimating total spore yield. Spores were collected by pipetting 750 µl of harvest

554 liquid (1M NAOH, 0.025% SDS) in the lid. Spores were then scraped off the lid using the pipette

555 tip. The liquid was then collected into a 2 ml Eppendorf tube. Another 750 µl of harvest liquid

556 was used to repeat the process to make sure most of the spores were collected from the lid.

557 The tubes were then heated for 4 hours at 85ºC, then shaken in a Qiagen Tissuelyser for 90

558 seconds at 30 Hz. After this, the tubes were stored at 4ºC overnight. The cooled tubes were

559 again shaken in a Qiagen Tissuelyser for 90 seconds at 30 Hz. This process prevents the

560 clumping of spores. Total yield was determined by counting the amount of spores in a volume of

561 5 µl of 50x diluted suspension pipetted on an object glass using a stereomicroscope. Counts

562 were taken five times for each replicate cross. Statistical analyses were conducted in base R v.

563 3.5.0 to determine significance and power.

Acknowledgements

References

Arkhipova, Irina R., and Irina A. Yushenova. 2019. "Giant Transposons in Eukaryotes: Is Bigger Better?" *Genome Biology and Evolution* 11 (3): 906–18.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19 (5): 455–77.

Bennett, E. A., H. Keller, R. E. Mills, S. Schmidt, J. V. Moran, O. Weichenrieder, and S. E. Devine. 2008. "Active *Alu* Retrotransposons in the Human Genome." *Genome Research* 18 (12): 1875-1883.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.

Boucher, Charlie, Tinh-Suong Nguyen, and Philippe Silar. 2017. "Species Delimitation in the *Podospora anserina/P. pauciseta/P. comata* Species Complex (Sordariales)." *Cryptogamie, Mycologie*. 38 (4): 485-506.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 421.

Cambareri, E., B. Jensen, E. Schabtach, and E. Selker. 1989. "Repeat-Induced G-C to A-T Mutations in *Neurospora*." *Science*. 244 (4912): 1571-1575.

Campbell, Michael S., Carson Holt, Barry Moore, and Mark Yandell. 2014. "Genome Annotation and Curation Using MAKER and MAKER-P." *Current Protocols in Bioinformatics* 48: 4.11.1–39.

Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon. 2009. "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25 (15): 1972-1973.

Chénais, Benoît, Aurore Caruso, Sophie Hiard, and Nathalie Casse. 2012. "The Impact of

603   Transposable Elements on Eukaryotic Genomes: From Genome Size Increase to Genetic
604   Adaptation to Stressful Environments." *Gene* 509 (1): 7–15.

605 Diepeningen, Anne D. van, Alfons J. M. Debets, S. Marijke Slakhorst, and Rolf F. Hoekstra.
606   2008. "Mitochondrial pAL2-1 Plasmid Homologs Are Senescence Factors in *Podospora*
607   *anserina* Independent of Intrinsic Senescence." *Biotechnology Journal* 3 (6): 791–802.

608 Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha,
609   Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal
610   RNA-Seq Aligner." *Bioinformatics* 29 (1): 15–21.

611 El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R. Eddy, Aurélien Luciani, Simon C. Potter,
612   Matloob Qureshi, et al. 2019. "The Pfam Protein Families Database in 2019." *Nucleic Acids*
613   *Research* 47 (D1): D427–32.

614 Espagne, Eric, Olivier Lespinet, Fabienne Malagnac, Corinne Da Silva, Olivier Jaillon, Betina M.
615   Porcel, Arnaud Couloux, et al. 2008. "The Genome Sequence of the Model Ascomycete
616   Fungus *Podospora anserina*." *Genome Biology* 9 (5): R77.

617 Esposito, D., and J. J. Scocca. 1997. "The Integrase Family of Tyrosine Recombinases:
618   Evolution of a Conserved Active Site Domain." *Nucleic Acids Research* 25 (18): 3605–14.

619 Gaag, M. van der, A. J. Debets, J. Oosterhof, M. Slakhorst, J. A. Thijssen, and R. F. Hoekstra.
620   2000. "Spore-Killing Meiotic Drive Factors in a Natural Population of the Fungus *Podospora*
621   *anserina*." *Genetics* 156 (2): 593–605.

622 Gilbert, Clément, and Cédric Feschotte. 2018. "Horizontal Acquisition of Transposable Elements
623   and Viral Sequences: Patterns and Consequences." *Current Opinion in Genetics &*
624   *Development* 49: 15–24.

625 Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X,
626   Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. (2014) "MycoCosm
627   portal: gearing up for 1000 fungal genomes." *Nucleic Acids Research* 42(1): D699-704.

628 Grognet, Pierre, Hervé Lalucque, Fabienne Malagnac, and Philippe Silar. 2014. "Genes That
629   Bias Mendelian Segregation." *PLoS Genetics* 10 (5): e1004387.

630 Haas, Brian J., Sophien Kamoun, Michael C. Zody, Rays H. Y. Jiang, Robert E. Handsaker,
631   Liliana M. Cano, Manfred Grabherr, et al. 2009. "Genome Sequence and Analysis of the
632   Irish Potato Famine Pathogen *Phytophthora infestans*." *Nature* 461 (7262): 393–98.

633 Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood,
634   Joshua Bowden, Matthew Brian Couger, et al. 2013. "De Novo Transcript Sequence
635   Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and
636   Analysis." *Nature Protocols* 8 (8): 1494–1512.

637 Holt, Carson, and Mark Yandell. 2011. "MAKER2: An Annotation Pipeline and Genome-
638   Database Management Tool for Second-Generation Genome Projects." *BMC*
639   *Bioinformatics* 12 : 491.

640 Inoue, Yusuke, Masahiko Kumagai, Xianbo Zhang, Tomonori Saga, Deshou Wang, Akihiko
641   Koga, and Hiroyuki Takeda. 2018. "Fusion of piggyBac-like Transposons and
642   Herpesviruses Occurs Frequently in Teleosts." *Zoological Letters* 4 (1): 6.

643 Inoue, Yusuke, Tomonori Saga, Takumi Aikawa, Masahiko Kumagai, Atsuko Shimada, Yasushi
644   Kawaguchi, Kiyoshi Naruse, Shinichi Morishita, Akihiko Koga, and Hiroyuki Takeda. 2017.
645   "Complete Fusion of a Transposon and Herpesvirus Created the *Teratorn* Mobile Element
646   in Medaka Fish." *Nature Communications* 8 (1): 551.

647 Jurka, Jerzy. 2004. "Evolutionary Impact of Human *Alu* Repetitive Elements." *Current Opinion in*
648   *Genetics & Development* 14 (6): 603–8.

649 Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K. F. Wong, Arndt von Haeseler, and Lars
650   S. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates."
651   *Nature Methods* 14 (6): 587–89.

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

Kojima, Kenji K., and Jerzy Jurka. 2011. "Crypton Transposons: Identification of New Diverse Families and Ancient Domestication Events." *Mobile DNA* 2 (1): 12.

Kordis, Dusan. 2005. "A Genomic Perspective on the Chromodomain-Containing Retrotransposons: Chromoviruses." *Gene* 347 (2): 161–73.

Köster, Johannes, and Sven Rahmann. 2012. "Snakemake – a Scalable Bioinformatics Workflow Engine." *Bioinformatics* 28 (19): 2520–22.

Krzywinski, Martin, Jacqueline Schein, Inanç Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. 2009. "Circos: An Information Aesthetic for Comparative Genomics." *Genome Research* 19 (9): 1639–45.

Kurtz, Stefan, Adam Phillippy, Arthur L. Delcher, Michael Smoot, Martin Shumway, Corina Antonescu, and Steven L. Salzberg. 2004. "Versatile and Open Software for Comparing Large Genomes." *Genome Biology* 5 (2): R12.

Li, Heng. 2016. "Minimap and Miniasm: Fast Mapping and de Novo Assembly for Noisy Long Sequences." *Bioinformatics* 32 (14): 2103–10.

———. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* 34 (18): 3094–3100.

Li, Heng, and Richard Durbin. 2010. "Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 26 (5): 589–95.

Linheiro, Raquel S., and Casey M. Bergman. 2012. "Whole Genome Resequencing Reveals Natural Target Site Preferences of Transposable Elements in *Drosophila melanogaster*." *PloS One* 7 (2): e30008.

Lomsadze, Alexandre, Vardges Ter-Hovhannisyan, Yury O. Chernoff, and Mark Borodovsky. 2005. "Gene Identification in Novel Eukaryotic Genomes by Self-Training Algorithm." *Nucleic Acids Research* 33 (20): 6494–6506.

Lowe, T. M., and S. R. Eddy. 1997. "tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence." *Nucleic Acids Research* 25 (5): 955–64.

Lyttle, T. W. 1991. "Segregation Distorters." *Annual Review of Genetics* 25: 511–57.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.journal*. 17 (1): 10-12.

McDonald, Megan C., Adam P. Taranto, Erin Hill, Benjamin Schwessinger, Zhaohui Liu, Steven Simpfendorfer, Andrew Milgate, and Peter S. Solomon. 2019. "Transposon-Mediated Horizontal Transfer of the Host-Specific Virulence Protein *ToxA* between Three Fungal Wheat Pathogens." *mBio* 10 (5): e01515-19.

Mieczkowski, Piotr A., Francene J. Lemoine, and Thomas D. Petes. 2006. "Recombination between Retrotransposons as a Source of Chromosome Rearrangements in the Yeast *Saccharomyces cerevisiae*." *DNA Repair* 5 (9-10): 1010–20.

Mikheenko, Alla, Gleb Valin, Andrey Prjibelski, Vladislav Saveliev, and Alexey Gurevich. 2016. "Icarus: Visualizer for de Novo Assembly Evaluation." *Bioinformatics* 32 (21): 3321–23.

Möller, Mareike, and Eva H. Stukenbrock. 2017. "Evolution and Genome Architecture in Fungal Plant Pathogens." *Nature Reviews. Microbiology* 15 (12): 771.

Nauta, M. J., and R. F. Hoekstra. 1993. "Evolutionary Dynamics of Spore Killers." *Genetics* 135 (3): 923–30.

NCBI Resource Coordinators. 2016. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 44 (D1): D7–19.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2014. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. "T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment." *Journal of Molecular Biology* 302 (1): 205–17.

Okonechnikov, Konstantin, Ana Conesa, and Fernando García-Alcalde. 2016. "Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data."

705   *Bioinformatics* 32 (2): 292–94.

706 Pei, Jimin, Bong-Hyun Kim, and Nick V. Grishin. 2008. "PROMALS3D: A Tool for Multiple
707   Protein Sequence and Structure Alignments." *Nucleic Acids Research* 36 (7): 2295–2300.

708 Quinlan, Aaron R. 2014. "BEDTools: The Swiss-Army Tool for Genome Feature Analysis."
709   *Current Protocols in Bioinformatics* 47: 11.12.1–34.

710 Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing
711   Genomic Features." *Bioinformatics* 26 (6): 841–42.

712 Ranwez, Vincent, Emmanuel J. P. Douzery, Cédric Cambon, Nathalie Chantret, and Frédéric
713   Delsuc. 2018. "MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for
714   Frameshifts and Stop Codons." *Molecular Biology and Evolution* 35 (10): 2582–84.

715 Ropars, Jeanne, Ricardo C. Rodríguez de la Vega, Manuela López-Villavicencio, Jérôme
716   Gouzy, Erika Sallet, Émilie Dumas, Sandrine Lacoste, et al. 2015. "Adaptive Horizontal
717   Gene Transfers between Multiple Cheese-Associated Fungi." *Current Biology: CB* 25 (19):
718   2562–69.

719 Schaack, Sarah, Clément Gilbert, and Cédric Feschotte. 2010. "Promiscuous DNA: Horizontal
720   Transfer of Transposable Elements and Why It Matters for Eukaryotic Evolution." *Trends in
721   Ecology & Evolution* 25 (9): 537–46.

722 Schnable, Patrick S., Doreen Ware, Robert S. Fulton, Joshua C. Stein, Fusheng Wei, Shiran
723   Pasternak, Chengzhi Liang, et al. 2009. "The B73 Maize Genome: Complexity, Diversity,
724   and Dynamics." *Science* 326 (5956): 1112–15.

725 Selker, E. U., E. B. Cambareri, B. C. Jensen, and K. R. Haack. 1987. "Rearrangement of
726   Duplicated DNA in Specialized Cells of *Neurospora*." *Cell* 51 (5): 741–52.

727 Silar, Philippe, Jean-Marc Dauget, Valérie Gautier, Pierre Grognet, Michelle Chablat, Sylvie
728   Hermann-Le Denmat, Arnaud Couloux, Patrick Wincker, and Robert Debuchy. 2019. "A
729   Gene Graveyard in the Genome of the Fungus *Podospora comata*." *Molecular Genetics
730   and Genomics: MGG* 294 (1): 177–90.

731 Slater, Guy St C., and Ewan Birney. 2005. "Automated Generation of Heuristics for Biological
732   Sequence Comparison." *BMC Bioinformatics* 6: 31.

733 Slot, Jason C., and Antonis Rokas. 2011. "Horizontal Transfer of a Large and Highly Toxic
734   Secondary Metabolic Gene Cluster between Fungi." *Current Biology: CB* 21 (2): 134–39.

735 Sun, Yu, Jesper Svedberg, Markus Hiltunen, Pádraic Corcoran, and Hanna Johannesson. 2017.
736   "Large-Scale Suppression of Recombination Predates Genomic Rearrangements in
737   *Neurospora tetrasperma*." *Nature Communications* 8 (1): 1140.

738 Talla, Venkat, Alexander Suh, Faheema Kalsoom, Vlad Dincă, Roger Vila, Magne Friberg,
739   Christer Wiklund, and Niclas Backström. 2017. "Rapid Increase in Genome Size as a
740   Consequence of Transposable Element Hyperactivity in Wood-White (*Leptidea*) Butterflies."
741   *Genome Biology and Evolution* 9 (10): 2491–2505.

742 Ter-Hovhannisyan, V., A. Lomsadze, Y. O. Chernoff, and M. Borodovsky. 2008. "Gene
743   Prediction in Novel Fungal Genomes Using an Ab Initio Algorithm with Unsupervised
744   Training." *Genome Research* 18 (12) 1979-1990.

745 Tomson, Brett N., Damien D'Amours, Brittany S. Adamson, Luis Aragon, and Angelika Amon.
746   2006. "Ribosomal DNA Transcription-Dependent Processes Interfere with Chromosome
747   Segregation." *Molecular and Cellular Biology* 26 (16): 6239–47.

748 Trapnell, Cole, Brian A. Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J. van
749   Baren, Steven L. Salzberg, Barbara J. Wold, and Lior Pachter. 2010. "Transcript Assembly
750   and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching
751   during Cell Differentiation." *Nature Biotechnology* 28 (5): 511–15.

752 Urquhart, A. S., Chong, N. F., Yang, Y., & Idnurm, A. (2020). "Eukaryotic transposable elements
753   as "cargo carriers": the forging of metal resistance in the fungus *Paecilomyces variotii*."
754   bioRxiv.

755 Van der Auwera, Geraldine A., Mauricio O. Carneiro, Chris Hartl, Ryan Poplin, Guillermo Del

Angel, Ami Levy-Moonshine, Tadeusz Jordan, et al. 2013. "From FastQ Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline." *Current Protocols in Bioinformatics* 43: 11.10.1–33.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 2017. "Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads." *Genome Research* 27 (5): 737–46.

Vogan, Aaron A., S. Lorena Ament-Velásquez, Alexandra Granger-Farbos, Jesper Svedberg, Eric Bastiaans, Alfons J. M. Debets, Virginie Coustou, et al. 2019. "Combinations of Spok Genes Create Multiple Meiotic Drivers in *Podospora*." *eLife* 8: e46454.

Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, et al. 2014. "Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement." *PloS One* 9 (11): e112963.

Wicker, Thomas, François Sabot, Aurélie Hua-Van, Jeffrey L. Bennetzen, Pierre Capy, Boulos Chalhoub, Andrew Flavell, et al. 2007. "A Unified Classification System for Eukaryotic Transposable Elements." *Nature Reviews. Genetics* 8 (12): 973–82.

Wisecaver, Jennifer H., and Antonis Rokas. 2015. "Fungal Metabolic Gene Clusters-Caravans Traveling across Genomes and Environments." *Frontiers in Microbiology* 6: 161.
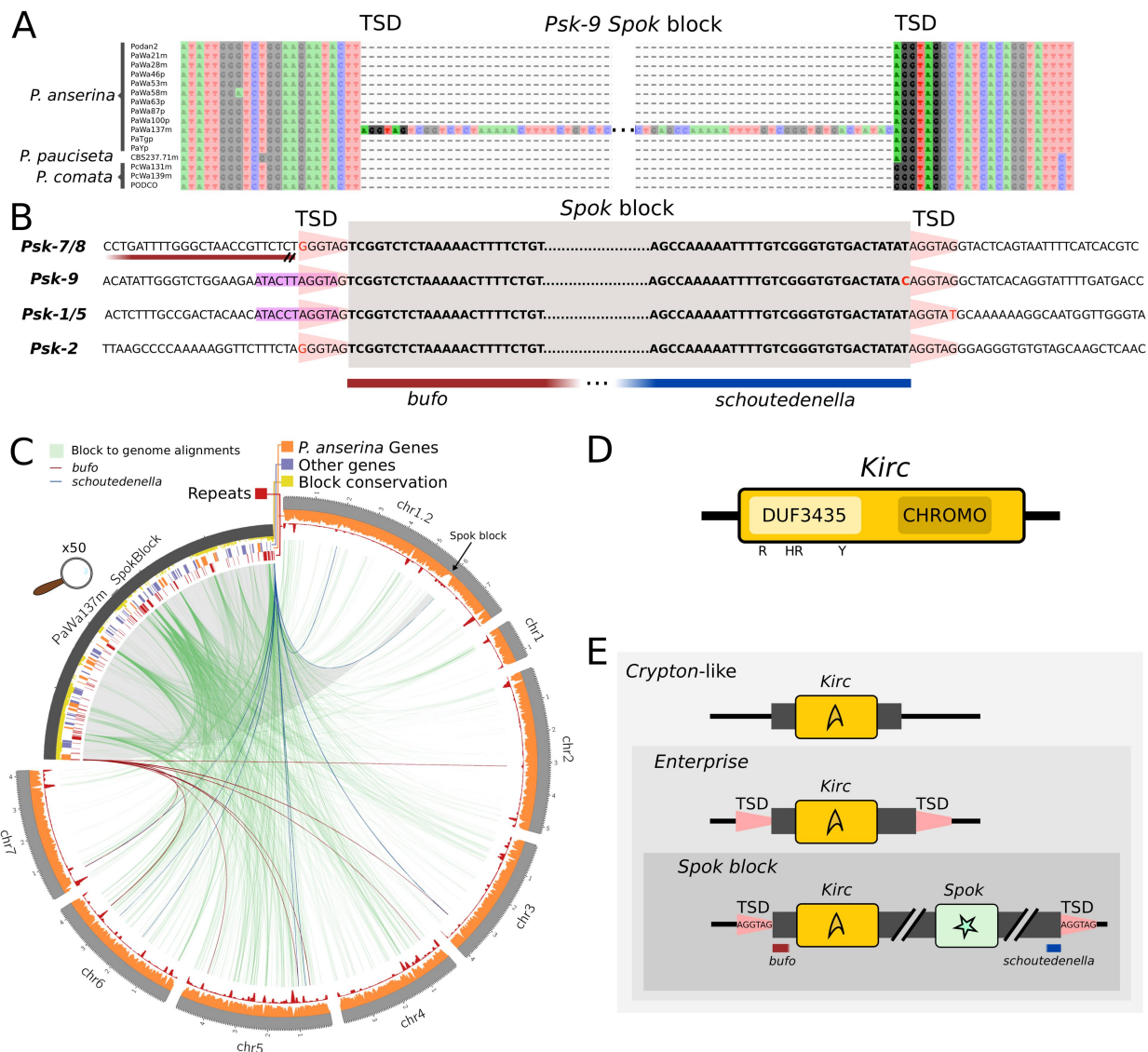
Figures

Figure 1. The genomic context of the *Spok* Block. **A** An alignment of the region where the *Spok* block is located in the strain Wa137 (*Psk*-9) from all *P. anserina* genomes. Putative target site duplication (TSD) is highlighted. Only the ends of the *Spok* block are shown for clarity. **B** An alignment of the ends of four versions of the *Spok* block displaying the TSD (pink trapezoid). The flanks of the *Spok* block that relate to the unclassified elements *bufo* and *schoutedenella* are underlined in red and blue, respectively. Note that for *Psk-7*, the *Spok* block is inserted immediately next to a truncated *bufo* element. **C** A Circos plot of the genome of Wa137 (light grey track, size in Mb) aligned against its own *Spok* block (dark grey, enlarged by 50x). The tracks on the *Spok* block from outside inward represent: the conservation of regions among the different iterations of the *Spok* block where the height of the conservation track is equivalent to

787   the number of blocks that have a given position (yellow); gene models from manual annotation

788   of the *Spok* block, where genes with homologs within the reference genome of *P. anserina* are

789   marked in orange and those without homologs are marked in purple; and annotated repeated

790   elements (red). The tracks on the chromosomal scaffolds show the coverage of genic regions

791   (orange) and repeats (red) calculated in sliding windows of 50 kb with steps of 10 kb. Green

792   lines connect homologous segments based on MUMmer alignments. The unclassified repetitive

793   elements *bufo* and *schoutedenella* are connected in dark red and blue segments, respectively,

794   based on BLASTn searches. **D** Cartoon representation of the predicted protein for the ORF *Kirc*.

795   The catalytic tetrad is marked below the DUF3435 Domain. **E** Cartoon model of the hierarchical

796   nomenclature of *Crypton*-like elements, *Enterprise* and the *Spok* block. See Supplementary

797   Figure 3 for more detailed depictions of the *Spok* block.
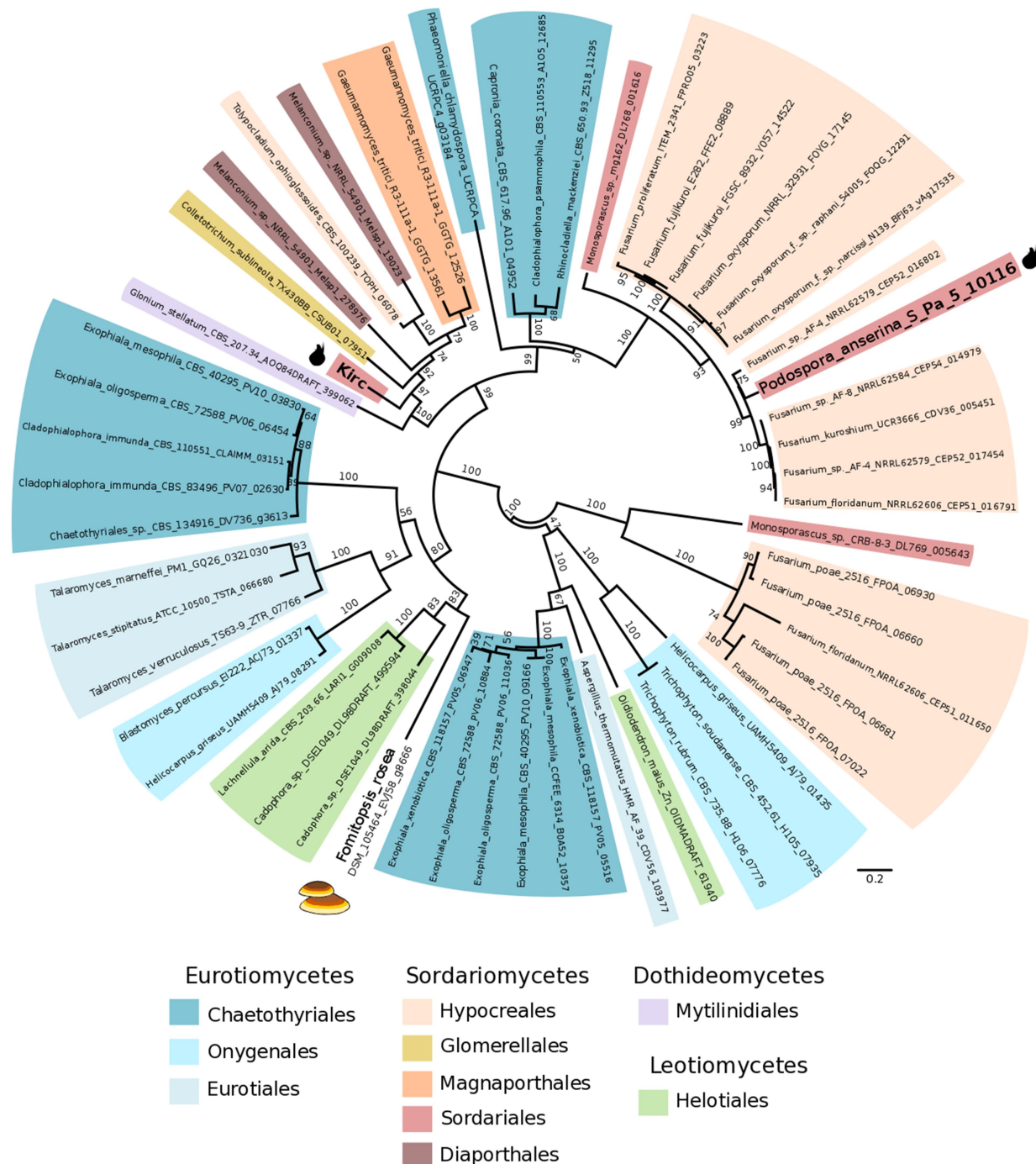
Figure 2. Unrooted Maximum likelihood phylogenetic tree showing the relationship between *Kirc* and other homologs from fungi. Bootstrap support values are shown above branches. Branch lengths are proportional to the scale bar (amino acid substitution per site). Taxonomic rank is indicated with coloured highlights. The sequences present in *Podospora anserina* and in the basidiomycete *Fomitopsis rosea* are marked with cartoons of the corresponding fruiting bodies.

805
806

807
808 **Figure 3.** Plot showing the total amount of ascospores collected from crosses of isogenic
809 strains possessing various iterations of the *Spok* block. Crosses where spore killing occurs are
810 in blue, crosses with no spore killing are in red. Letters mark results of a Tukeys HSD test. The
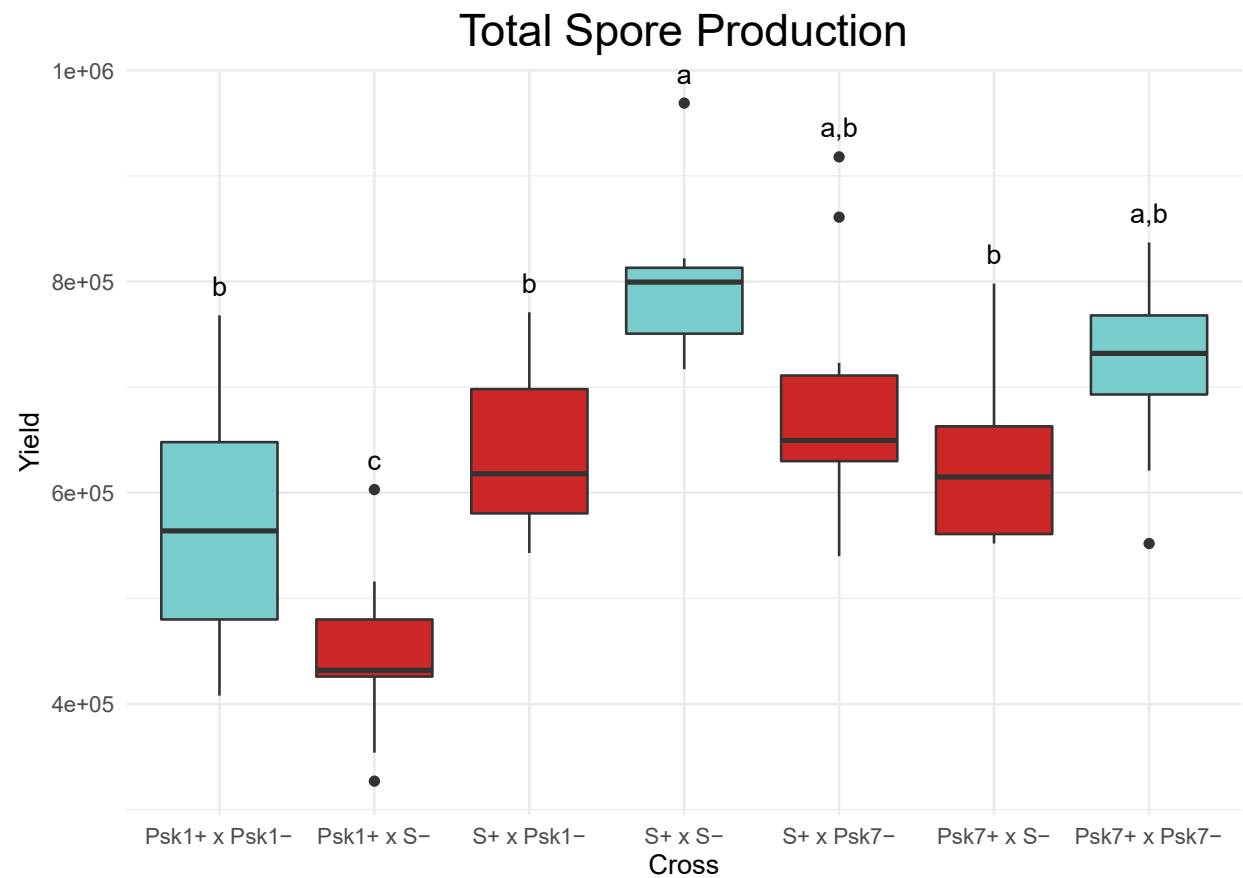811 Mating type is indicated by a + or – after the strain name. The *mat+* parent was always used as
812 the maternal strain.

Tables

Table 1. List of strains used in this study.

Strain	Species	Psk[a]	Genomic location	Size	Elements[b]	
Wa21	*P. anserina*	2	Chr5 R	142433	*Spok* block	*Spok2*
Wa28	*P. anserina*	2	Chr5 R	161297	*Spok* block	*Spok2*
Wa46	*P. anserina*	naïve	-	-	-	-
Wa53	*P. anserina*	1	Chr 3	113407	*Spok* block	*Spok2*
Wa58	*P. anserina*	7	Chr 5 L	167459	*Spok* block	*Spok2*
Wa63	*P. anserina*	S	-	-	-	*Spok2*
Wa87	*P. anserina*	1	Chr 3	113425	*Spok* block	*Spok2*
Wa100	*P. anserina*	8	Chr 5 L	127869	*Spok* block	*Spok2*
Wa137	*P. anserina*	9	Chr 1	247510	*Spok* block	*Spok2*
S	*P. anserina*	S	-	-	-	*Spok2*
TG	*P. anserina*	5	Chr 3	157028	*Spok* block	-
Y	*P. anserina*	5	Chr 3	113435	*Spok* block	-
Wa131	*P. comata*	C1	-	-	-	*Spok1*
Wa139	*P. comata*	C1	Chr5 L	38521	*Enterprise*	*Spok1*
T$_D$	*P. comata*	C1	-	-	-	*Spok1*
CBS 237.71	*P. pauciseta*	P1	Chr 4	74163	*Spok* block	-

[a]Notice that a given *Spok* block might have *Spok3*, *Spok4* or both, and this reflects the *Psk* designation.

[b]The location of the *Spok* block and *Enterprise* element of Wa139 are indicated in the "Genomic Location" column. *Spok1* and *Spok2* are not found within the *Spok* block.

Supplementary Figures
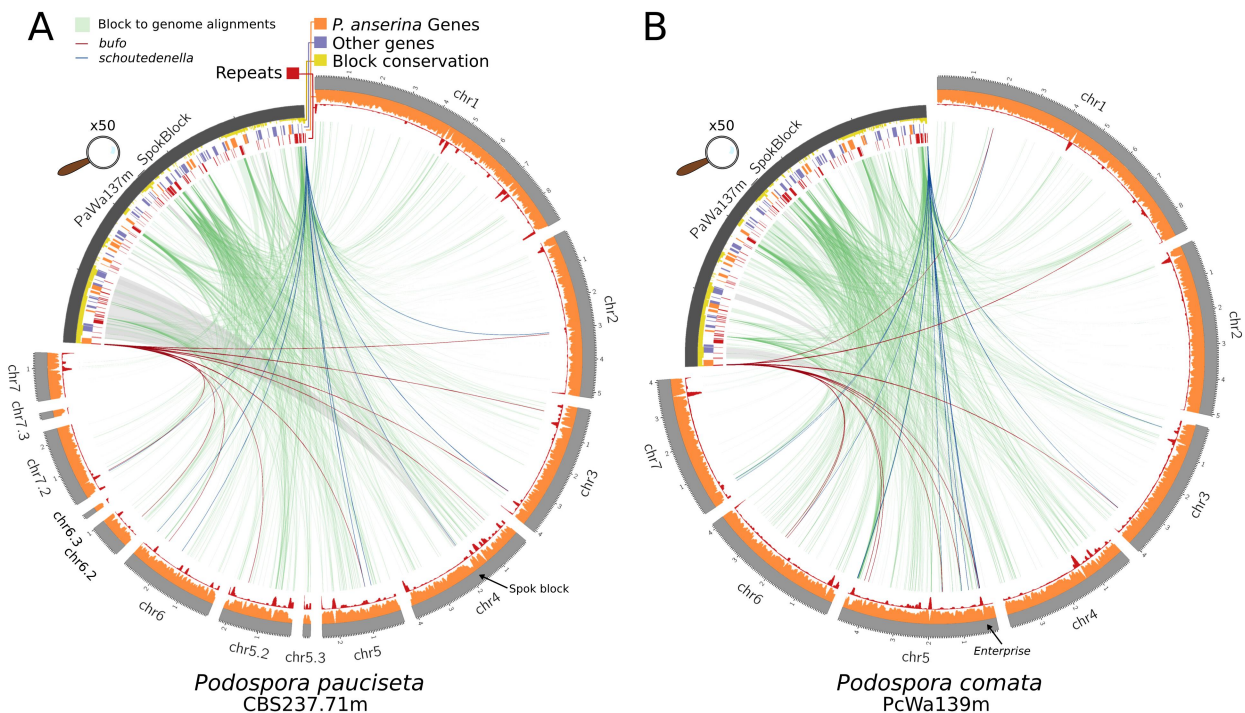
Supplementary Figure 1. Circos plots comparing the Wa137 *Spok* block to the genome of **A** the *P. pauciseta* strain CBS237.71 and **B** the *P. comata* strain Wa139. Colours follow Figure 1C.
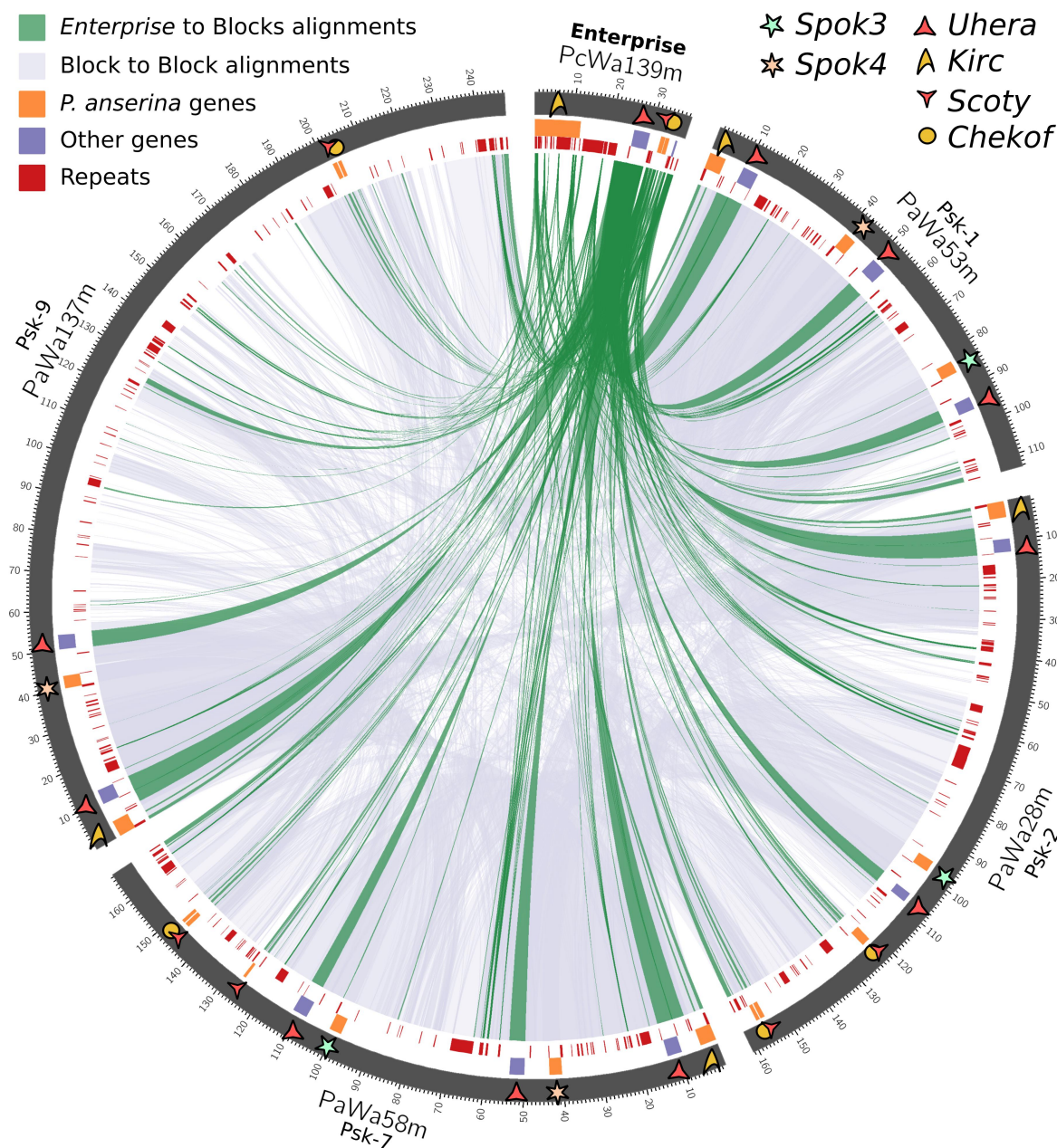
823
824 **Supplementary Figure 2.** A circos plot comparing the *Enterprise* of five different strains (four

825 representing different *Spok* blocks). Dark green lines connect homologous segments of the

826 Wa139 *Enterprise* to the various *Spok* blocks. Lilac lines show homologous regions among the

827 *Spok* blocks. Genes of interest are marked with symbols (See Supplementary Table 3); tracks

828 follow the colour scheme of Figure 1C, but highlighting only relevant genes.

829
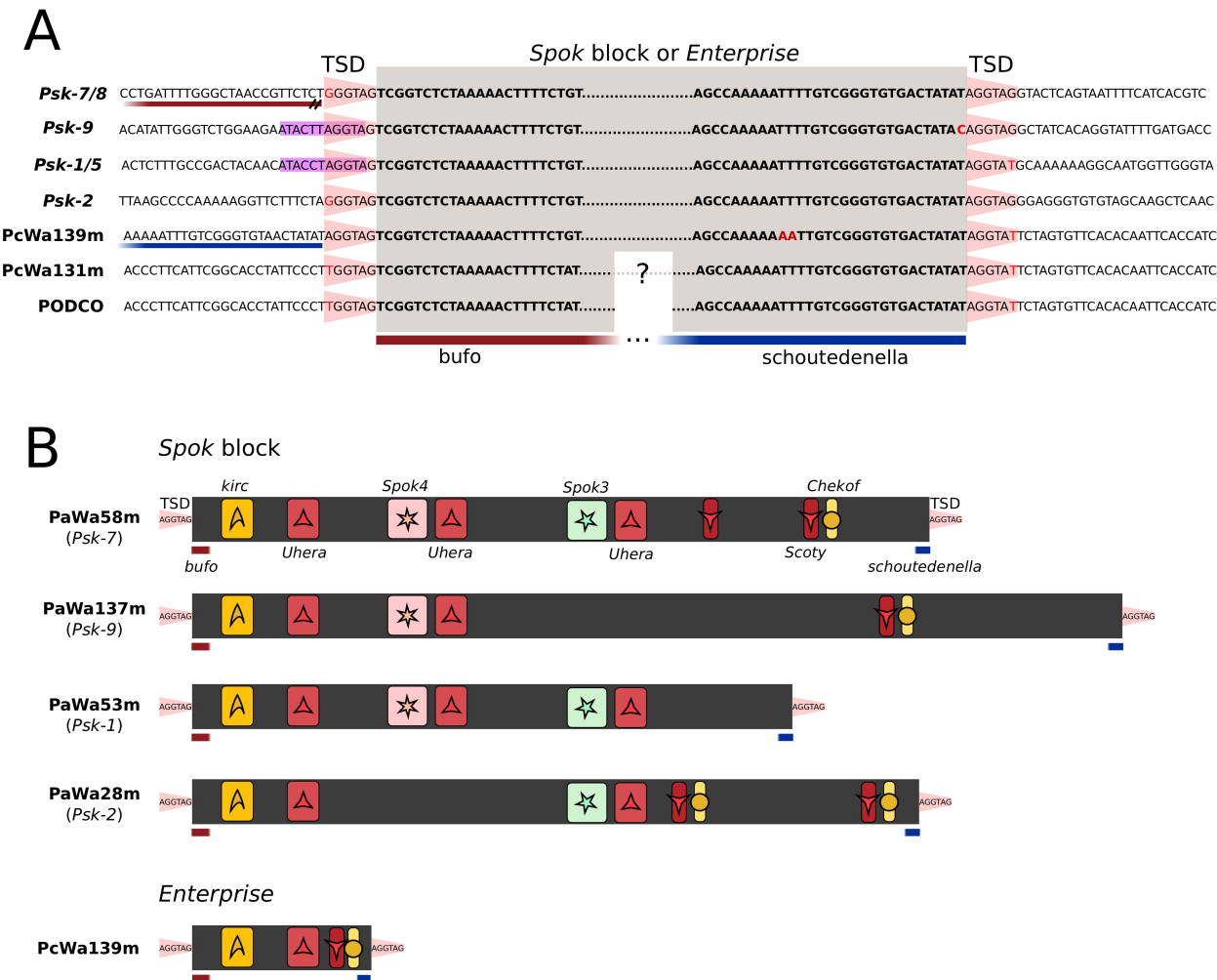
Supplementary Figure 3. Description of the *Spok* Block. **A** An alignment of the ends of four versions of the *Spok* block displaying the TSD (red trapezoid) plus the insertion site of *Enterprise* in three *P. comata* strains. The majority of *Enterprise* is deleted in PODCO and unassembled in Wa131. Additionally, Wa139 is inserted next to a copy of *schoutedenella* that is absent in both PODCO and Wa131. **B** Cartoon models representing the different structure and gene content of four *Spok* blocks and of *Enterprise* from Wa139. Relevant genes and features are annotated.

```
Conservation:                                      6        9 5
3mgv_chainA_p003 116 LA--------FERTDFDQVRSLMEN--SDRC-------QDIRNLAFLGIAYNTLLRIAEIAR-------- 160
Crypton-Cn1      120 GI--------ATTENMKKLFLHYLH--RDSV-------EGLRDLAAQAVGIHGLLRADDQLR-------- 164
1a0p_chainA_p001  92 KD--------LSEAQVERLLQAPL---IDQP-------LELRDKAMLEVLYATGLRVSELVG-------- 135
Kirc             107 RF--------GTKNHFLHLGRQLW---GNDWVVCDKPATRVYDWADLLAIVCSSARVGEYIESTCRAGSG 165
1flo_chainA_p002 132 MLKALLSEGESIWEITEKILNSFEYTSRFTK-------TKTLYQFLFLATFINCGRFSDIKN-------- 186
Consensus_aa:        .h...........p.h.pl.p.hb.....p........pch..bh.h.hhh.s..Rht-h.p........
Consensus_ss:                 hhhhhhhhhhh            h     hhhhhhhhhhhhhhhh


Conservation:                                      65          6
3mgv_chainA_p003 161 --IRVKDISRTD------------GGRMLIHIGRTKTLVSTAGVEKALS-------------------- 195
Crypton-Cn1      165 --ITLSSMSLRLFEDEGPTP----CRGVVFAIREGKTTHDGQIQYSTLLR------------------- 208
1a0p_chainA_p001 136 --LTMSDISL--------------RQGVVRV-IGKGNKE---RLVPLG------------------- 163
Kirc             166 RGLYYRNVTF-----------------GVFL-NEHGNAE---FAVQLVRDAKGMTDNPAKRPEHSLYEG 213
1flo_chainA_p002 187 --VDPKSFKL--------VQNKYLGVIIQCLVTETKTSVS---RHIYFF-------------------- 222
Consensus_aa:        ..l..pshp....................hhl.p.+s.hp.......h...................
Consensus_ss:           hhheee                     eeeee          eeeee


Conservation:                        6                                   6
3mgv_chainA_p003 196 ------------LGVTKLVERWISV--------------SGVAD----------------DPNNYLF-CR 222
Crypton-Cn1      209 -NKDVT------RCPVSFLVLYLFA-------------RFHFSEEPF------------INSDVSF-PS 244
1a0p_chainA_p001 164 ------------EEAVYWLETYLEHGRPW----------LLNGV----------------SIDVLF-PS 193
Kirc             214 LGEMPLI-----CNPMLPILAILIGTKAFKDYETIEDLLNIQPS----EGEMIHLQWKESVLDLPFF-KS 273
1flo_chainA_p002 223 -------SARGRIDPLVYLDEFLRN---------------------------------SEPVLKRVN 249
Consensus_aa:        .............ssh..l..hl.s.................s................p..hb..p
Consensus_ss:                    hhhhhhhhhh


Conservation:                                                9 9
3mgv_chainA_p003 223 VRKNG--VAA------PSAT-----SQLSTRALEGIFEATHRLIYGA-KDDSGQRYLAWSGHSARVGAAR 278
Crypton-Cn1      245 LKNRQDWYHI------PLFV-----SRQSNAVTRLKYDALNKSVRKA-LQSCNIHCR-ASTHTSRKWGAQ 301
1a0p_chainA_p001 194 ----------------QRA-----QQMTRQTFWHRIKHYAVLAGID-------SEK-LSPHVLRHAFAT 233
Kirc             274 ----------------MSARGTPGKIETATAFSKRLRLLGFRAGYSRP--------PTIHDFRAEGLY 317
1flo_chainA_p002 250 --------RTGNSSSNKQEY-----QLLKDNLVRSYNKALKKNAPYSIFA--IKNG-PKSHIGRHLMTS 302
Consensus_aa:        ....................h.....pbbpsphhp...chh...h..t.......p....p.H..Rh..h.
Consensus_ss:                 hhhhhhhhhhhhhhhh              hhhhhhhhhh


Conservation:              6                           9
3mgv_chainA_p003 279 DMARAGVSIPEIMQAGGWTNV--------------NIVMNYIRN------LDSETGAMVRL--------- 319
Crypton-Cn1      302 LAEDGGAPEEDIMRQGRWCTK--------------VMETVYLSKF-----PLKALRALAGFPKKKGSYYL 352
1a0p_chainA_p001 234 HLLNHGADLRVVQMLLSDL---------------STTQIYTHV------ATERLRQLHQ---------- 271
Kirc             318 -WIDKLYSVAQRMKHAGQKDPNTYNNHYQPNNSGTDGQGSYFGL-DVRNIANDLFRGLTL--------- 375
1flo_chainA_p002 303 FLSMKGLT-ELTNVVGNWSDK----------------TTYTHQ-------ITAIPDHY----------- 336
Consensus_aa:        .h....hs......htsb.s................s...sYh..........p.h.shh..........
Consensus_ss:        hhhh    hhhhhhhh               hhhhhhhh      hhhhhhhh
```

839
840 **Supplementary Figure 4.** Results of Promsal 3D showing the structural conservation of known

841 YRs to *Kirc.* Completely conserved residues are in bold. For a full legend see PROMSAL3D

842 documentation. Only the catalytic R-H-R-Y motif is completely conserved amongst all 5

843 sequences.

844

845 **Supplementary Figure 5.** Omitted from current version.

846

847 **Supplementary Table 1.** Assembly statistics for all strains with genomic data used in this study.

Species	Strain	Assembly	Technology	Assembler	Scaffolds[a]	N50	GC content	Size (bp)	Mean Depth long (x)
P. anserina	Wa21-	PaWa21m	PacBio	HGAP 3.0	9	5072363	51.49	35767411	80.3071
P. anserina	Wa28-	PaWa28m	PacBio	HGAP 3.0	7	5076631	51.41	35929333	76.2045
P. anserina	Wa46+	PaWa46p	PacBio	HGAP 3.0	9	4338573	51.37	35724019	117.8295
P. anserina	Wa53-	PaWa53m	PacBio	HGAP 3.0	7	4924837	51.52	35808992	83.821
P. anserina	Wa58-	PaWa58m	PacBio	HGAP 3.0	7	5037662	51.46	35985748	112.1709
P. anserina	Wa63+	PaWa63p	PacBio	HGAP 3.0	7	4907271	51.36	36002190	111.2245
P. anserina	Wa87+	PaWa87p	PacBio	HGAP 3.0	9	4338609	51.44	35896992	105.8792
P. anserina	Wa100+	PaWa100p	PacBio	HGAP 3.0	7	5096663	51.5	35953417	114.091
P. anserina	Wa137-	PaWa137m	MinION	Miniasm 0.2	9	4341106	51.08	36531367	49.924
P. anserina	Y+	PaYp	MinION	Miniasm 0.2	8	4821330	51.55	35719817	35.6112
P. anserina	T_G+	PaTgp	MinION	Miniasm 0.2	13	4230402	51.54	35778794	37.6743
P. pauciseta	CBS237.71-	CBS237.71m	MinION	Miniasm 0.2	13	4070182	51.6	35576233	24.3067
P. comata	Wa131-	PcWa131m	Illumina HiSeq	SPAdes 3.12.0	1119	258199	52.1	34680523	-
P. comata	Wa139-	PcWa139m	MinION	Miniasm 0.2	7	4603841	51.98	34819024	52.5779
P. anserina	**S+**	**Podan2**	**WGS**	**Arachne**	**7**	**4734292**	**52.17**	**35010595**	**-**
P. comata	**T_D+**	**PODCO**	**454 + Illumina**	**Newbler 2.3**	**7**	**4662433**	**52.42**	**34385490**	**-**

[a]The number of scaffolds correspond to those that map to the 7 chromosomes (excluding mitochondrial and rDNA contigs), but other columns include all scaffolds

[b]Numbers correspond to long-read data when available, otherwise to short-read data. For Wa139- and Wa137- the reads were filteread for QV>9 and to be larger than 1kb, resulting in 184301 and 202279 reads, respectively.

[c]Scaffolds larger than 50kb, excluding the mitochondrial and rDNA scaffolds

Species	Strain	Assembly	Mean Depth short (x)	No. Reads[b]	Mean Read Length (bp)	Repeat Content (%)[c]	Source
P. anserina	Wa21-	PaWa21m	78.0394	313285	11863	5.893	Vogan et al. 2019
P. anserina	Wa28-	PaWa28m	81.0651	327199	10105	5.961	Vogan et al. 2019
P. anserina	Wa46+	PaWa46p	107.4690	363721	12949	6.042	Vogan et al. 2019
P. anserina	Wa53-	PaWa53m	78.7383	327523	11382	5.924	Vogan et al. 2019
P. anserina	Wa58-	PaWa58m	88.3996	354055	13130	6.109	Vogan et al. 2019
P. anserina	Wa63+	PaWa63p	77.4007	351077	12562	6.223	Vogan et al. 2019
P. anserina	Wa87+	PaWa87p	74.3914	356924	12928	6.171	Vogan et al. 2019
P. anserina	Wa100+	PaWa100p	80.9098	361645	12857	5.914	Vogan et al. 2019
P. anserina	Wa137-	PaWa137m	143.6921	330780	7913	6.047	This study
P. anserina	Y+	PaYp	83.7234	697535	1952	5.873	Vogan et al. 2019
P. anserina	T_G+	PaTgp	155.9337	1286164	1384	6.009	Vogan et al. 2019
P. pauciseta	CBS237.71-	CBS237.71m	105.7037	240014	6188	6.056	Vogan et al. 2019
P. comata	Wa131-	PcWa131m	162.0216	39127244	150	3.555	This study
P. comata	Wa139-	PcWa139m	150.9355	236001	5786	3.769	This study
P. anserina	**S+**	**Podan2**	10	-	-	4.629	Espagne et al. 2008
P. comata	**T_D+**	**PODCO**	22	-	-	3.054	Silar et al. 2018

848

849 **Supplementary Table 2.** Hypothetical genes from within the *Spok* block of Wa137. Pfam domains inferred from NCBI conserved

850 domain database. Putative functions inferred from BLASTp to other genes with functional annotations. Blank cells indicate either no

851 pfam hits, or genes with BLASTp results to only hypothetical genes.

852

Gene code|name in Wa137	pfam				Putative function
gene01805|Kirc	DUF3435 superfamily	CD_CSD superfamily	CHROMO		
gene01806|Uhera1	DEAD-like_helicase_C superfamily	DEAD-like_helicase_N superfamily			
PaWa137m.02400|PpBr36_10652_Pyricularia_pennisetigena	Peptidase_C48 superfamily				
PaWa137m.02401|Pa_2_5490_homolog	NAP				
PaWa137m.02402|S7711_11618_Stachybotrys_chartarum_homolog					NAD-dependent protein deacetylase
PaWa137m_chromosome_1.2:hsp:120951:4.5.0.58|Pa_7_10380_fragment					
PaWa137m_CH35J_006576|CH35J_006576_Colletotrichum_higginsianum_fragment	Patatin_and_cPLA2 superfamily				Phospholipase A2
PaWa137m.02404|PODCO_502086_Fragment	DDE_Tnp_ISL3 superfamily				PKS-NRPS hybrid synthetase
PaWa137m.02405| Pa_4_1058_homolog					
gene00315|Spok4					Spore killer
PaWa137m.Pa_5_2010_fragment1|Pa_5_2010_fragment1					GPI anchored serine-threonine rich protein
gene01852|Uhera2					
PaWa137m_Pa_1_14730_fragment1|Pa_1_14730_fragment					EST/SMG-like protein 1

gene01815|Pa_6_9560_Fragment	p450 superfamily				pisatin demethylase
gene01816|Pa_6_9560_homolog	p450 superfamily				pisatin demethylase
PaWa137m_7_6640_fragment| Pa_7_6640_Fragment	Peptidase_M16_M superfamily				insulysin
PaWa137m_CTA1_5148| PaWa137m_CTA1_5148_ Colletotrichum_tanaceti_fragment	NK superfamily				Uridylate kinase
PaWa137m.gene.002245| Pa_6_9580_homolog					Satratoxin biosynthesis SC1 cluster protein 4
gene01818|Pa_4_3510_homolog	Peptidase_C48 superfamily	GumC superfamily			
PaWa137m.gene.002249| Pa_1_13070_fragment	zf-H2C2_2	zf-C2H2	COG5048		cutinase G-box binding protein
PaWa137m.gene.PODCO_102140| PODCO_102140_fragment	ABC_ATPase superfamily				Heavy metal tolerance protein
gene01820|PV05_03319_ Exophiala_xenobiotica_fragment	MdlB				Heavy metal tolerance protein
PaWa137m.02410| Pa_6_7965_homolog					
gene01823|Pa_1_13070_homolog	zf-C2H2	zf-H2C2_2	zf-C2H2	lambda-1 superfamily	cutinase G-box binding protein
PaWa137m_Pa_6_2490_fragment| Pa_6_2490_fragment					
gene01824| PODCO_511675_homolog					
PaWa137m_Pa_6_2440_fragment| Pa_6_2440_fragment	PRK09605 superfamily				
PaWa137m_MBR_09427_ Metarhizium_brunneum_fragment| PaWa137m_MBR_09427_ Metarhizium_brunneum_fragment					mitofusin
gene01824.2| PODCO_511675_homolog_2					

PaWa137m_Pa_1_14730_fragment				EST/SMG-like protein 1
PaWa137m.02414|S40285_10826_ Stachybotrys_chlorohalonata_ homolog				
PaWa137m.02415| Micbo1qcDRAFT_170078_ Microdochium_bolleyi_homolog				
PaWa137m.02416| Pa_5_2010_homolog				GPI anchored serine-threonine rich protein
PaWa137m.gene.002259| Pa_6_8910_homolog	AdoMet_MTases superfamily			Trans-aconitate 2-methyltransferase
gene01828|CHGG_04275_ Chaetomium_globosum				
gene01830|DL98DRAFT_554023_ Cadophora_sp_homolog				
gene01831|MBR_09427_ Metarhizium_brunneum_homolog				mitofusin
WaPa137m_4_1930| Pa_4_1930_fragment	Isoprenoid_Biosyn _C1 superfamily			Iridoid synthase
PaWa137m.gene.002264| CHGG_09441_ Chaetomium_globosum_homolog				
PaWa137m_NA56DRAFT_749014_ Pezoloma_ericae_fragment| PaWa137m_NA56DRAFT_749014_ Pezoloma_ericae_fragment				
PaWa137m.gene.002265|PV05_03319_ Exophiala_xenobiotica_fragment	ABC_ATPase superfamily			ABC transporter
gene01834|SPI_07993_homolog_ Metallophosphoesterase	MPP_Dcr2			Metallo-dependent phosphatase
PaWa137m_K432DRAFT_223816_ Lepidopterella_palustris_homolog| PaWa137m_K432DRAFT_223816_ Lepidopterella_palustris_homolog	TPR_12			

gene01835|Pa_1_1830_homolog	GT2_HAS				Hyaluronan synthase
gene01836|Pa_1_14730_homolog	EST1_DNA_bind superfamily				EST/SMG-like protein 1
PaWa137m.02425|S7711_10142_ Stachybotrys_chartarum_homolog	CD_CSD superfamily				
PaWa137m_HIM_12183|HIM_12183_ Hirsutella_minnesotensis_homolog	Pkinase superfamily				
PaWa137m_M441DRAFT_90628| M441DRAFT_90628_ Trichoderma_asperellum_homolog					
gene01838|S7711_11475_ Stachybotrys_chartarum_homolog					
gene01839|S40293_11413_ Stachybotrys_chartarum					patatin-like serine hydrolase
gene01840|NA56DRAFT_703776_ Pezoloma_ericae_homolog	DUF3723 superfamily	PTZ00121 superfamily			
gene01841|Micbo1qcDRAFT_129033_ Microdochium_bolleyi_homolog					Satratoxin biosynthesis SC1 cluster protein 4
gene01842|Pa_5_3190_homolog	MFS superfamily				Efflux pump FUS6
gene01843|Pa_7_6640_homolog	Ptr superfamily				insulysin
PaWa137m_SS1G_08066_Sclerotinia_ sclerotiorum_homolog|Scoty					
gene01844|Chekof	PKc_like superfamily				
PaWa137m_Pa_1_1830_fragment	Peptidase_C48 superfamily				
PaWa137m.gene.002278|PpBr36_10652 _ Pyricularia_pennisetigena_homolog					
gene01845|Pa_4_3510_fragment					
gene01847|GE09DRAFT_1263105_ Coniochaeta_sp_homolog	MdlB				Heavy metal tolerance protein

gene01849|CHGG_09572_ Chaetomium_globosum_fragment					Putative AC transposase
gene01850|MBR_09427_ Metarhizium_brunneum_homolog					
PaWa137m_AJ78_06594_ Emergomyces_pasteurianus_ homolog| PaWa137m_AJ78_06594_ Emergomyces_pasteurianus_ homolog					
PaWa137m.02432| Pa_5_3190_homolog	MFS superfamily				Efflux pump FUS6
PaWa137m_CSUB01_08519_ Colletotrichum_sublineola_homolog	CFEM superfamily				Satratoxin biosynthesis SC1 cluster protein 4
PaWa137m.gene.002286| MYCTH_64320_ Thermothelomyces_thermophilus_ homolog					
PaWa137m_DL765_007700_ Monosporascus_sp_fragment| PaWa137m_DL765_007700_ Monosporascus_sp_fragment					

853