bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

Ancestral admixture is the main determinant of global biodiversity in fission yeast

Sergio Tusso [1,2,*], Bart P.S. Nieuwenhuis [1], Fritz J. Sedlazeck [3] , John W. Davey [4], Daniel C. Jeffares [5,6], Jochen B. W. Wolf[1,2,*]

[1] Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany

[2] Science for Life Laboratories and Department of Evolutionary Biology, Norbyvägen 18D, 75236 Uppsala University, 75236 Uppsala, Sweden

[3] Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston TX 77030

[4] Bioscience Technology Facility, Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom

[5] Department of Biology, University of York, Wentworth Way, York YO10 5DD, United Kingdom

[6] York Biomedical Research Institute (YBRI), University of York, Wentworth Way, York YO10 5DD, United Kingdom

[*] Authors to whom correspondence should be addressed

Corresponding authors:

Sergio Tusso: situssog@gmail.com
Jochen B. W. Wolf: j.wolf@biologie.uni-muenchen.de

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mutation and recombination are key evolutionary processes governing phenotypic variation and reproductive isolation. We here demonstrate that biodiversity within all globally known strains of *Schizosaccharomyces pombe* arose through admixture between two divergent ancestral lineages. Initial hybridization occurred ~20 sexual outcrossing generations ago consistent with recent, human-induced migration at the onset of intensified transcontinental trade. Species-wide heritable phenotypic variation was explained near-exclusively by strain-specific arrangements of alternating ancestry components with evidence for transgressive segregation. Reproductive compatibility between strains was likewise predicted by the degree of shared ancestry. To assess the genetic determinants of ancestry block distribution across the genome, we characterized the type, frequency and position of structural genomic variation (SV) using nanopore and single-molecule real time sequencing, discovering over 800 SVs. Despite being associated with double-strand break initiation points, SV exerted overall little influence on the introgression landscape or on reproductive compatibility that exist between strains. In contrast, we find strongly increased statistical linkage between ancestral populations that is consistent with negative epistatic selection shaping genomic ancestry combinations during the course of hybridization. This study provides a detailed, experimentally tractable example that genomes of natural populations are mosaics reflecting different evolutionary histories. Exploiting genome-wide heterogeneity in the history of ancestral recombination and lineage-specific mutations sheds new light on the population history of *S. pombe* and highlights the importance of hybridization as a creative force in generating biodiversity.

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

61 # Introduction

62 Mutation is the ultimate source of biodiversity. In sexually reproducing organisms it is
63 assisted by recombination shuffling mutations of independent genomic backgrounds into
64 millions of novel combinations. This widens the phenotypic space upon which selection can
65 act and thereby accelerates evolutionary change (Muller, 1932; Fisher, 1999; McDonald et al.,
66 2016). This effect is enhanced for heterospecific recombination between genomes of
67 divergent populations (Abbott et al., 2013). Novel combinations of independently
68 accumulated mutations can significantly increase the overall genetic and phenotypic variation,
69 even beyond the phenotypic space of parental lineages (transgressive segregation
70 (Lamichhaney et al., 2017; Nolte and Sheets, 2005)). Yet, if mutations of the parental
71 genomes are not compatible to produce viable and fertile offspring, hybridization is a dead
72 end. Phenotypic variation then remains within the confines of genetic variation of each
73 reproductively isolated, parental lineage.
74
75 It is increasingly recognised that hybridization is commonplace in nature, and constitutes an
76 important driver of diversification (Abbott et al., 2013; Mallet, 2005). Ancestry components
77 of hybrid genomes can range from clear dominance of alleles from the more abundant species
78 (Dowling et al., 1989; Taylor and Hebert, 1993), over a range of admixture proportions
79 (Lamichhaney et al., 2017; Runemark et al., 2018) to the transfer of single adaptive loci (The
80 Heliconius Genome Consortium et al., 2012). The final genomic composition is determined
81 by a complex interplay of demographic processes, heterogeneity in recombination (e.g.
82 induced by genomic rearrangements) (Wellenreuther and Bernatchez, 2018) and selection
83 (Sankararaman et al., 2014; Schumer et al., 2016). Progress in sequencing technology, now
84 allows characterisation of patterns of admixture and the illumination of underlying processes
85 (Payseur and Rieseberg, 2016). Yet, research has largely focused on animals (Turner and Harr,

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

86 2014; Vijay et al., 2016; Meier et al., 2017; Jay et al., 2018) and plants (Twyford et al., 2015)
87 characterized by large genomes and long generation times. Relatively little attention has been
88 paid to natural populations of sexually reproducing micro-organisms (Leducq et al., 2016;
89 Stukenbrock, 2016; Peter et al., 2018; Steenkamp et al., 2018).
90
91 The fission yeast *Schizosaccharomyes pombe* is an archiascomycete haploid unicellular
92 fungus with a facultative sexual mode of reproduction. Despite of its outstanding importance
93 as a model system in cellular biology (Hoffman et al., 2015) and the existence of global
94 sample collections, essentially all research has been limited to a single isogenic strain isolated
95 by Leupold in 1949 (Leupold 972; JB22 in this study). Very little is known about the ecology,
96 origin, and evolutionary history of the species (Jeffares, 2018). Global population structure
97 has been described shallow with no apparent geographic stratification (Jeffares et al., 2015).
98 Genetic diversity ($\pi = 3\times10^{-3}$ substitutions/site) appears to be strongly influenced by genome-
99 wide purifying selection with the possible exception of region-specific balancing selection
100 (Fawcett et al., 2014; Jeffares et al., 2015). Despite the overall low genetic diversity, *S. pombe*
101 shows abundant additive genetic variation in a variety of phenotypic traits including growth,
102 stress responses, cell morphology, and cellular biochemistry (Jeffares et al., 2015). The
103 apparent worldwide lack of genetic structure in this species appears inconsistent with the large
104 phenotypic variation between strains and with evidence for post-zygotic reproductive
105 isolation between inter-strain crosses, ranging from 1% to 90 % of spore viability
106 (Kondrat'eva and Naumov, 2001; Teresa Avelar et al., 2013; Zanders et al., 2014; Jeffares et
107 al., 2015; Naumov et al., 2015; Marsellach, 2017).
108
109 In this study, we integrate whole-genome sequencing data from three different technologies -
110 sequencing-by-synthesis (Illumina technology data accessed from (Jeffares et al., 2015)),

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

111 single-molecule real-time sequencing (Pacific BioSciences technology, this study) and
112 nanopore sequencing (Oxford Nanopore technology, this study) - sourced from a mostly
113 human-associated, global sample collection to elucidate the evolutionary history of the *S.*
114 *pombe* complex. Using population genetic analyses based on single nucleotide polymorphism
115 (SNP) we show that global genetic variation and heritable phenotype variation of *S. pombe*
116 results from recent hybridization of two ancient lineages. 25 *de novo* assemblies from 17
117 divergent strains further allowed us to quantify segregating structural variation including
118 insertions, deletions, inversion and translocations. In light of these findings, we retrace the
119 global population history of the species, and discuss the relative importance of genome-wide
120 ancestry and structural mutations in explaining phenotypic variation and reproductive
121 isolation.
122

123 Results

124

125 **Global genetic variation in *S. pombe* is characterized by ancient admixture**
126 Genetic variation of the global *S. pombe* collection comprises 172,935 SNPs segregating in
127 161 strains. Considering SNPs independently, individuals can be sub-structured into 57 clades
128 that differ by more than 1900 variants, but are near-clonal within clades (Jeffares et al., 2015).
129 To examine population ancestry further, we divided the genome into 1925 overlapping
130 windows containing 200 SNP each and one representative from each clade (57 samples in
131 total). Principle component analysis conducted on each orthologous window showed a highly
132 consistent pattern along the genome (**Figure 1a, Supplementary Figure 1**): i) the major axis
133 of variation (PC1) split all samples into two clear discrete groups explaining 60% ± 13% of
134 genetic variance (**Figure 1b**). ii) All samples fell into either extreme of the normalized
135 distribution of PC1 scores ($PC1 \in [0; 0.3] \cup [0.7; 1]$) (**Supplementary Figures 2 & 3**)

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

136 with the only exception of strains with inferred changes in ploidy level (**Methods,**
137 **Supplementary Figure 4**). iii) PC2 explained 13% ± 6% of variation and consistently
138 attributed higher variation to one of the two groups. This strong signal of genomic windows
139 separating into two discrete groups suggested that the genomic diversity in this collection was
140 derived from two distinct populations. However, iv) group membership of strains changes
141 among windows moving along the genome, reflecting recombination between these two well
142 defined groups. This last point highlights the importance of considering haplotype structure
143 and explains the lack of observed population structure when disregarding non-independence
144 of SNPs (Jeffares et al., 2015).
145
146 The strong signal from the PCA that systematically differentiates between groups along the
147 genome were likewise reflected in population genetic summary statistics including
148 Watterson's theta (θ), pairwise nucleotide diversity (π), and Tajima's D (**Figure 1d and 2**).
149 Significant differences in these statistics (Kendall's τ p-value $\leq 2.2\times10^{-16}$), were also present
150 in mitochondrial genetic variation (**Figure 1a**), allowed polarising the two groups across
151 windows into a 'low-diversity' group (red) and a 'high-diversity group (blue) (**Figure 1a**,
152 **Supplementary Figure 5**). Genetic divergence between groups (D_{xy}) was 15 and 3 times
153 higher than mean genetic diversity (π) within each group, respectively, and thus supports a
154 period of independent evolution. Painting genomic windows by group membership revealed
155 blocks of ancestry distributed in sample specific patterns along the genome (**Figure 1c,**
156 **Supplementary Figure 6**). The sample corresponding to the reference genome isolated
157 originally from Europe (Leupold's 972; JB22) consisted almost exclusively of 'red' ancestry
158 (>96% red), whereas other samples were characterized near-exclusively by 'blue' ancestry
159 (>96% blue). The sample considered to be a different species from Asia, *S. kambucha*
160 (JB1180 (Singh and Klar, 2002)) had a large proportion of 'blue' windows (>70% blue).

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

161 Hereafter, we refer to the 'red' and 'blue' clade as *Sp* and *Sk*, for *S. pombe* and *S. kambucha*
162 respectively. Grouping samples by the pattern of genomic ancestry across the genome
163 revealed 8 discrete clusters (**Figure 1c**). Consistent with independent and/or recent
164 segregation of ancestral groups, cluster membership for several samples differed between
165 chromosomes (**Figure 1c**) and genome components (**Supplementary Figures 7 & 8**). This is
166 also reflected by low support in the relationship between the 8 discrete clusters.
167
168 The distribution of ancestry components was highly heterogeneous across the chromosome
169 (**Figure 2a, Supplementary Figure 6**). Most strains showed an excess of *Sp* ancestry in parts
170 of chromosome I, whereas several regions of chromosome III had an excess of *Sk* ancestry.
171 Failing to incorporate this genome-wide variation of admixture proportions can mimic
172 signatures of selection. For example, equal ancestry contributions for a certain genomic
173 region will yield high positive values of both Tajima's D (**Supplementary Figure 9**) and π
174 and may be mistaken as evidence for balancing selection. Strong skew in ancestry proportions
175 reduces both statistics to values of the prevailing ancestry and may appear as evidence for
176 selective sweeps (**Figure 2b**). Taking ancestry into account, however, there was no clear
177 signature of selection in either *Sp* or *Sk* genetic variation that could account for heterogeneity
178 in the genetic composition of hybrids (**Supplementary Figure 9)**. Signatures of selection
179 identified previously (cf. Fawcett et al., 2014) are likely artefacts due to skewed ancestry
180 proportions rather than events of positive or balancing selection in the ancestral populations.
181
182 Overall, our results provide strong evidence for the presence of at least two divergent
183 ancestral populations: one genetically diverse group (*Sk* clade) and a less diverse group (*Sp*
184 clade). We found a large range of ancestral admixture proportions between these two clades
185 broadly clustering samples into 8 weakly supported groups. These resemble clusters of strains

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

186 previously identified by *Structure* and *fineStructure* without explicit modelling of ancestral
187 admixture (Jeffares et al., 2015). Neglecting ancestry, Jeffares et al. (2015) argued that the
188 shallow population structure likely results from extensive gene flow between clusters. Yet,
189 considering the genome-wide distribution of *Sk* and *Sp* ancestry, and lack of geographical
190 structure, suggest that the 8 clusters are derived from one or a few centres of ancient
191 admixture (hybridization) without the need of subsequent or recent gene flow between them.
192

193 **Age of ancestral lineages and timing of hybridization**

194 To shed further light on the population history, we estimated the age of the parental lineages
195 and the timing of initial hybridization. Calibrating mitochondrial divergence by known
196 collection dates over the last 100 years, Jeffares et al. (2015) estimated that the time to the
197 most recent common ancestor for all samples was around 2,300 years ago. Current
198 overrepresentation of near-pure *Sp* and *Sk* in Europe or Africa / Asia, respectively, is
199 consistent with an independent history of the parental lineages on different continents for the
200 most part of the last millennia (**Supplementary Figure 8**). Yet, the variety of admixed
201 genomes bears testimony to the fact that isolation has been disrupted by heterospecific gene
202 flow. Using a theoretical model assuming secondary contact with subsequent hybridization
203 (Janzen et al., 2018) we estimated that hybridization occurred within the last 20 sexual
204 outcrossing generations (**Figure 3, Supplementary Figures 10 & 11**). Considering
205 intermittent generations of asexual reproduction, high rates of haploid selfing and dormancy
206 of spores (Farlow et al., 2015; Jeffares, 2018) it is difficult to obtain a reliable estimate of
207 time in years. This recent estimate of hybridization is consistent with hybridization induced
208 by the onset of regular trans-continental human trade between Europe with Africa and Asia
209 (~14th century) and with the Americas (~16th century), with fission yeast as a human
210 commensal (Jeffares, 2018). This fits with the observation that all current samples from the

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

211 Americas were hybrids, while samples with the purest ancestry stem from Europe, Africa and
212 Asia. Moreover, negative genome-wide Tajima's D estimates for both ancestral clades (mean
213 ± SD for Sp: -0.8 ± 0.9 and Sk: -0.7 ± 0.6) signal a period of recent expansion.
214

215 **Heritable phenotypic variation and reproductive isolation are governed by ancestry**
216 **components**

217 Hybridization can lead to rapid evolution due to selection acting on the genetic and
218 phenotypic variation emerging after admixture (Muller, 1932; Fisher, 1999; McDonald et al.,
219 2016). We assessed the consequences of hybridization on phenotypic variation making use of
220 a large data set including 228 quantitative traits collected from the strains under consideration
221 here (Jeffares et al., 2015). Contrary to genetic clustering of hybrid genomes (**Figure 1c**),
222 samples with similar ancestry proportions did not group in phenotypic space described by the
223 first two PC-dimensions capturing 31% of the total variance across traits (**Figure 4a**).
224 Moreover, phenotypic variation of hybrids exceeded variation of pure strains (>0.9 ancestry
225 for *Sp* or *Sk*). This was supported by trait specific analyses. We divided samples into three
226 discrete groups: pure *Sp*, pure *Sk* and hybrids with a large range of *Sp* admixture proportions
227 (0.1-0.9). 63 traits showed significant difference among groups (**Figure 4b, Supplementary**
228 **Figure 12**). In the vast majority of cases (50 traits), hybrid phenotypes were indistinguishable
229 from one of the parents, but differed from the other, suggesting dominance of one ancestral
230 background, consistent with some ecological separation of the backgrounds. In seven traits,
231 hybrid phenotypes were intermediate differing from both parents, indicative of an additive
232 contribution of both ancestral backgrounds. For six traits, hybrids exceeded phenotypic values
233 of both parents providing evidence for transgressive segregation. In all cases, the number of
234 significantly differentiated traits was found to be higher than under the null model (mean
235 number of significant traits after 10000 permutations: dominant Sk 4 +/- 2, dominant Sk 4 +/-

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

236 2, transgressive 0 +/- 0.3, intermediate 0 +/- 0.1; **Supplementary Figure 13**). Jeffares et al.
237 (2017) showed that for each trait the total proportion of phenotypic variance explained by the
238 additive genetic variance component (used as an estimated of the narrow-sense heritability)
239 ranged from 0 to around 90%. We found that across all 228 traits, considering *Sp* and *Sk*
240 ancestry components across the 1,925 genomic windows explained an equivalent amount of
241 phenotypic variance as all 172,935 SNPs segregating across all samples, being both highly
242 correlated (**Figure 4c, 4d;** $r = 0.82$, p-value $\leq 2.2 \cdot 10^{-16}$). Combinations of ancestral genetic
243 variation appear to be the main determinants of heritable phenotypic variation with only little
244 contribution from single-nucleotide mutations arising after admixture. In turn, this supports
245 that the formation of hybrids is recent (see above), and few (adaptive) mutations have
246 occurred after it.
247
248 Ancestry also explained most of the variation in postzygotic reproductive isolation between
249 strains. Previous work revealed a negative correlation between spore viability and genome-
250 wide SNP divergence between strains (Jeffares et al., 2015). The degree of similarity in
251 genome-wide ancestry had the same effect: the more dissimilar two strains were in their
252 ancestry, the lower the viability of the resulting spores (**Figure 4e**; Kendall correlation
253 coefficients, τ= -0.30, T= 259, p-value = $6.66 \cdot 10^{-3}$). This finding is consistent with
254 reproductive isolation being governed by many, genome-wide incompatibilities between the
255 *Sp* and *Sk* clade. Yet, in a number of cases spore survival was strongly reduced in strain
256 combinations with near-identical ancestry. In these cases, reproductive isolation may be
257 caused by few large effect mutations, including structural genomic changes that arose after
258 hybridization.
259
260 **Structural mutations do not determine the genome-wide distribution of ancestry blocks**

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

261 Structural genomic changes (structural variants or SVs hereafter) are candidates for large-
262 effect mutations governing phenotypic variation (Küpper et al., 2016; Jeffares et al., 2017),
263 reproductive isolation (Hoffmann and Rieseberg, 2008; Teresa Avelar et al., 2013) and
264 heterospecific recombination (Ortiz-Barrientos et al., 2016). They may thus importantly
265 contribute to shaping heterogeneity in the distribution of ancestry blocks observed along the
266 genome (Jay et al., 2018; Poelstra et al., 2014) (**Figure 2b**). However, inference of SVs in
267 natural strains of fission yeast has been primary based on short-read sequencing (Jeffares et al.,
268 2017). SV calls from short-read sequencing data are known to differ strongly by bioinformatic
269 pipeline, are prone to false positive inference and are limited in their ability to infer long-
270 range SV, in particular in repetitive regions of the genome (Jeffares et al., 2017).
271
272 To obtain a reliable and comprehensive account of SV segregating across strains and test for a
273 possible association of SVs with the skewed ancestry in the genome, we generated
274 chromosome-level *de novo* genome assemblies for 17 of the most divergent samples using
275 single-molecule real time sequencing (mean sequence coverage 105x; **Supplementary Table**
276 **7**). For the purpose of methodological comparison, we also generated *de novo* assemblies for
277 a subset of 8 strains (including the reference Leupold's 972) based on nanopore sequencing
278 (mean sequence coverage: 140x). SVs were called using a mixed approach combining
279 alignment of *de novo* genomes and mapping of individual reads to the reference genome
280 (Wood et al., 2002). Both approaches and technologies yielded highly comparable results
281 (**Methods, Supplementary Figure 14-17 and Supplementary Table 8).**
282
283 After quality filtering, we retained a total of 832 variant calls including 563 insertions or
284 deletions (indels), 118 inversions, 110 translocations and 41 duplications. The 17 strains we
285 examined with long reads could be classified into six main karyotype arrangements (**Figure**

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

286 **5a**). The previously reported list of SVs of the same strains using short reads consisted of only
287 52 SVs (Jeffares et al., 2017) of which only 8 were found to overlap with the 832 calls from
288 long-read data. The vast majority of SVs were smaller than 10 kb (**Figure 5b**). The size
289 distribution was dominated by elements of 6 kb and 0.5 kb in length corresponding to known
290 transposable elements (TEs) and their flanking long terminal regions (LTRs), respectively
291 (Kelly and Levin, 2005). Only a small number of SVs corresponded to large-scale
292 rearrangements (50 kb - 2.2 Mb) including translocations between chromosomes (**Figure 5a**).
293 A subset of these have been characterized previously as large-effect modifiers of
294 recombination promoting reproductive isolation (Brown et al., 2011; Teresa Avelar et al.,
295 2013; Jeffares et al., 2017).
296
297 Contrary to previous SV classification based on short reads (Jeffares et al., 2017), SV density
298 was not consistently increased in repetitive sequences such as centromeric and telomeric
299 regions illustrating the difficulty of short-read data in resolving SV in repetitive regions
300 (**Figure 5c**). Instead, we found that the frequency of SV was significantly elevated in close
301 proximity to developmentally programmed DNA double-strand breaks (DSB) associated with
302 recombination initiation (Fowler et al., 2014). The proportion of SV observed within [0, 0.5)
303 kb and [0.5, 1) kb of DSB was increased by 46% (p-value$<1x10^{-04}$) and 67% (p-
304 value$<1x10^{-04}$) relative to random expectations. On the contrary, regions more distant than 10
305 kb from DSB were relatively depleted of SV (**Supplementary Figure 18**).
306
307 Next, we imputed the ancestry of SV alleles from SNPs surrounding SV break points. We
308 calculated allele frequencies for SV in both ancestral clades and constructed a folded two-
309 dimensional site frequency spectrum (**Figure 5d**). The majority of variants (66 %) segregated
310 at frequencies below 0.3 in both ancestral genetic backgrounds. Very few SVs were

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

311 differentiated between ancestral populations (3 % of variants with frequency higher than 0.9
312 in one population and below 0.1 in the other). This pattern contrasted with the reference
313 spectrum derived from SNPs where the proportion of low frequency variants was similar at
314 60 %, but genetic differentiation between populations was substantially higher (21 % of SNP
315 variants with frequency higher than 0.9 in one population and below 0.1 in the other). The
316 difference was most pronounced for large SVs (larger than 10 kb) and TEs, for which we
317 estimated allele frequencies for all 57 strains by means of PCR and short-read data,
318 respectively. For TE's, 98 % of the total 1048 LTR variants segregated at frequencies below
319 0.3 in both ancestral populations without a single variant differentiating ancestral populations
320 (**Figure 5d**). Large SVs likewise segregated at low frequencies, being present at most in two
321 strains out of 57. This included the translocation reported for *S. kambucha* between
322 chromosome II and III (Zanders et al., 2014), which we found to be specific for that strain.
323 Only the large inversion on chromosome I segregated at higher frequency being present in
324 five strains out of 57, of which three were of pure *Sp* ancestry including the reference strain
325 (**Supplementary Table 10**). Additionally, SV segregating at high frequency (> 0.7) tended to
326 cluster in genomic regions with steep transitions in ancestry between *Sp* and *Sk* ancestry (p-
327 value > 0.1; **Supplementary Figure 19**).
328
329 In summary, long-read sequencing provided a detailed account of species-wide diversity in
330 structural genetic variation including over 800 high-quality variants ranging from small indels
331 to large-scale inter-chromosomal rearrangements. SV calls showed substantial overlap among
332 technologies (Pacific Biosciences, Nanopore) and approaches (de novo assembly vs.
333 mapping), but less than 1 % of this variation was inferred from short-read data. This finding
334 admonishes to caution when interpreting SV calls from short read data that is moreover
335 sensitive to genotyping methods. In contrast to genome-wide SNPs, SVs segregated near-

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

336 exclusively at low frequencies and were rarely differentiated by ancestral origin. This is
337 consistent with strong diversity-reducing purifying selection relative to SNPs. The fact that
338 SVs, including large-scale rearrangements with known effects on recombination and
339 reproductive isolation (Brown et al., 2011; Teresa Avelar et al., 2013; Zanders et al., 2014),
340 were often unique to single strains precludes a role of SVs in shaping patterns of ancestral
341 heterospecific recombination. Moreover, while being concentrated in proximity to double-
342 strand breaks, possibly due to improper repair upon recombination (Currall et al., 2013), SV
343 were not significantly associated with steep transitions in ancestry blocks. Summarizing the
344 evidence, SV appears to have had little influence in shape genome-wide patterns of ancestral
345 admixture and cannot explain the prevalence of reproductive isolation as a function on
346 ancestral similarity (**Figure 4e**).
347

348 **Negative epistasis shapes the distribution of ancestral blocks**

349 Alternatively, heterogeneity in the distribution and frequency of ancestry along the genome
350 may result from negative epistatic interactions of incompatible genetic backgrounds (Schumer
351 et al., 2016). An excess of homospecific combinations of physically distant loci can serve as
352 an indication of epistatic selection against genetic incompatibilities which can be segregating
353 at appreciable frequencies even within species (Corbett-Detig et al., 2013). We tested this
354 hypothesis by measuring ancestry disequilibrium (AD) between all possible pairs of genomic
355 windows within a chromosome. Specifically, we quantified linkage disequilibrium (LD)
356 between windows dominated by alleles from the same ancestral group (> 0.7) *Sp-Sp* or *Sk-Sk*
357 (reflecting positive AD) and contrasted it to the degree of linkage disequilibrium arising
358 between heterospecific allele combinations *Sp-Sk* (negative AD) (**Supplementary Figure 20**).
359 LD differed significantly between these two cases (**Figure 6**). While negative AD decreased
360 rapidly with genetic distance ($R^2 < 0.2$ after 66, 19 and 21 kb respectively for each

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

361 chromosome) positive LD was higher in magnitude and extended over larger distances ($R^2 <$
362 0.2 after 1.02, 0.54, and 0.18 Mb respectively for each chromosome in *Sk*-*Sk* comparisons and
363 1.59, 1.12, and 0.32 Mb for *Sp*-*Sp* comparisons). These results are consistent with a role of
364 epistatic selection during the course of hybridization shaping the ancestry composition of
365 admixed genomes.
366

367 ## Discussion

368 This study adds to the increasing evidence that hybridization plays an important role as a
369 rapid, 'creative' evolutionary force in natural populations (Seehausen, 2004; Mallet, 2007;
370 Soltis and Soltis, 2009; Abbott et al., 2013; Schumer et al., 2014; Abbott et al., 2016; Pennisi,
371 2016; Nieto Feliner et al., 2017). Recent heterospecific recombination between two ancestral
372 *S. pombe* populations shuffled genetic variation of genomes that diverged since classical
373 antiquity about 2,300 years ago. The timing of hybridization coincided with the onset of
374 intensified trans-continental human trade, suggesting an anthropogenic contribution. Several
375 samples showed similar distribution of ancestral blocks along the genome suggesting
376 comparable evolutionary histories, and allowing the identification of 8 discrete clusters. These
377 clusters, in general showed weak geographical grouping, initially interpreted as evidence for
378 reduced population structure with large recent world-wide gene flow (Jeffares et al., 2015). In
379 contrast, the world-wide distribution of the two ancestral linages suggests rapid and recent
380 global dispersion after hybridization followed by local differentiation. This study thus
381 highlights the importance of taking genomic non-independence into account. Allowing for the
382 fact that genomes are mosaics reflecting different evolutionary histories can fundamentally
383 alter inference on a species' evolutionary history.
384

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

385 Moreover, conceptualizing genetic variation as a function of ancestry blocks alternating along
386 the genome changes the view on adaptation. Admixture is significantly faster than
387 evolutionary change solely driven by mutation. Accordingly, phenotypic variation was near-
388 exclusively explained by ancestry components with only little contribution from novel
389 mutations. Importantly, admixture not only filled the phenotypic space between parental
390 lineages, but also promoted transgressive segregation in several hybrids. This range of
391 phenotypic outcomes opens the opportunity for hybrids to enter novel ecological niches
392 (Nolte and Sheets, 2005; Pfennig et al., 2016) and track rapid environmental changes
393 (Eroukhmanoff et al., 2013).
394
395 Structural mutations have been described as prime candidates for rapid large-effect changes
396 with implications on phenotypic variation, recombination and reproductive isolation (Faria
397 and Navarro, 2010; Ortiz-Barrientos et al., 2016; Wellenreuther and Bernatchez, 2018). This
398 study contributes to this debate providing a detailed account of over 800 high-quality
399 structural variants identified across 17 chromosome level *de novo* genomes sampled from the
400 most divergent strains within the species. On the whole, SVs had little effect. While large-
401 scale rearrangements in specific strains have been shown to affect fitness (Teresa Avelar et al.,
402 2013; Nieuwenhuis et al., 2018) and promote reproductive isolation between specific strains
403 in *S. pombe* (Brown et al., 2011; Teresa Avelar et al., 2013), reproductive isolation was
404 overall best predicted by the degree of shared ancestry with little contribution from SVs. SVs
405 segregated at low frequencies in both ancestral populations and, contrary to what has been
406 suggested for specific genomic regions in other systems (Jay et al., 2018), they did not
407 account for genome-wide heterogeneity in introgression among strains during hybridization.
408 Much rather, analyses of ancestry disequilibrium suggest a role for negative epistasis between
409 multiple ancestry-specific loci spread across the genome rather than single major effect

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

410 mutations such as selfish elements or meiotic drivers (Zanders et al., 2014; Hu et al., 2017;
411 Nuckolls et al., 2017). Functional work is needed to identify the genetic elements conveying
412 reproductive isolation.
413

414 Material and Methods

415 Strains

416 This study is based on a global collection of *S. pombe* consisting of 161 world-wide
417 distributed strains (see **Supplementary Table 1**) described in Jeffares *et al.* (2015).
418

419 Inferring ancestry components

420 To characterize genetic variation across all strains, we made use of publically available data in
421 variant call format (VCF) derived for all strains from Illumina sequencing with an average
422 coverage of around 80x (Jeffares et al., 2015). The VCF file consists of 172,935 SNPs
423 obtained after read mapping to the *S. pombe* 972 h^- reference genome (ASM294v264) (Wood
424 et al., 2003) and quality filtering (see **Supplementary Table 1** for additional information).
425 We used a custom script in R 3.4.3 (Team, 2014) with the packages *gdsfmt* 1.14.1 and
426 *SNPRelate* 1.12.2 (Zheng et al., 2017, 2012), to divide the VCF file into genomic windows of
427 200 SNPs with overlap of 100 SNPs. This resulted in 1925 genomic windows of 1 - 89 kb in
428 length (mean 13 kb). For each window, we performed principal component analyses (PCA)
429 using *SNPRelate* 1.12.2 (Zheng et al., 2017, 2012) (example in **Figure 1a** and
430 **Supplementary Figure 1**). The proportion of variance explained by the major axis of
431 variation (PC1) was consistently high and allowed separating strains into two genetic
432 groups/clusters, *Sp* and *Sk* (see main text, **Figure 1b**). We calculated population genetic
433 parameters within clusters including pairwise nucleotide diversity (π) (Nei and Li, 1979),
434 Watterson theta (θ_w) (Watterson, 1975), and Tajima's *D* (Tajima, 1989), as well as the

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

435 average number of pairwise differences between clusters (D_{xy}) (Nei and Li, 1979) using
436 custom scripts. Statistical significance of the difference in nucleotide diversity (*π)* between
437 ancestral clades was inferred using Kendall's τ as test statistic. Since values of adjacent
438 windows are statistically non-independent due to linkage, we randomly subsampled 200
439 windows along the genome with replacement. This was repeated a total of 10 times for each
440 test statistic, and we report the maximum p-value. Given the consistent difference between
441 clusters (**Figure 1** and **Supplementary Figure 2, 3 and 5**), normalised PC score could be
442 used to attribute either *Sp* (low-diversity) or *Sk* (high-diversity) ancestry to each window
443 (summary statistics for each window are given in **Supplementary Table 2)**. This was
444 performed both for the subset of 57 samples (**Figure 1c**) and for all 161 samples
445 (**Supplementary Figure 6**). Using different window sizes (100, 50 and 40 SNPs with overlap
446 of 50, 25 and 20 respectively) yielded qualitatively the same results. Intermediate values in
447 PC1 (between 0.25 and 0.75) were only observed in few, sequential windows where samples
448 transitioned between clusters (**Supplementary Figure 3**). The only exception was sample
449 JB1207, which we found to be diploid (for details see below).
450

451 **Population structure after hybridization**

452 To characterise the genome-wide distribution of ancestry components along the genome, we
453 ran a hierarchical cluster analysis on the matrix containing ancestry information (*Sp* or *Sk*) for
454 each window (columns) and strain (rows) using the R package *Pvclust 2.0.0* (Suzuki and
455 Shimodaira, 2006). *Pvclust* includes a multiscale bootstrap resampling approach to calculate
456 approximately unbiased probability values (p-values) for each cluster. We specified 1000
457 bootstraps using the Ward method and a Euclidian-based dissimilarity matrix. The analysis
458 was run both for the whole genome (**Figure 1c**) and by chromosome (**Figure 1c**,
459 **Supplementary Figure** 7).

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

460

461 **Phylogenetic analysis of the mitochondrial genome**

462 From the VCF file, we extracted mitochondrial variants for all 161 samples (Jeffares et al.,
463 2015) and generated an alignment in *.fasta* format by substituting SNPs into the reference *S.*
464 *pombe* 972 h− reference genome (ASM294v264) using the package *vcf2fasta*
465 (https://github.com/JoseBlanca/vcf2fasta/, version Nov. 2015). We excluded variants in
466 mitochondrial regions with SVs inferred from long reads. A maximum likelihood tree was
467 calculated using *RaxML* (version 8.2.10-gcc-mpi) (Stamatakis, 2014) with default parameters,
468 GTRGAMMAI approximation, final optimization with GTR + GAMMA + I and 1000
469 bootstraps. The final tree was visualised using *FigTree* 1.4.3
470 (http://tree.bio.ed.ac.uk/software/figtree/) (**Supplementary Figure 8**).

471

472 **Time of hybridization**

473 Previous work (Jeffares et al., 2015) has shown that the time to the most recent common
474 ancestor for 161 samples dates back to around 2300 years ago. This defines the maximum
475 boundary for the time of hybridization. We used the theoretical model by Janzen et al., (2018)
476 to infer the age of the initial hybridization event. The model predicts the number of ancestry
477 blocks and junctions present in a hybrid individual as a function of time and effective
478 population size (N_e). First, we obtained an estimate of N_e using the multiple sequential
479 Markovian coalescent (MSMC). We constructed artificial diploid genomes from strains with
480 consistent clustering by ancestry (**Figure 1c)** and estimated change in N_e as function across
481 time using *MSMC* 2-2.0.0 (Schiffels and Durbin, 2014). In total we took four samples per
482 group and produced diploid genomes in all possible six pairs for each group, except for one
483 cluster that had only two samples (JB1205 and JB1206). Bootstraps were produced for each
484 analysis, subsampling 25 genomic fragments per chromosome of 200 kb each. Resulting

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

485 effective population size and time was scaled using reported mutation rate of $2 \cdot 10^{-10}$
486 mutations site^{-1} generation^{-1} (Farlow et al., 2015). Although it is difficult to be certain of the
487 number of independent hybridization events, it is interesting to see that some clusters show
488 similar demographic histories (**Supplementary Figure 21**). Regardless of the demographic
489 history in each cluster, long-term N_e as estimated by the harmonic mean ranged between $1 \cdot$
490 10^5 and $1 \cdot 10^9$. N_e of the near-pure ancestral *Sp* and *Sk* cluster was $7 \cdot 10^5$ and $9 \cdot 10^6$,
491 respectively. These estimates of N_e are consistent with previous reports of $1 \cdot 10^7$ (Farlow et
492 al., 2015).
493 We then used a customised R script with the ancestral component matrix to estimate the
494 number of ancestry blocks (*Sp* or *Sk* clade) (**Supplementary Figure 10**). We used the R script
495 from Janzen et al., (2018), and ran the model in each sample and chromosome using: $N_e = 1 \cdot$
496 10^6, r = number of genomic windows per chromosome, $h0 = 0.298$ (mean heterogenicity ($h0$)
497 was estimated from the ancestral haplotype matrix) and c = 7.1, 5.6, and 4.1 respectively for
498 chromosome I, II and III (values taken from Munz et al. (1989)) (**Supplementary Figure 11**).
499 Given the large N_e, no changes in mean heterogenicity is expected over time after
500 hybridization due to drift (the proportion of ancestral haplotypes *Sp* and *Sk* in hybrids,
501 estimated as $2pq$, where p and q are the proportion of each ancestral clade in hybrids).
502 Accordingly, results did not change within the range of the large N_e values. For this analysis,
503 samples with proportion of admixture lower than 0.1 were excluded.
504

505 **Phenotypic variation and reproductive isolation**

506 We sourced phenotypic data of 229 phenotypic measurements in the 161 strains including
507 amino acid quantification on liquid chromatography (aaconc), growth and stress on solid
508 media (smgrowth), cell growth parameters and kinetics in liquid media (lmgrowth) and cell
509 morphology (shape1 and shape2) from Jeffares et al. (2015). Data on reproductive isolation

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

510 measured as the percentage of viable spores in pairs of crosses were compiled from Jeffares et
511 al. (2015) and Marsellach (2017). A summary of all phenotypic measurements and
512 reproductive data is provided in **Supplementary Table 4** and **5**, respectively.
513
514 First, we normalized each phenotypic trait y using rank-based transformation with the
515 relationship normal.y = qnorm(rank (y) / (1 + length(y))). We then conducted PCA on
516 normalized values of all phentoypic traits using the R package *missMDA* 1.12 (Josse and
517 Husson, 2016). We estimated the number of dimensions for the principal component analysis
518 by cross-validation, testing up to 30 PC components and imputing missing values. In addition
519 to PCA decomposing variance across all traits, we examined the effect of admixture on each
520 trait separately. Samples were divided into three discrete categories of admixture: two groups
521 including samples with low admixture proportions (proportion of *Sp* or *Sk* clades higher than
522 0.9), and one for hybrid samples (proportion of *Sp* or *Sk* clades between 0.1 to 0.9).
523 Significant differences in phenotypic distributions between groups were tested using *Tukey*
524 *Honest Significant Differences* as implemented in *Stats 3.4.2* (Team, 2014). **Supplementary**
525 **Figure 12** shows the distribution of phenotypic values by admixture category for each trait.
526 The number of traits with significant differences among groups was contrasted to values
527 obtained by randomising admixture categories without replacement (permutations of the *Sp*,
528 *Sk,* or *hybrid* category). Observed values were contrasted with distribution of the expected
529 number of significant traits after running 10000 independent permutations (**Supplementary**
530 **Figure 13**).
531
532 **Heritability**
533 Heritability was estimated for all normalized traits using *LDAK* 5.94 (Speed et al., 2012),
534 calculating independent kinship matrices derived from: 1) all SNPs and 2) ancestral

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

535 haplotypes. Both SNPs and haplotype data were binary encoded (0 or 1). Jeffares et al. (2015)
536 showed that heritability estimates between normalised and raw values are highly correlated (r
537 $= 0.69$, p-value $\leq 2.2 \cdot 10^{-16}$). Heritability estimated with SNP values were strongly correlated
538 with those from ancestral haplotypes ($r = 0.82$, p-value $\leq 2.2 \cdot 10^{-16}$). Heritability estimates
539 and standard deviation for each trait for both SNP and ancestral haplotypes are detailed in
540 **Supplementary Table 6**.
541

542 **Identification of ploidy changes**

543 *S. pombe* is generally considered haploid under natural conditions. Yet, for two samples
544 ancestry components did not separate on the principle component axis 1 (see above) for much
545 of the genome. Instead, these samples were intermediate in PC1 score. A possible explanation
546 is diploidisation of the two ancestral genomes. To establish the potential ploidy of samples,
547 we called variants for all 161 samples using the Illumina data from Jeffares *et al* (2015) .
548 Cleaned reads were mapped with *BWA* (version 0.7.17-r1188) in default settings and variants
549 were called using *samtools* and *bcftools* (version 1.8). After filtering reads with a QUAL
550 score > 25, the number of heterozygous sites per base per 20kb window were calculated.
551 Additionally the nuclear content (C) as measured by Jeffares et al. (2015) (Supplementary
552 Table S4 in Jeffares *et al* (2015)) were used to verify increased ploidy. Two samples showed
553 high heterozygosity along the genome (JB1169 and JB1207) of which JB1207 for which data
554 were available also showed a high C-value, suggesting that these samples are diploid
555 (**Supplementary Figures 4 & 22**). In JB1207, heterozygosity varies along the genome, with
556 regions of high and low diversity. Assigning ancestry (see **Supplementary Figure 6**), shows
557 that the haploid parents differed from each other and that both chromosomes stem from
558 hybrids between the *Sp* and *Sk* clades. Sample JB1110 showed genomic content similar to

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

559 JB1207, but did not show heterozygosity levels above that of haploid strains, suggesting the
560 increase in genome content occurred by autoploidization.
561
562 **High-weight genomic DNA extraction and whole genome sequencing**
563 To obtain high weight gDNA for long-read sequencing, we grew strains from single colonies
564 and cultured them in 200 mL liquid EMM at 32 °C shaking at 150 r.p.m. overnight. Standard
565 media and growth conditions were used throughout this work (Hagan et al., 2016) with minor
566 modifications: We used standard liquid Edinburgh Minimal Medium (EMM; Per liter:
567 Potassium Hydrogen Phthalate 3.0 g, Na $HPO_4{\cdot}2H_2O$ 2.76 g, NH_4Cl 5.0 g, D-glucose 20 g,
568 $MgCl_2{\cdot}6H2O$ 1.05 g, $CaCl_2{\cdot}2H_2O$ 14.7 mg, KCl 1 g, Na_2SO_4 40 mg, Vitamin Stock ×1000
569 1.0 ml, Mineral Stock ×10,000 0.1 ml, supplemented with 100 mg l^{-1} adenine and 225 mg l^{-1}
570 leucine) for the asexual growth. DNA extraction was performed with Genomic Tip 500/G or
571 100/G kits (Qiagen) following the manufacturer's instruction, but using Lallzyme MMX for
572 lysis (Flor-Parra et al 2014, doi:10.1002/yea.2994). For each sample, 20 kb libraries were
573 produced that were sequenced on one SMRT cell per library using the Pacific Biosciences
574 RSII Technology Platform (PacBio®, CA). For a subset of eight samples, additional
575 sequencing was performed using Oxford Nanopore (MinION). Sequencing was performed at
576 SciLifeLab, Uppsala, Gene centre LMU, Munich and The Genomics & Bioinformatics
577 Laboratory, University of York. We obtained on average 80x (SMRT) and 140x (nanopore)
578 coverage for the nuclear genome for each sample (summary in **Supplementary Table 7**).
579
580 Additionally, 2.5 µg of the same DNA was delivered to the SNP&SEQ Technology Platform
581 at the Uppsala Biomedical Centre (BMC), for Illumina sequencing. Libraries were prepared
582 using the TruSeq PCRfree DNA library preparation kit (Illumina Inc.). Sequencing was
583 performed on all samples pooled into a single lane, with cluster generation and 150 cycles

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

584 paired-end sequencing on the HiSeqX system with v2.5 sequencing chemistry (Illumina Inc.).
585 These data were used for draft genome polishing (see below).
586

587 ***De novo* assembly of single-molecule read data**

588 *De novo* genomes were assembled with *Canu* 1.5 (Koren et al., 2017) using default
589 parameters. BridgeMapper from the *SMRT* 2.3.0 package was used to polish and subsequently
590 assess the quality of genome assembly. Draft genomes were additionally polishing using short
591 Illumina reads, running four rounds of read mapping to the draft genome with *BWA* 0.7.15
592 and polishing with *Pilon* 1.22 (Walker et al., 2014). Summary statistics of the final assembled
593 genomes are found in **Supplementary Table 7**. *De novo* genomes were aligned to the
594 reference genome using *MUMmer* 3.23 (Kurtz et al., 2004). Contigs were classified by
595 reference chromosome to which they showed the highest degree of complementary. We used
596 customised python scripts to identify and trim mitochondrial genomes.
597

598 **Structural variant detection**

599 Structural variants (SVs) were identified by a combination of a *de novo* and mapping
600 approach. *De novo* genomes were aligned to the reference genome using *MUMmer*, and SVs
601 were called using the function show-diff and the package *SVMU* 0.2beta (Khost et al., 2016).
602 Then, raw long reads were mapped to the reference genome with *NGMLR* and genotypes were
603 called using the package *Sniffles* (Sedlazeck et al., 2018). We implemented a new function
604 within Sniffles “forced genotypes”, which calls SVs by validating the mapping calls from an
605 existing list of breaking points or SVs. This reports the read support per variant even down to
606 a single read. We forced genotypes using the list of *de novo* breaking points to generate a
607 multi-sample VCF file. SVs were merged using the package *SURVIVOR* (Jeffares et al., 2017)
608 option merge with a threshold of 1kbp and requiring the same type. In total, it resulted in a list

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

609 of 1498 SVs with 892 in common between the mapping and *de-novo* approaches
610 (**Supplementary Figure 14**).
611
612 Within the 892 common variants we compared the accuracy of genotyping between sample by
613 comparing genotypes obtained from *de novo* genomes and by mapping reads to reference
614 genome. Additionally, we compared genotypes in samples sequenced with both PacBio and
615 MinIon. In total we sequenced 8 samples with both technologies. We found high consistency
616 for variants called with both sequencing technologies and observed that allele frequencies
617 were highly correlated ($r = 0.98$, p-value $\leq 2.2 \times 10^{-16}$) (**Supplementary Figures 14 - 17**).
618 Only common SVs between the mapping and *de-novo* approach were considered, and variants
619 with consistency below 50% were removed. We manually checked large SVs (larger than
620 10kb) by comparing the list of SVs with the alignment of the *de novo* genomes to the
621 reference genome from MUMmer. This resulted in a final data set with 832 SVs
622 (**Supplementary Table 8**).
623
624 **Distribution of SVs around developmentally programmed DNA double-strand breaks**
625 **(DSB)**
626 We tested the association between DSB and SVs by comparing the physical genomic
627 coordinates of the final list of SV with DSB locations accessed from Fowler et al., (2014).
628 Maintaining the same number of SV per chromosome, we used a customized R script to
629 randomise SV coordinates and measure the distances to the closest DSB. We counted the
630 number of SV present within different intervals of physical genetic distance ([0,500), [500,
631 1000), [1000, 2000), [2000, 4000), [4000, 10000), [10000, 20000), [20000, 30000) bp).
632 Empirically observed values were contrasted with randomized distribution after running
633 10000 independent permutations. P-values of differences between randomization and

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

634 observed values were obtained from the fraction of expected values higher than the observed
635 value from the original data (**Supplementary Figures 18**).
636
637 **PCR validation of large SVs**
638 To test the frequency of large inversions and rearrangements observed from long read data,
639 we performed PCR verification over the breakpoints in the 57 non-clone samples. PCR was
640 performed for both sides of the breakpoints, with a combination of one primer 'outside' of the
641 inversion and both primers 'inside' the inversion (**Supplementary Figure 23**). PCR were
642 performed on DNA using standard *Taq* polymerase, with annealing temperature at 59°C. The
643 primers used, the coordinates in the reference and the expected amplicon length are given in
644 **Supplementary Table 9**.
645
646 **Distribution of structural variants in ancestral population – Two dimensional folded site**
647 **frequency spectrum**
648 We used the location of break points of SVs to identify whether a variant was located in the
649 *Sp* or *Sk* genetic background in each sample. Ancestral haplotypes are difficult to infer in
650 telomeric and centromeric regions given the low confidence in SNP calling in those regions,
651 resulting in low percentage of variance explained by PC1. Thus SVs with break points in
652 those regions were excluded from this analysis (19 SVs). SVs were grouped by ancestral
653 group and allele frequencies were calculated for each ancestral population. We used these
654 frequencies to build a two dimensional folded site frequency spectrum (2dSFS). In order to
655 compare this 2dSFS, we repeated the analysis using SNP data from all 57 samples.
656 Considering that the majority of identified SVs with long reads were transposable elements,
657 we also made use of LRT insertion-deletion polymorphism (indels) inferred from short reads.

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

658 For this additional data we produced a similar folded 2d SFS. LTR indel data were taken from
659 Jeffares et al., (2015) and are listed in **Supplementary Table 11**.
660
661 **Decay in linkage disequilibrium (LD)**
662 To contrast LD between alleles from alternative ancestral groups, we calculated LD between
663 all described genomic windows within chromosomes (**Supplementary Figure 20**). For this
664 analysis only hybrid samples were considered (strains with admixture proportion higher than
665 0.1). For each pair of windows, we polarized windows by ancestry (at a threshold of > 0.7)
666 and calculated standardized LD as the squared Pearson's correlation coefficient (R^2) (Hill and
667 Robertson, 1968; Weir, 1979). This measurement takes into consideration difference in allele
668 frequencies. The expected value of R^2 ($E(R^2)$) can be approximated by (Hill and Weir, 1988):

669 $$E(R^2) = \left(\frac{10 + C}{(2 + C) * (11 + C)}\right) * \left(1 + \frac{(3 + C) * (12 + 12C + C^2)}{n * (2 + C) * (11 + C)}\right)$$

670 Where C corresponds to product between the genetic distance (bp) and the population
671 recombination rate (ρ) in n number of haplotype sampled. The population recombination rate
672 was calculated as: $\rho = 4 * N_e * c$, where c is the recombination fraction between sites and Ne
673 is the effective population size. We fitted a nonlinear model to obtain least squares estimates
674 of ρ using a customized R script. The decay of LD with physical distance can be described
675 with this model (Remington et al., 2001). LD values were grouped in three categories: i)
676 comparison between windows with high proportion (*Sp*>0.7) of *Sp* ancestral group (*Sp*-*Sp*); ii)
677 high proportion (*Sk*>0.7) of *Sk* ancestral group (*Sk*-*Sk*); and iii) high proportion of opposite
678 ancestral groups (*Sp*-*Sk*). i) and ii) represent cases of positive ancestry disequilibrium, iii) will
679 be denoted as negative ancestry disequilibrium.
680
681 **Data availability**

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

682 Nanopore, single-molecule real time sequencing data and de-novo genomes are available at
683 NCBI Sequence Read Archive, BioProject ID XXX.
684

685 Acknowledgments

686
687 We thank Fidel Botero-Castro, Ana Catalán, Sebastian Höhna, Ulrich Knief, Claire Peart,
688 Joshua Peñalba, Ricardo Pereira, Matthias Weissensteiner, (LMU Munich) and S. Lorena
689 Ament-Velásquez (Uppsala University) for providing valuable intellectual input on the
690 various analyses, sharing scripts and critically comment on the manuscript. We are further
691 indebted to Bernadette Weissensteiner for extensive help with laboratory work and Saurabh
692 Pophaly for bioinformatics support (LMU Munich). We further acknowledge support for data
693 generation from the National Genomics Infrastructure, Uppsala, Sweden, the Gene Centre,
694 Munich, Germany, Sally James and Peter Ashton from the Bioscience Technology Facility,
695 Department of Biology, University of York, U.K, and James Chong, Department of Biology,
696 University of York, U.K. The computational infrastructure was provided by the UPPMAX
697 Next-Generation Sequencing Cluster and Storage (UPPNEX) project funded by the Knut and
698 Alice Wallenberg Foundation and the Swedish National Infrastructure for Computing and the
699 York Advanced Research Computing Cluster (YARCC), University of York, U.K. This study
700 was funded by LMU Munich to JW and NHGRI UM1 HG008898 to FS.
701

702 Contributions

703 ST, BN, DJ and JW conceived of the study; All analyses were performed by ST with
704 contributions from FS in structural variation calling, JD in de novo assembly, BN in ancestral
705 inference and population genetics parameters and DJ in phenotypic and heritability analyses;
706 ST and JD assembled de-novo genomes; BN designed primers for PCR validation of
707 structural variants; ST, BN, and JW wrote the manuscript with input from all other authors.
708

709 Competing Interests statement

710 The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

711
712 **References**
713

714 Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J.,
715 Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff,
716 F., Grill, A., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, A.G., Jiggins, C.,
717 Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M.,
718 Mullen, S., Nichols, R., Nolte, A.W., Parisod, C., Pfennig, K., Rice, A.M., Ritchie,
719 M.G., Seifert, B., Smadja, C.M., Stelkens, R., Szymura, J.M., Väinölä, R., Wolf,
720 J.B.W., Zinner, D., 2013. Hybridization and speciation. J. Evol. Biol. 26, 229–246.
721 https://doi.org/10.1111/j.1420-9101.2012.02599.x
722 Abbott, R.J., Barton, N.H., Good, J.M., 2016. Genomics of hybridization and its evolutionary
723 consequences. Mol. Ecol. 25, 2325–2332. https://doi.org/10.1111/mec.13685
724 Brown, W.R.A., Liti, G., Rosa, C., James, S., Roberts, I., Robert, V., Jolly, N., Tang, W.,
725 Baumann, P., Green, C., Schlegel, K., Young, J., Hirchaud, F., Leek, S., Thomas, G.,
726 Blomberg, A., Warringer, J., 2011. A Geographically Diverse Collection of
727 Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but
728 Extensive Karyotypic Diversity. G3 GenesGenomesGenetics 1, 615–626.
729 https://doi.org/10.1534/g3.111.001123
730 Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L., Ayroles, J.F., 2013. Genetic
731 incompatibilities are widespread within species. Nature 504, 135–137.
732 https://doi.org/10.1038/nature12678
733 Currall, B.B., Chiangmai, C., Talkowski, M.E., Morton, C.C., 2013. Mechanisms for
734 Structural Variation in the Human Genome. Curr. Genet. Med. Rep. 1, 81–90.
735 https://doi.org/10.1007/s40142-013-0012-8
736 Dowling, T. E., Smith, G. R. & Brown, W. M. Reproductive isolation and introgression
737 between *Notropis cornutus* and *Notropis chrysocephalus* (family Cyprinidae):
738 comparison of morphology, allozymes, and mitochondrial DNA. *Evolution.* **43,** 620–
739 634 (1989).
740 Eroukhmanoff, F., Hermansen, J.S., Bailey, R.I., Sæther, S.A., Sætre, G.-P., 2013. Local
741 adaptation within a hybrid species. Heredity 111, 286–292.
742 https://doi.org/10.1038/hdy.2013.47
743 Faria, R., Navarro, A., 2010. Chromosomal speciation revisited: rearranging theory with
744 pieces of evidence. Trends Ecol. Evol. 25, 660–669.
745 https://doi.org/10.1016/j.tree.2010.07.008
746 Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T.G., Nordborg, M., Lynch, M., 2015. The
747 Spontaneous Mutation Rate in the Fission Yeast Schizosaccharomyces pombe.
748 Genetics 201, 737–744. https://doi.org/10.1534/genetics.115.177329
749 Fawcett, J.A., Iida, T., Takuno, S., Sugino, R.P., Kado, T., Kugou, K., Mura, S., Kobayashi,
750 T., Ohta, K., Nakayama, J., Innan, H., 2014. Population Genomics of the Fission
751 Yeast Schizosaccharomyces pombe. PLOS ONE 9, e104241.
752 https://doi.org/10.1371/journal.pone.0104241
753 Fisher, R.A., 1999. The genetical theory of natural selection: a complete variorum edition.
754 Oxford University Press.
755 Fowler, K.R., Sasaki, M., Milman, N., Keeney, S., Smith, G.R., 2014. Evolutionarily diverse
756 determinants of meiotic DNA break and recombination landscapes across the genome.
757 Genome Res. 24, 1650–1664. https://doi.org/10.1101/gr.172122.114
758 Hagan, I.M., Carr, A.M., Grallert, A., Nurse, P., 2016. Fission yeast: a laboratory manual.
759 Cold Spring Harbor Laboratory Press.

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

760 Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl.
761 Genet. 38, 226–231. https://doi.org/10.1007/BF01245622
762 Hill, W.G., Weir, B.S., 1988. Variances and covariances of squared linkage disequilibria in
763 finite populations. Theor. Popul. Biol. 33, 54–78. https://doi.org/10.1016/0040-
764 5809(88)90004-4
765 Hoffman, C.S., Wood, V., Fantes, P.A., 2015. An Ancient Yeast for Young Geneticists: A
766 Primer on the Schizosaccharomyces pombe Model System. Genetics 201, 403–423.
767 https://doi.org/10.1534/genetics.115.181503
768 Hoffmann, A.A., Rieseberg, L.H., 2008. Revisiting the Impact of Inversions in Evolution:
769 From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? Annu.
770 Rev. Ecol. Evol. Syst. 39, 21–42.
771 https://doi.org/10.1146/annurev.ecolsys.39.110707.173532
772 Hu, W., Jiang, Z.-D., Suo, F., Zheng, J.-X., He, W.-Z., Du, L.-L., 2017. A large gene family
773 in fission yeast encodes spore killers that subvert Mendel’s law. eLife 6.
774 https://doi.org/10.7554/eLife.26057
775 Janzen, T., Nolte, A. W. & Traulsen, A. The breakdown of genomic ancestry blocks in hybrid
776 lineages given a finite number of recombination sites: breakdown of ancestry blocks
777 after hybridization. *Evolution.* **72,** 735–750 (2018).
778 Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.Á., Nowell, R.W., Mallet, J.,
779 Dasmahapatra, K.K., Joron, M., 2018. Supergene Evolution Triggered by the
780 Introgression of a Chromosomal Inversion. Curr. Biol. 28, 1839-1845.e3.
781 https://doi.org/10.1016/j.cub.2018.04.072
782 Jeffares, D.C., 2018. The natural diversity and ecology of fission yeast. Yeast 35, 253–260.
783 https://doi.org/10.1002/yea.3293
784 Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C.,
785 Bähler, J., Sedlazeck, F.J., 2017. Transient structural variations have strong effects on
786 quantitative traits and reproductive isolation in fission yeast. Nat. Commun. 8, 14061.
787 https://doi.org/10.1038/ncomms14061
788 Jeffares, D.C., Rallis, C., Rieux, A., Speed, D., Převorovský, M., Mourier, T., Marsellach,
789 F.X., Iqbal, Z., Lau, W., Cheng, T.M.K., Pracana, R., Mülleder, M., Lawson, J.L.D.,
790 Chessel, A., Bala, S., Hellenthal, G., O’Fallon, B., Keane, T., Simpson, J.T., Bischof,
791 L., Tomiczek, B., Bitton, D.A., Sideri, T., Codlin, S., Hellberg, J.E.E.U., van Trigt, L.,
792 Jeffery, L., Li, J.-J., Atkinson, S., Thodberg, M., Febrer, M., McLay, K., Drou, N.,
793 Brown, W., Hayles, J., Salas, R.E.C., Ralser, M., Maniatis, N., Balding, D.J., Balloux,
794 F., Durbin, R., Bähler, J., 2015. The genomic and phenotypic diversity of
795 Schizosaccharomyces pombe. Nat. Genet. 47, 235–241.
796 https://doi.org/10.1038/ng.3215
797 Josse, J., Husson, F., 2016. missMDA: A Package for Handling Missing Values in
798 Multivariate Data Analysis. J. Stat. Softw. 70. https://doi.org/10.18637/jss.v070.i01
799 Kelly, F.D., Levin, H.L., 2005. The evolution of transposons in *Schizosaccharomyces pombe*.
800 Cytogenet. Genome Res. 110, 566–574. https://doi.org/10.1159/000084990
801 Khost, D. E., Eickbush, D. G. & Larracuente, A. M. Single molecule long read sequencing
802 resolves the detailed structure of complex satellite DNA loci in. (2016).
803 Kondrat’eva, V.I., Naumov, G.I., 2001. The Phenomenon of Spore Killing in
804 Schizosaccharomyces pombe Hybrids. Dokl. Biol. Sci. 379, 385–388.
805 https://doi.org/10.1023/A:1011624918673
806 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017.
807 Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and
808 repeat separation. Genome Res. 27, 722–736. https://doi.org/10.1101/gr.215087.116

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

809 Küpper, C., Stocks, M., Risse, J.E., dos Remedios, N., Farrell, L.L., McRae, S.B., Morgan,
810 T.C., Karlionova, N., Pinchuk, P., Verkuil, Y.I., Kitaysky, A.S., Wingfield, J.C.,
811 Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D.B., Burke, T., 2016. A supergene
812 determines highly divergent male reproductive morphs in the ruff. Nat. Genet. 48, 79–
813 83.
814 Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg,
815 S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5,
816 R12.
817 Lamichhaney, S. *et al.* Rapid hybrid speciation in Darwin's finches. *Science* eaao4593 (2017).
818 Leducq, J.-B., Nielly-Thibault, L., Charron, G., Eberlein, C., Verta, J.-P., Samani, P.,
819 Sylvester, K., Hittinger, C.T., Bell, G., Landry, C.R., 2016. Speciation driven by
820 hybridization and chromosomal plasticity in a wild yeast. Nat. Microbiol. 1, 15003.
821 https://doi.org/10.1038/nmicrobiol.2015.3
822 Mallet, J., 2007. Hybrid speciation. Nature 446, 279–283.
823 https://doi.org/10.1038/nature05706
824 Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229–237.
825 Marsellach, X. A non-genetic meiotic repair program inferred from spore survival values in
826 fission yeast wild isolates: a clue for an epigenetic ratchet-like model of ageing?
827 (2017).
828 McDonald, M.J., Rice, D.P., Desai, M.M., 2016. Sex speeds adaptation by altering the
829 dynamics of molecular evolution. Nature 531, 233–236.
830 https://doi.org/10.1038/nature17143
831 Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L., Seehausen, O., 2017.
832 Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8,
833 ncomms14363. https://doi.org/10.1038/ncomms14363
834 Muller, H.J., 1932. Some genetic aspects of sex. Am. Nat. 66, 118–138.
835 Munz, P., Wolf, K., Kohli, J., Leupold, U., 1989. Genetics overview. Mol. Biol. Fission Yeast
836 1–30.
837 Naumov, G.I., Kondratieva, V.I., Naumova, E.S., 2015. Hybrid sterility of the yeast
838 Schizosaccharomyces pombe: Genetic genus and many species in statu nascendi?
839 Microbiology 84, 159–169. https://doi.org/10.1134/S0026261715010099
840 Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of
841 restriction endonucleases. Proc. Natl. Acad. Sci. 76, 5269–5273.
842 https://doi.org/10.1073/pnas.76.10.5269
843 Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F.,
844 Piñeiro, R., Riina, R., Rosselló, J.A., Soltis, P.S., Villa-Machío, I., 2017. Is homoploid
845 hybrid speciation that rare? An empiricist's view. Heredity 118, 513–516.
846 https://doi.org/10.1038/hdy.2017.7
847 Nieuwenhuis, B.P.S., Tusso, S., Bjerling, P., Stångberg, J., Wolf, J.B.W., Immler, S., 2018.
848 Repeated evolution of self-compatibility for reproductive assurance. Nat. Commun. 9.
849 https://doi.org/10.1038/s41467-018-04054-6
850 Nolte, A.W., Sheets, H.D., 2005. Shape based assignment tests suggest transgressive
851 phenotypes in natural sculpin hybrids (Teleostei, Scorpaeniformes, Cottidae). Front.
852 Zool. 2, 11. https://doi.org/10.1186/1742-9994-2-11
853 Nuckolls, N.L., Bravo Núñez, M.A., Eickbush, M.T., Young, J.M., Lange, J.J., Yu, J.S.,
854 Smith, G.R., Jaspersen, S.L., Malik, H.S., Zanders, S.E., 2017. wtf genes are prolific
855 dual poison-antidote meiotic drivers. eLife 6. https://doi.org/10.7554/eLife.26033
856 Ortiz-Barrientos, D., Engelstädter, J., Rieseberg, L.H., 2016. Recombination Rate Evolution
857 and the Origin of Species. Trends Ecol. Evol. 31, 226–236.
858 https://doi.org/10.1016/j.tree.2015.12.016

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

859 Payseur, B. A. & Rieseberg, L. H. A genomic perspective on hybridization and speciation.
860 *Mol. Ecol.* n/a-n/a (2016).
861 Pennisi, E., 2016. A shortcut to a species. Science 354, 818–818.
862 https://doi.org/10.1126/science.354.6314.818
863 Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A.,
864 Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B.,
865 Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G.,
866 Schacherer, J., 2018. Genome evolution across 1,011 Saccharomyces cerevisiae
867 isolates. Nature 556, 339–344. https://doi.org/10.1038/s41586-018-0030-5
868 Pfennig, K.S., Kelly, A.L., Pierce, A.A., 2016. Hybridization as a facilitator of species range
869 expansion. Proc. R. Soc. B Biol. Sci. 283, 20161329.
870 https://doi.org/10.1098/rspb.2016.1329
871 Poelstra, J.W., Vijay, N., Bossu, Christen, Lantz, Henrik, Ryll, Bettina, Müller, Inge,
872 Baglione, Vittorio, Unneberg, Per, Wikelski, Martin, Grabherr, Manfred, Wolf, Jochen
873 B. W., 2014. The genomic landscape underlying phenotypic integrity in the face of
874 gene flow in crows. Science 344, 1410–1414.
875 Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J.,
876 Kresovich, S., Goodman, M.M., Buckler, E.S., 2001. Structure of linkage
877 disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci.
878 98, 11479–11484. https://doi.org/10.1073/pnas.201394398
879 Runemark, A. *et al.* Variation and constraints in hybrid genome formation. *Nat. Ecol. Evol.*
880 (2018).
881 Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N.,
882 Reich, D., 2014. The genomic landscape of Neanderthal ancestry in present-day
883 humans. Nature 507, 354–357. https://doi.org/10.1038/nature12961
884 Schiffels, S., Durbin, R., 2014. Inferring human population size and separation history from
885 multiple genome sequences. Nat. Genet. 46, 919–925. https://doi.org/10.1038/ng.3015
886 Schumer, M., Cui, R., Powell, D.L., Rosenthal, G.G., Andolfatto, P., 2016. Ancient
887 hybridization and genomic stabilization in a swordtail fish. Mol. Ecol. 25, 2661–2679.
888 https://doi.org/10.1111/mec.13602
889 Schumer, M., Rosenthal, G. G. & Andolfatto, P. How common is homoploid hybrid
890 speciation?: Perspective. *Evolution* **68,** 1553–1560 (2014).
891 Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A.,
892 Schatz, M.C., 2018. Accurate detection of complex structural variations using single-
893 molecule sequencing. Nat. Methods 15, 461–468. https://doi.org/10.1038/s41592-018-
894 0001-7
895 Seehausen, O., 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19, 198–207.
896 https://doi.org/10.1016/j.tree.2004.01.003
897 Singh, G., Klar, A.J., 2002. The 2.1-kb inverted repeat DNA sequences flank the mat2, 3
898 silent region in two species of Schizosaccharomyces and are involved in epigenetic
899 silencing in Schizosaccharomyces pombe. Genetics 162, 591–602.
900 Soltis, P.S., Soltis, D.E., 2009. The Role of Hybridization in Plant Speciation. Annu. Rev.
901 Plant Biol. 60, 561–588. https://doi.org/10.1146/annurev.arplant.043008.092039
902 Speed, D., Hemani, G., Johnson, M.R., Balding, D.J., 2012. Improved Heritability Estimation
903 from Genome-wide SNPs. Am. J. Hum. Genet. 91, 1011–1021.
904 https://doi.org/10.1016/j.ajhg.2012.10.010
905 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
906 large phylogenies. Bioinformatics 30, 1312–1313.
907 https://doi.org/10.1093/bioinformatics/btu033

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

908 Steenkamp, E.T., Wingfield, M.J., McTaggart, A.R., Wingfield, B.D., 2018. Fungal species
909 and their boundaries matter – Definitions, mechanisms and practical implications.
910 Fungal Biol. Rev. 32, 104–116. https://doi.org/10.1016/j.fbr.2017.11.002
911 Stukenbrock, E.H., 2016. The Role of Hybridization in the Evolution and Emergence of New
912 Fungal Plant Pathogens. Phytopathology 106, 104–112.
913 https://doi.org/10.1094/PHYTO-08-15-0184-RVW
914 Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in
915 hierarchical clustering. Bioinformatics 22, 1540–1542.
916 https://doi.org/10.1093/bioinformatics/btl117
917 Tajima, F., 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA
918 Polymorphism. Genetics 123, 585–595.
919 Taylor, D.J., Hebert, P.D., 1993. Habitat-dependent hybrid parentage and differential
920 introgression between neighboringly sympatric Daphnia species. Proc. Natl. Acad. Sci.
921 90, 7079–7083. https://doi.org/10.1073/pnas.90.15.7079
922 Team, R.C., 2014. R: A language and environment for statistical computing.
923 Teresa Avelar, A., Perfeito, L., Gordo, I., Godinho Ferreira, M., 2013. Genome architecture is
924 a selectable trait that can be maintained by antagonistic pleiotropy. Nat. Commun. 4,
925 2235. https://doi.org/10.1038/ncomms3235
926 The Heliconius Genome Consortium, Dasmahapatra, K.K., Walters, J.R., Briscoe, A.D.,
927 Davey, J.W., Whibley, A., Nadeau, N.J., Zimin, A.V., Hughes, D.S.T., Ferguson, L.C.,
928 Martin, S.H., Salazar, C., Lewis, J.J., Adler, S., Ahn, S.-J., Baker, D.A., Baxter, S.W.,
929 Chamberlain, N.L., Chauhan, R., Counterman, B.A., Dalmay, T., Gilbert, L.E.,
930 Gordon, K., Heckel, D.G., Hines, H.M., Hoff, K.J., Holland, P.W.H., Jacquin-Joly, E.,
931 Jiggins, F.M., Jones, R.T., Kapan, D.D., Kersey, P., Lamas, G., Lawson, D., Mapleson,
932 D., Maroja, L.S., Martin, A., Moxon, S., Palmer, W.J., Papa, R., Papanicolaou, A.,
933 Pauchet, Y., Ray, D.A., Rosser, N., Salzberg, S.L., Supple, M.A., Surridge, A.,
934 Tenger-Trolander, A., Vogel, H., Wilkinson, P.A., Wilson, D., Yorke, J.A., Yuan, F.,
935 Balmuth, A.L., Eland, C., Gharbi, K., Thomson, M., Gibbs, R.A., Han, Y., Jayaseelan,
936 J.C., Kovar, C., Mathew, T., Muzny, D.M., Ongeri, F., Pu, L.-L., Qu, J., Thornton,
937 R.L., Worley, K.C., Wu, Y.-Q., Linares, M., Blaxter, M.L., ffrench-Constant, R.H.,
938 Joron, M., Kronforst, M.R., Mullen, S.P., Reed, R.D., Scherer, S.E., Richards, S.,
939 Mallet, J., Owen McMillan, W., Jiggins, C.D., 2012. Butterfly genome reveals
940 promiscuous exchange of mimicry adaptations among species. Nature 487, 94–98.
941 https://doi.org/10.1038/nature11041
942 Turner, L.M., Harr, B., 2014. Genome-wide mapping in a house mouse hybrid zone reveals
943 hybrid sterility loci and Dobzhansky-Muller interactions. eLife 3, e02504.
944 https://doi.org/10.7554/eLife.02504
945 Twyford, A.D., Streisfeld, M.A., Lowry, D.B., Friedman, J., 2015. Genomic studies on the
946 nature of species: adaptation and speciation in Mimulus. Mol. Ecol. 24, 2601–2609.
947 https://doi.org/10.1111/mec.13190
948 Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P., Wolf,
949 J.B.W., 2016. Evolution of heterogeneous genome differentiation across multiple
950 contact zones in a crow species complex. Nat. Commun. 7, 13195.
951 https://doi.org/10.1038/ncomms13195
952 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A.,
953 Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for
954 Comprehensive Microbial Variant Detection and Genome Assembly Improvement.
955 PLoS ONE 9, e112963. https://doi.org/10.1371/journal.pone.0112963

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

956 Watterson, G.A., 1975. On the number of segregating sites in genetical models without
957 recombination. Theor. Popul. Biol. 7, 256–276. https://doi.org/10.1016/0040-
958 5809(75)90020-9
959 Weir, B.S., 1979. Inferences about Linkage Disequilibrium. Biometrics 35, 235.
960 https://doi.org/10.2307/2529947
961 Wellenreuther, M., Bernatchez, L., 2018. Eco-Evolutionary Genomics of Chromosomal
962 Inversions. Trends Ecol. Evol. 33, 427–440. https://doi.org/10.1016/j.tree.2018.04.002
963 Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J.,
964 Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D.,
965 Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis,
966 P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J.,
967 Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K.,
968 James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P.,
969 Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S.,
970 Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D.,
971 Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K.,
972 Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward,
973 J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E.,
974 Rieger, M., Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Düsterhöft, A., Fritzc,
975 C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H.,
976 Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R.,
977 Purnelle, B., Goffeau, A., Cadieu, E., Dréano, S., Gloux, S., Lelaure, V., Mottier, S.,
978 Galibert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M.,
979 Rochet, M., Gaillardin, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado,
980 L., Jimenez, J., Sánchez, M., del Rey, F., Benito, J., Domínguez, A., Revuelta, J.L.,
981 Moreno, S., Armstrong, J., Forsburg, S.L., Cerutti, L., Lowe, T., McCombie, W.R.,
982 Paulsen, I., Potashkin, J., Shpakovski, G.V., Ussery, D., Barrell, B.G., Nurse, P., 2003.
983 Erratum: corrigendum: The genome sequence of Schizosaccharomyces pombe. Nature
984 421, 94–94. https://doi.org/10.1038/nature01203
985 Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J.,
986 Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D.,
987 Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis,
988 P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J.,
989 Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K.,
990 James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P.,
991 Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S.,
992 Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D.,
993 Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K.,
994 Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward,
995 J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E.,
996 Rieger, M., Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Fritzc, C., Holzer, E.,
997 Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R.,
998 Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B.,
999 Goffeau, A., Cadieu, E., Dréano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F.,
1000 Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M.,
1001 Gaillardin, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado, L.,
1002 Jimenez, J., Sánchez, M., del Rey, F., Benito, J., Domínguez, A., Revuelta, J.L.,
1003 Moreno, S., Armstrong, J., Forsburg, S.L., Cerrutti, L., Lowe, T., McCombie, W.R.,
1004 Paulsen, I., Potashkin, J., Shpakovski, G.V., Ussery, D., Barrell, B.G., Nurse, P., 2002.

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1005 Erratum: The genome sequence of Schizosaccharomyces pombe. Nature 415, 871–880.
1006 https://doi.org/10.1038/nature724
1007 Zanders, S.E., Eickbush, M.T., Yu, J.S., Kang, J.-W., Fowler, K.R., Smith, G.R., Malik, H.S.,
1008 2014. Genome rearrangements and pervasive meiotic drive cause hybrid infertility in
1009 fission yeast. eLife 3, e02630. https://doi.org/10.7554/eLife.02630
1010 Zheng, X., Gogarten, S.M., Lawrence, M., Stilp, A., Conomos, M.P., Weir, B.S., Laurie, C.,
1011 Levine, D., 2017. SeqArray—a storage-efficient high-performance data format for
1012 WGS variant calls. Bioinformatics 33, 2251–2257.
1013 https://doi.org/10.1093/bioinformatics/btx145
1014 Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-
1015 performance computing toolset for relatedness and principal component analysis of
1016 SNP data. Bioinformatics 28, 3326–3328.
1017 https://doi.org/10.1093/bioinformatics/bts606
1018
1019
1020
1021 **Figure legends**
1022

1023 **Figure 1 | Distribution of *Sp* (red) and *Sk* (blue) ancestry blocks along the *S. pombe*
1024 **genome.** (a) Example of principal component analysis (PCA) of a representative genomic
1025 window in chromosome I (top) and the whole mitochondrial DNA (bottom). Samples fall into
1026 two major clades, *Sp* (red square) and *Sk* (blue square). The proportion of variance explained
1027 by PC1 and PC2 is indicated on the axis labels. Additional examples are found in
1028 Supplementary Figure 1 (b) Proportion of variance explained by PC1 (black line) and PC2
1029 (grey line) for each genomic window along the genome. Centromeres are indicated with red
1030 bars. Note the drop in proximity to centromeres and telomeres where genotype quality is
1031 significantly reduced. (c) Heatmap for one representative of 57 near-clonal groups indicating
1032 ancestry along the genome (right panel). Samples are organised according to a hierarchical
1033 clustering, grouping samples based on ancestral block distribution (left dendrogram). Colours
1034 on the tips of the cladogram represent cluster membership by chromosome (see
1035 **Supplementary Figure 7**). Samples changing clustering group between chromosomes are
1036 shown in grey. (d) Estimate of D_{xy} between ancestral groups and genetic diversity (π) within
1037 the *Sp* (red) and *Sk* clade (blue) along the genome.

1038

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1039 **Figure 2 | Population genetic summary statistics.** (a) Proportion of *Sp* (red) and *Sk* (blue)
1040 ancestry across all 57 samples along the genome. (b) Tajima's D differentiated by *Sp* (red)
1041 and *Sk* (blue) ancestry and pooled across all samples irrespective of ancestry (grey line).
1042 Genomic regions previously identified under purifying selection (Fawcett et al., 2014) are
1043 shown with black triangles. Reported active meiotic drives (Zanders et al., 2014; Hu et al.,
1044 2017; Nuckolls et al., 2017) are indicated by yellow triangles. The third panel shows the
1045 difference between ancestry specific Tajima's D and the estimate from the pooled samples.
1046
1047 **Figure 3 | Inferred evolutionary history of contemporary *S. pombe* strains.** An ancestral
1048 population diverged into two major clades, *Sp* (red) and *Sk* (blue) since approximately 2300
1049 years ago (Jeffares et al., 2015). Recurrent hybridization upon secondary contact initiated
1050 around 20 sexual outcrossing generations ago resulted in admixed genomes with a range of
1051 admixture proportions (bottom) prevailing today.
1052
1053 **Figure 4 | Ancestry explains variation in phenotype and reproductive isolation.** (a) PCA
1054 of normalized phenotypic variation across 228 traits. The proportion of variance explained by
1055 PC1 and PC2 is indicated on the axis labels. Admixed samples (dots) are coloured coded by
1056 ancestry proportion (cf. **Figure 3**) ranging from pure *Sp* (red triangle) to pure *Sk* (blue
1057 triangle) ancestry. (b) Phenotypic distribution of example traits separated by the degree of
1058 admixture: admixed samples are shown in grey, pure ancestral *Sp* and *Sk* samples are shown
1059 in red and blue respectively. The number of traits corresponding to a dominant, additive and
1060 transgressive genetic architecture is indicated on the right hand side (c) Comparison of
1061 heritability estimates of all 228 traits based on 172,935 SNPs (abscissa) and on 1925 genomic
1062 windows polarized by ancestry (ordinate). Colours indicate statistical significance. *NS*:
1063 heritability values not significantly different from zero, *AncHap:* significant only using

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1064 ancestral blocks, *SNPs:* significant only using SNPs, *SB:* significant in both analyses.
1065 Diagonal (slope=1) added as reference. (d) Histogram of the difference between heritability
1066 estimates using SNPs and ancestry components for all 228 traits. (e) Correlation between the
1067 difference in ancestry proportions between two strains (cf. **Figure 2**) and spore viability of the
1068 cross. Red box shows samples with low spore viability but high genetic similarity.
1069

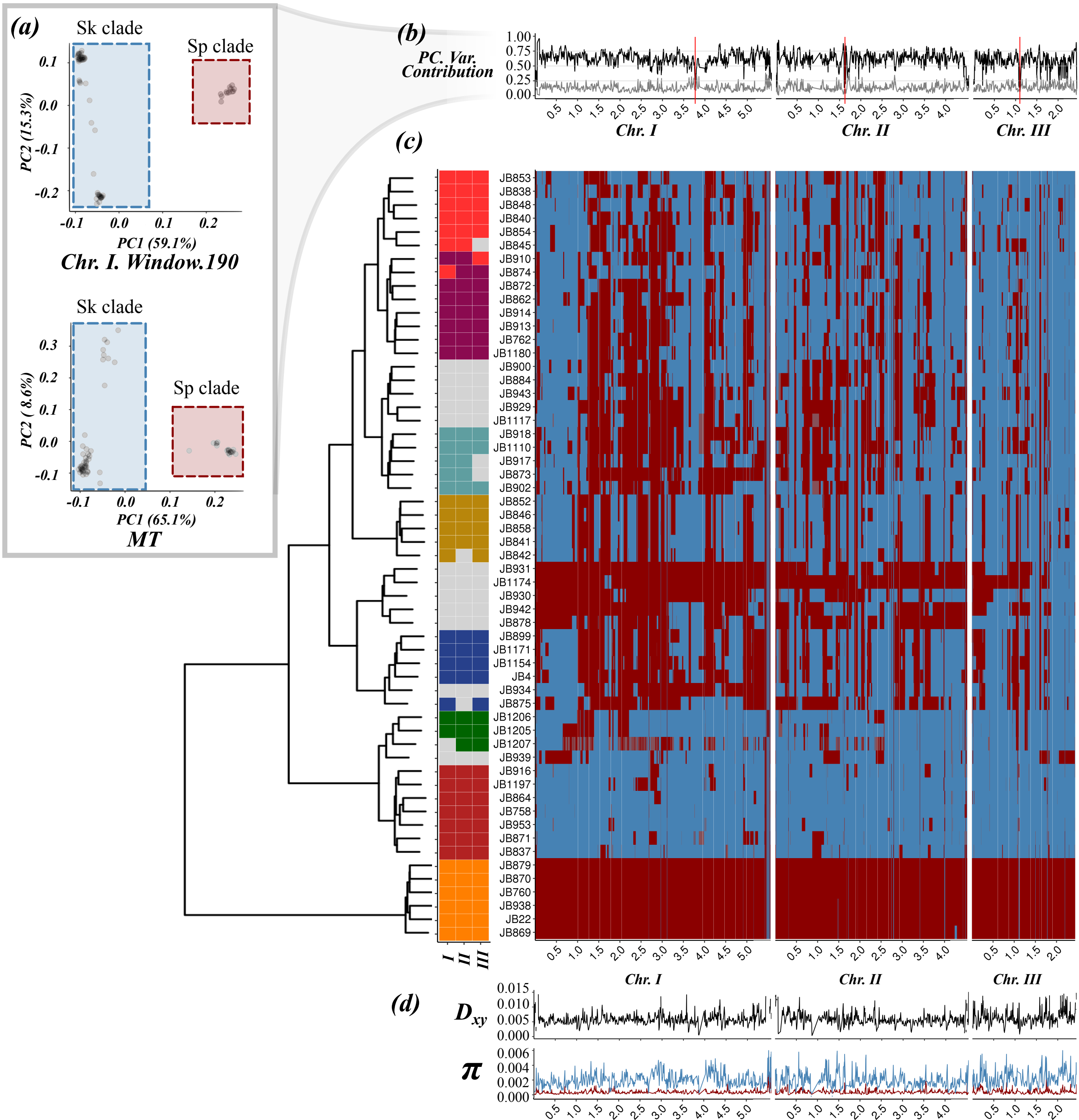
1070 **Figure 5 | Characterization of structural variation based on long-read, real-time**
1071 **sequencing.** (a) Schematic representation of the three chromosomes in different strains
1072 displaying SVs larger than 10kb relative to reference genome JB22 (left panel). Chromosome
1073 arms are differentiated by colour; orientation is indicated with arrows relative to the reference;
1074 black bars represent centromeres. In the second panel, additional SVs, their type and ID of the
1075 corresponding strain are illustrated in brackets. (b) Size distribution of SVs below 10 kb.
1076 Colours indicate the type of SV. (c) Distribution of SV density along the genome. Black bars
1077 represent centromeres. (d) Two-dimensional, folded site frequency spectrum between inferred
1078 ancestral populations for all SVs, SNPs and LTR INDELs. Numbers and colours show the
1079 percentage of the total number of variants in each category. Variants with low frequency in
1080 both populations are shown in the blue box. Variants highly differentiated between
1081 populations are show in red boxes with total in the upper right box. Fills with percentage
1082 lower than 0.01 are empty.
1083 *Abbreviations:* DEL: deletion; DUP: duplication; INS: insertion; INV: inversion
1084

1085 **Figure 6 | Decay in linkage disequilibrium (LD) with genetic distance.** Relationship
1086 between LD (R^2) and physical distance is depicted for each chromosome. Black points
1087 represent values for each window pair comparison. Lines show non-lineal regression model
1088 based on Hill & Weir (1988) and Remington et al. (2001). LD estimates were divided into

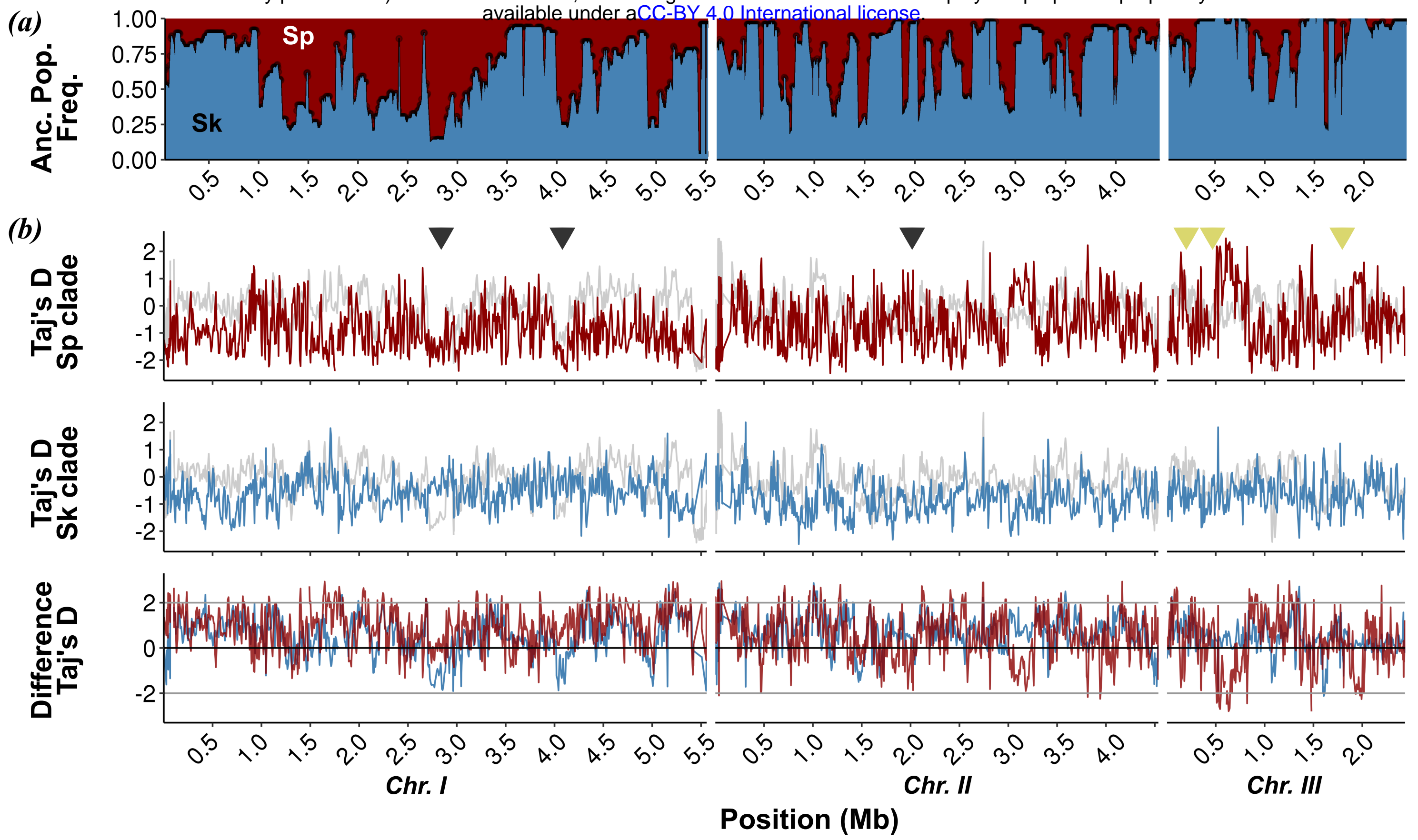
bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1089 three categories representing comparison between windows of shared ancestry (*Sp*-*Sp* or *Sk*-
1090 *Sk*) reflecting positive ancestry disequilibrium (AD) or of opposite ancestry (*Sk*-*Sp*) reflecting
1091 negative AD.
1092

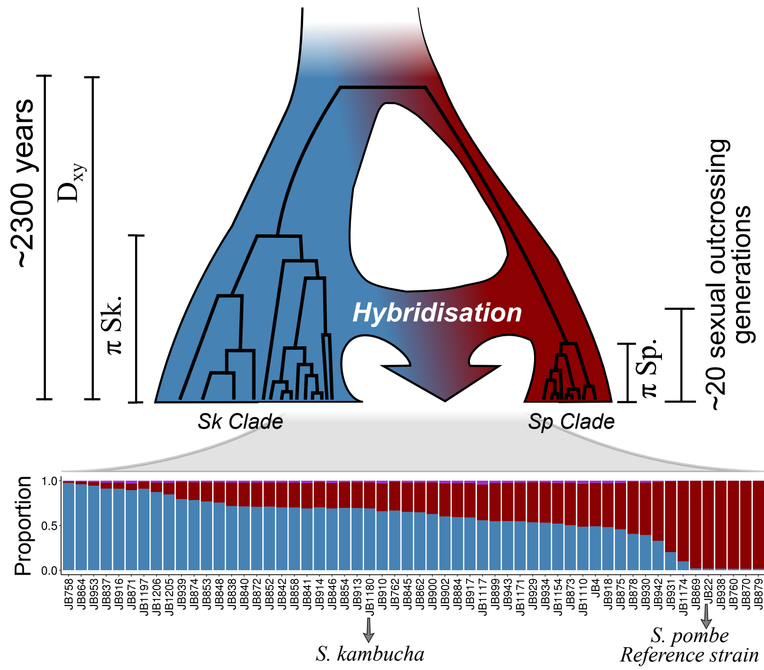
bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



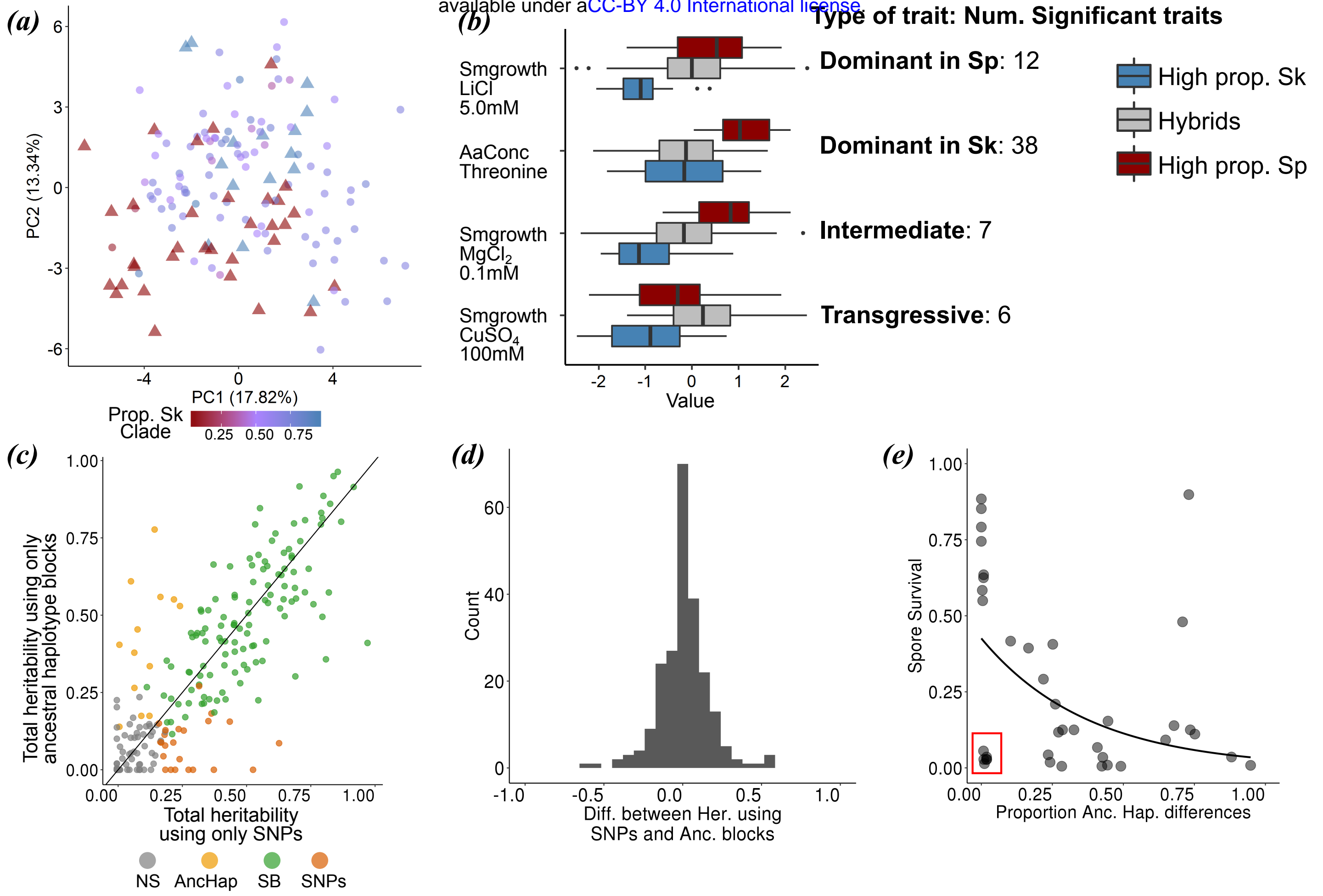
bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(a)
2 Mb
Chr. I
Chr. II
Chr. III
JB22 Reference
TR - Chr I
JB873
INV JB4
INV JB872
INV JB1174
JB1171
INV JB858
INV JB939
JB760
JB842
JB853
JB942
JB918
JB934
JB1197
JB1205
JB1206

(b)
Count
Length
DEL
INS
DUP
INV

(c)
Chr. I
Chr. II
Chr. III
Count
Position (Mb)

(d)
SVs
SNPs
LTR SVs
Allele Fq. Sk clade
Allele Fq. Sp clade
Percentage
3.2
20.8

bioRxiv preprint doi: https://doi.org/10.1101/415091; this version posted January 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.