Clustering of Inverted Triplications in Centromeric and Subtelomeric Chromosomal Regions of *Aspergillus flavus*

Short Title: Inverted Triplications in Centromeres and Subtelomeres

Jeffrey W. Cary[2], Amna Malik[1], Alexie Smith[1], Rose Ludecke[1], Erik Flemington[3], and Arthur J. Lustig[1,4]

[1]Department of Biochemistry and Molecular Biology
[3]Department of Pathology
Tulane School of Medicine and Tulane Cancer Center
New Orleans, LA 70112

[2]Food and Feed Safety Research Unit
USDA/ARS
Southern Regional Research Center
New Orleans, LA 70124

[4]Corresponding author:
Arthur J. Lustig, PhD
1430 Tulane Avenue SL-43
Department of Biochemistry and Molecular Biology
Tulane School of Medicine
New Orleans, LA 70112
alustig@tulane.edu
847-913-5968

ABSTRACT

The formation of inverted repeats frequently initiates rearrangements associated with gene amplification. Among these rearrangements are inverted triplications (TRP/INVs) that are associated with multiple disease states in humans. We have used the filamentous fungus Aspergillus flavus as a model system to analyze spontaneous DNA rearrangements using high-depth third-generation sequencing of DNA isolated from vegetatively cultured strains. Analysis of sequence data identified a class of infrequent and transient rearrangements having structures typical of TRP/INVs. These TRP/INVs form both unprocessed and processed species and have a variably sized deletion at the junction between direct and inverted sequences. We found that TRP/INVs are enriched in heterochromatic centromeric and subtelomeric A+T-rich regions, suggesting that this process is a source of genetic instability in these domains. Consistent with this finding, A+T-rich regions contain elevated levels of direct, inverted, and perfect palindromic repeats. Inverted junctions contain palindromic sequences that are typically associated with TRP/INVs. A closer examination of the class of genomic palindromes found at inverted junctions revealed a broad distribution, with the majority lying in AT-rich centromeric sequences. The distribution of TRP/INVs in centromeric and subtelomeric domains mirrors the frequency of palindromes, indicating that palindrome abundance is likely to be a driver of TRP/INVs. A further examination of predicted pairing patterns in palindrome-like structures suggests that these structures may function as sites of DNA resolution, strand transfer, and replicative stalling. These results are consistent with a replication-based model in which palindromes, under conditions of replication stress, facilitate the formation of TRP/INV structures. These studies also represent, to our knowledge, the first characterization of spontaneous TRP/INV formation, likely to underlie the complex rearrangements associated with higher eukaryotic genomic instability and disease states.

AUTHOR SUMMARY

Inverted triplications (TRP/INVs), coupled with gene amplification, are sources of genetic instability. However, little is known about how these species arise spontaneously prior to selective pressures. We have established a fungal model system that uses long-read third-generation sequencing to identify low-abundance rearrangements in the absence of an intentional selection. We identified and characterized multiple classes of TRP/INVs flanked by palindromic sequences. These rearrangements were concentrated in the AT-rich centromeric and subtelomeric regions, suggesting that these sites are regulated recombinational hotspots. Palindromes that are also associated with TRP/INVs are also most prevalent in A+T-rich centromeric regions, suggesting that the abundance of palindromes, both as a source of DNA damage and fork stalling, is a driver of TRP/INV formation. These studies provide the first evidence (to our knowledge) of the spontaneous formation of TRP/INVs during DNA replication, which is likely to precede more complex sources of rearrangements associated with oncogenesis and other disease states in higher eukaryotes.

INTRODUCTION

Gene duplication is a common theme in genetic change in all phyla ranging from bacterial to fungal to human cells and has been implicated in the genesis of disease states [1, 2]. Indeed, the spontaneous formation of large duplication inversions is an underlying mechanism contributing to oncogenic states [2, 3].

Model systems are valuable tools to select and characterize gene duplications. Duplications can occur with long sequence elements through transposition and homologous recombination. Rearrangements, however, also occur in the presence of minimal homology, particularly in the presence of perfect paired and imperfect (quasi-) palindromes [4]. Palindromes are particularly unstable due to their extrusion as either cruciform or hairpin structures. For example, breakage-fusion-bridge (BFB) cycles can involve either the loss of a telomere, or, alternatively the processing of palindromic cruciform

79 DNA to form inversions and subsequent DNA breakage, followed by fusion of DNA strands through
80 processes such as non-homologous end joining [3, 5-7].
81
82 Inverted triplications (TRP/INVs) are another class of species mediated by the rearrangement of
83 relatively short regions of homology [8-11]. These latter species are observed in bacteria and yeast and
84 can be initiated by palindromes under conditions of replicative stress and fork regression during bi-
85 directional DNA replication [12, 13]. Recent data has also suggested a role for TRP/INVs in the
86 rearrangements involved in a multiplicity of disease states [14]. In one model for the formation of these
87 structures, palindromes flanking a gene can create a hairpin that primes DNA replication and the synthesis
88 of the inverted copy. After reaching a second palindrome, template switching and DNA replication give
89 rise to the direct repeat. The junctions between direct and inverted repeats are thought to reform the
90 palindromes in the resulting TRP/INV, which may be resolved into asymmetric TRP/INVs containing
91 deletions surrounding the palindromic junctional sequence. On a statistical basis, short palindromes are
92 expected to form most easily in AT or GC-rich sequences. AT-rich palindromes form frequently in the
93 human genome and are involved in both chromosomal translocation and evolutionary hypervariability
94 [15, 16]. TRP/INVs and other rearrangements are typically observed only under selection. However, the
95 array of rearrangements that may occur in the absence of any selection or in regions of simple sequence
96 non-genic DNA have not been investigated.
97
98 Centromeric and telomeric regions of chromosomes are essential for the stable inheritance of
99 genetic information. At the microtubule attachment site, centromeric domains can be present as point or
100 regional structures, the latter including chromosome-specific satellite sequences [17]. In *Neurospora* and
101 other filamentous fungi, centromeres appear to be embedded within heterochromatic A+T-rich sequences,
102 modeled in part by a centromere-specific histone H3 variant [18, 19]. These centromeres are thought to be
103 generated through a process of repeat-induced point mutation (RIP) [20-22]. In another filamentous
104 fungus, *Aspergillus (A.) flavus*, centromeres consist of approximately 100 kb of 90-95% % A+T rich

105 sequence on each of the eight chromosomes [23, 24]. Generally, centromere sequences on homologs in
106 different strains share extensive (88-99%), but not complete, homology [24]. In contrast, the AT-rich
107 sequences of non-homologs share minimal homology. Thus, competing forces appear to maintain a
108 balance between variation and conservation.
109
110 Rearrangements near fungal telomeres are thought to drive genetic reshuffling, adaptive
111 polymorphisms, and elimination of gene sequences [23, 25]. The subtelomeric 100 kb, adjacent to the
112 telomeric $(TTAGGGTCAACA)_{9\text{-}11}$ repeats [26], is also interspersed with A+T-rich elements (ATEs), of
113 2-20 kb in length, that are distributed in different configurations among the chromosomal arms within and
114 between strains. Our previous studies have revealed remnant transposable elements in subtelomeres,
115 suggesting that repetitive cycles of transposition and RIP contribute to the evolution of these sequences
116 [24]. Nonetheless, we found highly conserved ATEs that lack any homology to transposable elements on
117 both the same chromosome of different strains and distinct chromosomes of the same strain. Hence,
118 subtelomeric ATEs are also maintained in a homeostasis of divergence and conservation, indicating a
119 process that acts to promote recombination. The processes acting to promote and to limit ATE
120 destabilization are not understood.
121
122 Identifying spontaneous rearrangements using standard genetic techniques has been challenging,
123 due to their low frequencies and the limited means to identify complex forms [27]. We have employed a
124 combination of two approaches to examine species arising from spontaneous events. First, the
125 multinucleate mycelial growth of *A. flavus* leads to the potential identification of rare deleterious
126 mutations and rearrangements [28]. Selection would tend to eliminate such species. Second, we used
127 analysis of high-coverage third-generation sequencing as a physical readout of these genetic
128 rearrangements. Here, we report on the identification of infrequent and transient TRP/INVs clustered in
129 heterochromatic centromeric and subtelomeric regions in the absence of overt selection; a process that
130 appears to be driven in part by the abundance of small palindromes.

131

132 **RESULTS**

133 **A+T Rich Elements in Centromeres and Subtelomeric Regions Have a Nonrandom Distribution of**

134 **Direct, Inverted, and Palindromic Repeats**

135 To ascertain the accumulation of short repeated sequences in *A. flavus* AT-rich centromeric

136 regions, we examined random 250 bp DNAs for BLAST self-homology and identified potential direct,

137 inverted, and palindromic repeats. We plotted the BLAST self-homology e-values [29] of each repeat

138 class in 129 randomly selected 250 bp, 90% A+T-rich sequences from centromeric regions. This result

139 was then compared to repeats identified among 500 computationally derived and random 250 bp, 90%

140 A+T-rich sequences (Table 1).

141

Sequence	Repeat	BLAST e value	≥ 1 Repeats	0 Repeats	P value (FET)
Centromeric	Direct	$<10^{-4}$	13	116	<0.0001
Subtelomeric	Direct	$<10^{-4}$	5	75	<0.02
Random	Direct	$<10^{-4}$	7	493	NA
Centromeric	Inverted	$<10^{-4}$	9	120	<0.001
Subtelomeric	Inverted	$<10^{-4}$	6	74	<0.003
Random	Inverted	$<10^{-4}$	6	494	NA
Centromeric	Palindrome	$<10^{-8}$	9	120	<0.003
Subtelomeric	Palindrome	$<10^{-8}$	6	74	<0.007
Random	Palindrome	$<10^{-8}$	8	492	NA

142 Table 1. **Direct, Inverted and Palindromic Repeats in Centromeric, Subtelomeric and Randomly**

143 **Generated A+T Rich tracts.** The sequence source, repeat type, BLAST e-value, and number of samples

144 assayed containing either ≥ 1 or 0 repeat are listed together with the p values from a comparison of the

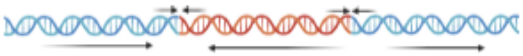
145 genomic and random sequences as determined by the Fisher Exact Test (FET).

146

147 The most significant inverted and direct repeats (e-values $<10^{-4}$) identified in either genomic or
148 randomly generated sequences were compared by the Fischer Exact Test. We determined that the direct
149 repeats within centromeres were at significantly higher values than those generated from random
150 sequences ($p<0.0001$). We further identified a significantly higher level of inverted repeats ($p<0.001$).
151 Due to the higher spontaneous level of palindromic sequences in AT-rich sequences, we compared
152 centromeric and random sequences at BLAST e-values of $<10^{-8}$ and found a significant increase in the
153 number of perfect palindromes ($p<0.003$) in centromeres. Palindromic sequences were found in 6.9% of
154 centromeres sequences that were tested.
155
156 We conducted an analogous test among eighty, 250 bp sequences extracted from ATEs within
157 subtelomeric regions, defined as the 100 kb adjacent to the telomeric $(TTAGGGTCAACA)_{9\text{-}11}$ repeats
158 [23, 25] (Table 1). Direct and inverted repeats were quantified in this region and compared with
159 randomized sequences. Both sets of repeats with e values $<10^{-4}$ were present at elevated levels in the
160 subtelomere, relative to random sequences (direct repeats: $p<0.002$; inverted repeats: $p<0.002$). Similarly,
161 perfect palindromic sequences were again present more frequently in subtelomeric sequence than in
162 random sequence datasets ($p<0.003$) and were present in 7.5% of the subtelomeric sequences. Hence, all
163 three classes of repeats were significantly elevated in centromeric and subtelomeric domains, suggesting
164 an active process that generates this nonrandom behavior.
165
166 **Qualitative Identification of a TRP/INV Class in Cultured Cells**
167 We were interested in identifying spontaneous rearrangements that may help to explain the
168 nonrandom characteristic of repeated elements in subtelomeric and centromeric ATE sequences. To this
169 end, we carried out high coverage (500-1500-fold) third-generation Oxford nanopore (ONT) sequencing
170 as a readout for infrequent structural variants (SVs) [30] after forced vegetative culturing. We optimized
171 the Sniffles 2 and minimap2 programs [31-33] for infrequent SVs and categorized them into apparent
172 deletions, insertions, inversions, duplications, and translocations. We specifically characterized individual

173 sequencing reads from the insertion class that predict sizes of ≥ 300 bp. BLAST analysis of these larger
174 insertion sequences revealed two major classes: intact tandem duplications and duplications interrupted
175 by a partial or complete inversion. This study focused on the latter class of structures that have structures
176 akin to the inverted triplications [TRP/INVs], previously described in gene amplification studies. [8-10]
177
178 Our initial qualitative analysis compared this subset of the SV insertion class that appeared to
179 arise in clusters from *A. flavus* genomic DNA prepared from a starter culture (G0) (generated from spores
180 isolated from conidiophores) after five (G5), ten (G10), and fifteen (G15) weeks of forced vegetative
181 (mycelial) growth. One of the major SV classes had the TRP/INV structures shown in Fig. 1 (top). Nine
182 TRP/INVs were present, five of which mapped in the ~1 Mb of ATE DNA and four dispersed in the
183 remaining ~ 36.5 Mb of genomic DNA. When compared to all single-read TRP/INV candidates derived
184 from other cultures (G5, G10, G15), each TRP/INV was found only at the indicated stage of growth.
185 Hence, within the limits of detection, TRP/INVs did not become fixed in the population and appeared to
186 arise only transiently during growth (Fig. 1, bottom).

TRP/INV Structure



CULTURE	CHR	COORDINATE	bp	bp	bp
G5	1	6193921	310	157	310
G5	2	2148241	210	200	210
G10	2	5040836	196	193	196
G10	4	1456506	7640	7789	7640
G10	4	2791817	2872	2916	2872
G10	6	422416	2213	2154	2213
G10	6	2287885	1336	1246	1336
G15	1	4565045	507	45	507
G15	7	692345	48	312	48

187
188 Fig. 1. **Initial Analysis of TRP/INV from Large Clustered Insertions in G5, G10 and G15 Cells.**

189 DNAs were isolated and sequenced after *A. flavus* NRRL3357 cells had been cultured for 5 (G5), 10
190 (G10), and 15 (G15) cycles, and subsequently screened for large (≥ 300 bp) insertions, using the Sniffles
191 2 algorithm as described in Materials and Methods with a variant read number of 3 and characterized
192 further by IGV analysis. Individual reads were then analyzed by BLASTing to the G0 assembly to
193 identify the TRP/INV shown here. The listed coordinates are the approximate sites of TRP/INV formation
194 relative to the chromosomal (CHR) orientation. The yellow boxes refer to centromeric TRP/INV, while
195 the orange boxes refer to subtelomeric TRP/INVs and TRP/INVs within ATE elements scattered at other
196 sites in the genome. Uncolored boxes refer to TRP/INVs dispersed at non-ATE sites. The structure of the
197 unprocessed (symmetric) TRP/INV is shown on top, with arrows below figure representing the relative
198 orientations of TRP/INV direct and inverted repeats. Arrows above figure represent the typical
199 palindromic sequences at these sites. The corresponding sizes of the direct (blue) and inverted (red)
200 repeats are shown below.
201

202 **Quantitative Characterization of the TRP/INV Class in G15 Cultured Cells**

203 To analyze the frequency of TRP/INVs quantitatively, we conducted an exhaustive study of all
204 G15 SVs identified as ≥ 300 bp insertions (relative to the G0 assembly) that also were associated with a
205 linked inversion. We characterized each of these reads and identified those having a TRP/INV
206 configuration. These 221 TRP/INV species were distributed over all eight chromosomes of *A. flavus*
207 (Table 2, SI_Table_S1). The low abundance and transitory nature of these species were further examined
208 by quantifying all chromosome 1 TRP/INV in DNA derived from G0, G5, G10 and G15. We found three
209 cases of heritable sites of TRP/INV. In each case, the TRP/INV structure was unique (SI_FigS1). We
210 cannot rule out, however, the possibility that each of these had a common origin that subsequently
211 underwent multiple rounds of rearrangement. These data reinforce the transient nature of TRP/INV.
212
213
214
215
216

Chromosome	# TRP/INV
1	33
2	23
3	44
4	27
5	34
6	14
7	20
8	26
total	221

Table 2**. Chromosomal Distribution of G15 TRP/INV Among All Chromosomes**

Direct BLAST search of individual sequencing reads relative to the G0 telomere-to-telomere assembly revealed three TRP/INV subtypes (Fig. 2). Eighty-nine species fall into the symmetric class (class 1); Fig. 2, top) and have intact head-to-tail configurations at both inversion junctions within the estimated margin of error of sequencing and sequence alignment (Δ0-50 bp). One hundred thirteen species fall into an asymmetric class (class 2; Fig. 2, middle), displaying an intact head-to-tail configuration (Δ0-50 bp) at one of the two junctions and a more extensive deletion at the other end. In comparison, 19 species belong to another asymmetric class (class 3; Fig. 2, bottom) consisting of TRP/INVs containing a > 50 bp loss of sequence at both inversion junctions. SI_Fig_S2 shows examples of each of these classes. We also identified a series of complex forms that may be derived from multiple rounds of TRP/INV formation (data not shown).

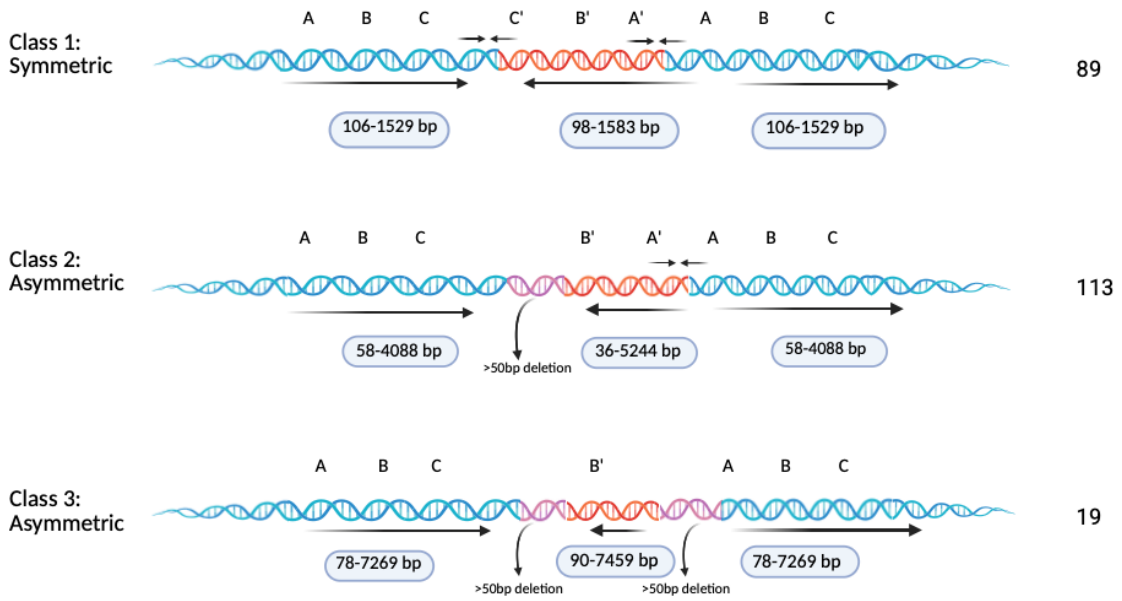


Fig. 2. **Three Structural Classes of TRP/INVs.** The three classes of TRP/INVs deduced by BLAST analyses of nanopore sequencing queries to genomic assemblies are presented here. The first class (Class 1, symmetric) has little or no deletion at the junction between direct (blue) and inverted (red) repeats within the error of sequencing, set at ≤ 50 bp. Class 2 (asymmetric) has a deletion at one of the two junctions >50 bp. Class 3 (asymmetric) has deletions of >50 bp at both junctions. The size ranges of both direct and inverted repeats are listed below for each class and the number of observed species in each class is listed on the right. The orientation of hypothetical sequences A, B, and C (and complementary sequence A', B' and C') as well as the directional arrows are shown for visual clarity as described in the legend to Fig. 1. Created in BioRender. Cary J. et al. (2025) https://BioRender.com/s70s676.

TRP/INV Clusters at Centromeric and Subtelomeric ATEs

We analyzed the TRP/INV distribution to determine their potential enrichment in specific chromosomal domains. We found a 17.5-22-fold higher frequency of TRP/INV in the AT-rich elements (ATEs) present at centromeres (1/14.3 kb), subtelomeres (1/11 kb), and other classes of ATEs (1/11.5 kb), compared to the overall frequency at other genomic sites (1/251.6 kb) (Table 3A). The latter class distributes to both coding (101/145) and non-coding (44/145) regions (Fig. S2A-S2D; Table S1).

249 A

Location	# TRP/INV	Frequency
Centromeric	56	1/14,285 bp
Subtelomeric ATE	14	1/11,023 bp
"Other" ATE	6	1/11, 500 bp
Dispersed	145	1/251,563 bp

250
251 B
252

Location	Direct Repeat Sequence (bp)	Inverted Repeat Sequence (bp)
Centromeres	425 ± 202	400 ± 242
Subtelomeres+ "Other" ATEs	455 ± 367	421 ± 349
Dispersed	471 ± 733	498 ± 771

253 Table 3**. Enrichment and Characterization of TRP/INVs in Centromeric and Subtelomeric ATEs**.
254 A) The chromosomal location, number, and fraction of TRP/INV in ATEs (centromeric, subtelomeric,
255 and ATEs located elsewhere in the genome) and non-ATE (dispersed) regions, scattered among all
256 chromosomes, are listed [24]. B) The direct and inverted repeat sizes (together with standard deviations)
257 of TRP/INVs, located in centromeric and non-centromeric ATEs (subtelomeres + other ATEs) or
258 dispersed non-ATEs are shown.
259

260 Analysis of the size of the direct and inverted TRP/INV sequences (Table 3B) revealed that the
261 range of values was reduced (as shown by the standard deviation) in centromeres and subtelomeric (and
262 other) ATEs, relative to dispersed non-ATE sequences. This reflects a tendency towards longer lengths in
263 both the direct and inverted duplications in non-ATE TRP/INVs. For example, the lengths of such
264 duplications extend up to 7.4 kb, compared to the maximum size of 0.92 kb at centromeres. Both the

265 direct and inverted duplications in TRP/INV also trended to slightly smaller sizes in centromeric,
266 compared to non-ATE, domains (Table 3B). These data suggest that selective forces may well restrict
267 centromeric TRP/INV lengths.
268
269 We compared the TRP/INV patterns obtained from the analysis of the nanopore sequencing with
270 the results from two independently PacBio (Revio) [34] sequenced G15 DNAs to ascertain the influence
271 of sequence methodologies on the TRP/INV generation (SI_Table_S2). We observed qualitatively similar
272 results with both sequencing methods, although TRP/INV levels were roughly 30-fold higher in ONT
273 than in PacBio sequencing. Seven TRP/INVs were present, all at unique centromeric and subtelomeric
274 sites. While substantiating the formation of TRP/INVs and their biased distribution towards AT-rich
275 regions, these results suggest the presence of uncharacterized methodological biases in the library
276 formation or sequencing of TRP/INV- containing DNA.
277
278 **Mitochondrial Genome Accumulates TRP/INVs Less Frequently Than the Nuclear Genome**
279 As a control for the establishment of transient TRP/INVs and to test the contribution of library or
280 sequencing anomalies that may generate these structures, we also measured SVs by the same methods in
281 the high copy AT-rich mitochondrial DNA present in the in the sequencing reactions of each culture.
282 The 30 kb circularly permuted and high copy mitochondrial genome is present in a sequencing depth of
283 20,000 to 30,000 copies. In contrast, centromeric sequences are in general covered in less depth (~250
284 copies) than other parts of the genome. The mitochondrial genome therefore can serve as a model for the
285 accumulation of TRP/INVs in centromeric sequences. In the G15 nanopore sequencing reads, we found
286 only two TRP/INVs at unique sites. Due to the small sample size, we also explored the frequency of
287 mitochondrial TRP/INVs in G0, G5 and G10 nanopore sequencing experiments. G0, G5, and G10 reads
288 gave rise to 1, 2, and 0 TRP/INVs, respectively. If mitochondrial and nuclear TRP/INVs were found at a
289 comparable frequency, we would have expected 210 TRP/INVs per 30 kb genome. Furthermore, the
290 frequency of TRP/INVs in dispersed sequences is > 25-fold higher than in the mitochondrial genome,

291 despite their similar palindrome density (Table 4). These results are inconsistent with a technical reason
292 for TRP/INV formation and additionally suggest that aspects of nuclear DNA replication and
293 recombination contribute to the formation of these species.

	TRP/INV (Observed)	Total Genome Sequenced (kb)	Sequencing Depth (Average)	TRP/INV Frequency (/kb/read)	Predicted Palindrome Density (/100kb)
Centromere	56	800 (G15)	278	2.5×10^{-4}	831
Mitochondria	5	120 (G0-G15)	26140	1.6×10^{-6}	140
Dispersed	145	36500 (G15)	1000	4.0×10^{-5}	100

294
295 Table 4. **Observed Frequencies of TRP/INVs at Centromeres and Mitochondrial Sequences**. Out of
296 the total of G0, G5, G10 and G15 DNAs (120 kb total) genomes sequenced, we found only TRP/INVs
297 despite the higher depth of mitochondrial sequencing. Taking the predicted palindromes density of the
298 78% AT-rich mitochondrial genome into account, we estimate at least a 30-fold enrichment of TRP/INVs
299 in the ~800 kb of centromeric sequence and a >25-fold enrichment in dispersed nuclear sequences.
300 Mitochondrial and centromeric palindromes were predicted using similar assumptions, as described in
301 **Materials and Methods**. The origin of the sequenced genome is given in the parentheses.
302

303 **Identification of Palindrome Sequences at the Inverted/Direct Repeat Junction**

304 Previous studies revealed that TRP/INVs of bacterial and yeast genes are formed via palindrome
305 formation and template switching during bi-directional DNA replication or DNA repair which recreates
306 the palindromic sequence in the symmetric product [8-10]. We tested the prediction that palindromes are
307 at the site of TRP/INV formation, by identifying junctional read sequences that were complimentary to
308 both strands of the genomic assembly by BLAST analyses (examples are shown in SI_Fig_S3A-S3D).
309 The palindromes were then screened further for their predicted thermodynamic stabilities using the
310 UNAfold mfold 4.2 program [35-37] (Table 5A- 5C, Table 6). Seventy five junctions had palindromic
311 characteristics and separate into those falling in centromeric (Table 5A), dispersed (Table 5B), and in
312 subtelomeric (and other ATE) (Table 5C) locations.

Chr	Coord	Sequence	Length)bp)	ΔG (kcalm-[1])	T_m (°C)
1	4510493	TAATTTTAATTATATAATAAAAATTA	26	-4	48.4
1	4615841	AGAATTATAATTATATATTAACTACTAACTACTAAATAATATTATAATTAAAAATA	56	-5	44.6
2	2064783	TTATAATAAATATTATATAATAATATTATTTAA	36	-3.94	51.7
3	2732726	CTATAACTTATAAGTTATAG	20	-7	60.6
3	2739625	TATTTAATTTATAAATAGTTTATAAAATTAAATAT	35[e]	-6.58	45.8
4	2818566	TTATATATTTTATCTATTATTATAATATTTTTTATTATTTTTTTATAGTAATATTATTTAATATTAAAATCTATAT	73	-6.8	39.2
4	2865046	TAATAAAATATAATATTTTTTAAAATATATAAATATCTACTATATTTATATAAAATTTTTAAAAAAATTATATTTATAGT	75	-17.3	52.1
5	2447757	TAAATAATAATTAATATATTAATATCTTAATAATATTAACTAGTAGTTA	50[e]	-5.32	45.4
5	2472736	TTAAATATTCTAGAATCTTTCTAGAATATT	18	-5.08	60.8
5	2482236	TTTAATTTATAATTCTATAATTTTATTATTATATATTAAA	33	-9.61	65.8
6	2149500	TTTAATTTATAATTCTATAATTTTATTATTATATATTAAA	40[e]	-3.7	41.7
7	2618350	ATCTTTATATTAATAATAATAAAATTAATTAATATTTTAATATAATTAAT	19	-2.65	44.1
7	2630477	AATATAATAATTTTATATTAAAATAAATATTAATAAAAAATAAAATTATT	46	-8.83	53.8
7	2636529	ATTTTTAATTTTAAAGATTTTTAATTAATTAGCTAGTTTAAAATATTTTATAAAAATTAATTTATATAATATTTACTATTAA ATTTAAATTAAAAAAAATAAATAAAAAAT	100	-13.09	41.3
7	2646080	TTAATAGTATAATTAATCTTAACTTATTTATACTTTTTTAA	21	-5.32	52.5
8	1814185	TTTATTATTTTATTTTATCAATTTAAAAAAATAATTATAA	33[m]	-8.53	41.8
8	1851716	AATAAAATATTTTTTATAAAAATAATTAATT	25	-3.18	45.5
8	1870116	TATAAGTAGACTAGCGCCGGGCGCTAGTCTTAATA	24	-12.28	79.7
8	1897025	ATATAAAATTAATTAATATAAAAAAGTAAAAATATTATTAATTTAAAATTTTAAA	21	-5.33	42.8

Table 5A. **Predicted Junctional Palindromes in Centromeric TRP/INVs**. The chromosomal coordinates, sequence coordinates, sequence, length and predicted ΔG (kcal/mol) and T_m values (°C) (calculated by UNAfold mfold 4.2 using the default settings: 25°C, 3 mM Mg, 50mM NaCl) of centromeric TRP/INV palindromes. Values of ΔG > -3 kcal/mole presented in red.

Palindromes ranged in size from 13-100 bp in non-ATE sequences and 18-100 bp in centromeric, subtelomeric (and other ATE) sequences. Based on the UNAfold DNA secondary structure algorithms, most putative palindromes have stabilities of -3 kcal/mmol or less (ranging from -3 to -20 kcal/mmol) and T_m values of at least 40°C (ranging from 40-84° C) (Table 5, right column). Based in part on the discussion below, -3 kcal/mol was used as the lowest stability cutoff if T_m were at or close to 40°C.

Chr	Coord	Sequence	Length (bp)	Δ G (kcalmol^{-1})	T_m (°C)
1	2266091	CCCAGAAATCCGCAGCAGCGGCTGGG	26^{c}	-9.21	65.8
1	2286600	TCCCATGACGTCATGGGA	18	-7.27	73.1
1	5016955	AACGTTATAACGTT	14	-4.96	64.8
2	454491	ATAAACCATCGCTTTGCATTGACTGCATCCAGGTTTAT	39^{c}	-6.82	50.1
2	454669	AAGCTTTGATTATACTACTGAATGAAAGCTT	31	-6.82	57.4
2	493210	CGGCCAAATTCGCCG	15^{e}	-4.72	68.6
2	1804865	CATTTCACCACGGTGAACTGTGGTGAATTGTGCTTG	30	-10.36	62.4
2	3365568	ATTAGTATCTGATACTAAT	21^{mp}	-4.84	57
2	3840022	GCTAGCACCATCTGTCAGTTTGGCTGCAGAAAGGGGCTTCATACAATAACCTACTCTTCATGGAACTCAT AGTCCAACGCCTTCCCCAGCCTGACCCGGA	100	-10.34	42
2	4116246	CCCGACAGTTACATACATTAGTAGTACTCGTACTGTACTCCCTGTATGTAACTTTCGGG	59	-17.44	59.7
2	4273901	GTAATGTCCTCAGCGGTGCGGACGTTAC	27	-7.05	59.7
2	5407963	CGACGGCGACAGCCTCGCCGTCG	23	-13.94	83.9
3	567045	TGTCGACGTCGACA	16	-4.75	61.1
3	573282	CACGCAAGAGCTTGGCCGGTTTGCGTGATCG	31	-8.35	69.9
3	616220	CAAAGTCCAGTGAGACGGCTCCGGGACTTTG	31	-9.18	59.7
3	1526699	ATTCCGTATCAATACGGAAT	20	-8.37	71.4
3	1572702	GATCGACGGAAGATGTAACGTCGATC	26	-9.18	68
3	1734350	TTTTTTTCTTTAAAAAAAAAAAAAAG	27	-6.42	50.8
3	3287741	TTCATTGAGGCAATGAA	17^{e}	-4.78	56.9
3	3921549	TTACCTTATCACGGTAAAT	19	-1.84	40.5
3	4939733	GAATTCTAGATCC	14	-2.24	45.1
4	568414	AATTGCAGTCACTGATGTTCGCAATT	25^{1}	-4.05	59.2
4	794920	AACGAGACATGATCCATGTCGCGTT	24	-9.95	70
4	1194263	CAGGCGAAGCGCTGCAGCCTG	21^{e}	-6.12	69
4	1273121	AAACCTCTCAGAGGTTT	17	-5.97	66.8
4	1274984	GCAGGACCATGAGCTCATGCGACATGTTGTGCA	31	-5.22	45.3
4	2528116	ATACTCCGGTATGCAATGCAATGCGGAGTAT	31	-7.88	55.4
4	4238747	ATCCCCCAATGCGGGGAT	18	-5.65	73.4
4	4238730	GTAGACTGGGCGCCCAGTCTAC	22	-9.27	69.6
4	4239051	TCTTGCGCGAGCGTCTGCAGGCGCTGCAGCAGA	34	-7.62	61.1
4	4240170	GTAGACTGGGCGCCCAGTCTAC	24	-9.54	68.3
4	4414478	TCGGTAGAGCCTCTACCGA	20^{e}	-7.39	72.1
5	618917	TGTCAAAATTTCATGAAATTTTTGACA	27^{e}	-6.54	56.3
5	2331951	AGAATGCTAAGAAAGACAATATCTATTTCTTAATTCTC	37^{e}	-4.94	46.5
5	3445422	CATCGCGACGATGATG	16	-4.6	61.5
5	3772422	CAGGGCGCCTGCGACCTG	18^{e}	-7.07	71.7
5	4111968	GACATCATCAAGTGATG	17	-4.91	62.2
6	1167663	TTTATAAATTTTGATATTTATAAA	26	-4.21	45.7
6	1661313	GTCCTCGCAATGTTACATCGAGGAC	25^{e}	-7.71	66.7
6	2486550	ACGGAGCCGCAACGGCTCCGT	21	-12.1	87.5
6	2486789	TTGTCGCAGACACGGCAA	18	-4.13	62.2
6	2523428	ACTAGGAGTACCTAGT	16	-4.68	57.6
7	1656247	GTACCGTTTTCGGCTATGGGTTTCAAATCCCCCAAGTCCGAAAACCGTAC	50	-12.75	53.2
7	1814658	GCGGCGTTTGCACAGCCGCGT	19^{e}	-4.36	67.9
8	256532	GTATGTCAACAGGTAACATAC	21^{e}	-4.3	52
8	1428359	GAAGAACCAGACGATAGTTCGGTTCTTC	34	-7.82	63.6

Table 5B**. Predicted Palindromes at Dispersed Non-ATE TRP/INV Junctions.** The chromosomal coordinates, sequences, length and predicted ΔG and T_m values of palindromes at inverted repeat junctions of non-ATE TRP/INVs. Stabilities were determined as in Table 5A. Those marked in red are > -3 kcal/mole, calculated as in Table 5A. The superscript e refers to palindromes associated with extended regions of pairing; mp refers to multi-palindromic structures, and c refers to more complex pairing structures.

Chr	Coord	Sequence	Length	ΔG (kcalm-1)	Tm (°C)
1	75341	AAATTAGAATTATAAGATTAATTTTTAAGCAGCTTTTTAAATATAGATTCTAATTTC	57	-5.76	38.3
3	2693002	TTATAATTATATAGAATTATAA	22	-3.22	44.3
3	2705424	ATATTTTTTTAATTAAAATAT	21	-2.99	44.2
4	2071	TAATATTAATAATTATTATTTTTATAATATTAAATTCTATA	41[e]	-3.98	44.3
5	969	ATTAATTTAAATATTATTAATTTTATAATTTAATTAATATTATATTTAAAATACTTTAAT	60	-7.92	42.7
5	1720	ATAAATATAAATTATATTTAT	21	-4.88	52.1
5	4510716	AATTATAAATACTTTTTTAATATATTATTA	21	-1.21	34
7	690480	CTATAAATTTAAATAGTTAATAAAAATTTTAATAATATATATTAAATTTTTAAAATATTATTATTAATATTTAAATAAATAGATAA	89	-13.45	44
8	3238765	ATAAAATAGTTTAGTAAATATATATATAATTATTACTAGAAAATAAATAATTTATTTTCTAGATATATAATTTTATTAGACTATTTTAT	90	-18.55	51.7

Table 5C. **Predicted Palindromes at \Subtelomeric and Dispersed ATE TRP/INV Junctions.** The chromosomal coordinates, sequences, length and predicted ΔG and T_m values of palindromes at inverted repeat junctions of subtelomeric (and other ATE) TRP/INVs of all stabilities were determined as in Fig. 5A. Those marked in red are > -3 kcal/mole (as calculated in Fig. 5A).

Location	Sequenced Palindromes (Total)	Sequenced Palindromes (≤ -3 kcal/mole; $T_m \geq 40°C$)
Centromeres	**20**	**18**
Subtelomeres + Other ATEs	**9**	**7**
Dispersed	**46**	**44**
Total	**75**	**69**

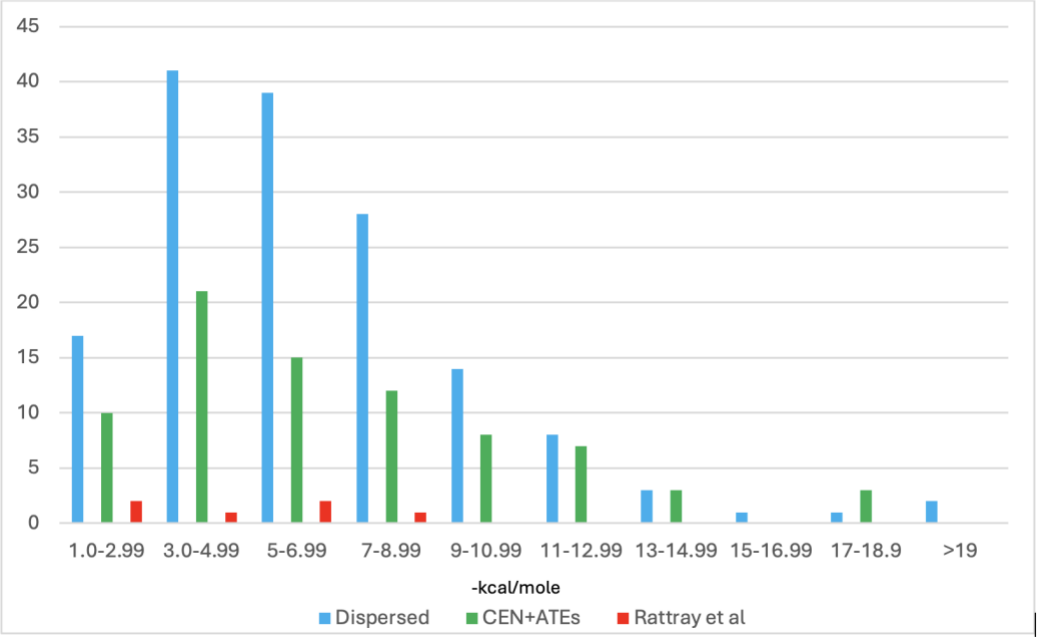
Table 6. **Palindrome Distribution Identified by BLAST and UNAfold Analyses.** The distribution of palindromes from ATEs (centromeric and subtelomeric + ATEs located elsewhere in the genome) and non-ATE (Dispersed) regions are listed together with the distribution of palindromes predicted to have ΔG values of ≤ -3kcal and $T_m \geq 40°C$. The palindromes of lower stabilities may still form secondary structures but do not meet the set limitations set in this study.

347 As a control, we compared the predicted stability of a previously isolated set of short palindromic
348 junctions deduced to be at the sites of inverted sequence formation in yeast [7] (Table 7). These
349 palindromes have ΔG values ranging from -2.31 to -7.06 kcal/mole and T_m values from 42-67°C,
350 substantiating the validity of the choice of parameters in this study.

Sequence	Length (bp)	Δ G ($K^{-1}mol^{-1}$)	T_m (°C)
GTACTATTGGTAC	13	-2.31	47.1
CGCCGGCCCAGTGGGCG	17	-5.44	72.3
TTGGTGCGGCCAA	13	-2.89	69.7
GGGCAATCACTTTTGCCC	18	-7.06	67
CTGGGGTTACCGCCAG.	16	-3.86	58.7
CCCAGGTGCCTGGG	14	-6.67	72.9

351
352 Table 7. **Predicted Stability of Palindromes Identified at Junctions of Inverted Repeats by Rattray**
353 **et al (2005)** [7].
354
355 Not all junctional sequences contained identifiable palindromic sequences within the threshold
356 defined in this study. Out of 75 sequenced junctions with no significant deletion, four fell below the
357 threshold value. This could either be an indication that the threshold is somewhat arbitrary or that a stable
358 palindrome may not always be necessary for TRP/INV formation.
359
360 **Palindromes Are Present in Genomic Sequences at Sites Adjacent to the TRP/INV Junctions**
361 We sought to test more generally whether palindromes are present in the genomic sequences that
362 predict the formation of TRP/INVs. To this end, we examined the 200 bp genomic sequence surrounding
363 the junction in unrearranged sequences at TRP/INVs sites that had junctional deletions of 40 bp or less,

364 for the presence of simple palindromes, complex paired structures and multi-palindromic sequences.
365 TRP/INV sites that contain larger deletions can obscure the actual site of palindromic sequences and were
366 excluded from this study. We found that most junctions contain simple palindromes with or without
367 extended associated pairing (147 [88%] in non-centromeric dispersed regions] 52 in centromeric regions
368 [93%]; and 24 [100%] in subtelomeric domains). The palindrome lengths (stem plus loop) averaged 33 bp
369 in dispersed regions, and 41-43 bp in AT-rich subtelomeric and centromeric regions. All stabilities
370 averaged -6.7 kcal/mol, while T_m values averaged higher in dispersed (56.4°C) than in AT rich regions
371 (48°C), as expected due to differences in base composition [Fig. 3; (SI Table 3A-C)]. A fraction (10%) of
372 junctions had stabilities with $\Delta G > 3.0$ kcal/mol. Some junctions were associated with long regions of
373 extensive pairing (13%), and a fraction (12%) had more complex pairing patterns or multi-palindromic
374 structures (SI Fig. 4 for examples). These findings raise the possibility that some homologous pairing
375 patterns may have a role in perturbing localized structure with or without strand exchange. These
376 predicted structures are unlikely to be static. Rather, they are likely to represent one among many in an
377 equilibrium of multiple palindromic structures.



378

379

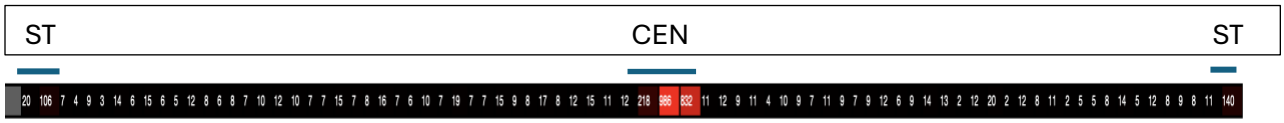
380 Fig. 3. **Palindrome Stabilities in Centromeric, Dispersed ATE and Non-ATE Sequences.** Palindrome
381 Stabilities Predicted by UNAfold Analyses in TRP/INVs with a deletion of 40 bp or less. The distribution
382 of TRP/INV palindromes in centromeric, subtelomeric, and other ATEs (CEN + ATEs) (green) and from
383 dispersed non-centromeric (Dispersed) sequences (blue) are shown as a function of stability (-kcal/mol)
384 of as determined by UNAfold analysis. Also shown are the stabilities found at sites of inversion in yeast
385 (Rattray et al MCB 2006) (red) for comparison.
386

387 **Palindrome Formation Mirrors TRP/INV Centromere Enrichment**

388 Regions of high AT or GC content produce a greater number of palindromes on a simple statistical
389 basis, and we observed an even greater frequency of palindromes in centromeric, subtelomeric and other
390 ATEs (Table 1). We wanted to test whether the frequency of a palindrome, having the predicted
391 characteristics of TRP/INV-associated palindromes, could be sufficient to explain the higher density of
392 TRP/INVs at centromeres. We therefore estimated the relative number of palindromes that would be
393 likely to generate a TRP/INV in centromeric and dispersed regions (see **Materials and Methods**). To this
394 end, we calculated the most frequently observed stem and loop sizes, as well as ΔG and ΔG_{cf}-ΔG_{lin} values,
395 the latter a measure of the energetic favorability of cruciform formation [38]. Due to the complexity of the
396 centromeric palindrome patterns, we first determined that the most frequent classes of the terminal portions
397 of the TRP/INV-associated palindromes in the centromere of chromosome 3 fell into a stem size of 12-100
398 bp, a loop size of 4-8 bp with 0-2 stem mismatches. The values were input into a program, Palindrome
399 Analyzer [39], designed to estimate the frequency and stabilities of palindromes at the chromosomal level,
400 and the sites of the TRP/INVs were confirmed to be present in the output. The distribution of utilized
401 palindromes had ΔG_{cf}-ΔG_{lin} values that fell between 10 and 16 kcal/mol, which were applied proportionally
402 to the output, since the distribution of stabilities in the observed and predicted palindromes were quite
403 similar. The frequency of the predicted palindromes was then corrected to account the number of false
404 positives. Based on this analysis, we estimate an average 100 kb region of centromere contained on average
405 over 800 palindromes (Fig. 4**)**. However, these values could not be applied to non-centromeric sequence
406 due to the differing TRP/INV-associated palindrome parameters in AT-rich and dispersed regions.

407 Examination of non-centromeric TRP/INV-associated palindromes in chromosomes 1, 3 and 4 revealed
408 stem lengths of 5-12 bp, loop sizes of 4-19 bp, and stem mismatches of 0-1. In this case, the observed and
409 predicted patterns were significantly different, with ΔG_{cf}-ΔG_{lin} values of the observed palindromes
410 distributed as follows: 15 at 3.1-6.0 kcal/mol (Class 1), 10 at 6.1-9.0 kcal/mol (Class 2) and 14 from at -
411 9.1-12.0 kcal/mol (Class 3). Since, the predicted TRP/INV-associated palindrome output had a different
412 relative distribution within a 60 kb region of centromere 3 (24 in class 1, 105 in class 2, and 507 in class
413 3), we normalized the observed recombination frequencies using a weighted eligibility model [40, 41] (**see**
414 **Materials and Methods**). Under these conditions, Palindrome Analyzer predicted ~100 palindromes/100
415 kb. Although these values reflect the characteristics of most, but not all, palindromes, the overall frequency
416 of palindromes clearly mirrors the enrichment of TRP/INVs in centromeric regions. We propose therefore
417 that the abundance of short palindromes is a major driver of TRP/INV formation.

418 A.

419


420
421 B.
422

Chromosome	**Centromeric Palindromes (100 kb)**	**Dispersed Palindromes (per 100kb)**
1	888	
2	807	
3	1038	
4	875	
5	761	
6	798	
7	698	
8	781	
Density/100kb	831	100

423
424 Fig. 4**. Predicted Distribution of TRP/INV-Associated Palindromes. (**A) Heat Map of Predicted
425 Distribution of Centromeric TRP/INV-Associated Palindromes (before ΔG filtering) (B) Predicted
426 Centromeric TRP/INV-Associated Centromeric Palindromes are Present at Higher Densities than
427 Dispersed TRP/INV Associated Palindromes (measured on a segment of ca 60 kb segment of

428 chromosome 3). The assumptions used for calculations of these values are described in the text. CEN,
429 centromere; ST, subtelomeric.
430

431 **DISCUSSION**

432 Here, using the *A. flavus* model system, we report the first evidence, to our knowledge, for
433 spontaneously formed TRP/INVs, a process is driven by the frequency of palindrome formation. Several
434 lines of evidence support the identification of spontaneous but infrequent inverted triplications in
435 *Aspergillus flavus* through structural variant analysis following high-coverage third-generation
436 sequencing. Although we did not apply any intentional selection during growth, selective pressures
437 induced by vegetative growth are possible. Indeed, in studies to be presented elsewhere, we found
438 reductions in conidiation levels between the fifth and tenth weeks of culturing (data not shown).
439

440 First, BLAST searches of individual reads of presumptive large insertions generated after two
441 independent methods of genome sequencing (ONT and PacBio) predicted TRP/INV structures consistent
442 with unprocessed symmetric and processed asymmetric TRP/INV forms [8-10], albeit differing in
443 abundance. Asymmetric products with deletions at one or more junctions are consistent with a role for a
444 classical or alternative non-homologous end joining (see below). A large fraction of sequenced inversion
445 junctions and sequences in short, deleted regions of asymmetric junctions revealed palindromes and
446 quasi-palindromes at the site of genomic rearrangement, arguing for a mechanism that favors the use of a
447 hairpin structure to generate the inverted sequence.
448

449 Second, we have shown that TRP/INVs are clustered at centromeres and centromeric AT-rich
450 DNA. Interestingly, the abundance of palindromes in centromeric regions mirror the clustering of
451 TRP/INVs. In addition, these studies have found that palindrome formation (as well as direct and inverted
452 repeats) is also significantly elevated in genomic, compared to randomly generated DNA having the same
453 AT content. These data argue for a central role of palindromes in driving TRP/INV formation and

454 clustering in centromeric and subtelomeric regions. However, palindromes and AT-rich DNA are likely
455 insufficient for TRP/INVs since the mitochondrial genome has a significantly decreased TRP/INV
456 frequency per unit size, despite being AT-rich. In addition, mitochondrial TRP/INVs were present at a
457 lower abundance than in dispersed sequences despite having a comparable density of predicted TRP/INV-
458 associated palindromes. These data argue for a preference for inverted triplications in the nuclear genome.
459
460 Third, we found that some palindromes were predicted to be associated with more extended
461 regions of homology or had more complex pairing structures. Although it is likely that predicted
462 algorithms only reflect a snapshot of likely stable forms, the data does raise the interesting possibility that
463 palindromes may have functions beyond or in addition to strand transfer [required for template switching
464 (see below)], such as in structural perturbations contributing to fork stalling.
465
466 Fourth, we found that the TRP/INVs are infrequent and do not become fixed in the population
467 after 15 cycles of forced vegetative growth. For the G15 cultures sequenced by ONT, we identified 221
468 reads out of 3.6×10^6 reads of sizes (> 2kb) that were used in this analysis. Since each read is derived
469 from an independent DNA molecule, this suggests an overall frequency of 6×10^{-5} TRP/INV/genome.
470 This estimate is only an approximation, since the coverage of different genomic regions is not identical
471 using either sequence method. These structures arise at every stage of subculturing and are in low
472 abundance, possibly due to high rates of TRP/INV processing.
473
474 Such TRP/INV instability may arise via distinct mechanisms of homologous and non-
475 homologous recombination (Fig. 3 left). Recombination between the direct repeats would result in
476 excision of a TRP/INV extrachromosomal circle, restoring the original structure (Fig. 3, left, pathway 1).
477 On the other hand, recombination between inverted repeats would generate a novel pattern of direct
478 repeats and inversions (Fig.3, left, pathway 2). Indeed, the high frequency of direct and inverted repeats
479 that we have found in centromeric and subtelomeric regions are among the expected products of

TRP/INV recombination. In contrast, unequal recombination between sister chromatids would result in both the loss and duplication of these regions.

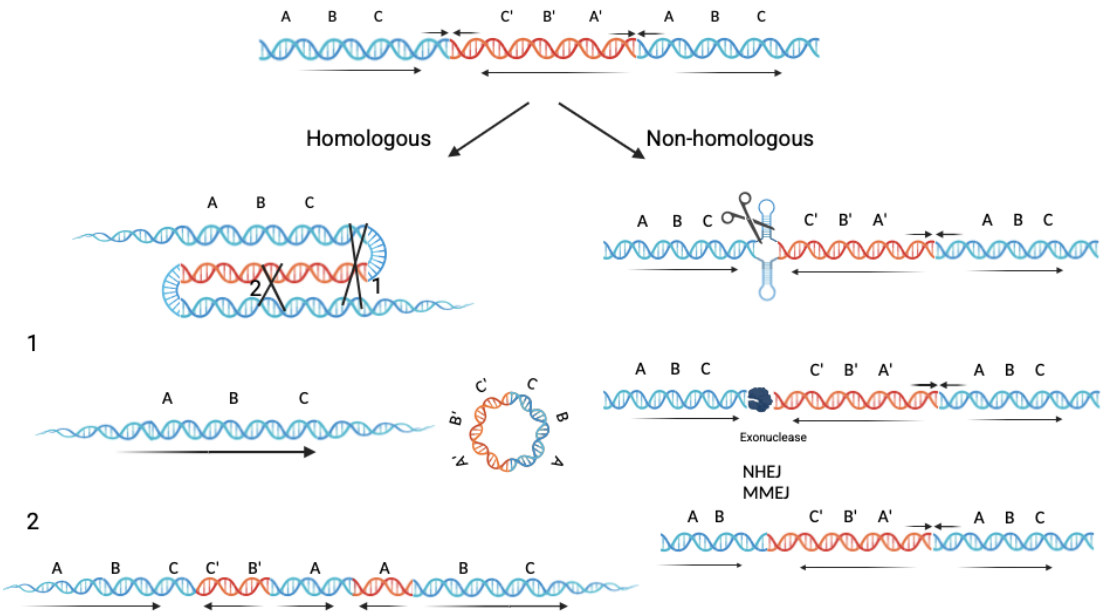
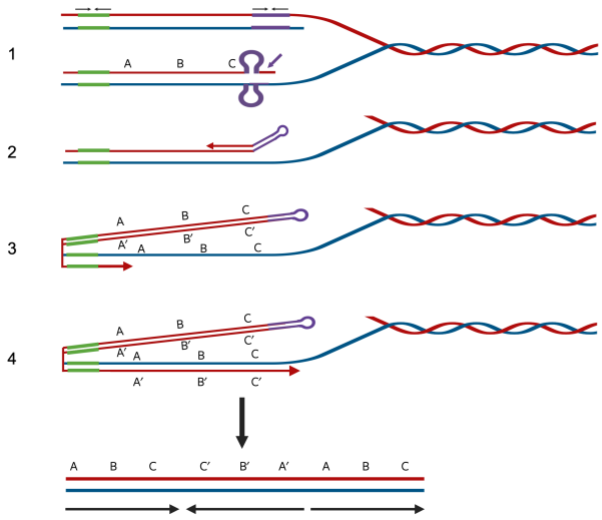


Fig. 5. **Models for the Resolution of TRP/INV Structures**. Homologous and non-homologous pathways for resolution of TRP/INVs. (Top) The structure of a TRP/INV. Color designations are described in Fig. 1. (Left) Recombination between direct repeats (1) gives rise to excision of a dimeric inverted circle that has the potential to reintegrate into the genome, creating repeat amplifications, while recombination between direct and inverted repeats (2) creates additional inversions that undergo further rearrangements. (Right) The extrusion of a palindrome is expected to undergo resolution or cleavage, resulting in a double strand break susceptible to exonuclease digestion. Microhomology mediated end joining (MMEJ) or non-homology end joining (NHEJ) can then produce Class 2 and Class 3 TRP/INV deletions. The orientation of hypothetical sequences A, B, and C and the directional arrows are shown for visual clarity. Created in BioRender. Cary, J.W. et al (2025) https://BioRender.com/r14e12.8

The instability of palindrome DNA that leads to cruciform and hairpin extrusion and promotes double strand break formation has been well-documented in both prokaryotes and eukaryotes [4]. Several pathways may be responsible for the resolution of the double-strand break, including exonucleolytic degradation coupled with microhomology-dependent non-homologous end-joining (MMEJ) [42, 43] (Fig. 3 right), or alternatively, Ku-dependent non-homologous end-joining (NHEJ) [43]. MMEJ is indeed

499 found in yeast and filamentous fungi, although by mechanisms independent from the pol θ-dependent
500 process in multicellular organisms and is likely to involve, in part, subunit 3 *(S. pombe cdc27*, S.
501 *cerevisiae pol32*) of polymerase ∂ [42, 44]. The frequent processing of palindromic DNA in AT-rich
502 sequences coupled with MMEJ or NHEJ is also expected to generate small deletions within these regions.
503 This means of resolution is likely to account for characteristics of asymmetric class 2 and 3 TRP/INVs.
504
505 The transient presence of TRP/INV could also be due to a strong selection against the rearranged
506 sequences leading to diminished growth rate or cell death. This explanation seems unlikely, however,
507 given the multinucleate nature of mycelial growth in *A. flavus*, which would permit the retention of
508 deleterious rearrangements [28]. If such a selection is present, however, it is more likely to be mediated
509 through TRP/INV disruption of coding regions.
510
511 While TRP/INV-like structures were identified in the centromeric and subtelomeric regions using
512 nanopore and PacBio sequencing, the frequencies of TRP/INV formation differ between the two methods
513 [30, 34]. Quantitative analysis using such different methodologies is indeed problematic. Although
514 TRP/INVs are unlikely to be the consequence of any known sequencing artifacts [45], as supported by the
515 reduced frequence of TRP/INV formation in mitochondrial DNA, several methodological distinctions
516 could contribute to these quantitative differences. For example, nanopore sequencing results in a direct
517 readout of the structural variant, while PacBio amplifies each extracted read to provide greater
518 consistency and potentially selects against TRP/INV amplification. In addition, the pre-treatment methods
519 to repair damaged DNA before the construction of the appropriate library may bias the size of DNA
520 fragments in distinct ways. We cannot, unfortunately, test these speculative possibilities at present.
521
522 Our data suggest that both structural and strand transfer properties of palindromes may be
523 involved in the formation of TRP/INV structures. The mechanism of formation is likely to be akin to the
524 models from Roth's laboratory [8] for TRP/INVs in *Salmonella* (Fig. 6). We propose that the structural

525 constraints of palindromes result in fork stalling, followed by subsequent palindrome resolution forming a
526 foldback that primes the replication of the inverted repeat. Processing of the second palindrome is then
527 responsible for the synthesis of the direct repeat. Replication stalling has indeed been observed in
528 experimentally manipulated constructs in both prokaryotes and eukaryotes [12, 13]. Following
529 palindrome-mediated template switching, microhomology-mediated pathway, such MMEJ or "micro"
530 single-strand annealing (SSA) processes, would prime the synthesis of the direct repeat [46]. This model
531 is appealing, since we can experimentally test the genetic dependence on proteins involved in
532 homologous recombination, MMEJ and micro-SSA, as well as the inhibition of these processes by Ku
533 binding and protection of damaged DNA. The presence of palindromes formed in the TRP/INV product
534 can also give rise to multiple cycles of inversion and duplication, leading to the observed complex
535 products (data not shown).
536



537

538 Fig. 6**. A Mechanism for Palindromic and Direct Repeat Strand Exchange in TRP/INV Formation.**
539 A four-step model modified from Roth and colleagues [8], based on their extensive studies in
540 *Salmonella* is presented: 1. Palindrome extrusion and processing in replicating DNA and resolution
541 by nucleolytic activities. 2. Formation of hairpin foldback that primes synthesis of the inverted
542 repeat. 3. DNA synthesis reaches the second palindrome (green) followed by Ku-independent (or

543 inhibited) mediated strand transfer and the priming of the direct repeat synthesis(4). Created in
544 BioRender. Cary J.W. et al. (2025) https://BioRender.com/m61z558
545

546 One of the exciting aspects of this study is the accumulation of TRP/INV in AT- rich
547 subtelomeric and centromeric sequences which is due, in part, to the critical role of palindromes
548 abundance in TRP/INV formation. Indeed, this study is the first characterization of short repeat-mediated
549 recombination within centromeric and subtelomeric regions. This finding may be related to our previous
550 observation that AT-rich regions at the centromeres of non-homologs (and in many cases, even
551 homologs) are highly divergent. We also found that specific subtelomeric ATE sequences are present at
552 different sites within and among strains. This observation may be related to the role of subtelomeres in the
553 elimination or the evolution of secondary metabolite gene clusters and their participation in other selected
554 rearrangements in fungi [25]. Both results support an equilibrium model, whereby deleterious functions
555 such as perturbation of kinetochore function or rampant subtelomeric rearrangement may counteract
556 evolutionary advantageous roles for centromeric divergence and subtelomeric sequence shuffling. We
557 propose that mechanisms that regulate such an equilibrium must be present at centromeres and
558 subtelomeres, possibly modulated by recombination activities. In addition, heterochromatin formation
559 may be regulated. The fungal centromeric (and probably subtelomeric) AT-rich sequences are
560 heterochromatic. In yeast, as in most eukaryotes, histone modification and chromatin condensation lead to
561 reduced levels of transcription, recombination and DNA repair in heterochromatin [47-49]. For example,
562 the histone deacetylase protein Sir2 is responsible for telomeric heterochromatin condensation and
563 position effects [50]. The use of *A. flavus sir2* (*sirE*) mutants with wide-ranging phenotypes [51] is an
564 attractive possibility to test this hypothesis. Such regulation may be the underlying basis for the selection
565 against larger sizes of centromeric TRP/INVs. In addition, subtelomeric genes that are bounded by
566 heterochromatic regions may serve to regulate transcription from these regions.
567

568 Our findings suggest that centromeres and subtelomeres can be more recombinationally active
569 than anticipated. ATEs are maintained in all filamentous fungi and are indeed sites of evolutionary
570 polymorphism [20]. However, any region that is high in AT or GC content has a statistically higher
571 probability of palindrome formation and may be subject to a similar process, a possibility that requires
572 further exploration. The TRP/INVs reported here are yet another example of the reported association of
573 AT-rich sequences with genetic instability [52-54], which may form hot spots for rearrangement in the
574 eukaryotic genome in normal and diseased cells. The exploration of spontaneous TRP/INVs in other
575 genetically manipulatable organisms using the methods described above, will be a valuable future
576 direction to test the generality of this finding.
577
578 TRP/INVs are associated with genetic diseases. Brewer et al. (2024) conducted an extensive study
579 that demonstrated that many human disease-related loci had structures consistent with the involvement of
580 TRP/INVs, and these loci were often flanked by direct repeats [14]. Other laboratories have also
581 demonstrated that these structures are one of the most prevalent mutational signatures in cancer and a
582 common source of disease-related genomic instability [55, 56]. Understanding the fundamental
583 mechanisms of TRP/INV formation in the absence of selection will be critical to understand the underlying
584 basis of genetic instability that precedes the chromosomal amplifications found in these disease states.
585
586 **MATERIALS AND METHODS**
587 **Strains, media, and growth conditions**
588 G0 spore stocks of *Aspergillus flavus* NRRL 3357 were generated following nine-point
589 inoculations of a diluted $1x10^3$/ml spore stock onto the surface of double-strength 5/2 agar [(per liter: 50
590 mL V8 juice, 40 g agar, pH 5.2 [57, 58]] and incubation at 30 °C in the light for five days, a condition
591 that promotes conidiation. NRRL 3357 conidia were collected from plates by gentle scraping of the agar
592 surface with 10 ml sterile 0.01% Triton X-100, then transferred to a sterile 15 ml screwcap polypropylene
593 tube, and stored at 4°C. For forced vegetative growth studies, addition of 1 ml (about $1x10^8$) of the spore

594 suspension into 100 ml of A&M medium (per liter: sucrose, 50 g; $(NH_4)_2SO_4$, 3 g; KH_2PO_4, 10 g; $MgSO_4$
595 $\cdot 7H_2O$, 2 g; $Na_2B_4O_7 \cdot 10H_2O$, 0.7 mg; $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 0.5 mg; $Fe_2(SO_4)_3 \cdot 6H_2O$, 10 mg; $CuSO_4 \cdot$
596 $5H_2O$, 0.3 mg; $MnSO_4 \cdot H_2O$, 0.11 mg; and $ZnSO_4 \cdot 7H_20$, 17.6 mg.; pH 6.5) and incubation at 28 °C, at
597 200 rpm, in the dark for seven days, led to generation 1 of the experiment (G1). After 7 days (d), 20 ml of
598 the G1 culture was removed from the 100 ml A&M shake culture and placed in a sterile 50 ml screw cap,
599 polypropylene tube, then macerated for 30 sec using the Tekmar Tissuemizer (Tekmar Co., Cincinnati,
600 OH). One ml of macerated G1 mycelia was transferred into fresh 100 ml A&M broth in a sterile 250 ml
601 Erlenmeyer flask and incubated for another 7 d. This culture represented the second "generation" (G2) of
602 vegetative growth. This process was continued for 15 cycles (15 weeks; 105 d). We diluted a 10 µl
603 aliquot of macerated mycelia in 990 µl sterile water, spread 10 and 100 µl aliquots onto the surface of
604 double-strength 5/2 medium, and incubated for three days to observe colony phenotype. Mycelia for
605 genomic DNA preparation from 5 (G5), 10 (G10), and 15-week (G15) cultures of NRRL 3357 mycelia
606 were collected as follows. At the end of the desired growth period, 50 ml of potato dextrose broth (PDB,
607 Difco) was inoculated with 0.5 ml of macerated mycelia and incubated for 24 h at 28 °C with agitation
608 (200 rpm). For NRRL 3357 generation 0 (G0), genomic DNA was prepared from a 24 h 50 ml PDB
609 culture that had been inoculated with 0.5 ml of a $1x10^8$ spores/ml NRRL 3357 stock. Mycelia were
610 collected from the 50 ml PDB shake cultures by vacuum filtration through sterile miracloth (EMD
611 Millipore, Temecula, CA). The remaining liquid was squeezed from the collected mycelial mat by
612 pressing it between paper towels. Mycelia were lyophilized for at least 24 h and stored at -80 °C until
613 needed.
614
615 **Preparation of genomic DNA**
616 Genomic DNA was prepared from lyophilized mycelia using the Qiagen Plant Mini kit (Qiagen,
617 Beverly, MA) following the manufacturer's protocol. Lyophilized mycelia were ground to a fine powder
618 in liquid nitrogen and stored at -80 °C until needed. Briefly, 15 mg of each ground sample was
619 processed in duplicate. After a spin column, the DNA was eluted with two 100 µl elutions using Qiagen

620 AE buffer. The DNA suspension was further enriched by the addition of an equal volume of 25:24:1
621 phenol: chloroform: isoamyl alcohol (IAA). The upper layer was collected and extracted with an equal
622 volume of 24:1 chloroform: IAA, and then precipitated by the addition of one-tenth volume of 3M
623 sodium acetate pH 5.2 solution and two volumes of cold 95% ethanol followed by centrifuging at 10K
624 rpm for 10 min. The DNA pellet was washed with 70 % ethanol, dried in a speed vac, and resuspended in
625 50 µl TE. DNA quantity and quality were determined using a NanoDrop OneC (Thermo Scientific). The
626 quality of genomic DNA was also determined by electrophoresis on a 0.7% agarose gel.
627
628 **Sequencing**
629 CD Genomics (Shirley, NY) carried out all sequencing. Both sequencing methods gave an average of
630 500-1500-fold coverage, although not equally distributed over all parts of the genome.
631
632 **Nanopore sequencing:** The library was prepared using the SQK-LSK112 ligation kit under
633 standard conditions. The purified library was loaded onto primed R10.4 Spot-On Flow Cells and
634 sequenced using a PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK) with 48-h runs
635 [30]. Base-calling of raw data used the Oxford Nanopore Dorado v 0.7.3
636 (https://github.com/nanoporetech/dorado/releases).
637
638 **PacBio HiFi Sequencing:** The SMRTbell library was constructed using the SMRTbell Express
639 Template Prep kit 2.0 (Pacific Biosciences). Briefly, ~2µg of the genomic DNA was carried into the first
640 enzymatic reaction to remove single-stranded overhangs followed by treatment with repair enzymes to
641 repair any damage that may be present on the DNA backbone. After DNA damage repair, ends of the
642 double-stranded fragments were polished and subsequently tailed with an A-overhang. Ligation with T-
643 overhang SMRTbell adapters was performed at 20° C for 60 minutes. Following ligation, the SMRTbell
644 library was purified with 1X AMPure PB beads. The size distribution and concentration of the library
645 were assessed using the FEMTO Pulse automated pulsed-field capillary electrophoresis (Agilent

646 Technologies, Wilmington, DE) and the Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA).
647 Following library characterization, 3µg was subjected to a size selection step using the BluePippin system
648 (Sage Science, Beverly, MA) to remove SMRTbells ≤15 kb. After size selection, the library was purified
649 with 1X AMPure PB beads. Library size and quantity were assessed using the FEMTO Pulse and the
650 Qubit dsDNA HS reagents Assay kits. The sequencing primer and Sequel II DNA Polymerase were
651 annealed and bound, respectively, to the final SMRTbell library. The library was sequenced using 8M
652 SMRT Cell on the Sequel II System with Sequel II Sequencing Kit [34].
653
654 **Assemblies**
655 We bootstrapped G0 and G15 assemblies using both *A. flavus* strain NRRL-3357
656 (GCA_009017415.1_ASM901741v1) [23] and mitochondrial (*A. flavus* SRRC1009) [59] sequences that
657 were concatenated to prepare a genome fasta file. This file served as input for G0 ONT or G15 PacBio
658 reads to Rebaler (https://github.com/rrwick/Rebaler) [60]. G0 and G15 assemblies, stored in the Sequence
659 Server Cloud, are available upon request.
660
661 **Structural variant determination**
662 Structural variants were identified relative to the G0 assembly using Sniffles 2 (version 2.2) [32].
663 Alignments used minimap 2 [33] (version 2.26-r1175) with a gap open and gap extension set at 16 and 2,
664 respectively. This version was determined empirically to be only ~70% efficient for low abundance
665 sequences in simple sequence regions, leading to both false positives and negatives. We therefore often
666 coupled the algorithm with visual confirmation using IGV (version 2. (10.0)) [61, 62].
667
668 For qualitative analyses, structural variants were identified as ≥ 300 bp insertions with variant
669 read numbers (DV) of 3 and confirmed by IGV analysis. Individual reads from these nanopore sequenced
670 DNA were then used as BLAST queries to the G15 or G0 assemblies using SequenceServer (London,
671 UK) [62]. We analyzed simple sequence organization for AT-rich sequences by removing the dust filter.

672

673 For quantitative analysis, we selected all reads that had a DV of 1 and individual reads from these
674 nanopore sequenced DNAs were then analyzed by BLAST as noted above. A similar analysis was carried
675 out for PacBio insertions of ≥ 200 bp. For all studies, the variant frequency (VAF) calculated as (variant
676 read number (DV)/reference read number (DR) x 100). VAF values ranged from 0.1 to 0.2 %,consistent
677 with a high coverage.

678

679 The most frequent products identified by BLAST analysis among large "insertions" were
680 duplications and tandem inverted duplications. The TRP/INV species were categorized based on the
681 extent of deletion at each junction between duplicated and inverted sequences. We established 50 bp as
682 the maximum error at a junction, likely due to sequencing and alignment defects.

683

684 **Palindrome analysis**

685 Palindromes were identified first from BLAST analysis of single reads with the G15 assembly by
686 the presence of sequences at inverted sequence junctions homologous to both strands of the genomic
687 sequence, representing the palindromes that initiate TRP/INV formation. We determined the predicted
688 relative thermodynamic stability (T_m, ΔG, ΔH, and ΔS) and secondary structure predictions of the
689 palindromes by the application of UNAfold mfold 4.2 program [35-37] using our default settings of 25°C,
690 3 mM $MgCl_2$ and 50mM NaCl. The stability of palindrome structures was routinely quantified in the
691 presence of short flanking single strand regions for uniform alignment, although the listed coordinates
692 exclude these tails. As a control, we characterized the palindromes at the junctions of inverted repeats
693 formed in a yeast model system [7]. The UNAFold® software (RPI 2017-042) is under license 80-00840
694 from Rensselaer Polytechnic Institute. We also analyzed the potential palindromes in the 200 bp of
695 genomic sequence surrounding the incipient site of recombination in TRP/INVs exhibiting deletions of ≤
696 40 bp.

697

698 The UNAfold algorithm most often generated simple hairpin predictions. However, three types of
699 other pairing patterns were also observed. Structures in which the hairpin was closely adjacent to more
700 extensive base pairing were termed extended forms, while predicted patterns containing two or more
701 adjacent hairpins were termed multi-palindromic. Structures that were paired but appeared to contain
702 some form of distortion were termed complex forms. Examples of these forms are shown in SI Fig. 4.
703
704 Overall palindrome predictions for centromeric and non-centromeric were determined by
705 Palindrome Analyzer [39], using parameters exhibited by the majority of palindromes associated with the
706 TRP/INVs from centromeric or dispersed domains of chromosomes. The same parameters were then used
707 for the entire genome. We first determined that the most frequent classes of the terminal portions of
708 centromere 3 TRP/INV-associated palindromes. This class had stem sizes of 12-100 bp, loop sizes of 4-8
709 bp, with 0-2 stem mismatches. We then identified these partial structures in the output of the Palindrome
710 Analyzer [39] and analyzed them for the presence of part or all of the input palindrome structure. These
711 ΔG_{cf}-ΔG_{lin} values generally ranged from 10-16 kcal/mol. Since the distribution was similar for the observed
712 and predicted palindromes, we determined the number of predicted palindromes as the sum of all candidates
713 within this range. However, this number was corrected for that structures that had stabilities lower than the
714 study threshold. For each centromere, we added candidates in the 10-16 kcal/mole range and adjusted these
715 for the expected number of false positives.
716
717 We could not apply to non-centromeric sequence due to the differing conditions identified in
718 TRP/INV associated in palindromes in AT-rich and dispersed regions. Examination non-centromeric
719 palindromes from chromosomes 1, 3 and 4 revealed stem lengths ranging from 5-12 bp, loops ranging from
720 4-19 bp with stem mismatches from 0-1. We empirically converted ΔG into ΔG_{cf}-ΔG_{lin} values by first
721 remeasuring ΔG at 37°C, 50mM NaCl, and 1mM $MgCl_{2.,}$ The values, representing the stability of a hairpin,
722 were then doubled to generate ΔG_{cf}-ΔG_{lin}, a measure of cruciform stability. ΔG_{cf}-ΔG_{lin} values ranging from
723 3-12 kcal/mol distributed as follows: 15 at 3-6 kcal/mol (Class 1), 10 at 6.1-9 kcal/mol (Class 2) and 14

724 from at kcal/mol (Class 3). Since the predicted output from Palindrome Analyzer for a 60 kb region of
725 centromere 3 had a different distribution (24 in class 1, 105 in class 2, and 507 in class 3) than the TRP/INV-
726 associated palindromes, we normalized the output values using a weighted eligibility model [40, 41].
727 Specifically, to estimate how many palindrome sites in a genomic interval are actually competent to initiate
728 recombination, we compared the observed distribution across palindrome classes with the relative
729 abundance of candidates. For each class i, we calculated a per-candidate usage rate $r_i = U_i/n_i$, where U_i is
730 the number of observed TRP/INV-associated palindromes and n_i is the number of candidate palindromes
731 in the interval. To account for the unequal usage across classes, we normalized by the maximum observed
732 rate, defining a relative eligibility fraction $h_i = r_i/\max_j r_j$. The estimated pool of competent candidates in
733 the interval is described by

734 $$A = \sum_i h_i n_i.$$

735

736 **Sequence Structure in Centromeric and Subtelomeric ATEs**

737 To ascertain the frequency of short repeats within ATEs, we identified 250 bp DNAs from
738 centromeric and subtelomeric ATEs having a 90% A+T content and assayed for self-homology by
739 BLAST analysis to identify direct, inverted, and palindromic repeats. We rated ungapped regions of
740 homology as a function of the BLAST e-value. We compared this distribution with 250 bp sequences
741 derived from an Excel random sequence generator (Microsoft) set at 90%A+T DNA. Both perfect and
742 mismatched repeats at each value were combined for the analysis of direct and inverted repeats. To
743 quantify the statistical significance of the deviation from nonrandom behavior, we evaluated the number
744 of sequences containing at least one event present at e-values of $<10^{-4}$ for genomic and random datasets
745 by the Fisher Exact Test. Perfectly matched palindromes were similarly analyzed at BLAST e values of
746 $<10^{-8}$ and evaluated by the Fisher Exact Test.

747

748 **DATA ACCESS**

The raw data from nanopore and PacBio sequencing are publicly available at NCBI SRA listed as BioProject PRJNA1193636 with the following BioSample accession numbers: SAMN45150363 (G0 Nanopore), SAMN45150364 (G5 Nanopore), SAMN45150365 (G10 Nanopore), SAMN45150366 (G15 Nanopore), SAMN45150367 (G15 PacBio-1), SAMN45150368 (G15 PacBio-2). The G0 and G15 assemblies are present in the Sequence Cloud Server and are available on request.

ACKNOWLEDGEMENTS

We would like to thank and Bonnie Brewer (U of Washington) Drs. Prescott Deininger (Tulane), Matthew Lebar (USDA), Brian Mack (USDA), and Bill Wimley (Tulane), for stimulating discussions during these studies. We would also like to give our appreciation to Melody Baddoo and for bioinformatic assistance. Drs. Mary J Clancy (UNO) and Bibo Li (Cleveland State University), and Bonnie Hoffman are thanked for critical reading of the manuscript.

Funding: A.J.L was funded by USDA NCA 58-6054-0-014. J.W.C was funded by USDA. E.F was funded by R01CA262090 and R01CA272142. The Next Generation Sequencing Analysis core is supported by the National Institutes of Health National Cancer Institute, LSUHSC Tumor Virology COBRE, and the Louisiana Cancer Research Consortium (LCRC).

Author contributions:

Conceptualization: AJL
Methodology: AJL, JWC
Investigation: JWC, AM, AS, RL, AJL
Visualization: AJL, AS
Supervision: AJL, EF
Writing—original draft: AJL
Writing—review & editing: AJL, JWC, EF, AM, AS

Competing Interests:

The authors declare no competing interests.

Data and materials availability

All data presented herein is available to the research community.

REFERENCES

1. Brown RE, Freudenreich CH. Structure-forming repeats and their impact on genome stability. Curr Opin Genet Dev. 2021;67:41-51. Epub 20201203. doi: 10.1016/j.gde.2020.10.006. PubMed PMID: 33279816; PubMed Central PMCID: PMCPMC8084909.
2. Reams AB, Roth JR. Mechanisms of gene duplication and amplification. Cold Spring Harb Perspect Biol. 2015;7(2):a016592. Epub 20150202. doi: 10.1101/cshperspect.a016592. PubMed PMID: 25646380; PubMed Central PMCID: PMCPMC4315931.
3. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. Annu Rev Genet. 2011;45:247-71. Epub 2011/09/14. doi: 10.1146/annurev-genet-110410-132435. PubMed PMID: 21910633.
4. Svetec Miklenic M, Svetec IK. Palindromes in DNA-A Risk for Genome Stability and Implications in Cancer. Int J Mol Sci. 2021;22(6). Epub 20210311. doi: 10.3390/ijms22062840. PubMed PMID: 33799581; PubMed Central PMCID: PMCPMC7999016.
5. Al-Zain AM, Nester MR, Ahmed I, Symington LS. Double-strand breaks induce inverted duplication chromosome rearrangements by a DNA polymerase delta-dependent mechanism. Nat Commun. 2023;14(1):7020. Epub 20231102. doi: 10.1038/s41467-023-42640-5. PubMed PMID: 37919272; PubMed Central PMCID: PMCPMC10622511.
6. Narayanan V, Mieczkowski PA, Kim HM, Petes TD, Lobachev KS. The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. Cell. 2006;125(7):1283-96. doi: 10.1016/j.cell.2006.04.042. PubMed PMID: 16814715.

801 7. Rattray AJ, Shafer BK, Neelam B, Strathern JN. A mechanism of palindromic gene
802 amplification in Saccharomyces cerevisiae. Genes Dev. 2005;19(11):1390-9. doi:
803 10.1101/gad.1315805. PubMed PMID: 15937224; PubMed Central PMCID: PMCPMC1142561.
804 8. Kugelberg E, Kofoid E, Andersson DI, Lu Y, Mellor J, Roth FP, et al. The tandem
805 inversion duplication in Salmonella enterica: selection drives unstable precursors to final
806 mutation types. Genetics. 2010;185(1):65-80. Epub 20100309. doi:
807 10.1534/genetics.110.114074. PubMed PMID: 20215473; PubMed Central PMCID:
808 PMCPMC2870977.
809 9. Reams AB, Kofoid E, Kugelberg E, Roth JR. Multiple pathways of duplication formation
810 with and without recombination (RecA) in Salmonella enterica. Genetics. 2012;192(2):397-415.
811 Epub 20120803. doi: 10.1534/genetics.112.142570. PubMed PMID: 22865732; PubMed Central
812 PMCID: PMCPMC3454872.
813 10. Brewer BJ, Payen C, Raghuraman MK, Dunham MJ. Origin-dependent inverted-repeat
814 amplification: a replication-based model for generating palindromic amplicons. PLoS Genet.
815 2011;7(3):e1002016. Epub 20110317. doi: 10.1371/journal.pgen.1002016. PubMed PMID:
816 21437266; PubMed Central PMCID: PMCPMC3060070.
817 11. Martin R, Espinoza CY, Large CRL, Rosswork J, Van Bruinisse C, Miller AW, et al.
818 Template switching between the leading and lagging strands at replication forks generates
819 inverted copy number variants through hairpin-capped extrachromosomal DNA. PLoS Genet.
820 2024;20(1):e1010850. Epub 20240104. doi: 10.1371/journal.pgen.1010850. PubMed PMID:
821 38175823; PubMed Central PMCID: PMCPMC10766183.
822 12. Zeman MK, Cimprich KA. Causes and consequences of replication stress. Nat Cell Biol.
823 2013;16(1):2-9. doi: 10.1038/ncb2897. PubMed PMID: 24366029.

824 13. Voineagu I, Narayanan V, Lobachev KS, Mirkin SM. Replication stalling at unstable
825 inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. Proc Natl Acad
826 Sci U S A. 2008;105(29):9936-41. Epub 20080715. doi: 10.1073/pnas.0804510105. PubMed
827 PMID: 18632578; PubMed Central PMCID: PMCPMC2481305.
828 14. Brewer BJ, Dunham MJ, Raghuraman MK. A unifying model that explains the origins of
829 human inverted copy number variants. PLoS Genet. 2024;20(1):e1011091. Epub 20240104. doi:
830 10.1371/journal.pgen.1011091. PubMed PMID: 38175827; PubMed Central PMCID:
831 PMCPMC10766186.
832 15. Inagaki H, Ohye T, Kogo H, Yamada K, Kowa H, Shaikh TH, et al. Palindromic AT-rich
833 repeat in the NF1 gene is hypervariable in humans and evolutionarily conserved in primates.
834 Hum Mutat. 2005;26(4):332-42. doi: 10.1002/humu.20228. PubMed PMID: 16116616; PubMed
835 Central PMCID: PMCPMC2818517.
836 16. Kato T, Kurahashi H, Emanuel BS. Chromosomal translocations and palindromic AT-
837 rich repeats. Curr Opin Genet Dev. 2012;22(3):221-8. Epub 20120306. doi:
838 10.1016/j.gde.2012.02.004. PubMed PMID: 22402448; PubMed Central PMCID:
839 PMCPMC3378763.
840 17. Bloom K, Costanzo V. Centromere Structure and Function. Prog Mol Subcell Biol.
841 2017;56:515-39. doi: 10.1007/978-3-319-58592-5_21. PubMed PMID: 28840251; PubMed
842 Central PMCID: PMCPMC6535225.
843 18. Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. Centromeres of
844 filamentous fungi. Chromosome Res. 2012;20(5):635-56. doi: 10.1007/s10577-012-9290-3.
845 PubMed PMID: 22752455; PubMed Central PMCID: PMCPMC3409310.

846 19. Smith KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M. Heterochromatin is
847 required for normal distribution of Neurospora crassa CenH3. Mol Cell Biol. 2011;31(12):2528-
848 42. Epub 20110419. doi: 10.1128/MCB.01285-10. PubMed PMID: 21505064; PubMed Central
849 PMCID: PMCPMC3133421.
850 20. Testa AC, Oliver RP, Hane JK. OcculterCut: A Comprehensive Survey of AT-Rich
851 Regions in Fungal Genomes. Genome Biol Evol. 2016;8(6):2044-64. Epub 20160703. doi:
852 10.1093/gbe/evw121. PubMed PMID: 27289099; PubMed Central PMCID: PMCPMC4943192.
853 21. Cambareri EB, Aisner R, Carbon J. Structure of the chromosome VII centromere region
854 in Neurospora crassa: degenerate transposons and simple repeats. Mol Cell Biol.
855 1998;18(9):5465-77. doi: 10.1128/MCB.18.9.5465. PubMed PMID: 9710630; PubMed Central
856 PMCID: PMCPMC109131.
857 22. Lewis ZA, Honda S, Khlafallah TK, Jeffress JK, Freitag M, Mohn F, et al. Relics of
858 repeat-induced point mutation direct heterochromatin formation in Neurospora crassa. Genome
859 Res. 2009;19(3):427-37. Epub 20081217. doi: 10.1101/gr.086231.108. PubMed PMID:
860 19092133; PubMed Central PMCID: PMCPMC2661801.
861 23. Skerker JM, Pianalto KM, Mondo SJ, Yang K, Arkin AP, Keller NP, et al. Chromosome
862 assembled and annotated genome sequence of Aspergillus flavus NRRL 3357. G3 (Bethesda).
863 2021;11(8):1-5. doi: 10.1093/g3journal/jkab213. PubMed PMID: 34849826; PubMed Central
864 PMCID: PMCPMC8496237.
865 24. Lustig AJ. Investigating the origin of subtelomeric and centromeric AT-rich elements in
866 Aspergillus flavus. PLoS One. 2023;18(2):e0279148. Epub 20230209. doi:
867 10.1371/journal.pone.0279148. PubMed PMID: 36758027; PubMed Central PMCID:
868 PMCPMC9910759.

869 25. Rahnama M, Wang B, Dostart J, Novikova O, Yackzan D, Yackzan A, et al. Telomere
870 Roles in Fungal Genome Evolution and Adaptation. Front Genet. 2021;12:676751. Epub
871 20210809. doi: 10.3389/fgene.2021.676751. PubMed PMID: 34434216; PubMed Central
872 PMCID: PMCPMC8381367.
873 26. Kusumoto KI, Suzuki S, Kashiwagi Y. Telomeric repeat sequence of Aspergillus oryzae
874 consists of dodeca-nucleotides. Appl Microbiol Biotechnol. 2003;61(3):247-51. Epub 20030114.
875 doi: 10.1007/s00253-002-1193-3. PubMed PMID: 12698283.
876 27. Kondrashov FA, Kondrashov AS. Measurements of spontaneous rates of mutations in the
877 recent past and the near future. Philos Trans R Soc Lond B Biol Sci. 2010;365(1544):1169-76.
878 doi: 10.1098/rstb.2009.0286. PubMed PMID: 20308091; PubMed Central PMCID:
879 PMCPMC2871817.
880 28. Roper M, Ellison C, Taylor JW, Glass NL. Nuclear and genome dynamics in
881 multinucleate ascomycete fungi. Curr Biol. 2011;21(18):R786-93. doi:
882 10.1016/j.cub.2011.06.042. PubMed PMID: 21959169; PubMed Central PMCID:
883 PMCPMC3184236.
884 29. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
885 architecture and applications. BMC Bioinformatics. 2009;10:421. Epub 20091215. doi:
886 10.1186/1471-2105-10-421. PubMed PMID: 20003500; PubMed Central PMCID:
887 PMCPMC2803857.
888 30. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sorensen EA, Wollenberg RD, et
889 al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished
890 bacterial genomes from pure cultures and metagenomes without short-read or reference

891 polishing. Nat Methods. 2022;19(7):823-6. Epub 20220704. doi: 10.1038/s41592-022-01539-7.
892 PubMed PMID: 35789207; PubMed Central PMCID: PMCPMC9262707.
893 31. Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al.
894 Publisher Correction: Detection of mosaic and population-level structural variants with Sniffles2.
895 Nat Biotechnol. 2024. Epub 20240122. doi: 10.1038/s41587-024-02141-2. PubMed PMID:
896 38253882; PubMed Central PMCID: PMCPMC11260898.
897 32. Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, et al.
898 Detection of mosaic and population-level structural variants with Sniffles2. Nat Biotechnol.
899 2024. Epub 20240102. doi: 10.1038/s41587-023-02024-y. PubMed PMID: 38168980; PubMed
900 Central PMCID: PMCPMC11217151.
901 33. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics.
902 2018;34(18):3094-100. doi: 10.1093/bioinformatics/bty191. PubMed PMID: 29750242; PubMed
903 Central PMCID: PMCPMC6137996.
904 34. Hon T, Mars K, Young G, Tsai YC, Karalius JW, Landolin JM, et al. Highly accurate
905 long-read HiFi sequencing data for five complex genomes. Sci Data. 2020;7(1):399. Epub
906 20201117. doi: 10.1038/s41597-020-00743-4. PubMed PMID: 33203859; PubMed Central
907 PMCID: PMCPMC7673114.
908 35. Peyret N. Prediction of nucleic acid hybridization: Parameters and Algorithms PhD
909 Thesis, Department of Chemistry, Wayne State University, Detroit, MI. 2020.
910 36. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic
911 Acids Res. 2003;31(13):3406-15. doi: 10.1093/nar/gkg595. PubMed PMID: 12824337; PubMed
912 Central PMCID: PMCPMC169194.

913 37. SantaLucia J, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA
914 nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A. 1998;95(4):1460-5. doi:
915 10.1073/pnas.95.4.1460. PubMed PMID: 9465037; PubMed Central PMCID: PMCPMC19045.
916 38. Courey AJ, Wang JC. Cruciform formation in a negatively supercoiled DNA may be
917 kinetically forbidden under physiological conditions. Cell. 1983;33(3):817-29. doi:
918 10.1016/0092-8674(83)90024-7. PubMed PMID: 6871994.
919 39. Brazda V, Kolomaznik J, Lysek J, Haronikova L, Coufal J, St'astny J. Palindrome
920 analyser - A new web-based server for predicting and evaluating inverted repeats in nucleotide
921 sequences. Biochem Biophys Res Commun. 2016;478(4):1739-45. Epub 20160904. doi:
922 10.1016/j.bbrc.2016.09.015. PubMed PMID: 27603574.
923 40. Johnson DH. The Comparison of Usage and Availability Measurements for Evaluating
924 Resource Preference. Ecology. 1980;61(1):65-71. doi: Doi 10.2307/1937156. PubMed PMID:
925 WOS:A1980JY71100011.
926 41. Manly B, McDonald L, Thomas D, McDonald T, Erickson W. Resource Selection by
927 Animals: Statistical Design and Analysis for Field Studies2002.
928 42. Sfeir A, Symington LS. Microhomology-Mediated End Joining: A Back-up Survival
929 Mechanism or Dedicated Pathway? Trends Biochem Sci. 2015;40(11):701-14. Epub 20151001.
930 doi: 10.1016/j.tibs.2015.08.006. PubMed PMID: 26439531; PubMed Central PMCID:
931 PMCPMC4638128.
932 43. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining
933 and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017;18(8):495-
934 506. Epub 20170517. doi: 10.1038/nrm.2017.48. PubMed PMID: 28512351; PubMed Central
935 PMCID: PMCPMC7062608.

936 44. Lee K, Lee SE. Saccharomyces cerevisiae Sae2- and Tel1-dependent single-strand DNA
937 formation at DNA break promotes microhomology-mediated end joining. Genetics.
938 2007;176(4):2003-14. Epub 20070611. doi: 10.1534/genetics.107.076539. PubMed PMID:
939 17565964; PubMed Central PMCID: PMCPMC1950609.
940 45. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. PLoS
941 One. 2021;16(10):e0257521. Epub 20211001. doi: 10.1371/journal.pone.0257521. PubMed
942 PMID: 34597327; PubMed Central PMCID: PMCPMC8486125.
943 46. Decottignies A. Microhomology-mediated end joining in fission yeast is repressed by
944 pku70 and relies on genes involved in homologous recombination. Genetics. 2007;176(3):1403-
945 15. Epub 20070504. doi: 10.1534/genetics.107.071621. PubMed PMID: 17483423; PubMed
946 Central PMCID: PMCPMC1931558.
947 47. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function.
948 Nat Rev Mol Cell Biol. 2018;19(4):229-44. Epub 20171213. doi: 10.1038/nrm.2017.119.
949 PubMed PMID: 29235574; PubMed Central PMCID: PMCPMC6822695.
950 48. Stavenhagen JB, Zakian VA. Internal tracts of telomeric DNA act as silencers in
951 Saccharomyces cerevisiae. Genes Dev. 1994;8(12):1411-22. PubMed PMID: 7926741.
952 49. Sandell LL, Zakian VA. Telomeric position effect in yeast. Trends Cell Biol.
953 1992;2(1):10-4. Epub 1992/01/01. PubMed PMID: 14731632.
954 50. Rybchuk J, Xiao W. Dual activities of a silencing information regulator complex in yeast
955 transcriptional regulation and DNA-damage response. mLife. 2024;3(2):207-18. Epub 20240515.
956 doi: 10.1002/mlf2.12108. PubMed PMID: 38948145; PubMed Central PMCID:
957 PMCPMC11211678.

958 51. Wen M, Lan H, Sun R, Chen X, Zhang X, Zhu Z, et al. Histone deacetylase SirE
959 regulates development, DNA damage response and aflatoxin production in Aspergillus flavus.
960 Environ Microbiol. 2022;24(11):5596-610. Epub 20220916. doi: 10.1111/1462-2920.16198.
961 PubMed PMID: 36059183.
962 52. Zhang H, Freudenreich CH. An AT-rich sequence in human common fragile site
963 FRA16D causes fork stalling and chromosome breakage in S. cerevisiae. Mol Cell.
964 2007;27(3):367-79. doi: 10.1016/j.molcel.2007.06.012. PubMed PMID: 17679088; PubMed
965 Central PMCID: PMCPMC2144737.
966 53. Kaushal S, Wollmuth CE, Das K, Hile SE, Regan SB, Barnes RP, et al. Sequence and
967 Nuclease Requirements for Breakage and Healing of a Structure-Forming (AT)n Sequence
968 within Fragile Site FRA16D. Cell Rep. 2019;27(4):1151-64 e5. doi:
969 10.1016/j.celrep.2019.03.103. PubMed PMID: 31018130; PubMed Central PMCID:
970 PMCPMC6506224.
971 54. Irony-Tur Sinai M, Salamon A, Stanleigh N, Goldberg T, Weiss A, Wang YH, et al. AT-
972 dinucleotide rich sequences drive fragile site formation. Nucleic Acids Res. 2019;47(18):9685-
973 95. doi: 10.1093/nar/gkz689. PubMed PMID: 31410468; PubMed Central PMCID:
974 PMCPMC6765107.
975 55. Grochowski CM, Bengtsson JD, Du H, Gandhi M, Lun MY, Mehaffey MG, et al.
976 Inverted triplications formed by iterative template switches generate structural variant diversity
977 at genomic disorder loci. Cell Genom. 2024;4(7):100590. Epub 20240621. doi:
978 10.1016/j.xgen.2024.100590. PubMed PMID: 38908378; PubMed Central PMCID:
979 PMCPMC11293582.

980 56. Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of
981 somatic structural variation in human cancer genomes. Nature. 2020;578(7793):112-21. Epub
982 20200205. doi: 10.1038/s41586-019-1913-9. PubMed PMID: 32025012; PubMed Central
983 PMCID: PMCPMC7025897.
984 57. Chang PK, Cary JW, Bhatnagar D, Cleveland TE, Bennett JW, Linz JE, et al. Cloning of
985 the Aspergillus parasiticus apa-2 gene associated with the regulation of aflatoxin biosynthesis.
986 Appl Environ Microbiol. 1993;59(10):3273-9. doi: 10.1128/aem.59.10.3273-3279.1993. PubMed
987 PMID: 8250554; PubMed Central PMCID: PMCPMC182448.
988 58. Adye J, Mateles RI. Incorporation of Labelled Compounds into Aflatoxins. Biochim
989 Biophys Acta. 1964;86:418-20. doi: 10.1016/0304-4165(64)90077-7. PubMed PMID: 14171025.
990 59. Park J, Lee MK, Yu JH, Zhu B, Kim JH, Han KH. Complete mitochondrial genome
991 sequence of Aspergillus flavus SRRC1009: insight of intraspecific variations on A. flavus
992 mitochondrial genomes. Mitochondrial DNA B Resour. 2020;5(3):3585-7. Epub 20201021. doi:
993 10.1080/23802359.2020.1771226. PubMed PMID: 33367020; PubMed Central PMCID:
994 PMCPMC7594760.
995 60. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for
996 Oxford Nanopore sequencing. Genome Biol. 2019;20(1):129. Epub 20190624. doi:
997 10.1186/s13059-019-1727-y. PubMed PMID: 31234903; PubMed Central PMCID:
998 PMCPMC6591954.
999 61. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-
1000 performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.
1001 Epub 20120419. doi: 10.1093/bib/bbs017. PubMed PMID: 22517427; PubMed Central PMCID:
1002 PMCPMC3603213.

1003 62. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: A
1004 Modern Graphical User Interface for Custom BLAST Databases. Mol Biol Evol.
1005 2019;36(12):2922-4. doi: 10.1093/molbev/msz185. PubMed PMID: 31411700; PubMed Central
1006 PMCID: PMCPMC6878946.
1007
1008
1009
1010