Clustering of Tandem Inverse Duplications in Centromeric and Subtelomeric Chromosomal Regions of *Aspergillus flavus*

Short Title: Duplication/Inversions in Centromeres and Subtelomeres

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

[1]Department of Biochemistry and Molecular Biology

[3]Department of Pathology

Tulane School of Medicine

Tulane Cancer Center

New Orleans, LA 70118

[2]Food and Feed Safety Research Unit

USDA/ARS

Southern Regional Research Center

New Orleans, LA, United States

ABSTRACT

Rearrangements associated with gene amplification are frequently initiated by the formation of inverted repeats. Among these rearrangements are breakage-fusion-bridge cycles and tandem inverted duplications (TIDs), both of which can be mediated through palindromic sequences. We have investigated spontaneous rearrangements within the genome of the filamentous fungus *Aspergillus flavus* by high-coverage third-generation sequencing of vegetatively cultured strains. Analysis of sequence data identified a class of rare and transient rearrangements having structures typical of TIDs. These putative TIDs (pTIDs) form both unprocessed and processed species and have a variably sized deletion at the junction between direct and inverted sequences. Most intact junctions contain the predicted palindromic sequences, supporting the TID model. However, palindrome-free junctions are also infrequently observed. pTIDs are highly 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. We propose a replication-based model in which palindromes and short inverted repeats, under condition of replication stress, are responsible for pTID structures.

TEASER

The dynamic formation of duplication/inversions clustered in fungal chromosomes reveals the genetic instability of centromeric and subtelomeric regions**.**

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 the oncogenic state (*2, 3*).

Model systems are a valuable tool 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 perfectly and imperfectly paired 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 the loss of a telomere, followed by fusion of DNA strands through processes such as non-homologous end joining (*3, 5-7*). Alternatively, BFBs can form through processing of palindromic cruciform or hairpin DNA structures. Tandem inversion duplications (TIDs) are another class of events mediated by the rearrangement of relatively short regions of homology (*8-10*). These latter events are observed in bacteria and yeast and can form through origin- or DNA repair-mediated replication. After DNA damage, palindromes that flank a gene can create a hairpin that primes DNA replication. After reaching a second palindrome, template switching and DNA replication give rise to the duplication. The junctions between direct and inverted repeats reform the palindromes to form symmetric TIDs, which may be resolved into asymmetric forms containing deletions at the palindromic junctional sequence. This palindrome-initiated process can also create replicative stress and fork regression during bi-directional DNA replication (*11, 12*). On a statistical basis, short palindromes are expected to form most easily in AT or GC-rich sequences. AT-rich palindromes form frequently

in the human genome and are involved in both chromosomal translocation and evolutionary hypervariability (*13, 14*). Such events have been observed typically under selection. However, the array of rearrangements that may occur in the absence of any selection or in regions of simple sequence non-genic DNA have not been investigated.

Centromeric and telomeric regions of chromosomes are essential for the stable inheritance of genetic information. At the microtubule attachment site, centromeric domains can be present as point or regional structures, the latter including chromosome-specific satellite sequences (*15*). In *Neurospora* and probably other filamentous fungi, centromeres appear to be embedded within heterochromatic A+T-rich sequences, modeled in part by a centromere-specific histone H3 variant (*16, 17*). These centromeres are thought to be generated through repeat-induced point mutation (RIP) (*18-20*). In another filamentous fungus, *Aspergillus (A.) flavus*, centromeres consist of approximately 100 kb of 90-95% % A+T rich sequence on each of the eight chromosomes (*21, 22*). Generally, centromere sequences on homologs in different strains share extensive (88-99%), but not complete, homology (*22*). In contrast, the AT-rich sequences of non-homologs share minimal homology. Competing forces appear to maintain a balance between variation and conservation.

Rearrangements near fungal telomeres are thought to drive genetic reshuffling and adaptive polymorphisms in, and potentially the elimination of, gene sequences (*21, 23*). The subtelomeric 100 kb, adjacent to the telomeric $(TTAGGGTCAACA)_{9\text{-}11}$ repeats (*24*), is also interspersed with A+T-rich elements (ATEs), from 2-20 kb in length, that are distributed in different configurations among the chromosomal arms within and between strains. Our previous

studies have revealed remnant transposable elements in subtelomeres, suggesting that repetitive cycles of transposition and RIP contribute to the evolution of these sequences (*22*). Nonetheless, we found highly conserved ATEs that lack any homology to transposable elements on both the same chromosome of different strains and distinct chromosomes of the same strain. Hence, subtelomeric ATEs are also maintained in a homeostasis of divergence and conservation indicating a process that acts to promote recombination. The processes acting to promote and to limit ATE destabilization are not understood.

Identifying spontaneous rearrangements using standard genetic techniques has been challenging, due to their low frequencies and the limited means to identify complex forms (*25*). We have employed a combination of two approaches to examine spontaneous events. First, the multinucleate mycelial growth of *A. flavus* leads to the potential identification of rare deleterious mutations and rearrangements (*26*). Selection would tend to eliminate such species. Second, we used analysis of high-coverage third-generation sequencing as a physical readout of rare genetic rearrangements. Here, we report on the identification of novel structural variants that have a structure consistent with tandem inverted duplications clustered in centromeric and subtelomeric regions. These results suggest a dynamic equilibrium between factors governing the stability of these simple sequences.

RESULTS

A+T Rich Elements in Centromeres and Subtelomeric Regions Have a Nonrandom Distribution of Direct, Inverted, and Palindromic Repeats

To ascertain the frequency with which *A. flavus* AT-rich centromeric regions contain short repeated sequences, we examined random 250 bp DNAs for BLAST self-homology and identified potential direct, inverted, and palindromic repeats. We plotted the BLAST e-values (*27*) of each class of repeats in each of 129 randomly selected 250 bp, 90% A+T-rich sequences from centromeric regions (*27*). This result was then compared to repeats identified among 500 computationally derived and random 250 bp, 90% A+T-rich sequences (Table 1).

Sequence	**Repeat**	**BLAST e value**	Events **≥ 1**	**Events 0**	**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

Table 1. **Direct, Inverted and Palindromic Repeats in Centromeric, Subtelomeric and Randomly Generated A+T Rich tracts.** The sequence source, repeat type, BLAST e-value, and number of samples assayed containing either ≥ 1 or 0 repeat are listed together with the p values from a comparison of the genomic and random sequences as determined by the Fisher Exact Test (FET).

The most significant inverted and direct repeats (e-values $<10^{-4}$) identified in either genomic or randomly generated sequences were compared by the Fischer Exact Test. We determined that the direct repeats within centromeres were at significantly higher values than those generated from random sequences ($p<0.0001$). We further identified a significantly higher level of inverted repeats ($p<0.001$). Due to the higher spontaneous level of palindromic sequences in AT-rich sequences, we compared centromeric and random sequences at BLAST e-values of $<10^{-8}$ and found a significant increase in the number of perfect palindromes ($p<0.003$) in centromeres. Palindromic sequences were found in 6.9% of centromeres sequences that were tested.

We conducted an analogous test among eighty, 250 bp sequences extracted from subtelomeric ATEs defined as the 100 kb adjacent to the telomeric $(TTAGGGTCAACA)_{9\text{-}11}$ repeats (*21*, *23*) (Table 1). Direct and inverted repeats were quantified in this region and compared with randomized sequences. Both sets of repeats with e values $<10^{-4}$ were present at elevated levels in the subtelomere compared with random sequences (direct repeats: $p<0.002$; inverted repeats: $p<0.002$). Similarly, perfect palindromic sequences were again present more frequently in subtelomeric sequence than in random sequence datasets ($p<0.003$) and were present in 7.5% of the subtelomeric sequences. Hence, all three classes of repeats were significantly elevated in centromeric and subtelomeric domains, suggesting a selective process that generates this nonrandom behavior.

Identifications of the pTID Class

We were interested in identifying spontaneous rearrangements that may help to explain the nonrandom characteristic of repeated elements in subtelomeric and centromeric ATE sequences. To this end, we carried out high coverage (500-1500 fold) third-generation Oxford nanopore (ONT) sequencing as a readout for rare structural variants (*28*) after forced vegetative culturing. We optimized the Sniffles 2 and minimap2 programs (*29-31*) for rare structural variants and categorized them into apparent deletions, insertions, inversions, duplications, and translocations. We specifically characterized individual sequencing reads from the insertion class that predict sizes of ≥ 300 bp. BLAST analysis of these larger insertion sequences revealed two structural classes: intact tandem duplications and tandem duplications interrupted by a partial or complete inversion. This study focused on structures consistent with those described in gene amplification studies as tandem inversion duplication [(TID, (*8-10*)]. These events will be termed here as putative tandem inversion duplications (pTIDs).

Our initial qualitative analysis compared this subset of the insertion class of structural variants (SVs) that appeared to arise in clusters from *A. flavus* genomic DNA prepared from a starter culture (G0) (generated from spores isolated from conidiophores) after five (G5), ten (G10), and fifteen (G15) weeks of forced vegetative (mycelial) growth. One of the major SV classes had the pTID structures shown in Fig. 1 (top). Nine events were present in both ATEs and at dispersed locations throughout the genome. When compared to all single-read pTID candidates derived from other cultures (G5, G10, G15), we found each pTID was only found at the indicated stage of growth. Hence, within the limits of detection, pTIDs did not become fixed in the population, and appeared to arise only transiently during growth (Fig. 1, bottom).

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

Fig. 1. **Initial Analysis of pTIDs from Large Clustered Insertions in G5, G10 and G15 Cells.** DNAs were isolated and sequenced after *A. flavus* NRRL3357 cells had been cultured for 5 (G5), 10 (G10), and 15 (G15) cycles, and subsequently screened for large (≥300 bp) insertions using the Sniffles 2 algorithm as described in Materials and Methods with a variant read number of 3 and characterized further by IGV analysis. Individual reads were then analyzed by BLASTing to the G0 assembly to identify the pTIDs shown here. The listed coordinates are the approximate sites of pTID formation relative to the chromosomal (CHR) orientation. The yellow boxes refer to centromeric pTIDs, while the orange boxes refer to subtelomeric pTIDs and pTID within ATE elements scattered at other sites in the genome. Uncolored boxes refer to pTIDs dispersed at non-ATE sites. The structure of the unprocessed (symmetric) pTID is shown on top, with arrows below figure representing the relative orientations of pTID direct and inverted repeats and arrows above figure representing the typical palindromic sequences at these sites. The corresponding sizes of the direct (blue) and inverted (red) repeats are shown below.

To analyze these events quantitatively, we conducted an exhaustive study of all G15 SVs identified as ≥ 300 bp insertions (compared to the G0 assembly) that also were associated with a linked inversion. We characterized each of these reads and identified those having a pTID configuration. These 221 pTID events were distributed over all eight chromosomes of *A. flavus* (Table 2, Table S1). The low abundance and transitory nature of these events were further examined by quantifying all the chromosome 1 pTIDs in DNA derived from G0, G5, G10 and G15. We found three cases of heritable sites of pTID. In each case, the pTID structure was unique (Fig S1). We cannot rule out, however, the possibility that each of these had a common origin that subsequently underwent multiple rounds of rearrangement. These data reinforce the transient nature of pTIDs.

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

Table 2**. Chromosomal Distribution of G15 pTIDs Among Chromosomes**

Direct BLAST search of individual sequencing reads to the G0 telomere-to-telomere assembly revealed three pTID subtypes (Fig. 2). Ninety one events fall into the first class (termed “symmetric”; Fig. 2, top) and have intact head-to-tail configurations at both inversion junctions within the margin of error of sequencing and sequence alignment (Δ0-50 bp). One hundred six events fall into the second class (termed “asymmetric 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, 23 events belong to a third class (termed “asymmetric class 3”; Fig. 2, bottom) consisting of pTIDs containing a > 50 bp loss of sequence at both inversion junctions. 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 pTID formation (data not shown).

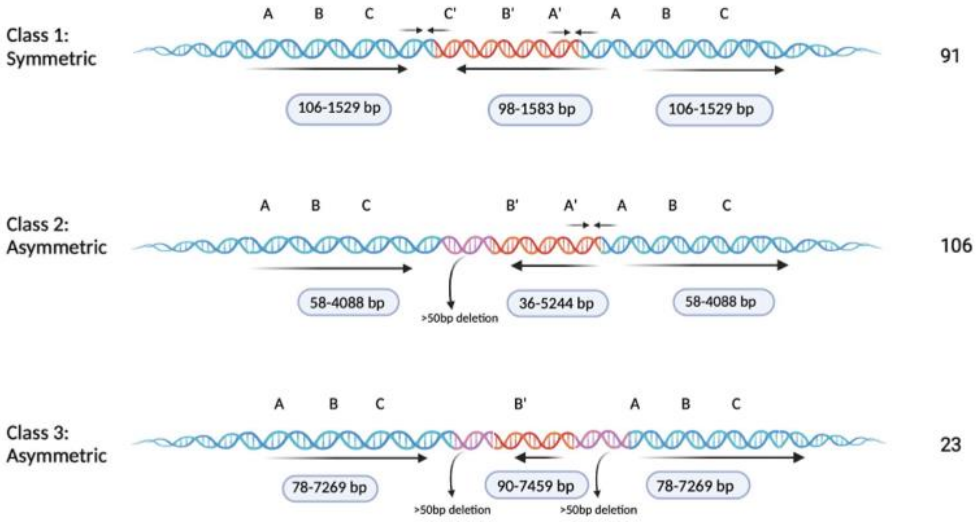

Fig. 2. **Three Structural Classes of pTIDs.** The three classes of pTIDs deduced by BLAST analyses of nanopore sequencing queries to genomic assemblies are presented here. The first

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 events 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.

Clustering of pTID at Centromeric and Subtelomeric ATEs

We analyzed the pTID distribution to determine their potential enrichment in specific chromosomal domains. We found a 17.5-22-fold higher frequency of pTIDs 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 appears to distribute to both coding (101/145) and non-coding (44/145) regions (Fig. S2A-S2D; Table S1).

A

Location	# pTIDs	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

B

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

Table 3**. Enrichment and Characterization of pTIDs in Centromeric and Subtelomeric ATEs**. A) The chromosomal location, number, and fraction of pTIDs in ATEs (centromeric, subtelomeric, and ATEs located elsewhere in the genome) and non-ATE (dispersed) regions, scattered among all chromosomes, are listed (*22*). B) The direct and inverted repeat sizes (together with standard deviations) of pTIDs located in centromeric and non-centromeric ATEs (subtelomeres + other ATEs) or dispersed non-ATEs are shown.

Analysis of the size of the direct and inverted pTID sequences (Table 3B) revealed that the range of values was reduced significantly (as shown by the standard deviation) in centromeres and subtelomeric (and other) ATEs, relative to dispersed non-ATE sequences. This reflects a tendency towards longer lengths in both the direct and inverted duplications in non-ATE pTIDs. For example, the lengths of such duplications extend up to 7.4 kb, compared to the maximum size of 0.92 kb at centromeres. Both the direct and inverted duplications in pTIDs also trended to slightly smaller sizes in centromeric compared to non-ATE domains. These data suggest that selective forces may well restrict centromeric pTID lengths.

We compared the pTID patterns obtained from the analysis of the nanopore sequencing with the results from two independently PacBio (ResBio) (*32*) sequenced G15 DNAs to ascertain the influence of sequence methodologies on the pTID generation (Table S2). We observed qualitatively similar results with both sequencing methods, although pTID levels were roughly 30-fold higher in ONT than in PacBio sequencing. Seven pTIDs were present, all at unique centromeric and subtelomeric sites. While substantiating the formation of pTIDs and their biased distribution towards AT-rich regions, these results suggest the presence of uncharacterized methodological biases in the library formation or sequencing of DNA containing pTIDs (see Discussion).

Identification of Palindrome Sequences at the Inverted/Direct Repeat Junction

Previous studies revealed that TIDs of bacterial and yeast genes are formed via palindrome formation and template switching during DNA repair or bi-directional DNA replication that recreates the palindromic sequence in the symmetric product (*8-10*). To test the prediction that pTIDs retain palindromes at inversion junctions, we sought to identify junctional read sequences that were complimentary to both strands of the genomic assembly sequence by BLAST analyses (examples are shown in Fig. S3A-S3D). The palindromes were then screened further for their predicted thermodynamic stabilities using the UNAfold mfold 4.2 program (*33-35*) (Table 4A, 4B). Seventy two of the 142 Class 1 and Class 2 junctions (Δ0-50bp) had palindromic characteristics.

Chr	Coord	Sequence	Length (bp)	Δ G ($K^{-1}mol^{-1}$)	T_m (°C)
1	2265912	CCGCTGGCTGCGG	13	-4.52	71.2
1	2286434	TCCCATGACGTCATGGGA	18	-7.27	73.1
1	5016740	AACGTTATAACGTT	14	-4	60
2	365195	ATTAGTATCTGATACTAAT	19	-3.98	51.6
2	4274226	GTAATGTCCTCAGCGGTGCGGACGTTAC	28	-5.82	55
2	5407930	CGACGGCGACAGCCTCGCCGTCG	23	-12.7	85.9
2	3840594	CCGGGTAATCAGATATCCGG	20	-3.3	54.9
2	1804843	CATTTCACCACGGTGAACTGTGGTGAATTG	30	-9.3	61.8
2	454494	AAGCTTTCATTCAGTAGTATAATCAAAGCTT	30	-4.05	50
2	493401	CGGCGAATTTGGCCG	15	-3.47	62.2
2	4115827	AGTTACATACAGGGAGTACAGTAGAGTACTAATGTAACT	39	-4.05	50
3	1734385	GAGAGAAGTTATTTTTCTTTTTTTTTTTTTAAGAAAAAAAGTAAGATAA	51	-5.66	45.2
3	1572690	GATCGACGGAAGATGTAACGTCGATC	26	-8.38	68.8
3	3837633	ACCTTATCACGGTAAGGT	18	-4	52.2
3	3287705	TTCATTGAGGCAATGAA	17	-4.21	55
3	567261	TGTCGACGTCGACA	14	-4.75	61.1
3	1526908	ATTCCGTATCAATACGGAAT	20	-7.4	66.7
3	615936	CAAAGTCCCGGAGCCGTCTCACTGGACTTTG	31	-7.85	56.8
4	4239136	GTAGACTGGGCGCCCAGTCTAC	22	-9.27	69.6
4	794961	AACGAGACTGATCCATGTCGCGTT	24	-4.74	55.4
4	1275195	GGACCATGAGCTCATGCGACATGTTGTGCAGCTAGGTCC	41	-5.58	45.9
4	4414808	TCGGTAGAGCCTCTACCGA	19	-7.39	72.1
4	1194513	CAGGCGAAGCGCTGCAGCCTG	21	-4.15	60.3
4	2527891	ATACTCCGGTATGCAATGCAATGCGGAGTAT	31	-7.83	56.5
4	4238843	L:ATCCCCCAATGCGGGGAT	18	-4.9	78
		R:TCTGCTGCAGCGCCTGCAGACGCTCGCGCAAGA	33	-7.62	61.1
5	618809	TGTCAAAATTTCATGAAATTTTTGACA	27	-6.54	53.3
5	4111924	TCATCACTTGATGA	14	-3.28	54.9
5	3446394	CATCGCGACGATGATG	16	-3.79	59.4
6	1661374	GTCCTCGCAATGTTACATCGAGGAC	25	-6.56	65.8
6	2523037	ACTAGGAGTACCTAGT	16	-4.58	57.6
6	1167658	TTTATAAATTTTGATATTTATAAA	24	-4.1	46.7
7	1656240	GTACGGTTTTCGGACTTGGGGGATTTGAAACCCATAGCGAAAACGGTAC	49	-11.39	56.2
8	1462561	CCCTGCAAGATGTGGCAGGG	20	-5.99	70.6
8	1428248	GAAGAACCAGACGATAGTTCGGTTCTTC	28	-7.51	59.3

Table 4A. **Predicted Non-ATE Palindromic pTID Junctions.** The chromosomal coordinates, sequences, length and predicted ΔG and T_m values of palindromes at inverted repeat junctions of non-ATE pTIDs that have a minimal defined stability (≤ -3 kcal/mole; $T_m \geq 40$°C) are presented. L and R refer to the two inverted junctions at a single pTID.

The identified palindromes were sorted among subtelomeric, centromeric, and dispersed non-ATE sequences (Table 5, middle column). Palindromes ranged in size from 13-51 bp in non-ATE sequences and 21-143 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 -55 kcal/mmol) and T_m values of at least 40°C (ranging from 40-78° C) (Table 5, right column).

Chr	Coord	Sequence	Length	ΔG ($K^{-1}m^{-1}$)	T_m (°C)
1	77041	TTAATATTTTAGTTATTTATATAATTATATAAAAATAATAATATTAA	47bp	-5.67	41.9
1	4510493	TAATTTTAATTATATAATAAAAATTA	26	-4	48.4
2	2064565	TTATAATAAATATTATATAATAATATTATTTAA	33	-3.94	51.7
3	2693002	TTATAATTCTATATAATTATAA	22	-3.2	44.2
3	2739469	TAGCTTTAAAAAATAAAATATTTAATTTTATAAACTA	37	-4.11	54
4	2732735	L: CTATAACTTATAAGTTATAG	20	-5.31	54.8
4	2818293	TAGATAAAATATATAATTAAAATTAATATATATTAAATCTA	41	-5.32	44.5
4	2865070	ATATTTATATATTTTAAAAAATATTATATTTTATTATAAAAAAATAATTAAATATAATTTAATATAATTTATAATATAT	79	-9.4	44.8
5	2482236	AATATTCTAGAATCTTTCTAGAATATT	27	-8.8	62.2
5	1790	ATAAATATAAATTATATTTAT	21	-3.92	48.1
5	527	TATTTTTAAATATAATATTAATTAAATTATAAAATTAATAATATTTAAATTAA TATAGATTTAAAAAGTACTCTAAAATTTTTAATCTATATTAATTTAAATATT ATTAATTTTATAATTTAATTAATATTATATTTAAAATA	143	-53.06	64.5
5	2447902	TAAATAATAATTAATATATTAATATCTTAATAATATTAACTAGTAGTTA	49	-3.93	45.4
5	2472764	TAGTAGATCTATAAAGATCTTCTA	24	-5.02	53.7
6	2149517	TTTATAAATTATAATAAATTTAATTTATAATTCTATAATTTTATTATTATATATTAAA	57	-6	43
7	690328	TTAATATATTATTAAAATTTTTATTAACTATTTAAATTTATAGATTTTTATAAATAAATTAACTAAATAATATTAA	78	-5.08	39.1
7	2630527	TTTTTAATATAATAATTTTATATTAAAA	28	-6.68	53.8
8	1814495	ATTATTTTATTTTTATCAATTTAAAAAATAATTATAAATTAAAATTAATTTTTAATCTTATAAATATATATAAT	74	-7.52	41.8
8	1869802	GACTAGCGCCCGGCGCTAGTC	21	-10.09	75.5
8	3239040	AAATATATATATAATTATTACTAGAAATAAATAATTTATTTTCTAGATATATAATTTTATTAGACTATTT	70	-5.79	42.6
8	1852346	AATAAAATATTTTTTATAAAAATAATTAATT	31	-3.22	50.2
8	1896688	TTTATATATATATATAATATATATATATATAAA	33 bp	-8.72	54.2

Table 4B. **Predicted ATE Junctional Palindromic pTID DNA**. The chromosomal coordinates, sequences, length and predicted ΔG and T_m values of pTID palindromes in AT-rich elements as described in Table 4A.

Location	All Palindromes (142 ends)	Palindromes (≤ -3 kcal/mole; T_m ≥ 40°C)
Centromeres	22	16
Subtelomeres + Other ATEs	8	6
Dispersed	42	35
Total	72	57

Table 5. **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 stabilities may still form secondary structures but do not meet the limitations set in this study.

As a control, we compared the predicted stability of a previously isolated set of short palindromic junctions deduced to be at the sites of inverted sequence formation in yeast (Table 6) (*7*). These palindromes had ΔG values ranging from -1.6 to -5.5 kcal/mole and T_m values from 42-67°C, 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	-1.69	42.9
CGCCGGCCCAGTGGGCG	17	-4.55	65.2
TTGGTGCGGCCAA	13	-2.32	53.4
GGGCAATCACTTTTGCCC	18	-5.5	67
CTGGGGTTACCGCCAG.	16	-1.88	45.3
CCCAGGTGCCTGGG	14	-4.64	64.9

Table 6. **Predicted Stability of Palindromes Identified at Junctions of Inverted Repeats by Rattray et al (2005) .** The sequences, length and predicted ΔG and T_m values of palindromes derived from published sequences (*7*).

We analyzed 29 junctions with no more than ten base pairs of deleted sequences, not identified as palindromes by BLAST analysis, to examine the structures of sequences at the inverted repeat junctions. Events fell into three classes, including a) hairpins not initially identified by BLAST (n=11), b) multiple palindromic sequences adjacent to junctions (n=6), and c) regions that appeared to lack any junctional palindromic sequences (n=12). These data suggest that mechanisms independent from simple hairpin priming of DNA synthesis can also play a role in generating the pTID structures.

DISCUSSION

Here, using the *A. flavus* model system, we report to our knowledge the first evidence for short repeat-mediated recombination within centromeric and subtelomeric regions in any organism. Several lines of evidence support the identification of spontaneous but rare tandem inverted duplications in *Aspergillus flavus* through structural variant analysis following high-coverage third-generation sequencing. Although we did not apply any intentional selection during growth, selective pressures induced by vegetative growth are possible. Indeed, in studies to be presented elsewhere, we found reductions in conidiation levels between the fifth and tenth weeks of culturing (data not shown).

First, BLAST searches of individual reads of presumptive large insertions generated after two independent methods of genome sequencing (ONT and PacBio) predicted pTID structures consistent with unprocessed symmetric and processed asymmetric TID forms (*8-10*), albeit differing in abundance. Although asymmetric products were not previously observed in filamentous fungi or yeast, they did arise in this study. The presumptive processed forms had a

deletion at one or more junctions, consistent with a role for a mechanism of classical or alternative non-homologous end joining (see below). A large fraction of testable inversion junctions contained palindromes and quasi-palindromes, arguing for a mechanism favoring the use of a hairpin structure to generate the inverted sequence, as in the models for TID formation based on gene amplification studies.

Second, we have shown that the frequencies of direct and inverted repeats are elevated in subtelomeric and centromeric AT-rich DNA, relative to randomly generated DNA having the same AT content. Palindromic formation was also significantly elevated in genomic compared to random AT-rich sequences. This high rate of palindrome formation may well play a role in pTID clustering in centromeric and subtelomeric regions.

Third, we found that the pTIDs are rare and do not become fixed in the population after 15 cycles of forced vegetative growth. For the G15 cultures sequenced by ONT, we identified 221 reads out of 3.6×10^6 reads of sizes (> 2kb) that were used in this analysis. Since each read is derived from an independent DNA molecule, this suggests an overall frequency of 6×10^{-5} pTID/genome. This estimate is only an approximation, since the coverage of different genomic regions is not identical using either sequence method. These structures arise at every stage of subculturing. The low abundance may be due to high rates of pTID processing.

Instability of pTIDs may arise via distinct mechanisms of homologous and non-homologous recombination (Fig. 3 left). Recombination between the direct repeats would result in complete loss of the pTID and restore the original structure (Fig. 3, left, pathway 1). On the

other hand, recombination between inverted repeats would generate a novel pattern of direct repeats inversions (Fig.3, left, pathway 2). Indeed, the direct and inverted repeats found in centromeric and subtelomeric regions are among the expected products of pTID recombination. In contrast, unequal recombination between sister chromatid pTIDs would result in both the loss and duplication of these regions.

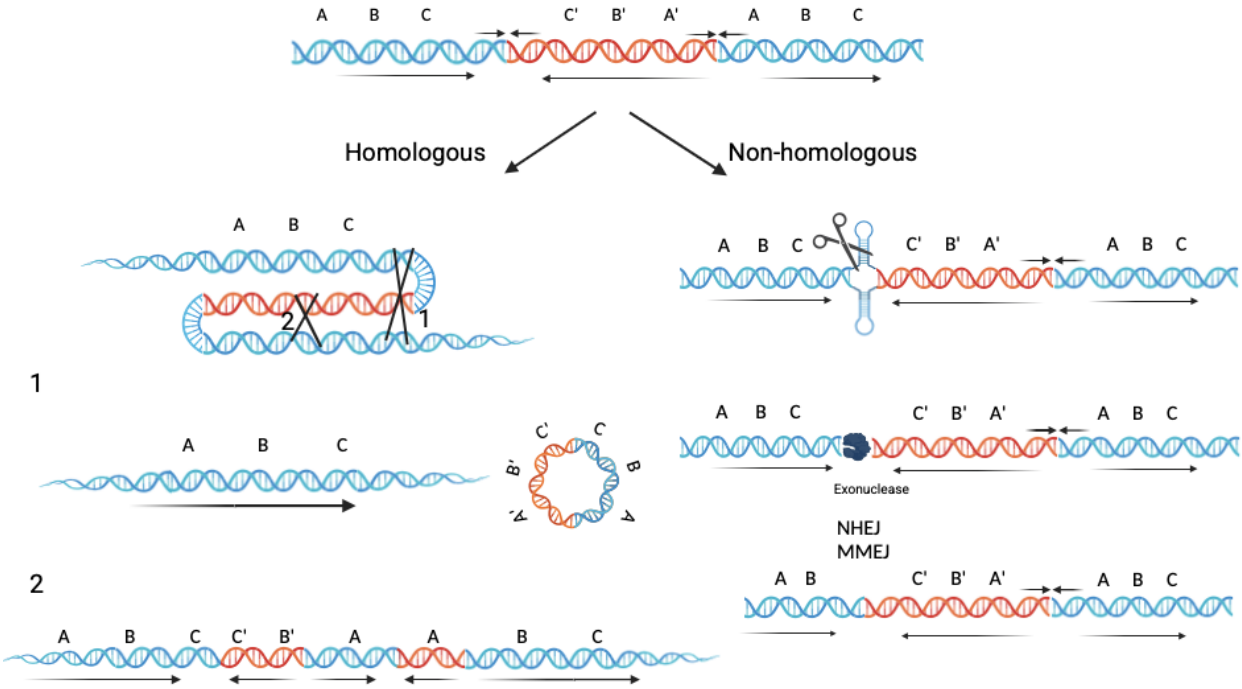

Fig. 3. **Models for the Resolution of pTID Structures**. Homologous and non-homologous pathways for resolution of pTIDs. (Top) The structure of pTIDs. Color designations are described in Fig. 1. (Left) Recombination between direct repeats (1) gives rise to excision of a dimeric inverted circle that 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 pTID 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) (*36*, *37*) (Fig. 3 right), or alternatively, Ku-dependent non-homologous end-joining (NHEJ) (*37*). The prevalence of short, inverted repeats found in centromeric and subtelomeric ATEs may facilitate MMEJ. MMEJ is indeed found in yeast and filamentous fungi, although by mechanisms independent from the pol θ-dependent process in multicellular organisms. The frequent processing of palindromic DNA in AT-rich sequences coupled with MMEJ or NHEJ are also expected to generate small deletions within these regions.

The transient presence of pTIDs could also be due to a strong selection against the rearranged sequences leading to diminished growth rate or cell death. This explanation seems unlikely, however, given the multinucleate nature of mycelial growth in *A. flavus*, which would permit the retention, but not necessarily expansion of rare deleterious rearrangements (*26*). If such a selection is present, however, it is more likely to be mediated through pTID disruption of coding regions. However, we cannot exclude the possibility that specific areas of centromeric DNA responsible for kinetochore binding proteins are lost.

While pTID-like structures were identified in the centromeric and subtelomeric regions using nanopore and PacBio sequencing, the frequencies of pTID formation differ between the two methods (*28, 32*). Quantitative analysis using such different methodologies is indeed problematic. Although pTIDs are unlikely to be the consequence of any known sequencing artifacts (*38*), several distinctions between the methods may contribute to these quantitative differences. First, nanopore sequencing results in a direct readout of the structural variant, while PacBio amplifies each extracted read to provide greater consistency and potentially selects against pTID amplification. In addition, the pre-treatment methods to repair damaged DNA preceding the construction of the appropriate library may bias the size of DNA fragments in distinct ways.

One difference between the reported TIDs formed in gene amplification and those observed in this study is the presence of a fraction of non-palindromic junctions. This fraction may well be underestimated since palindromes (especially those not readily forming a hairpin structure) could cause replication stalling without resolving the stalled fork (Fig. 4). In the replication-mediated model for TID structures in yeast (*10*), random fork regressions in the presence of palindromes promote template switching and pairing with the opposite fork. Following the resolution of the palindromic end and the merging of forks from different origins, the extruded inverted dimer circle replicates and integrates to generate the TID. We propose that palindromic secondary structures initiate the fork regression that leads to strand exchange. Replication stalling has indeed been observed in experimentally manipulated constructs in both prokaryotes and eukaryotes (*11, 12*) . In addition to palindrome-mediated template switching, we postulate that during fork convergence, extruded strands that either cannot pair or can pair via

short, inverted repeats with the complementary strand may be subject to NHEJ or MMEJ, respectively. A non-homologous end-joining process has been found to be present under conditions of replicative stress (*39-41*); in this case, caused primarily by fork collision. The prevalence of short, inverted repeats within AT-rich sequences that we report may assist a microhomology mediated process (Fig. 4). This model is appealing, since we can experimentally test for the presence of extruded extrachromosomal circles and, if present, their genetic dependence on proteins involved in homologous recombination, NHEJ, and MMEJ. The presence of palindromes formed in the pTID product can also give rise to multiple cycles of inversion and duplication, producing complex products. Indeed, we have identified such complex products, consistent with such a process (data not shown).

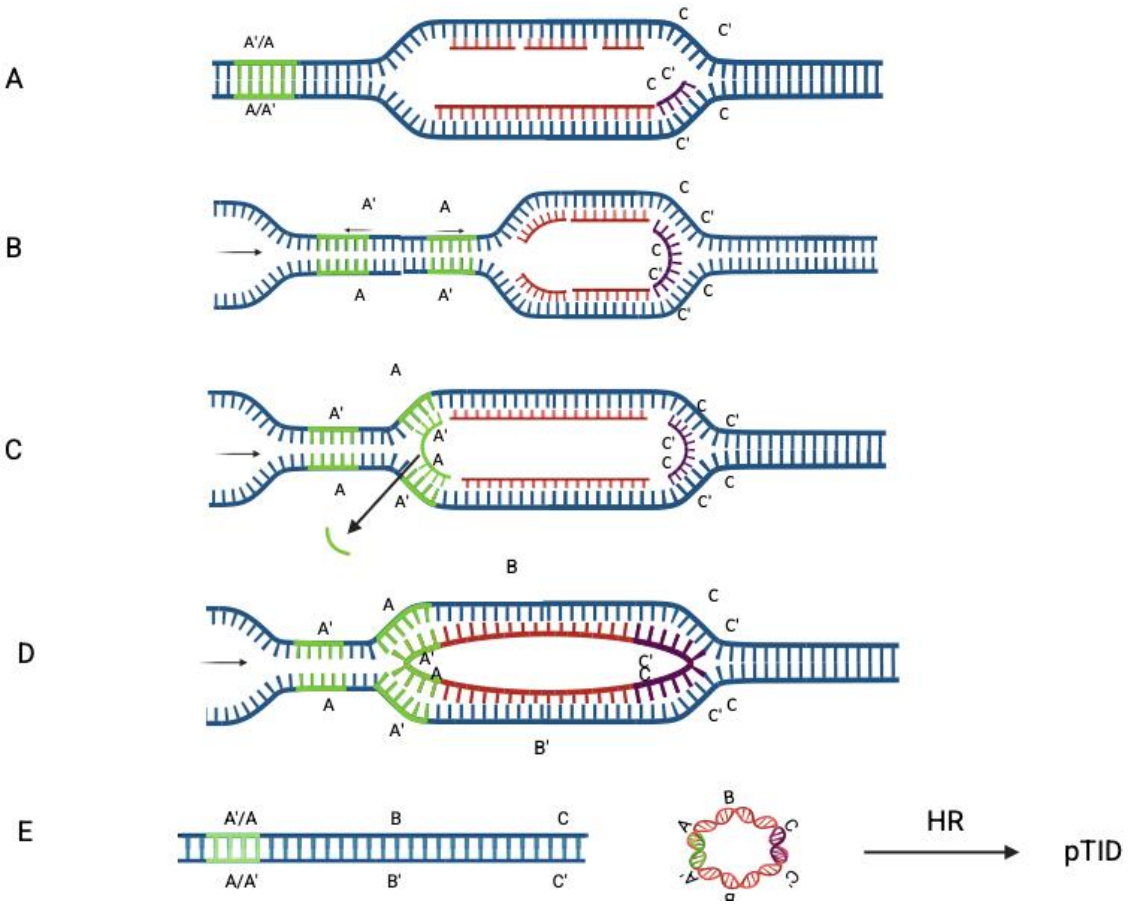

Fig. 4. **A Mechanism for Palindromic and Direct Repeat Strand Exchange in pTID Formation.** This modification of the model by Brewer et al (*10*) proposes:

A. The presence of a palindrome (purple; C, C' where the prime mark indicates the complementary sequence) during DNA replication (newly synthesized strands in red, replicating strands in blue) results in replication stress. On the left is shown an inverted repeat A/A' (green) detailed in B.

B. Because of the stress, the fork regresses which results in the pairing of palindromic sequences to the opposing strand. After completion of replication at this end, an intact palindrome is reformed at the junction of an inverted sequence. The structure of an inverted repeat A/A' present to the left of the forks is detailed in green. Arrows indicate the relative orientation of the inverted sequences.

C. At the left end, in the absence of a palindrome formation, replication forks may converge in the vicinity of an inverted repeat (A/A', shown in green) which are frequent in centromeric and subtelomeric sequences. As the replication fork moves to the left, strand transfer and extrusion is followed by deletion of the intervening sequence (green sequence released) in a process mediated by NHEJ or MMEJ. Unresolved palindromic sequences may also result in replication stress and a similar series of events even in the absence of an immediate fork collision.

D. An intact single strand is formed after repair on the left and ligation of palindromic sequences on the right.

E. As in the Brewer model, this circle is replicated to form a dimeric circle. After DNA repair, the pre-replicative structure was regenerated, which can then recombine with the dimeric circle to create a pTID structure. We note that palindromes reformed at undeleted

junctions allow for further cycles of pTID to occur during subsequent rounds of replication.

Created in BioRender. Cary J.W. et al. (2025) https://BioRender.com/m61z558

One of the exciting aspects of this study is the accumulation of pTIDs in AT- rich subtelomeric and centromeric sequences, which may be more subject to pTID formation if the rate-limiting factor is the frequency of palindrome formation. This clustering is likely the secondary consequence of the AT-rich sequences generated evolutionarily by repeat-induced point mutation. We have previously noted that the AT-rich regions at the centromeres of non-homologs are divergent, while centromeres of homologs are 88-99% identical. We cannot determine whether specific regions of the centromere are conserved, since the *A. flavus* centromere has not been well characterized and has been defined by a high A + T-content, the absence of transcription units, and its similarity to the *Neurospora crassa* centromere (*16*).

We also found that specific subtelomeric ATE sequences exist at different sites within and among strains. This observation may be related to the role of subtelomeres in elimination or the evolution of secondary metabolite gene clusters and their participation in other selected rearrangements in fungi (*23*). Both results support an equilibrium model, whereby deleterious functions such as perturbation of kinetochore function or rampant subtelomeric rearrangement may counteract advantageous roles for centromeric divergence and subtelomeric sequence shuffling. Hence, mechanisms that regulate such an equilibrium must be present, possibly modulated by recombination activities, at centromeres and subtelomeres. Alternatively, heterochromatin formation may be regulated. The fungal centromeric (and probably

subtelomeric) AT-rich sequences are heterochromatic. In yeast, as in most eukaryotes, histone modification and chromatin condensation lead to reduced levels of transcription, recombination and DNA repair in heterochromatin (*42-44*). For example, the histone deacetylase protein Sir2 is responsible for telomeric heterochromatin condensation and position effects (*45*). The use of *A. flavus sir2* (*sirE*) mutants with wide-ranging phenotypes (*46*) is an attractive possibility to test this hypothesis. Such regulation may be the underlying basis for the selection against larger sizes of pTIDs in centromeric regions. In addition, subtelomeric genes that are bounded by heterochromatic regions may serve to regulate transcription from these regions.

Our findings suggest that centromeres and subtelomeres can be more recombinationally active than anticipated. ATEs are maintained in all filamentous fungi and are indeed sites of evolutionary polymorphism (*18*). However, any region that is high in AT or GC content has a statistically higher probability of palindrome formation and may be subject to a similar process, a possibility that requires further exploration. The pTIDs of *A. flavus*, hence, are yet another example of the reported association of AT-rich sequences with genetic instability (*47-49*) which may form hot spots for rearrangement in the eukaryotic genome in normal and diseased cells.

MATERIALS AND METHODS

Strains, media, and growth conditions

G0 spore stocks of *Aspergillus flavus* NRRL 3357 were generated following nine-point inoculations of a diluted $1x10^3$/ml spore stock onto the surface of double-strength 5/2 agar [(per liter: 50 mL V8 juice, 40 g agar, pH 5.2 (*50, 51*)] and incubation at 30 °C in the light for five days, a condition that promotes conidiation. NRRL 3357 conidia were collected from plates by

gentle scraping of the agar surface with 10 ml sterile 0.01% Triton X-100, then transferred to a sterile 15 ml screwcap polypropylene tube, and stored at 4°C. For forced vegetative growth studies, addition of 1 ml (about $1x10^8$) of the spore suspension into 100 ml of A&M medium (per liter: sucrose, 50 g; $(NH_4)_2SO_4$, 3 g; KH_2PO_4, 10 g; $MgSO_4 \cdot 7H_2O$, 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; $CuSO_4 \cdot 5H_2O$, 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 200 rpm, in the dark for seven days, led to generation 1 of the experiment (G1). After 7 days (d), 20 ml of the G1 culture was removed from the 100 ml A&M shake culture and placed in a sterile 50 ml screw cap, polypropylene tube, then macerated for 30 sec using the Tekmar Tissuemizer (Tekmar Co., Cincinnati, OH). One ml of macerated G1 mycelia was transferred into fresh 100 ml A&M broth in a sterile 250 ml Erlenmeyer flask and incubated for another 7 d. This culture represented the second "generation" (G2) of vegetative growth. This process was continued for 15 cycles (15 weeks; 105 d). We diluted a 10 µl aliquot of macerated mycelia in 990 µl sterile water, spread 10 and 100 µl aliquots onto the surface of double-strength 5/2 medium, and incubated for three days to observe colony phenotype. Mycelia for genomic DNA preparation from 5 (G5), 10 (G10), and 15-week (G15) cultures of NRRL 3357 mycelia were collected as follows. At the end of the desired growth period, 50 ml of potato dextrose broth (PDB, Difco) was inoculated with 0.5 ml of macerated mycelia and incubated for 24 h at 28 °C with agitation (200 rpm). For NRRL 3357 generation 0 (G0), genomic DNA was prepared from a 24 h 50 ml PDB culture that had been inoculated with 0.5 ml of a $1x10^8$ spores/ml NRRL 3357 stock. Mycelia were collected from the 50 ml PDB shake cultures by vacuum filtration through sterile miracloth (EMD Millipore, Temecula, CA). The remaining liquid was squeezed from the collected mycelial mat

by pressing it between paper towels. Mycelia were lyophilized for at least 24 h and stored at -80 °C until needed.

Preparation of genomic DNA

Genomic DNA was prepared from lyophilized mycelia using the Qiagen Plant Mini kit (Qiagen, Beverly, MA) following the manufacturer's protocol. Lyophilized mycelia were ground to a fine powder in liquid nitrogen and stored at -80 °C until needed. Briefly, 15 mg of each ground sample was processed in duplicate. After a spin column, the DNA was eluted with two 100 µl elutions using Qiagen AE buffer. The DNA suspension was further enriched by the addition of an equal volume of 25:24:1 phenol: chloroform: isoamyl alcohol (IAA). The upper layer was collected and extracted with an equal volume of 24:1 chloroform: IAA, and then precipitated by the addition of one-tenth volume of 3M sodium acetate pH 5.2 solution and two volumes of cold 95% ethanol followed by centrifuging at 10K rpm for 10 min. The DNA pellet was washed with 70 % ethanol, dried in a speed vac, and resuspended in 50 µl TE. DNA quantity and quality were determined using a NanoDrop OneC (Thermo Scientific). The quality of genomic DNA was also determined by electrophoresis on a 0.7% agarose gel.

Sequencing

CD Genomics (Shirley, NY) carried out all sequencing. Both sequencing methods gave an average of 500-1500 fold coverage, although not equally distributed over all parts of the genome.

PacBio HiFi Sequencing: The SMRTbell library was constructed using the SMRTbell Express Template Prep kit 2.0 (Pacific Biosciences). Briefly, ~2µg of the genomic DNA was carried into the first enzymatic reaction to remove single-stranded overhangs followed by treatment with repair enzymes to repair any damage that may be present on the DNA backbone. After DNA damage repair, ends of the double-stranded fragments were polished and subsequently tailed with an A-overhang. Ligation with T-overhang SMRTbell adapters was performed at 20° C for 60 minutes. Following ligation, the SMRTbell library was purified with 1X AMPure PB beads. The size distribution and concentration of the library were assessed using the FEMTO Pulse automated pulsed-field capillary electrophoresis (Agilent Technologies, Wilmington, DE) and the Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA). Following library characterization, 3µg was subjected to a size selection step using the BluePippin system (Sage Science, Beverly, MA) to remove SMRTbells ≤15 kb. After size selection, the library was purified with 1X AMPure PB beads. Library size and quantity were assessed using the FEMTO Pulse and the Qubit dsDNA HS reagents Assay kits. The sequencing primer and Sequel II DNA Polymerase were annealed and bound, respectively, to the final SMRTbell library. The library was sequenced using 8M SMRT Cell on the Sequel II System with Sequel II Sequencing Kit (*32*).

Nanopore sequencing: We prepared the library using the SQK-LSK112 ligation kit under standard conditions. We loaded the purified library onto primed R10.4 Spot-On Flow Cells and sequenced using a PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK) with 48-h runs (*28*). Base-calling of raw data used the Oxford Nanopore Dorado v 0.7.3 (https://github.com/nanoporetech/dorado/releases).

Assemblies

We bootstrapped G0 and G15 assemblies using both *A. flavus* strain NRRL-3357 (GCA_009017415.1_ASM901741v1) (*21*) and mitochondrial (*A. flavus* SRRC1009) (*52*) sequences that were concatenated to prepare a genome fasta file. This file served as input for G0 ONT or G15 PacBio reads to Rebaler (https://github.com/rrwick/Rebaler) (*53*). G0 and G15 assemblies, stored in the Sequence Server Cloud, are available upon request.

Structural variant determination

Structural variants were identified relative to the G0 assembly using Sniffles 2 (version 2.2) (*30*). Alignments used minimap 2 (*31*) (version 2.26-r1175) with a gap open and gap extension set at 16 and 2, respectively. This version was determined empirically to be only ~70% efficient for low abundance sequences in simple sequence regions, leading to both false positives and negatives. We therefore often coupled the algorithm with visual confirmation using IGV (version 2.(10.0)) (*54, 55*).

For qualitative analyses, structural variants were identified as ≥ 300 bp insertions with variant read numbers (DV) of 3 and confirmed by IGV analysis. Individual reads from these nanopore sequenced DNA were then used as BLAST queries to the G15 or G0 assemblies using SequenceServer (London, UK) (*55*). We analyzed simple sequence organization for AT-rich sequences by removing the dust filter.

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

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

Palindrome analysis

Palindromes were identified from BLAST analysis of single reads with the G15 assembly by the presence of sequences at inverted sequence junctions homologous to both strands of the genomic sequence. We determined the predicted relative thermodynamic stability (T_m, ΔG, ΔH, and ΔS) and secondary structure predictions of the palindromes by the application of UNAfold mfold4.2 program (*33-35*). We set cutoffs of ΔG ≤ -3.0 kcal/mole and a Tm ≥ 40°C for palindrome identification. As a control, we characterized the palindromes at the junctions of an inverted repeats formed in a yeast model system developed by Rattray et al, 2005 (*7*) using the UNAfold program. These included the sequences GTACTATTGGTA, CGCCGGCCCAGTGGGCG, TTGGTGCGGCCAA, GGGCAATCACTTTTGCCC, CTGGGGTTACCGCCAG, and CCCAGGTGCCTGGG. [A license for The UNAFold® software (RPI 2017-042) from Rensselaer Polytechnic Institute has been pending and will be provided upon receipt.]

Sequence Structure in Centromeric and Subtelomeric ATEs

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

ACKNOWLEDGEMENTS

We would like to thank Drs. Matt Lebar, Bill Wimley, Brian Mack, and Prescott Deininger for stimulating discussions during these studies. We would also like to give our appreciation to Melody Baddoo 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. Tulane Cancer Center Cancer Crusaders Next Generation Sequence Analysis Core for assistance with PacBio and ONT long-read analysis. The NGS Analysis core is supported by the National Institutes of Health National Cancer Institute, Award Number, LSUHSC Tumor Virology COBRE, and the Louisiana Cancer Research Consortium (LCRC).

Author contributions:

Conceptualization: AJL
Methodology: AJL, JWC
Investigation: JWC, AM, AS, 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.

Data Accession Information

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 is available on request.

REFERENCES

1. R. E. Brown, C. H. Freudenreich, Structure-forming repeats and their impact on genome stability. *Curr Opin Genet Dev* **67**, 41-51 (2021).
2. A. B. Reams, J. R. Roth, Mechanisms of gene duplication and amplification. *Cold Spring Harb Perspect Biol* **7**, a016592 (2015).
3. L. S. Symington, J. Gautier, Double-strand break end resection and repair pathway choice. *Annu Rev Genet* **45**, 247-271 (2011).
4. M. Svetec Miklenic, I. K. Svetec, Palindromes in DNA-A Risk for Genome Stability and Implications in Cancer. *Int J Mol Sci* **22**, (2021).
5. A. M. Al-Zain, M. R. Nester, I. Ahmed, L. S. Symington, Double-strand breaks induce inverted duplication chromosome rearrangements by a DNA polymerase delta-dependent mechanism. *Nat Commun* **14**, 7020 (2023).
6. V. Narayanan, P. A. Mieczkowski, H. M. Kim, T. D. Petes, K. S. Lobachev, The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* **125**, 1283-1296 (2006).
7. A. J. Rattray, B. K. Shafer, B. Neelam, J. N. Strathern, A mechanism of palindromic gene amplification in Saccharomyces cerevisiae. *Genes Dev* **19**, 1390-1399 (2005).
8. E. Kugelberg *et al.*, The tandem inversion duplication in Salmonella enterica: selection drives unstable precursors to final mutation types. *Genetics* **185**, 65-80 (2010).

9. A. B. Reams, E. Kofoid, E. Kugelberg, J. R. Roth, Multiple pathways of duplication formation with and without recombination (RecA) in Salmonella enterica. *Genetics* **192**, 397-415 (2012).

10. B. J. Brewer, C. Payen, M. K. Raghuraman, M. J. Dunham, Origin-dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons. *PLoS Genet* **7**, e1002016 (2011).

11. M. K. Zeman, K. A. Cimprich, Causes and consequences of replication stress. *Nat Cell Biol* **16**, 2-9 (2013).

12. I. Voineagu, V. Narayanan, K. S. Lobachev, S. M. Mirkin, Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci U S A* **105**, 9936-9941 (2008).

13. H. Inagaki *et al.*, Palindromic AT-rich repeat in the NF1 gene is hypervariable in humans and evolutionarily conserved in primates. *Hum Mutat* **26**, 332-342 (2005).

14. T. Kato, H. Kurahashi, B. S. Emanuel, Chromosomal translocations and palindromic AT-rich repeats. *Curr Opin Genet Dev* **22**, 221-228 (2012).

15. K. Bloom, V. Costanzo, Centromere Structure and Function. *Prog Mol Subcell Biol* **56**, 515-539 (2017).

16. K. M. Smith, J. M. Galazka, P. A. Phatale, L. R. Connolly, M. Freitag, Centromeres of filamentous fungi. *Chromosome Res* **20**, 635-656 (2012).

17. K. M. Smith, P. A. Phatale, C. M. Sullivan, K. R. Pomraning, M. Freitag, Heterochromatin is required for normal distribution of Neurospora crassa CenH3. *Mol Cell Biol* **31**, 2528-2542 (2011).

18. A. C. Testa, R. P. Oliver, J. K. Hane, OcculterCut: A Comprehensive Survey of AT-Rich Regions in Fungal Genomes. *Genome Biol Evol* **8**, 2044-2064 (2016).

19. E. B. Cambareri, R. Aisner, J. Carbon, Structure of the chromosome VII centromere region in Neurospora crassa: degenerate transposons and simple repeats. *Mol Cell Biol* **18**, 5465-5477 (1998).

20. Z. A. Lewis *et al.*, Relics of repeat-induced point mutation direct heterochromatin formation in Neurospora crassa. *Genome Res* **19**, 427-437 (2009).

21. J. M. Skerker *et al.*, Chromosome assembled and annotated genome sequence of Aspergillus flavus NRRL 3357. *G3 (Bethesda)* **11**, 1-5 (2021).

22. A. J. Lustig, Investigating the origin of subtelomeric and centromeric AT-rich elements in Aspergillus flavus. *PLoS One* **18**, e0279148 (2023).

23. M. Rahnama *et al.*, Telomere Roles in Fungal Genome Evolution and Adaptation. *Front Genet* **12**, 676751 (2021).

24. K. I. Kusumoto, S. Suzuki, Y. Kashiwagi, Telomeric repeat sequence of Aspergillus oryzae consists of dodeca-nucleotides. *Appl Microbiol Biotechnol* **61**, 247-251 (2003).

25. F. A. Kondrashov, A. S. Kondrashov, Measurements of spontaneous rates of mutations in the recent past and the near future. *Philos Trans R Soc Lond B Biol Sci* **365**, 1169-1176 (2010).

26. M. Roper, C. Ellison, J. W. Taylor, N. L. Glass, Nuclear and genome dynamics in multinucleate ascomycete fungi. *Curr Biol* **21**, R786-793 (2011).

27. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

28. M. Sereika *et al.*, Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods* **19**, 823-826 (2022).

29. M. Smolka *et al.*, Publisher Correction: Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol*, (2024).

30. M. Smolka *et al.*, Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol*, (2024).

31. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).

32. T. Hon *et al.*, Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci Data* **7**, 399 (2020).

33. N. Peyret, Prediction of nucleic acid hybridization: Parameters and Algorithms *PhD Thesis, Department of Chemistry, Wayne State University, Detroit, MI*, (2020).

34. M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).

35. J. SantaLucia, Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95**, 1460-1465 (1998).

36. A. Sfeir, L. S. Symington, Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci* **40**, 701-714 (2015).

37. H. H. Y. Chang, N. R. Pannunzio, N. Adachi, M. R. Lieber, Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**, 495-506 (2017).

38. C. Delahaye, J. Nicolas, Sequencing DNA with nanopores: Troubles and biases. *PLoS One* **16**, e0257521 (2021).

39. C. Audoynaud, S. Vagner, S. Lambert, Non-homologous end-joining at challenged replication forks: an RNA connection? *Trends Genet* **37**, 973-985 (2021).

40. Z. Wang *et al.*, DNA polymerase theta (POLQ) is important for repair of DNA double-strand breaks caused by fork collapse. *J Biol Chem* **294**, 3909-3919 (2019).

41. Z. Kais *et al.*, FANCD2 Maintains Fork Stability in BRCA1/2-Deficient Tumors and Promotes Alternative End-Joining DNA Repair. *Cell Rep* **15**, 2488-2499 (2016).

42. R. C. Allshire, H. D. Madhani, Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**, 229-244 (2018).

43. J. B. Stavenhagen, V. A. Zakian, Internal tracts of telomeric DNA act as silencers in Saccharomyces cerevisiae. *Genes Dev* **8**, 1411-1422 (1994).

44. L. L. Sandell, V. A. Zakian, Telomeric position effect in yeast. *Trends Cell Biol* **2**, 10-14 (1992).

45. J. Rybchuk, W. Xiao, Dual activities of a silencing information regulator complex in yeast transcriptional regulation and DNA-damage response. *mLife* **3**, 207-218 (2024).

46. M. Wen *et al.*, Histone deacetylase SirE regulates development, DNA damage response and aflatoxin production in Aspergillus flavus. *Environ Microbiol* **24**, 5596-5610 (2022).

47. H. Zhang, C. H. Freudenreich, An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in S. cerevisiae. *Mol Cell* **27**, 367-379 (2007).

48. S. Kaushal *et al.*, Sequence and Nuclease Requirements for Breakage and Healing of a Structure-Forming (AT)n Sequence within Fragile Site FRA16D. *Cell Rep* **27**, 1151-1164 e1155 (2019).

49. M. Irony-Tur Sinai *et al.*, AT-dinucleotide rich sequences drive fragile site formation. *Nucleic Acids Res* **47**, 9685-9695 (2019).

50. P. K. Chang *et al.*, Cloning of the Aspergillus parasiticus apa-2 gene associated with the regulation of aflatoxin biosynthesis. *Appl Environ Microbiol* **59**, 3273-3279 (1993).

51. J. Adye, R. I. Mateles, Incorporation of Labelled Compounds into Aflatoxins. *Biochim Biophys Acta* **86**, 418-420 (1964).

52. J. Park *et al.*, Complete mitochondrial genome sequence of Aspergillus flavus SRRC1009: insight of intraspecific variations on A. flavus mitochondrial genomes. *Mitochondrial DNA B Resour* **5**, 3585-3587 (2020).

53. R. R. Wick, L. M. Judd, K. E. Holt, Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20**, 129 (2019).

54. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192 (2013).

55. A. Priyam *et al.*, Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. *Mol Biol Evol* **36**, 2922-2924 (2019).