MDPI

Article

Gross Chromosomal Rearrangements in *Kluyveromyces Marxianus* Revealed by Illumina and Oxford Nanopore Sequencing

Lin Ding[1], Harrison D. Macdonald[1†], Hamilton O Smith[1,2], Clyde A. Hutchison III[1], Chuck Merryman[1†], Todd P. Michaels[1†], Bradley W. Abramson[1†], Krishna Kannan[2], Joe Liang[2†], John Gill[2], Daniel G. Gibson[1,2], John I. Glass[1*]

[1] J. Craig Venter Institute; 4120 Capricorn Lane, La Jolla, CA 92037 USA; lding@jcvi.org (L.D.); hsmith@jcvi.org (H.O.M.); chutchis@jcvi.org (C.A.H.); jglass@jcvi.org (J.I.G)

[2] Codex DNA; 9535 Waples St #100, San Diego, CA 92121; kkannan@codexdna.com (K.K); jgill@codexdna.com (J.G.); dan@codexdna.com (D.G.G.)

Current email: harrison.macdonald01@gmail.com (H.D.M.); chuckmerryman@gmail.com (C.M.); toddpmichael@gmail.com (T.P.M.); theabramson@gmail.com (B.W.A.); madonjoe@gmail.com (J.L.)

* Correspondence: jglass@jcvi.org; Tel.: +1-858-200-1856

Abstract: *Kluyveromyces marxianus* (*K. marxianus*) is a newly emerging industrially relevant yeast. It is known to possess a highly efficient Non-Homologous End Joining (NHEJ) pathway that promotes random integration of non-homologous DNA fragments into its genome. The nature of the integration events was traditionally analyzed by Southern blot hybridization. However, the precise DNA sequence at the insertion sites were not fully explored. We transformed a PCR product of the *Saccharomyces cerevisiae URA3* gene (*ScURA3)* into an uracil auxotroph *K. marxianus* wildtype strain and picked 24 stable Ura+ transformants for sequencing analysis. We took advantage of rapid advances in DNA sequencing technologies and developed a method using a combination of Illumina MiSeq and Oxford Nanopore sequencing. This approach enables us to uncover the Gross Chromosomal Rearrangements (GCRs) that are associated with the *ScURA3* random integration. Moreover, it will shine a light on understanding DNA repair mechanisms in Eukaryotes, which could potentially provide insights for cancer research.

Keywords: gross chromosomal rearrangements; non-homologous end joining; translocation; Illumina MiSeq; Oxford Nanopore; *kluyveromyces marxianus*; *saccharomyces cerevisiae*; *URA3* gene

1. Introduction

Kluyveromyces marxianus (*K. marxianus*) is thermotolerant yeast [1] and it the fastest-growing Eukaryote identified to date [2]. It has many other physiological features that the conventional yeasts, such as *Saccharomyces cerevisiae*, are lacking. As a result, *K. marxianus* is becoming a potentially valuable industrial yeast. Therefore, many molecular tools [3-7] have been developed for its genetic engineering. Among these are many that harness the power of the Homologous Recombination (HR) pathway, which requires homologous flanking sequences, for targeted gene editing. However, similar to its cousin, *Kluyveromyces lactis* [8], *K. marxianus* also embraces a robust Non-Homologous End Joining (NHEJ) pathway in the absence of homology. It can be transformed with non-homologous DNA fragments by illegitimate recombination (IR) that involves the NHEJ pathway [9]. In the resulting transformants, the DNA fragments insert at various chromosomal locations. These two modes exist with various frequencies in different yeast strains [10, 11].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202008.0687.v1

Kegel et al., [8] reported genome wide IR insertion events in *Kluyveromyces lactis* that were strongly biased toward intergenic regions. If they induced ectopic DSBs using restriction cleavage, then there was no bias for genic versus intergenic sequence, suggesting that genome wide IR occurred at spontaneous mitotic DSBs that are preferentially distributed in intergenic sites. Nonklang et al., [12] used a PCR product of the *Saccharomyces cerevisiae URA3* (Sc*URA3)* gene to transform a *K. marxianus* DMKU3-1042 *ura3Δ* mutant. They analyzed 9 Ura+ transformants by Southern blot hybridization. The Sc*URA3* insertions were in different genome locations and one transformant had multiple insertions. Abdel-Banet et al., [9] reported very high frequency insertion of Sc*URA3* DNA into the genome of *K. marxianus* DMKU3-1042 ($2x10^6$ transformants/µg DNA) by the NHEJ pathway. They suggested that a high density *ScURA3* insertion map analogous to that obtained by global transposon mutagenesis [13, 14] could be generated. In order to fully understand the NHEJ repair mechanism in *K. marxianus*, it is desired to have the precise sequence at the junctions of insert sites in the genome. This is unlikely to be achieved by Southern blot hybridization due to its low resolution, not to mention its laboriousness.

We generated our own Ura+ transformants by transforming a 1121 bp PCR product of the *Saccharomyces cerevisiae URA3* gene (*ScURA3*) and took advantages of Illumina MiSeq and whole genome Oxford Nanopore sequencing to produce a detailed analysis of 24 Ura+ transformant clones. We found a surprising variety of insertion events. In addition to 3 tandem dimer and 2 trimer Sc*URA3* concatemer insertions, two insertions produced inversions, one produced a large deletion and another 3 Ura+ transformants produced chromosomal translocations in which each end of Sc*URA3* were inserted in a different chromosome. Sequencing analysis, especially the long reads obtained with Nanopore sequencing was instrumental in detecting and analyzing GCRs that are a hallmark of cancers. Therefore, our results and method may provide insights to understanding the basic mechanisms of DNA repair and cancer biology.

2. Results

2.1. Determination of ScURA3 genomic insertion sites by Illumina MiSeq analysis.

We constructed two separate libraries containing either Sc*URA3*-head or tail junctions (Figure 1). Illumina MiSeq 100 bp sequences from the head library were scanned for Sc*URA3* head sequence matches followed by at least 20 bp of *K. marxianus* sequence. This *K. marxianus* sequence identified the chromosome and the site of a Sc*URA3* insertion. Similarly, Sc*URA3* tail junctions were identified. In total, we identified 23 head and 23 tail genome junctions (Figure 2). If head and tail junctions were very close on the same chromosome, they were paired and assumed to be produced by a single Sc*URA3* insertion event. This could then be confirmed by PCR using primers designed from the flanking *K. marxianus* sequences (Table 1S). Testing all the clones with a given primer set will point to the specific clone carrying the Sc*URA3* insert.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202008.0687.v1

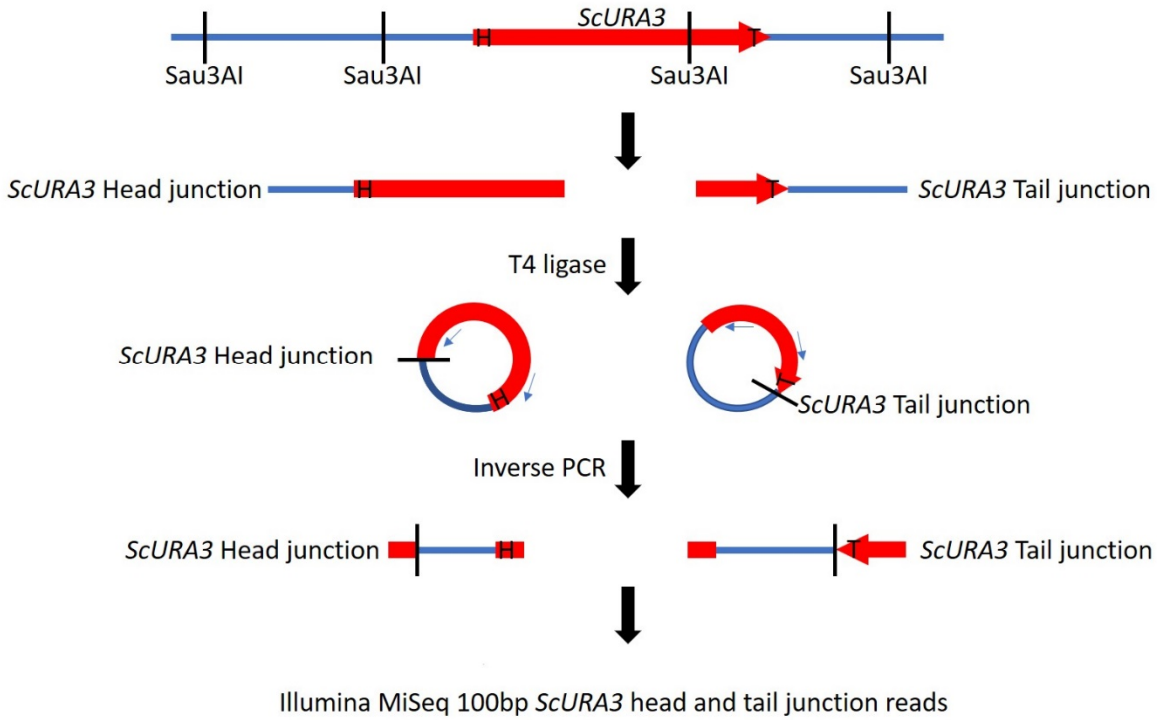

Figure 1. Construction of *ScURA3* head and tail junction libraries. DNA from a *ScURA3* transformant clone is cleaved at GATC sites with Sau3AI and the fragments are ligated under dilute conditions to yield DNA circles with separate *ScURA3* head and tail junctions. The ligated DNA is then divided into two aliquots. Circles containing the head junctions are amplified in a PCR reaction with primers ura3+61c and ura3 TestF. The tail junctions are amplified with ura3+720c and ura3 R1-1-DN primers. The result is separate head and tail junction libraries.

A. Head junctions

Clone #	*K. marxianus* genome	*ScURA3* Head	Chr	Location	Method
1	GGAATTTGCCAAGTCCAGAGGTAT	TGAGAGTGCACCACGCTTTTCAA	1	397946	MiSeq
2	CTCTATATCATTAAAGAGCCTG	***tgagagtg***CACCACGCTTTTCAA	1	1128870	MiSeq
3	TCTTCAATACCAAATTCGTCGTCTAGAG	TGAGAGTGCACCACGCTTTTCAA	2	628004	MiSeq
4	TTCAAAGTGGAAAAAAACACCA	TGAGAGTGCACCACGCTTTTCAA	4	1364429	MiSeq
5	TAACTTCGGAATATTGAATTAATCGTCTCC	TGAGAGTGCACCACGCTTTTCAA	7	24526	MiSeq
6	ACCAAAAAAAAACCATAAAAACCAAAACA	TGAGAGTGCACCACGCTTTTCAA	4	714164	MiSeq
7	ATCTTGAAGAGGAAAAGCATACCA	TGAGAGTGCACCACGCTTTTCAA	3	953765	MiSeq
8	TCCAAAGTCAAACCTTCCAAGTCGACG	TG***ag***AGTGCACCACGCTTTTCAA	1	377332	MiSeq
9	TGTACTTCTACGAAGGGCCAAGGCC	TGAGAGTGCACCACGCTTTTCAA	7	185317	MiSeq
10	TGTGGCTAAGGTAACGGCAAGC	TGAGAGTGCACCACGCTTTTCAA	2	1256630	MiSeq
11	CCCTGCCGTATATACCTGAAAGTTGATTTA	TGAGAGTGCACCACGCTTTTCAA	5	525553	MiSeq
12	GATGATGATGATAAGATAGTGAAGAAGG	TGAGAGTGCACCACGCTTTTCAA	6	166286	MiSeq
14	GACCTCGGGGCGTCGGGGTAACTA	TGAGAGTGCACCACGCTTTTCAA	6	392781	MiSeq
15	AAATTATTATGGAACAATTTGTGTGATTA	***tgagagtgcacca***CGCTTTTCAA	1	1098288	MiSeq
16	GCCGTAACCTCTTTTTTATTGCCAC	TGAGAGTGCACCACGCTTTTCAA	3	718057	MiSeq
17	GATTTAATTAATATATTATATAACAAC	TGAGAGTGCACCACGCTTTTCAA	1	560632	MiSeq
18	CATTAGCACCAAACCCACGGAAATTCCTG	TGAGAGTGCACCACGCTTTTCAA	5	821611	MiSeq
19	CCTATTGAGAAAAAAACCGAAACTATTT	TGAGAGTGCACCACGCTTTTCAA	5	1339373	MiSeq
20	ACTTCTAAATCGGATAGATTCGAGGT	TGAGAGTGCACCACGCTTTTCAA	8	866469	MiSeq
21	TTTATTTTAGTGTTTTATTG	TGAGAGTGCACCACGCTTTTCAA	1	926450	Nanopore
22	TGAAAAAATGATCCTGAAAAGA	TCAGAGTGCACCACGCTTTTCAA	2	826489	MiSeq
23	TATACACGAGATATTAGTCGAGTGGGA	***GTCCCA***TGCACCACGCTTTTCAA	4	101355	MiSeq
24	CATGCCGGCCGAGTAACCGAGAGAA	TGAGAGTGCACCACGCTTTTCAA	1	1643311	MiSeq

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202008.0687.v1

B. Tail junctions

Clone #	*ScURA3* Tail	*K. marxianus* genome	Chr	Location	Method
1	CAATTTAATTATATCAGTTATTACCCTG	GGGCAGCAGTAGTGGATATGCAGC	1	397947	Nanopore
2	CAATTTAATTATATCAGTTATTACCCTG	TCTATCAGGGTCGAATCATCGCATC	1	1128881	MiSeq
3	CAATTTAATTATATCAGTTATTACCC***tg***	CCTTCACCAAAGGAGCCAAACAA	2	628003	MiSeq
4	CAATTTAATTATATCAGTTATTACCC***tg***	CGTTAGATAATCCTGTGAAATCGT	4	1364423	MiSeq
5	CAATTTAATTATATCAGTTATTACCC***tg***	TATATTATATTAAATAAAAATAAT	6	375003	MiSeq
6	CAATTTAATTATATCAGTTATTACCCTG	AACGACTTGATGATGTACATAG	4	762150	MiSeq
7	CAATTTAATTATATCAGTTATTACCCTG	ATGCTGGCATGAGGGGGGGAAGTC	3	824416	MiSeq
8	CAATTTAATTATATCAGTTATTACCCTG	GTAATGCAAATTTATACAGGTCAAA	1	375438	MiSeq
9	CAATTTAATTATATCAGTTATTACCCTG	AGCTTTTGAGAAAAAAAAGTTTTA	7	185307	MiSeq
10	CAATTTAATTATATCAGTTATTACCCTG	TTGGCGCCAGGTCACTTGGT	2	1256656	MiSeq
11	CAATTTAATTATATCAGTTATTACGGCT	TACGCCTCTTAAATCCCAGATC	5	525567	MiSeq
12	CAATTTAATTATATCAGTTATTACCCTG	AGAATCTAGTCTTGTTTGGAAAGTAAA	6	166289	MiSeq
14	CAATTTAATTATATCAGTTATTACC***G***TG	GCGGTCTGAGGGCGTTTCTTTTTCG	6	392769	MiSeq
15	CAATTTAATTATATCAGTTATTACCC***tg***	CATACGCGAAACTCAGGTGCTGCA	5	1165171	Nanopore
16	CAATTTAATTATATCAGTTATTACCC***tg***	TATTTTTTTTTTTTTTTTCGTTTTTTTCCG	3	718026	MiSeq
17	CAATTTAATTATATCAGTTATTACCCT***g***	CTAAAGATACCAAGACGATAGTTTG	1	560631	MiSeq
18	CAATTTAATTATATCAGTTATTACCCTG	TCCTAATAAAACACCAGGTCTCAA	5	821620	MiSeq
19	G***GACttaattatatcagttattaccctg***	AAATTCTTGACTAAATAACACACTC	5	1339385	MiSeq
20	CAATTTAATTATATCAGTTATTACCCTG	GACTCAAAATTAAATGCCAGAGT	8	866482	MiSeq
21	CAATTTAATTATATCAGTTATTACCCTG	TTTTTCCATTATTTTTTTATATTATATT	1	926445	MiSeq
22	CAATTTAATTATATCAGTTATTACCCTG	TATCAGCATCTACGTCACATGCAACACC	2	826479	MiSeq
23	CAATTTAATTATATCAGTTATTACCCTG	GTATCTTCGCTTTGTCTCTTAGCTTCC	6	7292	MiSeq
24	CAATTTAATTATATCAGTTATTACCCTG	TCGGATCGATTCTTATGCG	1	1643296	MiSeq

Figure 2. Head and tail junction sequences from Illumina MiSeq reads of the inverse PCR 24-clone libraries constructed as shown in Figure 1. Chromosome (Chr) and insertion sites are indicated. Lower case, italicized, bold letters indicate deleted bases. Upper case, italicized, bold letters indicate base substitutions.

2.2 Concatemer dimer and trimer ScURA3 inserts.

The 24-clones were screened for the presence of URA3+ concatemers by performing PCR with the primers ura3+61c and ura3+720 (Table 1S). If Sc*URA3* head to tail joints are present, a PCR product of 463 bp should result. Five clones 2, 3, 5, 12, and 19 gave the expected product. In addition, PCRs performed with primers flanking the *ScURA3* concatemers yielded products of the expect sizes. Clones 3, 5, and 19 contained dimers and clones 2 and 12 contained trimers. These PCRs also showed ladders of bands corresponding to monomers, dimers, and higher bands as expected since during the annealing step of each PCR cycle, some of the *ScURA3* concatemers may hybridize out of phase (Figure S4). These results were subsequently confirmed by Oxford Nanopore sequence reads that spanned the entire concatemer and flanking sequence.

The head to tail junctions in the dimer and trimer inserts might be expected to show the effects of NHEJ repair. These junctions are readily observed in both Oxford Nanopore and Illumina MiSeq reads. Among the 7 junctions, we observed only 3 different types: CCCTG|TGAGA, CCC***tg***|TGAGA, and CC***A***T***g***|TGAGA, where the bold lowercase italicized letters are deleted bases and the bold uppercase italicized letters are substitutions or new bases. Note that we cannot distinguish whether the ***tg*** in the second junction type is deleted from the tail as shown, or from the head.

2.3 Sixteen ScURA3 insertion events were either precise or resulted in small deletions of genome sequence.

Sixteen clones appeared to be simple events that involved insertion of Sc*URA3* into a single break in the *K. marxianus* genome (Table 1). Two of the clones (3 and 17) contained precise Sc*URA3* insertions, that is, no loss of genome sequence occurred at the insertion site. In both cases the Sc*URA3* terminal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202008.0687.v1

117 head sequences were unaltered, however, there was a loss of the terminal G of the tail sequence in
118 both cases. In fourteen clones (2, 4, 9, 10, 11, 12, 14, 16, 18, 19, 20, 21, 22, and 24), insertions were
119 accompanied by small deletions of genome sequence, ranging from 2 to 30 bp. Several of these also
120 involved alterations of the terminal head or tail sequences at the junctions. In clone 19, The Sc*URA3*
121 tail had lost 25 terminal bases followed by 3 base substitutions (Figure 2). The observations are
122 compatible with DSB repair by the error prone NHEJ pathway [15].
123

clone #	primer set	Chr	head junction location	tail junction location	type of insertion and base pair involved		*ScURA3*	gene or intergenic insertion
1*	none	1	397946	397947	**precise**	**0 bp**	monomer	*SG4EUKG585063* (*ADE1*)
3	2	2	628004	628003	**precise**	**0 bp**	dimer	*SG4EUKG585526*
17	16	1	560632	560631	**precise**	**0 bp**	monomer	intergenic
2	1	1	1128870	1128881	**del**	**10 bp**	trimer	intergenic
4	26	4	1364429	1364423	**del**	**5 bp**	monomer	*SG4EUKG587507*
9	37	7	185317	185307	**del**	**9bp**	monomer	intergenic
10	11	2	1256630	1256656	**del**	**25 bp**	monomer	intergenic
11	21	5	525553	525567	**del**	**13 bp**	monomer	intergenic
12	4	6	166286	166289	**del**	**2 bp**	trimer	intergenic
13*	none	?	?	?	**?**	**?**	monomer	?
14	20	6	392781	392769	**del**	**11 bp**	monomer	intergenic
16	25	3	718057	718026	**del**	**30 bp**	monomer	intergenic
18	27	5	821611	821620	**del**	**8 bp**	monomer	*SG4EUKG588009*
19	3	5	1339373	1339385	**del**	**11 bp**	dimer	intergenic
20	chr8	8	866469	866482	**del**	**12 bp**	monomer	*SG4EUKG589280*
21	32	1	926450	926445	**del**	**4 bp**	monomer	intergenic
22	12	2	826489	826479	**del**	**9bp**	monomer	*SG4EUKG585988*
24	17	1	1643311	1643296	**del**	**14 bp**	monomer	intergenic
6	9	4	714164	762150	**inversion**	**47,986 bp**	monomer	intergenic-intergenic
7	7	3	953765	824416	**del**	**129,349 bp**	monomer	intergenic-intergenic
8	5	1	377332	375438	**inversion**	**1892 bp**	monomer	*SG4EUKG584656*-intergenic
5	cl5	7--6	24526(7)	375003(6)	**Translocation**	**Translocation**	dimer	intergenic-intergenic
15	10	1--5	1098288(1)	1165171(5)	**Translocation**	**Translocation**	monomer	intergenic-intergenic
23	8	4--6	101355(4)	7292(6)	**Translocation**	**Translocation**	monomer	*SG4EUKG586913*-intergenic

124 **Table 1. Types of *ScURA3* insertion events**. The asterisk next to clones 1 and 13 indicates incomplete
125 information (See text). Chr: Chromosome
126
127

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202008.0687.v1

2.4 Two ScURA3 insertion events were simple insertions but involved corruption of the ScURA3 termini.

Clone 1 has an insertion in the ade1p gene at position 397,946 on chromosome 1 which is in the Open Reading Frame of (*SG4EUKG585063*, homologous to *ScADE1*). As a result of the disruption of the *ADE1* gene, *K. marxianus* colonies are pink on low adenine medium. This is consistent with Ade mutants in other yeasts. It appeared to be precise except for an inserted 65 bp DNA sequence between the Sc*URA3* tail junction and the genome sequence. This DNA did not convincing match any of the genome sequence. Its source is not known. It seems unlikely, because of its length, to have been inserted by an NHEJ polymerase not requiring a template. One result of this insertion was the inability to identify a tail junction sequence for clone 1 in the MiSeq data. The head junction sequence was readily identified. Only by Oxford Nanopore sequencing was the insertion detected and the point of insertion of Sc*URA3* DNA into the genome identified.

In clone 13, about 20 bp of the Sc*URA3* tail sequence and 78 bp of the head sequence were missing, thus there were no junction sequences in Table 1 and the chromosome location of the insert was not found.

2.5 Three ScURA3 insertions resulted in large chromosomal inversions or deletions.

Clone 6 produced an inversion of the genomic segment between 2 DSBs 47,986 bp apart on chromosome 4 (Table 1 and Figure 3). Oxford Nanopore reads were necessary to solve the structure (Figure S5). Each DSB produces two ends, and these can be labeled 1 and 2 for the first break, and 3 and 4 for the second break (Figure 3A). Sc*URA3* DNA has 2 ends labeled H and T. Thus, several possible repair reactions can occur. Clone 6 exhibited 1H—T3 joining followed by 2—4 joining, resulting in inversion of the 2—3 segment. CHEF gel analysis of this clone was identical to wild type (Figure 4).

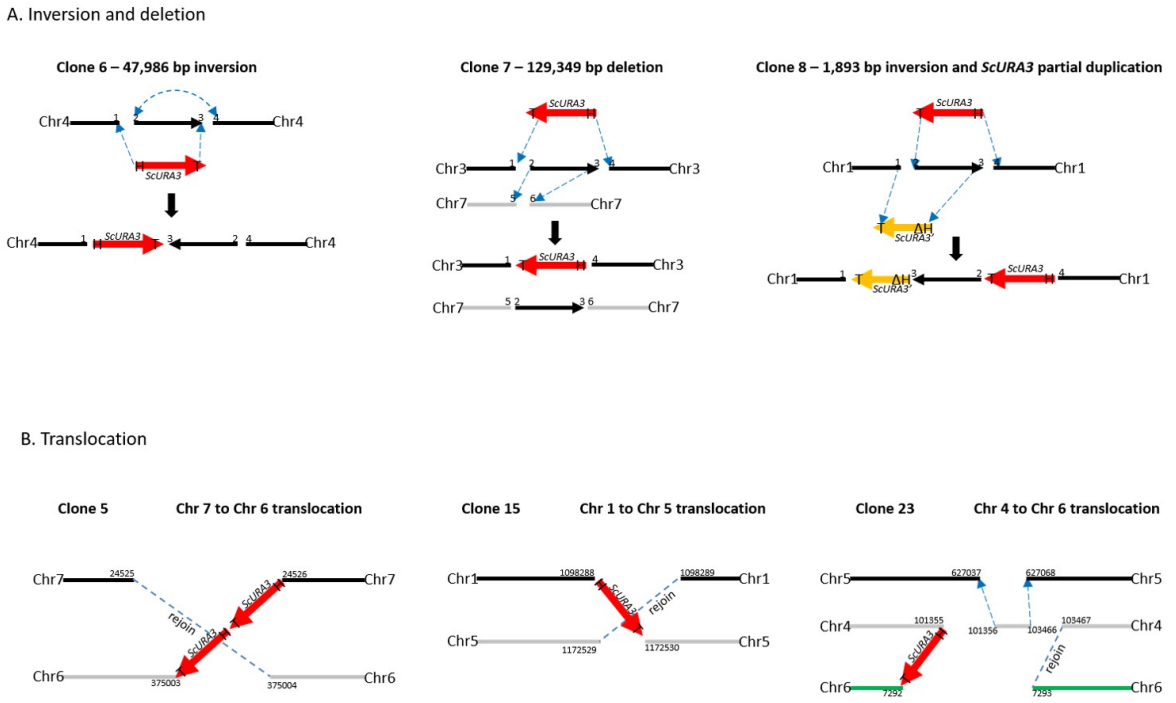

Figure 3. Diagrams of the structures of the *ScURA3* insertion events revealed from Illumina MiSeq and Oxford Nanopore sequencing analysis. A. Inversion and deletion events observed in clone 6, 7, and 8. B Translocation events observed in clone 5, 15, and 23.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202008.0687.v1

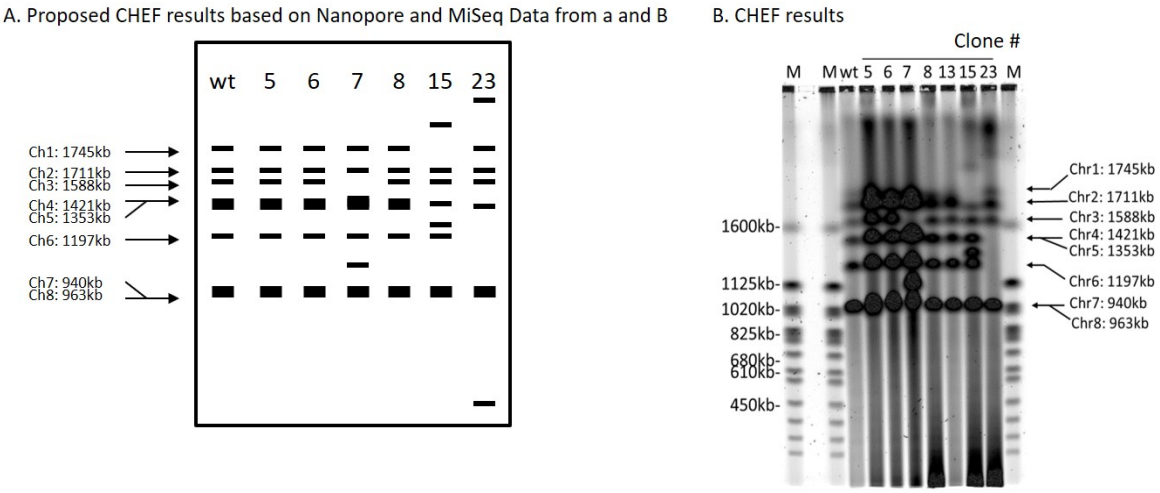

Figure 4. CHEF analysis of *ScURA3* Inversion and deletion as well as translocation events. A. A schematic of expected band patterns of each clone. B. CHEF results of clone 5, 6, 7, 8, 13, 15, and 23.

Clone 7 involved 2 DSBs in chromosome 3 and one DSB in chromosome 7. First a 129,349 bp 2—3 segment was deleted from chromosome 3 in a 1T—H4 Sc*URA3* insertion reaction. This was followed by insertion of the 2—3 segment at a third DSB in chromosome 7 (Table 1, Figure 3a, and Figure S6). Thus, it is predicted that chromosome 7 will increase from 940 kb to 1070 kb and chromosome 3 will decrease from 1588 kb to 1458 kb as confirmed by CHEF gel analysis (Figure 4).

The clone 8 insertion is more complex. In this case the DSBs were 1893 bp apart on chromosome 1 and Sc*URA3* integration involved 2T—H4 joining resulting in an inversion of the 1893 bp segment. The 1 and 3 ends then interacted with another Sc*URA3* sequence yielding the inverted 1892 bp segment separating the two Sc*URA3s*. Furthermore about 200 bp of the head of the second Sc*URA3* is missing. The overall structure is thus 1T—△H3—2T—H4 (Figure 3A). Several Oxford Nanopore reads confirm this structure both from the 24-clone pool and from the barcoded clone 8 reads (Figure S7). The CHEF gel chromosome banding pattern was wildtype as expected (Figure 4).

2.6 Three ScURA3 insertions resulted in inter-chromosomal translocations. For clones 5, 15, and 23, Oxford Nanopore reads located the Sc*URA3* head and tail junctions on separate chromosomes, indicating translocation events (Table 1, Figures S8, S9, and S10). In clone 5, a Sc*URA3* dimer inserted to produce a reciprocal exchange between the arms of chromosomes 7 and 6 (Figure S8) as diagrammed in Figure 3B. One pair of arms is bridged by the Sc*URA3* dimer. The nature of the rejoin by the other two arms has not been determined. The CHEF gel bands are similar to wild type (Figure 4).

Clone 15 contained a Sc*URA3* monomer insert with one end in chromosome 1 and the other in chromosome 5 (Figure S9). The Sc*URA3* tail junction in chromosome 5 falls in a region of repetitious sequence making the exact location uncertain. However, *K. marxianus* assembly based on Oxford Nanopore reads gives a single location in chromosome 5. The rejoin of chromosome 1 and 5 was not successfully located in the Oxford Nanopore reads. Chromosome 5 is predicted to decrease from 1353 kb to 1279 kb while chromosome 1 should increase from 1745 kb to 1819 kb (Figure 4).

Clone 23 contained a Sc*URA3* insert with one end in chromosome 4 and the other in chromosome 6. The translocation is interesting in that the short ends of chromosome 4 and 6 are joined to yield a new short chromosome (Figure 3B and Figure S10). Rejoin of the long arms occurs

 doi:10.20944/preprints202008.0687.v1

at positions 7293 and 103467 to yield an extra-long chromosome. In addition, a short piece of chromosome 1 is released and integrates into chromosome 5 at approximately position 627037 (Figure 3B). Bands 4 and 6 disappear, while the short and very long new chromosomes are not identifiable on the CHEF gel (Figure 4). Interestingly, the band for chromosome 5 also seems to have disappeared.

3. Discussion

During transformation of *K. marxianus*, there is typically a 50 to 100-fold excess of Sc*URA3* DNA molecules to yeast cells. It is expected that many cells will take up several *ScURA3* molecules. The NHEJ pathway proteins, Ku70 [9], Ku80 [6], DNA ligase 4 (data not shown), are required for production of Ura+ transformants while DNA polymerase IV is not required (data not shown) in *K. marxianus*. NHEJ proteins act on the free Sc*URA3* ends resulting in monomer circles as well as tandem concatemers. The probability of these different events is not known, but it seems likely that free unreacted Sc*URA3* ends would not persist for long. Proximity of ends to each other probably determines how likely a pair of ends are to react. However, if endogenous DSBs are simultaneously present in the genome of the same cell, these would be expected to occasionally join to free Sc*URA3* ends to yield Ura+ transformants. This is a rare event leading to only a few thousand transformants among the millions of *K. marxianus* cells present per transformation reaction. By plating on CAA-U plates, insertion of Sc*URA3* DNA into the genome is specifically selected for since non-inserted Sc*URA3* DNA does not independently replicate and is diluted out among the progeny cells. To discover the types of insertion events that might occur, we isolated 24 Ura+ clones for detailed study.

We developed a method using a combination of Illumina MiSeq and Oxford Nanopore sequencing analysis to reveal the precise nucleotide sequence right at the junction of Sc*URA3* random insertion sites, instead of using the laborious Southern blot hybridization. We found that for certain insertions only long-read Nanopore sequencing is capable of resolving the new structure. Our analysis showed that 3 clones contained dimer inserts and two were trimers. The rest were monomer inserts. Eighteen events involved simple insertion into a single DSB in the *K. marxianus* genome while 6 involved 2 or more DSBs. Three events produced translocations in which the two ends of the *ScURA3* cassette inserted into different chromosomes. Two of the latter events produced inversion of DNA between the two DSBs and one produced a large deletion in which the 129 kb deleted segment reinserted into another chromosome. Interestingly, no notable growth defects were observed in these clones even for clone 7 that lost 129kb of genetic material. This indicates the elasticity that *K. marxianus* genome has. Further investigation may uncover DNA repair mechanisms that could be key to understanding cancer biology. s

4. Materials and Methods

4.1 Yeast strains.

K. marxianus NRRL Y-6860 was obtained from the U. S. Department of Agriculture Agricultural Research Service Culture Collection. We sequenced the genomic DNA and identified 8 chromosomes comprising the 10,837,618 genome, and 4963 genes were identified (GenBank number GCA_002356615.1). Strain G13 (*ura3Δ*) was constructed as in Figure S1. The *KmURA3* ORF in *K. marxianus* NRRL Y-6860 was removed by Homologous Recombination (HR)-mediated 5-FOA counter selection. Primers used for strain construction and confirmation are listed in Table S1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 31 August 2020 doi:10.20944/preprints202008.0687.v1

4.2 Preparation of ScURA3 cassette DNA.

The *ScURA3* cassette contained in the plasmid pRS316 (ATCC® 77145™) was PCR-amplified using the two primers 5′-tgagagtgcaccacgcttttcaattc and 5′-cagggtaataactgatataattaaattg. The 5′ OH PCR product (1121 bp, Figure S2) was purified using the QIAquick PCR purification kit. For purposes of calculation, 1 µg of *ScURA3* DNA contains approximately 10^{12} molecules. For convenience, the 5′ end of Sc*URA3* is called the "head" and the 3′ end is the "tail" (Figure 1).

4.3 Isolation of K. marxianus ScURA3 transformants and preparation of transformant DNAs.

Transformation buffer (TFB) consists of 9 parts PEG/Li acetate solution (20 ml of 60% polyethylene glycol 3350, 1.5 ml of 4M lithium acetate and 5.5 ml of sterile water) and 1 part of fresh 1M dithiothreitol. For transformation, *K. marxianus* cells were grown for 24 h in 30 ml of YPD medium at 30°C, centrifuged at 3000 rpm for 5 min and resuspended in 900 µl of TFB. The cell suspension was transferred to a 1.5 ml Eppendorf tube and centrifuged at 3000 rpm for 5 min. The supernatant was removed, and the cells were resuspended in 600 µl TFB. *ScURA3* DNA (70 ng) was then mixed with 50 µl of the *K. marxianus* cell suspension (containing approximately 10^9 cells) and incubated at 42°C for 30 min. 100 µl of CAA-U medium (2% glucose, 0.6% casamino acid, 25 µg/ml adenine 50 µg/ml, tryptophan, and 0.67% YNB without amino acids) was added and the cells were spread on a CAA-U/2% agar plate and incubated at 30°C for 2 days. 24 isolated colonies were picked and patched onto a YPD plate. After incubation for a day at 30°C, the 24 patches were re-patched on a CAA-U plate and grown another day. Twenty-four 50 ml tubes containing 10 ml of CAA-U medium were inoculated from the patches and grown at 30°C for 24h on a shaker. Cells were harvested from each of the cultures by centrifugation and resuspended in 200 µl of P1 cocktail (5 ml of P1 solution (Qiagen), 5 µl of 14M β-mercaptoethanol, and 125 µl of Zymolyase 20mg/ml) in 1.5ml Eppendorf tubes. The cells were incubated at 37°C for 30 min followed by addition of 20 µl of 3M sodium acetate, and extraction with an equal volume of phenol. After centrifugation, the supernatants were harvested and precipitated with 2 volumes of ethanol. The precipitates were washed with ethanol and dissolved in 200 µl of TE buffer (10 Mm Tris-Cl, 1 mM Na EDTA, pH8). The 24 transformant DNAs ranged in concentration from approximately 50 ng/µl to 200 ng/µl.

In addition to the individual transformant DNAs, cells were pooled from a plate of 24 patches and extracted as above to yield 24-clone pool DNA at approximately 200 ng/µl.

4.4 Preparation of 24-clone pool DNA libraries for Illumina MiSeq sequencing.

The 10.8 Mb *K. marxianus* genome contains approximately 37,500 Sau3AI restriction sites (5'GATC) occurring on average every 280 bp along the genome. There is a single site in the 1121 bp *ScURA3* sequence (Figure S2) that cleaves the sequence into a 918 bp left fragment and a 203 bp right fragment. The 24-clone pool DNA was digested with Sau3AI to produce fragments with 5'GATC overhangs in a reaction mixture (50 µl) containing 5 µl of 10X NEB 1.1 buffer, 5 µl of 24-clone pool DNA (∼200 µg/µl, ∼ 10^8 genome equivalents), 2 µl Sau3AI (10U/µl), and 38 µl water. Incubation was at 37°C for 2h followed by inactivation of Sau3AI enzyme at 65°C for 30 min. A 5µl aliquot of the fragments was mixed with 20 µl of 10X T4 ligase buffer (NEB), 5 µl T4 ligase (400u/µl, NEB), and 170 µl water. Incubation was at 23°C for 17 h followed by 72°C for 10 min to inactivate the ligase. Two PCR reactions were performed. The first contained 5 µl of the ligated DNA, 16 µl water, 25 µl of 2X Q5 master mix (NEB), and 2 µl of each of the primers ura3+720c and ura3 R1-1-DN at 25 µM. The second PCR was the same except that the primers were ura3+61c and ura3 TestF (Table S1). PCR settings

 doi:10.20944/preprints202008.0687.v1

were 98°C 10sec, 55°C 20sec, 72°C 2 min for 30 cycles. The first PCR reaction contained *ScURA3* head to *K. marxianus* genome junctions and the second contained the tail junctions (Figure 1). PCR product yields were 83 ng/µl and 109 ng/µl, respectively. Illumina MiSeq 100 nucleotide sequencing reads were done on the two libraries.

4.5 Oxford Nanopore sequencing.

Oxford Nanopore DNA sequencing of the 24-clone pool DNA was performed as described by Oxford Nanopore Technologies. It yielded 2.1 Gb of sequence with a mean read length of 4,658 bp and a maximum read length of 194,297 bp (Figure S3). Single Nanopore reads gave a 10-15% base calling error rate including base deletions. However, alignment of multiple Nanopore reads and *K. marxianus* genome assembly yielded sequence that was > 95% accurate compared to the MiSeq assembly. We generally relied on Illumina MiSeq data for design of primers.

In addition to Oxford Nanopore sequencing of the 24-pool DNA, each clone DNA was bar-coded and sequenced in two 12-clone pools. Reads for each clone were then collected into 24 files ranging from 21 Mb to 827 Mb in size and averaging 370 Mb per clone.

4.6 CHEF Genomic DNA Plug preparation.

A single colony from Clone 5, 6, 7, 8, 13, 15, and 23 out of the 24 clones was inoculated into 50 ml YPD and grown at 30°C overnight. The 1% agarose plugs were prepared using CHEF Yeast Genomic DNA Plug Kit (Bio-Rad, 170-3593) following manufacturer's manual. The plugs were inserted into wells of 1% 0.5X TBE agarose gel and sealed with 1% 0.5X TBE agarose.

4.7 Pulsed-field gel electrophoresis (PFGE).

CHEF-DR® III Pulsed Field Electrophoresis Systems was used for running PFGE with the following settings: initial switch time: 26.3 s, final switch time: 228 s, gradient: 6 V/cm, angle: 120°, temperature: 14°C, and total time: 36h. Gel was stained with 0.5 µg/ml ethidium bromide solution in water for 30 minutes and de-stained in distilled water for 1 hour. Genomic DNA was visualized using Typhoon 9410 Variable Mode Imager.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, Hamilton O Smith, Clyde A. Hutchison III, Chuck Merryman, Krishna Kannan, Daniel G. Gibson and John I. Glass; Data curation, Lin Ding and Hamilton O Smith; Formal analysis, Lin Ding, Hamilton O Smith, Clyde A. Hutchison III, Todd P. Michaels, Bradley W. Abramson and John I. Glass; Funding acquisition, Daniel G. Gibson and John I. Glass; Investigation, Lin Ding, Harrison D. Macdonald, Clyde A. Hutchison III, Bradley W. Abramson, Krishna Kannan, John Gill and Daniel G. Gibson; Methodology, Lin Ding, Bradley W. Abramson and John Gill; Project administration, Daniel G. Gibson and John I. Glass; Resources, Todd P. Michaels; Supervision, Lin Ding and John I. Glass; Writing – original draft, Lin Ding, Hamilton O Smith, Clyde A. Hutchison III and Krishna Kannan; Writing – review & editing, Lin Ding, Hamilton O Smith and John I. Glass. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Defense Advanced Research Projects Agency's Biocontrol program (contract HR0011-16-0010) and the Intelligence Advanced Research Projects Activity's Finding Engineering-Linked Indicators (FELIX) program (contract N6600118C4506) for funding this work. C.A.H., D.G.G., J.G., K.K., and the J. Craig Venter Institute (JCVI) hold SGI stock and/or stock options.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Banat, I. M.; Nigam, P.; Marchant, R., Isolation of thermotolerant, fermentative yeasts growing at 52 degrees C and producing ethanol at 45 degrees C and 50 degrees C. *World journal of microbiology & biotechnology* **1992,** 8, (3), 259-63.
2. Groeneveld, P.; Stouthamer, A. H.; Westerhoff, H. V., Super life--how and why 'cell selection' leads to the fastest-growing eukaryote. *The FEBS journal* **2009,** 276, (1), 254-70.
3. Rajkumar, A. S.; Varela, J. A.; Juergens, H.; Daran, J. G.; Morrissey, J. P., Biological Parts for Kluyveromyces marxianus Synthetic Biology. *Frontiers in bioengineering and biotechnology* **2019,** 7, 97.
4. Lee, M. H.; Lin, J. J.; Lin, Y. J.; Chang, J. J.; Ke, H. M.; Fan, W. L.; Wang, T. Y.; Li, W. H., Genome-wide prediction of CRISPR/Cas9 targets in Kluyveromyces marxianus and its application to obtain a stable haploid strain. *Scientific reports* **2018,** 8, (1), 7305.
5. Lobs, A. K.; Engel, R.; Schwartz, C.; Flores, A.; Wheeldon, I., CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in Kluyveromyces marxianus. *Biotechnology for biofuels* **2017,** 10, 164.
6. Choo, J. H.; Han, C.; Kim, J. Y.; Kang, H. A., Deletion of a KU80 homolog enhances homologous recombination in the thermotolerant yeast Kluyveromyces marxianus. *Biotechnology letters* **2014,** 36, (10), 2059-67.
7. Hoshida, H.; Murakami, N.; Suzuki, A.; Tamura, R.; Asakawa, J.; Abdel-Banat, B. M.; Nonklang, S.; Nakamura, M.; Akada, R., Non-homologous end joining-mediated functional marker selection for DNA cloning in the yeast kluyveromyces marxianus. *Yeast* **2014,** 31, (1), 29-46.
8. Kegel, A.; Martinez, P.; Carter, S. D.; Astrom, S. U., Genome wide distribution of illegitimate recombination events in kluyveromyces lactis. *Nucleic acids research* **2006,** 34, (5), 1633-45.
9. Abdel-Banat, B. M.; Nonklang, S.; Hoshida, H.; Akada, R., Random and targeted gene integrations through the control of non-homologous end joining in the yeast kluyveromyces marxianus. *Yeast* **2010,** 27, (1), 29-39.
10. Maassen, N.; Freese, S.; Schruff, B.; Passoth, V.; Klinner, U., Nonhomologous end joining and homologous recombination DNA repair pathways in integration mutagenesis in the xylose-fermenting yeast pichia stipitis. *FEMS yeast research* **2008,** 8, (5), 735-43.
11. Cormack, B. P.; Falkow, S., Efficient homologous and illegitimate recombination in the opportunistic yeast pathogen candida glabrata. *Genetics* **1999,** 151, (3), 979-87.
12. Nonklang, S.; Abdel-Banat, B. M.; Cha-aim, K.; Moonjai, N.; Hoshida, H.; Limtong, S.; Yamada, M.; Akada, R., High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast kluyveromyces marxianus dmku3-1042. *Applied and environmental microbiology* **2008,** 74, (24), 7514-21.
13. Hutchison, C. A., 3rd; Chuang, R. Y.; Noskov, V. N.; Assad-Garcia, N.; Deerinck, T. J.; Ellisman, M. H.; Gill, J.; Kannan, K.; Karas, B. J.; Ma, L.; Pelletier, J. F.; Qi, Z. Q.; Richter, R. A.; Strychalski, E. A.; Sun, L.; Suzuki, Y.; Tsvetanova, B.; Wise, K. S.;

 doi:10.20944/preprints202008.0687.v1

Smith, H. O.; Glass, J. I.; Merryman, C.; Gibson, D. G.; Venter, J. C., Design and synthesis of a minimal bacterial genome. *Science* **2016,** 351, (6280), aad6253.

14. Hutchison, C. A.; Peterson, S. N.; Gill, S. R.; Cline, R. T.; White, O.; Fraser, C. M.; Smith, H. O.; Venter, J. C., Global transposon mutagenesis and a minimal mycoplasma genome. *Science* **1999,** 286, (5447), 2165-9.
15. Lieber, M. R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **2010,** 79, 181-211.