Non-tandem repeat polymorphisms at microsatellite loci in wine yeast species

María Laura Raymond Eder & Alberto Luis Rosa

Molecular Genetics and Genomics

ISSN 1617-4615

Mol Genet Genomics
DOI 10.1007/s00438-020-01652-2

Springer

Molecular Genetics and Genomics
https://doi.org/10.1007/s00438-020-01652-2

ORIGINAL ARTICLE

Non-tandem repeat polymorphisms at microsatellite loci in wine yeast species

María Laura Raymond Eder[1] · Alberto Luis Rosa[1]

Received: 4 December 2019 / Accepted: 7 February 2020

Abstract

Yeast microsatellite loci consist of short tandem-repeated DNA sequences of variable length. The high mutational rate at these loci generates a remarkable repertoire of alleles, useful for strain differentiation and population genetic studies. In this work, we analyze the DNA sequences of thirteen alleles from each of ten microsatellite loci described for the yeast *Starmerella bacillaris*. Our results show that polymorphic variants of some informative alleles are dependent on SNPs and *indels* rather than on length variation at their originally defined tandem-repeated motifs. The analysis was extended to 55 previously described hypervariable microsatellite loci from a total of 26 sequenced genomes of yeast species that dominate the microbiota of spontaneously fermenting grape musts (i.e., *Hanseniaspora uvarum*, *Saccharomyces cerevisae*, *Saccharomyces uvarum*, and *Torulaspora delbrueckii*) or lead to wine spoilage (*Brettanomyces bruxellensis* and *Meyerozyma guilliermondii*). We found that allelic variants for some microsatellite loci of these yeast species are also dependent on SNPs and/or *indels* flanking their tandem-repeated motifs. For some loci, the number of units at their tandem repeats was found to be identical among the various characterized alleles, with allelic differences being dependent exclusively on flanking polymorphisms. Our results indicate that allele sizing of microsatellite loci using PCR, although valid for strain differentiation and population genetic studies, does not necessarily score the number of units at their tandem-repeated motifs. Sequence analysis of microsatellite loci alleles could provide relevant information for evolutionary and phylogeny studies of yeast species.

Keywords Yeast · Microsatellite · Tandem repeat · Polymorphism · Mutation

Introduction

The characterization of indigenous yeast populations in oenological ecosystems is of considerable interest in ecological and evolutionary studies (Capece et al. 2016; Combina et al. 2005; Vigentini et al. 2016). Several molecular methods, including RAPD-PCR fingerprinting, mtDNA-RFLP, AFLP, delta elements, and microsatellite loci analyses, are widely used to genotype wine yeasts (Masneuf-Pomarede et al. 2016a). Microsatellite loci, consisting of tandem repeats of a variable number of short DNA motifs (i.e., 1–6 bp) (Guillamón and Barrio 2017), are useful markers for yeast strain differentiation and population studies of yeast species (Albertin et al. 2014a, b, 2016; Estoup et al. 2002; Hranilovic et al. 2017; Legras et al. 2005; Masneuf-Pomarede et al. 2015, 2016a). Allele variants of microsatellite loci primarily originate on DNA polymerase slippage, during DNA replication, and/or unequal crossover events at their tandem-repeated motifs (TRM) (Ellegren 2004; Guillamón and Barrio 2017). In addition to the variable lengths of their TRM, single-nucleotide polymorphisms (SNPs) and/or insertions/deletions (*indels*) flanking their TRM have been documented in yeast (Sampaio et al. 2007).

Here, we report the sequence analysis of 65 hypervariable microsatellite loci from 31 sequenced genomes of selected wine yeast species. Eight *Brettanomyces bruxellensis* (Albertin et al. 2014b), ten *Hanseniaspora uvarum* (Albertin et al. 2016), four *Meyerozyma guilliermondii* (Wrent et al. 2016), sixteen *Saccharomyces cerevisiae* (Legras et al. 2005), nine

Communicated by Stefan Hohmann.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00438-020-01652-2) contains supplementary material, which is available to authorized users.

✉ Alberto Luis Rosa
alrosa@ucc.edu.ar

[1] Laboratorio de Genética y Biología Molecular, IRNASUS-CONICET, Facultad de Ciencias Químicas, Universidad Católica de Córdoba, Córdoba, Argentina

Published online: 22 February 2020

Saccharomyces uvarum (Masneuf-Pomarede et al. 2016b), ten *Starmerella bacillaris* (Masneuf-Pomarede et al. 2015), and eight *Torulaspora delbrieckii* (Albertin et al. 2014a) microsatellite loci were analyzed. Additional information was obtained by sequencing indigenous alleles of two microsatellite loci (i.e., CZ11 and CZ54) from *Starm. bacillaris*. Our results show that SNPs and *indels* at sequences flanking the TRM largely contribute to the allelic repertoire (i.e., size and/or sequence) of some wine yeast microsatellite loci. As a consequence of *indels*, allele sizing using PCR does not necessarily reflect the number of repeated units at the TRM of a given microsatellite locus.

Materials and methods

In silico analysis of wine yeast microsatellite loci

BLAST software (https://blast.ncbi.nlm.nih.gov) was used for sequence analyses of previously described microsatellite loci from *B. bruxellensis* (Albertin et al. 2014b), *H. uvarum* (Albertin et al. 2016), *M. guilliermondii* (Wrent et al. 2016), *S. cerevisiae* (Legras et al. 2005), *S. uvarum* (Masneuf-Pomarede et al. 2016b), *Starm. bacillaris* (Masneuf-Pomarede et al. 2015), and *T. delbrueckii* (Albertin et al. 2014a) (Table 1; Table S1). Microsatellite sequences were obtained, using forward (F**S**) and reverse (R**S**) primer sequences (Fig. 1; Table S1) as queries, from five sequenced genomes each of *B. bruxellensis* (AWRI1499, LAMAP2480, CBS 2796, CBS 2499, UMY321), *H. uvarum* (DSM 2768, AWRI3581, 34-9, AWRI3580, CBA6001), *S. cerevisiae* (AWRI1631, JAY291, W303, M22, YPS163), *S. uvarum* (MCYC 623, U1, U2, U3, U4) and *Starm. bacillaris* (FRI751, PAS13, PYCC 3044, NP2, CBS 9494), as well as four and two sequenced genomes of *M. guilliermondii* (RP-YS-11, SO, W2, ATCC6260) and *T. delbrueckii* (CBS 1146, SRCM101298) (Table 2). Missing alleles for some microsatellite loci correspond to BLAST searches where a single contig including both F**S** and R**S** sequences, was not found. For each microsatellite locus, the analyzed sequence included five additional nucleotides upstream and downstream of the genomic sequence for F**S** and R**S** (Fig. 1) (Albertin et al. 2014a, b, 2016; Legras et al. 2005; Masneuf-Pomarede et al. 2015, 2016b; Wrent et al. 2016). Sequence alignments were performed using the Clustal Omega Multiple Sequence Alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo).

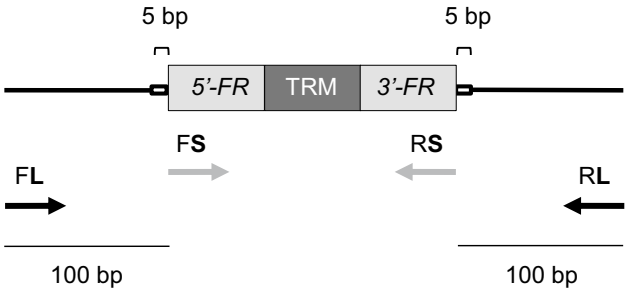

Fig. 1 Schematic representation of a yeast microsatellite locus. The figure shows the location of the internal tandem-repeated motif (TRM) and the upstream (5′-FR) and downstream (3′-FR) flanking regions. The hybridization positions are indicated for the primers forward (F**S**) and reverse (R**S**), originally described for the wine yeast microsatellites, as well as the forward (F**L**) and reverse (R**L**) primers used in this work for *Starm. bacillaris* microsatellite loci CZ11 and CZ54. The primers F**L** and R**L** hybridize ~100 bp upstream and downstream, respectively, the originally described CZ11 and CZ54 DNA sequences (Masneuf-Pomarede et al. 2015). Short 5′ and 3′ (5 bp) regions, flanking the originally described microsatellite sequences, were included in the characterization of all the analyzed wine yeast microsatellite loci

Microsatellites CZ11 and CZ54 from *Starm. bacillaris*

PCR amplification of microsatellite loci CZ11 and CZ54 from *Starm. bacillaris* was performed using the originally described forward and reverse primers (Masneuf-Pomarede et al. 2015) as well as primers CZ11-F**L**/CZ11-R**L** (5′CAACAAAGAGGTTCTGACG-C3′/5′AAGAGCTTCTCTGCTCCTTC3′) and CZ54-F**L**/CZ54-R**L** (5′AATGGAA-TTG

Table 1 Yeast species and microsatellite loci

Yeast species	Microsatellite loci	References
B. bruxellensis	B101, B122, B135, B174, B22, B224, B273, B301	Albertin et al. (2014b)
H. uvarum	HU292, HU440, HU508, HU593, HU467, HU620, HU409, HU853, HU68, HU594	Albertin et al. (2016)
M. guilliermondii	sc15, sc22, sc32, sc72	Wrent et al. (2016)
S. cerevisiae	C3, C4, C5, C6, C8, C9, C11, SCAAT1, SCAAT2, SCAAT3, SCAAT5, SCAAT6, YKR072CS, SCYOR267C, YKL172, YPL009	Legras et al. (2005)
S. uvarum	SuARS409, SuYBR049C, SuYKL017C, SuYKR045C, SuYGC170W, SuYHR042-043, SuHTZ1PLB3, SuYHR102W, SuYIL130W	Masneuf-Pomarede et al. (2016b)
Starm. bacillaris	CZ1, CZ4, CZ11, CZ13, CZ15, CZ20, CZ33, CZ45, CZ54, CZ59	Masneuf-Pomarede et al. (2015)
T. delbrueckii	TD1A, TD1B, TD1C, TD2A, TD6A, TD7A, TD5A, TD8A	Albertin et al. (2014a)

Table 2 Yeast strains and yeast sequenced genomes used for microsatellite loci sequence analyses

Species	Strain	Description	Accession #	References
B. bruxellensis	AWRI1499	Australian wine isolate	PRJNA78661	Curtin et al. (2012)
	LAMAP2480	Chilean wine isolate	PRJNA231184	Valdes et al. (2014)
	CBS 2796	Sparkling wine	PRJNA335438	Cheng et al. (2017)
	CBS 2499	French wine isolate	PRJNA76499	Piškur et al. 2012
	UMY321	Italian red wine isolate	PRJEB21262	Founier et al. (2017)
H. uvarum	DSM 2768	Type strain	PRJNA178141	Langenberg et al. (2017)
	AWRI3581	Australian Chardonnay grape must isolate	PRJNA305659	Sternes et al. (2017)
	34–9	Chinese isolate from epiphytes of citrus roots	PRJNA254213	Liu et al. (2016)
	AWRI3580	Australian Chardonnay fermenting grape must isolate	PRJNA325557	Sternes et al. (2016)
	CBA6001	South Korean "kimchi" isolate	PRJNA434537	Kim et al. (2019)
M. guilliermondii	RP-YS-11	Soil isolate	PRJNA559974	Singh et al. (2019)
	SO	Spoilage orange isolate	PRJNA547962	Zainudin et al. (2018)
	W2	Retrieved activated sludge	PRJEB27464	Yang et al. (2019)
	ATCC6260	Culture from type material of *Candida guilliermondii*	PRJNA12729	Butler G et al. (2009)
S. cerevisiae	AWRI1631	Haploid derivative of South African commercial wine strain N96	PRJNA30553	Borneman et al. (2008)
	JAY291	Haploid derivative of Brazilian industrial bioethanol strain PE-2	PRJNA32809	Argueso et al. (2007)
	W303	Laboratory strain	PRJNA83445	Ralser et al. (2012)
	M22	Italian vineyard isolate	PRJNA28815	Doniger et al. (2008)
	YPS163	Pennsylvania woodland isolate	PRJNA260311	Fay et al. (2004)
S. uvarum	MCYC 623	Microbiology Collection of Yeasts Cultures, Spain	PRJNA1441	Kellis et al. (2003)
	U1	Monosporic clone of grape must isolate PM12	PRJNA388544	Albertin et al. (2018)
	U2	Monosporic clone of PJP3	PRJNA388544	Albertin et al. (2018)
	U3	Monosporic clone of apple juice fermentation isolate BR6	PRJNA388544	Albertin et al. (2018)
	U4	Monosporic clone of grape must isolate RC4-15	PRJNA388544	Albertin et al. (2018)
Starm. bacillaris	CBS 9494	*C. zemplinina* Type strain	PRJNA476215	Rosa et al. (2018)
	FRI751	Italian dried Raboso grapes isolate from fermenting must	PRJNA376556	Lemos Junior et al. (2017b)
	PAS13	Italian isolate from destemmed dried grapes	PRJNA376556	Lemos Junior et al. (2017a)
	PYCC 3044	Portuguese yeast culture collection	PRJNA416493	Gonçalves et al. (2018)
	NP2	Isolate from peach peels	PRJNA397468	Ko et al. (2017)
	MT017-011	Isolate from Malbec fermenting must	MK733726/MK733731	Raymond Eder et al. (2018)
	MT017-020	Isolate from Malbec fermenting must	MK733730	Raymond Eder et al. (2018)
	MT017-021	Isolate from Malbec fermenting must	MK733733	Raymond Eder et al. (2018)
	MT017-022	Isolate from Malbec fermenting must	MK733729	Raymond Eder et al. (2018)
	MT017-029	Isolate from Malbec fermenting must	MK733727/MK733732	Raymond Eder et al. (2018)
	MT117-003	Isolate from Malbec fermenting must	MK733728	Raymond Eder et al. (2018)
	L13	Isolate from *V. labrusca* grapes	MK733724	Drumonde-Neves et al. (2016)
	L14	Isolate from *V. labrusca* grapes	MK733725	Drumonde-Neves et al. (2016)
T. delbrueckii	CBS 1146	Type strain	PRJNA79345	Gordon et al. 2011
	SRCM101298	Isolate from food	PRJNA388014	M.I.F.I. (South Korea)

TGCACGCAAG3′/5′ACAAGTGGAGGAATGGTCAG3′), designed on the basis of the genome sequence of *C. zemplinina* (syn. *Starm. bacillaris*) strain CBS 9494 (Rosa et al. 2018). Primers CZ11-F**L**/CZ11-R**L** and CZ54-F**L**/CZ54-R**L** recognize genome sequences ~100 bp upstream and downstream (Fig. 1) of the loci CZ11 (GenBank #LN864678.1) and CZ54 (GenBank #LN864684.1), respectively (Manseuf-Pomarede et al. 2015). For sequencing purposes, universal

M13 primer sequences (i.e., F: 5′GTAA-AACGACGGC CAGT3′; R: 5′CAGGAAACAGCTATGAC3′) were added at the 5′ end of the CZ11 and CZ54 F**L**/R**L** primers. PCR mixtures contained 100 ng DNA, 1.5 mM $MgCl_2$, *Taq* polymerase buffer 1X (Invitrogen, USA), 200 µM dNTPs, 10 pmol of each forward, and reverse primer and 1.25 units of *Taq* polymerase (Invitrogen, USA). Amplification reactions were performed in a MJ Mini Bio-Rad thermocycler (Bio-Rad, USA) using an initial denaturation step at 93 °C for 3 min, followed by 35 cycles of 93 °C for 30 s, annealing at 53 °C for 45 s, extension at 72 °C for 60 s followed by a final extension at 75 °C for 5 min.

Results

Apparent microsatellite heterozygosity in rare *Starm. bacillaris* strains

Initial studies of microsatellites CZ15 and CZ59 in indigenous isolates of *Starm. bacillaris* suggested allele heterozygosis for a single rare strain (i.e., strain 11-6) (Masneuf-Pomarede et al. 2015). Additional apparent heterozygosis for microsatellites CZ11, CZ13, and CZ54 was recently recognized (Raymond Eder et al. 2019).

To study if these findings reveal rare diploid *Starm. bacillaris* strains or represent technical artifacts, we first analyzed in silico the sequences of loci CZ11, CZ13, CZ15, CZ54, and CZ59 using the available draft genome sequence of *C. zemplinina* (syn. *Starm. bacillaris*) type strain CBS 9494 (Rosa et al. 2018). Although no explanation was found for the apparent heterozygosity previously observed for microsatellites CZ15, CZ54 and CZ59, analyses of loci CZ11 and CZ13 showed the presence of potential alternative genomic binding sites for the forward and reverse primers, respectively, originally described for PCR amplification (Fig. S1 and S2). A potential alternative genomic site for the binding of the reverse primer was also recognized in microsatellite CZ20 (Fig. S2). These alternative primer binding sites at loci CZ11, CZ13 and CZ20 were also found in the genomes of *Starm. bacillaris* strains FRI751, PAS13, PYCC 3044 and NP2 (Fig. S1 and S2), indicating that they do not represent rare polymorphisms present in strain CBS 9494. Based on these observations, we hypothesized that alternative genomic sites for annealing of the primers used for PCR may generate more than one PCR product, resulting in apparent heterozygosity at these loci.

To test this hypothesis, we analyzed locus CZ11 using PCR and primers CZ11-F**L**/CZ11-R**L**, designed to bind genome sequences 100-bp upstream and downstream of the originally described CZ11-F**S**/CZ11-R**S** primers (Fig. 1; Table S1). Single PCR amplification products were obtained when primers CZ11-F**L**/CZ11-R**L** were used to genotype either strain CBS 9494 or indigenous *Starm. bacillaris* strains L13, L14, MT017-011, MT017-029, and MT117-003 (not shown). However, at least two major CZ11-derived PCR fragments were obtained when the same strains were genotyped using primers CZ11-F**S**/CZ11-R**S** (Table 3). Sequence analysis of PCR products (Fig. S1) obtained with primers CZ11-F**L**/CZ11-R**L** confirmed that all the analyzed strains have the alternative annealing site for the originally described forward primer.

A detailed analysis of the various CZ11 alleles characterized (i.e., five alleles from indigenous strains and five alleles from the available *Starm. bacillaris* sequenced genomes—Fig. S1) showed that primer CZ11-F**S** overlaps a highly polymorphic A + T rich, TRM of 16-nt (Fig. 2; Fig. S1). The number of units of this 16-nt TRM, among the analyzed CZ11 alleles, was: 3 (strain L13), 4 (strain L14), 5 (strains MT017-011, MT017-029 and PYCC 3044), 6 (strains CBS 9494, PAS13 and MT117-003) and 8 (strain NP2) (Fig. S1). Interestingly, size differences among CZ11 alleles genotyped using primer CZ11-F**S** are consistent with their different number of 16-nt repeated units, as determined by DNA sequencing (Table 3). Thus, annealing of primer CZ11-F**S** at alternative linked genomic sites may generate more than one PCR product, resulting in apparent heterozygosity at locus CZ11.

Multiple sequence alignment of the CZ11 alleles characterized (Fig. S1) showed that the recognized polymorphic 16-nt TRM, rather than the various lengths of its originally recognized GT/TA/GA TRM, is a major determinant of CZ11 informativity (Masneuf-Pomarede et al. 2015). Thus, we concluded that the apparent heterozygosity previously observed in *Starm. bacillaris* strains for microsatellite CZ11 may result from more than one PCR product, originated from alternative annealing of the forward primer to the recognized 16-nt TRM (Fig. 2; Fig. S1).

Table 3 Apparent heterozygosity of CZ11 and CZ54 microsatellite loci

Starm. bacillaris strain	Microsatellite locus[a]	
	CZ11	CZ54
MT017-011	276/325	265/277
MT017-029	275/310	265/280
MT117-003	276/325	265
L13	273/289	289
L14	287/319	292
MT017-020	310	265/286
MT017-021	273	271/277
MT017-022	275	277/280

[a]Alleles for the CZ11 and CZ54 loci correspond to the size of the PCR products obtained with primers CZ11-F**S**/CZ11-R**S** and CZ54-F**S**/CZ54-R**S**, respectively (see Fig. 1)

A

R_{01}	TATTTTGAGCTTATAC	R_{10}	TAGTTTGAGTTTATAT
R_{02}	TATTTTGCGATTATAC	R_{11}	TATTTTGCGATTTATA
R_{03}	TATTTTGAATTTATAC	R_{12}	TATTTTGAGAATCTGT
R_{04}	TAGTTTGCGATTATAC	R_{13}	TATTTTGAGTTTATAG
R_{05}	TATTTTGAGTTTATAT	R_{14}	TGTTTTGGGTTTATAT
R_{06}	TATTTTGAGATTATAT	R_{15}	TAGTTTGAATTTATAC
R_{07}	TATTTTGAGATTTATA	R_{16}	TATTTTGAGATTATAC
R_{08}	TATTTTGAGAATTTCT	R_{17}	TAGTTTGCGATTATAT
R_{09}	TATTTTGCGATTATAC	R_{18}	TATTTTGAGTTTATAC

B

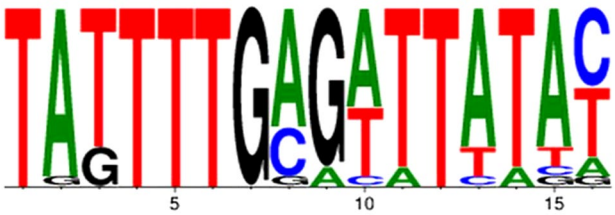

Fig. 2 CZ11 associated 16-nt polymorphic tandem repeat. **a** Diversity of 18 16-nt repeat DNA sequences (R_{01} to R_{18}) recognized among ten analyzed CZ11 alleles. **b** Consensus sequence for the 16-nt repeated motif. The nucleotide positions 5, 10, and 15 are indicated

Sequence analyses of *Starm. bacillaris* microsatellite CZ54

The lack of explanation for the apparent heterozygosity observed in some *Starm. bacillaris* strains for microsatellite CZ54 prompted us to study this locus using PCR with primers CZ54-F**L**/CZ54-R**L** (Fig. 1), followed by sequence analysis. A remarkable repertoire of CZ54 alleles, dependent on its TRM (i.e., 6 to 20 AGA units), was observed in *Starm. bacillaris* strains MT017-011, MT017-020, MT017-021, MT017-022 and MT017-029, as well as CZ54 sequences corresponding to the four *Starm. bacillaris* draft genome sequences analyzed (Fig. S3). In addition, two different *indel* polymorphisms were recognized upstream of the CZ54 TRM: (i) a variable number of units (i.e., 2 to 3) of the sequence AGACCAAGA (Fig. S3), and (ii) a deletion of 18 nt in strain PYCC 3044 (Fig. S3). Thus, non-TRM polymorphisms also contribute to the repertoire of allele size variations at locus CZ54. Interestingly, SNPs observed among CZ54 alleles were all located in the region upstream of its TRM (i.e., 6 SNPs in a region of 94 nt), while no SNPs were present downstream of the TRM (i.e., no SNPs in a region of 96 nt).

Tandem and non-tandem repeat polymorphisms at other *Starm. bacillaris* loci

The finding that non-TRM polymorphisms largely contribute to allele size variations of loci CZ11 and CZ54 prompted us to explore the molecular bases of allele differences at other *Starm. bacillaris* microsatellite loci. Six loci (i.e., CZ1, CZ4, CZ15, CZ33, CZ45, and CZ59) were characterized (Table 1). The in silico analysis of these loci was performed using the draft genome sequences from *Starm. bacillaris* strains CBS 9494, FRI751, PAS13, PYCC 3044, and NP2. Taken together the results showed that, in addition to the expected size variations associated with the polymorphic TRM at these loci, allele differences among the various loci analyzed were also dependent on *indels* and/or SNPs, both at their TRM as well as in upstream (5′-FR) and downstream (3′-FR) regions (Fig. 1) flanking the TRM (Table 4; Fig. S2).

Microsatellite loci from *B. bruxellensis*, *H. uvarum*, *M. guilliermondii*, *S. cerevisiae*, *S. uvarum*, and *T. delbrueckii*

The analysis of microsatellite loci sequences was extended to the wine yeast species *H. uvarum*, *S. cerevisiae*, *S. uvarum* and *T. delbrueckii* as well as the wine spoilage yeasts *B. bruxellensis* and *M. guilliermondii* (Table 1). These species were selected based on the availability of their previously characterized microsatellite loci as well as their relevance in oenological ecosystems. For this purpose, the available genome sequences of five strains each of *B. bruxellensis*, *H. uvarum*, *S. cerevisiae*, and *S. uvarum* as well as four strains of *M. guilliermodii* and two strains of *T. delbrueckii* (Table 2), were analyzed.

In silico study of DNA sequences corresponding to microsatellite loci HU292, HU409, HU467, HU508, HU593, HU594, HU620, and HU853 from *H. uvarum* (Table 1) did not show allele size variants dependent on their repeated units at the TRM (Fig. S4). Moreover, as it was observed for some alleles of *Starm. bacillaris* microsatellite loci, allele size variants for some *H. uvarum* loci (i.e., HU440, HU593, and HU594) were mostly dependent on *indels* at the 5′-FR or 3′-FR (Fig. S4). Interestingly, most of the alleles of *H. uvarum* microsatellite loci contain a remarkable repertoire of SNPs and/or *indels* at their TRMs (Fig. S4). With the exception of loci HU68 and HU292, SNPs and/or *indels* were also recognized at the 5′-FR and/or 3′-FR of all the *H. uvarum* analyzed loci (Fig. S4). Thus, allelic variants of *H. uvarum* microsatellite loci, at least for the analyzed genomes, may be better recognized by sequence analyses than by allele sizing using PCR.

Sixteen *S. cerevisiae* microsatellite loci (i.e., C3, C4, C5, C6, C8, C9, C11, SCAAT1, SCAAT2, SCAAT3, SCAAT5, SCAAT6, YKR072CS, SCYOR267C, YKL172,

Table 4 Non-tandem repeat polymorphisms in yeast microsatellite loci

Yeast species	SNP		*indels*	
	TRM	5′-FR/3′-FR	TRM	5′-FR/3′-FR
H. uvarum	HU68, HU292, HU594	HU409, HU440, HU467, HU508, HU593, HU594, HU620, HU853	HU68, HU292, HU409, HU440, HU467, HU508, HU593, HU853	HU440, HU593, HU594
S. cerevisiae	C4, C6, C8, C11, SCAAT1, SCAAT2, SCAAT3, YPL009, YKR072CS, SCYOR267C	C3, C4, C6, C8, C11, SCAAT1, SCAAT2, SCAAT3, SCAAT5, SCAAT6, YKR072CS, SCYOR267C	C6, C8, SCAAT2, SCAAT3, SCAAT6, YKL172	C3, C4, C8, C9, C11, SCAAT5, SCAAT6
Starm. bacillaris	CZ1, CZ4, CZ11, CZ15, CZ33, CZ45, CZ59	CZ1, CZ4, CZ11, CZ13, CZ15, CZ45, CZ54, CZ59	CZ1, CZ11, CZ45, CZ59	CZ11, CZ45, CZ54
T. delbrueckii	TD1A, TD1B, TD5A, TD6A, TD8A	TD1A, TD1C	TD7A	TD1A
S. uvarum	SuARS409, SuHTZ1PLB3, SuYHR042-043, SuYHR102W, SuYIL130W, SuYKR045C	SuYHR042-043, SuYHR102W, SuYIL130W, SuYKR045C	–	SuYHR042-043, SuYKL017C, SuYKR045C
M. guilliermondii	SC15, SC22, SC32	SC15, SC32, SC72	SC15, SC22	–
B. bruxellensis	B22, B101, B122, B224, B273, B301	B22, B122, B135, B174, B273, B301	–	B273

Single-nucleotide polymorphisms (SNP) and/or insertion/deletions (*indels*) are present in the tandem-repeated motifs (TRM) and/or the TRM-flanking regions (5′-FR and 3′-FR) of the indicated wine yeast microsatellite loci

and YPL009) (Table 1) were analyzed using the genome sequences of five haploid strains (i.e., AWRI1631, JAY291, W303, M22, and YPS163) (Table 2). It was found that only allele variants of locus C5 were exclusively dependent on the polymorphic length of its TRM. All other analyzed *S. cerevisiae* loci showed, in addition to polymorphic lengths of their TRMs, sequence variations dependent on *indels* and/or SNPs, either at their TRMs or at the 5′-FR and/or 3′-FR (Table 4; Fig. S5).

Microsatellite loci SuARS409, SuYBR049C, SuYKL017C, SuYKR045C, SuYGC170W, SuYHR042-043, SuHTZ1PLB3, SuYHR102W and SuYIL130W, described for *S. uvarum* (Masneuf-Pomarede et al. 2016b), were analyzed using genome sequences from strains MCYC 623, U1, U2, U3, and U4 (Table 2). No allelic variants were observed in these genomes for locus SuYBR049C (Fig. S6). Microsatellite loci SuHTZ1PLB3 and SuYGC170W showed informative alleles based on the number of AG and AAG units, respectively, at their TRMs (Fig. S6). As observed for the above yeast species analyzed, SNPs interrupting the TRM and/or at the 5′-FR and 3′-FR, were present at loci SuARS409, SuYHR042-043, SUYIL130W, and SuYHR102W (Fig. S6). The remaining microsatellite loci reveled more complex alleles (Fig. S6). Interestingly, in addition to a variable number of units at their TRMs, allele variants of loci SuYKL017C and SuYKR045C from strain U3 showed large duplications and insertions (Fig. S6).

The analysis of microsatellite loci sequences of *T. delbrueckii* (TD1A, TD1B, TD1C, TD2A, TD6A, TD7A, TD5A, and TD8A) (Table 1) showed that some loci have SNPs (i.e., TD1A and TD1C) and/or *indels* (i.e., TD1A) upstream and downstream of their TRMs, contributing to allele size differences. For this yeast species, at least considering the two analyzed genomes, the allele differences observed among loci were mostly dependent on their polymorphic TRMs (Fig. S7).

Microsatellite loci from the wine spoilage yeast species *B. bruxellensis* (i.e., B101, B122, B135, B174, B22, B224, B273, and B301) and *M. guilliermondii* (i.e., sc15, sc22, sc32, and sc72) were also analyzed (Table 1). All microsatellite loci from *B. bruxellensis* showed allele size variations dependent on their polymorphic TRMs (Table 4; Fig. S8). The only locus showing an *indel* at its TRM was B273 (Fig. S8). SNPs, however, resulted in sequence variation in all the alleles characterized (Fig. S8). SNPs were present at the TRM in loci B101 and B224 as well as at the 5′-FR and 3′-FR in loci B174 and B135, respectively. On the other hand, loci B22, B273, and B301 showed SNPs in their 5′-FR, TRM and 3′-FR (Table 4; Fig. S8). In the case of *M. guilliermondii*, similar complex structures were found for most of the microsatellite loci analyzed, including *indels* at the TRM (sc15 and sc22), length variations at the TRM (sc32 and sc72) and SNPs at the 5′-FR (sc15, sc32 and sc72), 3′-FR (sc15) and TRM (sc32) (Table 4; Fig. S9).

Discussion

Wine yeast microsatellite loci are widely used for strain genotyping and analysis of yeast populations (Albertin et al. 2014a, b, 2016; Hranilovic et al. 2017; Legras et al. 2005; Masneuf-Pomarede et al. 2015, 2016a; Raymond Eder et al. 2019). Recent studies of informative *Starm. bacillaris* microsatellite loci suggested heterozygosis for loci CZ15 and CZ59 in a single European isolate (Masneuf-Pomarede et al. 2015), as well as for loci CZ11, CZ13 and/or CZ54 in some isolates from Argentina and Portugal (Raymond Eder et al. 2019). In this work, we propose that alternative genomic annealing sites for the primers used for PCR may explain the apparent heterozygosity observed for loci CZ11 and CZ13. No explanation was obtained from in silico analysis of genomic DNA sequences for the previously apparent heterozygosity of loci CZ15, CZ59, and CZ54.

Following an extensive sequence analysis of all the currently used *Starm. bacillaris* microsatellite loci (Masneuf-Pomarede et al. 2015; Raymond Eder et al. 2019), we show in this work that SNPs and/or *indels* upstream and downstream of the microsatellite TRM largely contribute to the sequence and size polymorphisms, respectively, at some of these loci.

Our DNA sequence studies of microsatellite loci were extended to the wine yeast species *B. bruxellensis*, *H. uvarum*, *M. guilliermondii*, *S. cerevisiae, S. uvarum* and *T. delbrueckii*. In these studies, we found that 3 out of 8 loci of *B. bruxellensis*, 3 out of 10 loci of *H. uvarum*, 1 out of 4 loci of *M. guilliermondii*, 6 out of 16 loci of *S. cerevisisae,* 3 out of 9 loci of *S. uvarum* and 1 out of 8 loci of *T. delbrueckii* have SNPs and/or *indels* that markedly contribute to both sequence and size differences between alleles. For some microsatellite loci, multiple SNPs and/or *indels* were recognized within the same allele (e.g., HU440, HU467, HU508, HU593, and HU594 of *H. uvarum*; sc32 of *M. guilliermondii*; C4, C11, SCAAT3, and SCAAT6 of *S. cerevisiae*; TD1A of *T. delbrueckii*). SNPs were also found at the TRM of loci SCAAT1, C6 and SCYOR267C of *S. cerevisiae* (Table 3; Fig. S5), which were previously considered "perfect" (i.e., pure motifs at their TRM) microsatellite loci (Legras et al. 2005). Finally, although there were some examples of a bias in the distribution of SNPs among regions upstream and downstream of the TRMs, most of the SNPs at the microsatellite loci analyzed were evenly distributed along their sequences.

Studies conducted in other organisms, such as salmonid fishes (Angers and Bernatchez 1997), insect species (Behura and Severson 2015), and clinical isolates from the yeast species *Candida albicans* (Sampaio et al. 2007), revealed that *indels* and SNPs also contribute to the allelic variation of microsatellite loci. To our knowledge, there are no previous studies addressing the structural complexity of non-tandem repeat polymorphisms at microsatellite loci in wine yeast species. Taken together, our results show that SNPs and *indels*, located upstream and downstream of TRM sequences, largely contribute to allele sequence and allele size variations of wine yeast microsatellite loci. The sequence analyses of these allelic variants could provide useful information for the evolutionary and phylogenetic analyses of wine yeast populations.

Acknowledgements MLRE held a fellowship from the National Research Council of Argentina (CONICET). ALR is a Career Research Investigator from CONICET. This work was supported by PICT-2014-3113 from Fondo para la Investigación Científica y Tecnológica (FONCyT) (Argentina) and SUV2015 (Universidad Católica de Córdoba) to ALR.

Author contributions MLRE and ALR contributed to the study conception and design, material preparation, data collection, analysis, and writing of the manuscript. Both authors read and approved the final manuscript.

Funding This work was supported by PICT-2014-3113 from FONCYT (Argentina) and SUV2015 (UCC) to ALR.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Albertin W, Chasseriaud L, Comte G, Panfili A, Delcamp A, Salin F, Marullo P, Bely M (2014a) Winemaking and bioprocesses strongly shaped the genetic diversity of the ubiquitous yeast *Torulaspora delbrueckii*. PLoS ONE. https://doi.org/10.1371/journal.pone.0094246

Albertin W, Panfili A, Miot-Sertier C, Goulielmakis A, Delcamp A, Salin F, Lonvaud-Funel A, Curtin C, Masneuf-Pomarede I (2014b) Development of microsatellite markers for the rapid and reliable genotyping of *Brettanomyces bruxellensis* at strain level. Food Microbiol. https://doi.org/10.1016/j.fm.2014.03.012

Albertin W, Setati ME, Miot-Sertier C, Mostert TT, Colonna-Ceccaldi B, Coulon J, Girard P, Moine V, Pillet M, Salin F, Bely M, Divol B, Masneuf-Pomarede I (2016) *Hanseniaspora uvarum* from winemaking environments show spatial and temporal genetic clustering. Front Microbiol. https://doi.org/10.3389/fmicb.2015.01569

Albertin W, Chernova M, Durrens P, Guichoux E, Sherman DJ, Masneuf-Pomarede I, Marullo P (2018) Many interspecific chromosomal introgressions are highly prevalent in Holarctic *Saccharomyces uvarum* strains found in human-related fermentations. Yeast. https://doi.org/10.1002/yea.3248

Angers B, Bernatchez L (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Mol Biol Evol 14:230–238

Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OV, Missawa SK, Galzerani F, Costa GG, Vidal RO, Noronha MF, Dominska M, Andrietta MG, Andrietta SR, Cunha AF, Gomes LH, Tavares FC, Alcarde AR, Dietrich FS, McCusker JH, Petes TD, Pereira GA (2007) Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. Genome Res. https://doi.org/10.1101/gr.091777.109

Behura SK, Severson DW (2015) Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. DNA Res. https://doi.org/10.1093/dnares/dsu036

Borneman AR, Forgan AH, Pretorius IS, Chambers PJ (2008) Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. FEMS Yeast Res. https://doi.org/10.1111/j.1567-1364.2008.00434

Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, Agrafioti I, Arnaud MB, Bates S, Brown AJ, Brunke S, Costanzo MC, Fitzpatrick DA, de Groot PW, Harris D, Hoyer LL, Hube B, Klis FM, Kodira C, Lennard N, Logue ME, Martin R, Neiman AM, Nikolaou E, Quail MA, Quinn J, Santos MC, Schmitzberger FF, Sherlock G, Shah P, Silverstein KA, Skrzypek MS, Soll D, Staggs R, Stansfield I, Stumpf MP, Sudbery PE, Srikantha T, Zeng Q, Berman J, Berriman M, Heitman J, Gow NA, Lorenz MC, Birren BW, Kellis M, Cuomo CA (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature. https://doi.org/10.1038/nature08064

Capece A, Granchi L, Guerrini S, Mangani S, Romaniello R, Vicenzini M, Romano P (2016) Diversity of *Saccharomyces cerevisiae* strains isolated from two italian wine-producing regions. Front Microbiol. https://doi.org/10.3389/fmicb.2016.01018

Cheng J, Guo X, Cai P, Cheng X, Piškur J, Ma Y, Jiang H, Gu Z (2017) Parallel evolution of chromatin structure underlying metabolic adaptation. Mol Biol Evol. https://doi.org/10.1093/molbev/msx220

Combina M, Elía A, Mercado L, Catania C, Ganga A, Martinez C (2005) Dynamics of indigenous yeast populations during spontaneous fermentation of wines from Mendoza, Argentina. Int J Food Microbiol 99:237–243

Curtin CD, Borneman AR, Chambers PJ, Pretorius IS (2012) De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast *Dekkera bruxellensis* AWRI1499. PLoS ONE. https://doi.org/10.1371/journal.pone.0033840

Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, Fay JC (2008) A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet. https://doi.org/10.1371/journal.pgen.1000183

Drumonde-Neves J, Frando-Duarte R, Lima T, Schuller D, Pais C (2016) Yeast biodiversity in Vineyard environment is increased by human intervention. PLoS ONE 11:e0160579

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604

Fay JC, McCullough HL, Sniegowski PD, Eisen MB (2004) Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. Genome Biol 5:R26

Fournier T, Gounot JS, Freel K, Cruaud C, Lemainque A, Aury JM, Wincker P, Schacherer J, Friedrich A (2017) High-quality de novo genome assembly of the *Dekkera bruxellensis* yeast using nanopore MinION sequencing. G3: Genes, Genomes, Genetics. https://doi.org/10.1534/g3.117.300128

Gonçalves C, Wisecaver JH, Kominek J, Oom MS, Leandro MJ, Shen XX, Opulente DA, Zhou X, Peris D, Kurtzman CP, Hittinger CT, Rokas A, Gonçalves P (2018) Evidence for loss and reacquisition of alcoholic fermentation in a fructophilic yeast lineage. Elife. https://doi.org/10.7554/eLife.33034

Gordon JL, Armisén D, Proux-Wéra E, ÓhÉigeartaigh SS, Byrne KP, Wolfe KH (2011) Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. PNAS. https://doi.org/10.1073/pnas.1112808108

Guillamón JM, Barrio E (2017) Genetic polymorphism in wine yeasts: mechanisms and methods for its detection. Front Microbiol. https://doi.org/10.3389/fmicb.2017.00806

Hranilovic A, Bely M, Masneuf-Pomarede I, Jiranek V, Albertin W (2017) The evolution of *Lachancea thermotolerans* is driven by geographical determination, anthropisation and flux between different ecosystems. PLoS ONE. https://doi.org/10.1371/journal.pone.0184652

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423(6937):241–254

Kim JY, Kim J, Cha IT, Jung MY, Song HS, Kim YB, Lee C, Kang SY, Bae JW, Choi YE, Kim TW, Roh SW (2019) Community structures and genomic features of undesirable white colony-forming yeasts on fermented vegetables. J Microbiol. https://doi.org/10.1007/s12275-019-8487-y

Ko HJ, Park HJ, Lee SH, Jeong H, Bae JH, Sung BH, Choi IG, Sohn JH (2017) Draft genome sequence of an acid-tolerant yeast, *Candida zemplinina* NP2, a potential producer of organic acids. Genome Announc. https://doi.org/10.1128/genomeA.01052-17

Langenberg AK, Bink FJ, Wolff L, Walter S, von Wallbrun C, Grossmann M, Heinisch JJ, Schmitz HP (2017) Glycolytic functions are conserved in the genome of the wine yeast *Hanseniaspora uvarum*, and pyruvate kinase limits its capacity for alcoholic fermentation. Appl Environ Microbiol. https://doi.org/10.1128/AEM.01580-17

Legras JL, Ruh O, Merdinoglu D, Karst F (2005) Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. Int J Food Microbiol 102:73–83

Lemos Junior WJF, Treu L, da Silva DV, Carlot M, Nadai C, Campanaro S, Giacomini A, Corich V (2017a) Whole-genome sequence of *Starmerella bacillaris* PAS13, a nonconventional enological yeast with antifungal activity. Genome Announc. https://doi.org/10.1128/genomeA.00788-17

Lemos Junior WJF, Treu L, Duarte VDS, Campanaro S, Nadai C, Giacomini A, Corich V (2017b) Draft genome sequence of the yeast *Starmerella bacillaris* (syn., *Candida zemplinina*) FRI751 isolated from fermenting must of dried raboso grapes. Genome Announc. https://doi.org/10.1128/genomeA.00224-17

Liu P, Chen K, Li G, Yang X, Long CA (2016) Comparative transcriptional profiling of orange fruit in response to the biocontrol yeast *Kloeckera apiculata* and its active compounds. BMC Genomics. https://doi.org/10.1186/s12864-015-2333-3

Masneuf-Pomarede I, Juquin E, Miot-Sertier C, Renault P, Laizet Y, Salin F, Alexandre H, Capozzi V, Cocolin L, Colonna-Ceccaldi B, Englezos V, Girard P, Gonzalez B, Lucas P, Mas A, Nisiotou A, Sipiczki M, Spano G, Tassou C, Bely M, Albertin W (2015) The yeast *Starmerella bacillaris* (synonym *Candida zemplinina*) shows high genetic diversity in winemaking environments. FEMS Yeast Res. https://doi.org/10.1093/femsyr/fov045

Masneuf-Pomarede I, Bely M, Marullo P, Albertin W (2016a) The genetics of non-conventional wine yeasts: current knowledge and future challenges. Front Microbiol. https://doi.org/10.3389/fmicb.2015.01563

Masneuf-Pomarede I, Salin F, Börlin M, Coton E, Coton M, Jeune CL, Legras JL (2016b) Microsatellite analysis of *Saccharomyces uvarum* diversity. FEMS Yeast Res. https://doi.org/10.1093/femsyr/fow002

Piškur J, Ling Z, Marcet-Houben M, Ishchuk OP, Aerts A, LaButti K, Copeland A, Lindquist E, Barry K, Compagno C, Bisson L, Grigoriev IV, Gabaldón T, Phister T (2012) The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related

properties. Int J Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2012.05.008

Ralser M, Kuhl H, Ralser M, Werber M, Lehrach H, Breitenbach M, Timmermann B (2012) The *Saccharomyces cerevisiae* W303–K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt. Open Biol. https://doi.org/10.1098/rsob.120093

Raymond Eder ML, Conti F, Rosa AL (2018) Differences between indigenous yeast populations in spontaneously fermenting musts from *V. vinifera L.* and *V. labrusca L.* grapes harvested in the same geographic location. Front Microbiol 9:1320

Raymond Eder ML, Conti F, Bely M, Masneuf-Pomarède I, Albertin W, Rosa AL (2019) *Vitis* species, vintage, and alcoholic fermentation do not drive population structure in *Starmerella bacillaris* (synonym *Candida zemplinina*) species. Yeast. https://doi.org/10.1002/yea.3385

Rosa AL, Miot-Sertier C, Laizet Y, Salin F, Sipiczki M, Bely M, Masneuf-Pomarede I, Albertin W (2018) Draft genome sequence of the *Starmerella bacillaris* (syn., *Candida zemplinina*) type strain CBS 9494. Microbiol Res Announc. https://doi.org/10.1128/MRA.01116-18

Sampaio P, Correia A, Gusmão L, Amorim A, Pais C (2007) Sequence analysis reveals complex mutational processes for allele length variation at two polymorphic microsatellite loci in *Candida albicans*. In: Méndez-Vilas A (ed) Communicating current research and educational topics and trends inapplied microbiology. FORMATEX 2007, Spain, pp 926–935

Singh RP, Manchanda G, Maurya IK, Rai AR (2019) Isolation, identifcation and whole genome sequencing of *Meyerozyma guilliermondii* strain RP-YS-11. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/1723551440. Accessed 23 Jan 2020

Sternes PR, Lee D, Kutyna DR, Borneman AR (2016) Genome sequences of three species of *Hanseniaspora* isolated from spontaneous wine fermentations. Genome Announc. https://doi.org/10.1128/genomeA.01287-16

Sternes PR, Lee D, Kutyna D, Borneman AR (2017) A combined metabarcoding and shotgun metagenomic analysis of spontaneous wine fermentation. Gigascience. https://doi.org/10.1093/gigascience/gix040

Valdes J, Tapia P, Cepeda V, Varela J, Godoy L, Cubillos FA, Silva E, Martinez C, Ganga MA (2014) Draft genome sequence and transcriptome analysis of the wine spoilage yeast *Dekkera bruxellensis* LAMAP2480 provides insights into genetic diversity, metabolism and survival. FEMS Microbiol Lett 361:104–106

Vigentini I, Maghradze D, Petrozziello M, Bonello F, Mezzapelle V, Valdetara F, Failla O, Foschino R (2016) Indigenous gregorian wine-associated yeasts and grape cultivars to edit the wine quality in a precision oenology perspective. Front Microbiol. https://doi.org/10.3389/fmicb.2016.00352

Wrent P, Rivas EM, Peinado JM, de Silóniz MI (2016) Development of an affordable typing method for *Meyerozyma guilliermondii* using microsatellite markers. Int J Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2015.10.008

Yang HL, Liao YY, Zhang J, Wang XL (2019) Comparative transcriptome analysis of salt tolerance mechanism of *Meyerozyma guilliermondii* W2 under NaCl stress. Biotech. https://doi.org/10.1007/s13205-019-1817-2

Zainudin RA, Oslan SN, Sabri S, Salleh AB, Abu A (2018) *Meyerozyma guilliermondii* strain SO, whole genome shotgun sequencing project. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/1727427301. Accessed 23 Jan 2020