doi 10.18699/bgrs2024-1.1-17

Absence of 5mC in thermo-tolerant yeast and telomere-to-telomere assembly by nanopore technology

Eremin A.*, Sergeev A., Malyushev D., Rodin V., Panova T., Polyakov I., Zvereva M.
Lomonosov Moscow State University, Moscow, Russia
** andreyeremin9903@gmail.com*

Key words: long-read sequencing; promethion; ogataea parapolymorpha dl-1; assembly improvement; comparative genomics; methylation

Motivation and Aim: A thermotolerant methylotrophic yeast (Ogataea parapolymorpha DL-1) has wide-ranging applications in biotechnology, from protein production to methanol metabolism studies [1, 2]. Despite its utility, gaps exist in our understanding of its genome and methylome, particularly in 5mC DNA modification and genome structure. This study seeks to resolve these gaps using long-read sequencing technology, aiming to provide a comprehensive genome assembly and an in-depth methylome analysis for 5mC. We hypothesize that this approach will reveal novel insights into the yeast's genomic architecture and regulation, which could enhance its biotechnological applications [3].
Methods and Algorithms: Nanopore sequencing technology (long-read sequencing technology of PromethION) was used to obtain T2T assembly of Ogataea parapolymorpha DL-1 genome and its methylation patterns. The process involved extracting DNA, preparing libraries, and sequencing on a PromethION device. The raw data underwent basecalling, trimming, quality checks, genome assembly and polishing. The resulting assembly was compared with previously reported draft [4] and methylation patterns of 6mA, 5mC and 5hmC were analyzed.
Results: The study resulted in a comprehensive T2T assembly of the Ogataea parapolymorpha DL-1 genome. The final assembly includes seven chromosomes and one mitochondrial genome. No plasmids were assembled from the reads. Methylation analysis revealed an absence of 5mC across the entire genome of Ogataea parapolymorpha DL-1, with only 0.01 % modifications noted, which aligns with potential sequencing errors rather than true methylation events due to 5hmC impossibly exceeding it's precursor (Table 1).

Table 1. Comparative analysis of total number modifications (5mC, 5hmC and 6mA) found with dorado. Percentage in brackets shows the part from all of the A or C in the genome

	5mC	5hmC	6mA
Ogataea parapolymorpha DL-1	123 (0.0057 %)	614 (0.0286 %)	21050 (0.9 %)
Escherichia coli	24181 (1.68 %)	1669 (0.12 %)	82863 (5.92 %)

Sequence motif TCCACCA was found in the regions ±10 pb from 6mA sites in Ogataea parapolymorpha DL-1 (Table 2). The site itself wasn't methylated suggesting that it is used by methyltransferases to modify adenosines nearby. In the validation experiment with *E. coli* known motifs GmATC and CmC(T/A)GG were found as previously reported [5].

Table 2. Modifications analysis. Motifs found with MEME in regions ±10 pb from methylated sites

	Sequence of motif	Number of motifs in regions ±10 pb from methylated sites	E-value
Ogataea parapolymorpha DL-1	TCCACCA	875	4.2e-021
Escherichia coli	GmATC	41833	1.2e-384
Escherichia coli	CmC(T/A)GG	21357	1.9e-1156

Table 3. Count of non-telomeric dT addition derived from reads aligned to the 3'end of the chromosomes

Region	A_end	T_end	perc_A	perc_T	N_total_reads
chr1:990461-990594	3	37	5.17	63.79	58
chr2:994069-994173	3	78	2.8	72.9	107
chr3:1276403-1276589	6	117	3.87	75.48	155
chr4:1299816-1299982	13	126	7.3	70.79	178
chr5:1332697-1332871	6	107	4.05	72.3	148
chr6:1517270-1517420	4	68	4.26	72.34	94
chr7:1515437-1515541	10	69	8.7	60	115

Note. A_end – number of Adenosines at the 3'end, T_end – number of Timidines respectively. perc_A – percentage of 3'end aligned reads containing dA at the end. N_total_reads – total reads, aligned to the 3'end.

Uniform distribution of 6mA modifications was observed across the chromosomes, with an average density of approximately 2 modifications per kilobase. Telomere analysis indicated that non-telomeric dT additions were present at the 3' ends of all chromosomes showing an unusual regulatory mechanism (Table 3).

The study also confirmed the presence of the previously reported [4] GGTGGCGG telomere motif at the 3' ends and identified a CCACCCCG motif at the 5' ends, which hasn't been reported before. The lengths of telomeric regions were found to be extended compared to earlier draft assembly.

Conclusion: Our comprehensive investigation into the Ogataea parapolymorpha DL-1 genome through long-read sequencing technology has advanced our understanding of this organism's genomic architecture and methylation patterns. By achieving a telomere-to-telomere assembly, we have closed previous gaps and also extended the known lengths of chromosomes, offering a complete and continuous genomic map. The absence of 5-methylcytosine across the genome suggests unique methylation dynamics and an evolutionary divergence from typical eukaryotic methylation patterns. These findings underscore the power of long-read sequencing to reveal detailed and accurate genomic features that are often obscured by short-read sequencing technologies. This work not

only enriches our genomic knowledge of Ogataea parapolymorpha DL-1 but also contributes to the broader understanding of genomic structure and stability in eukaryotes, paving the way for future studies on genomic and epigenetic diversity. Genomic assembly derived in this study as well as the sequencing reads are available at NCBI under the BioProject ID PRJNA1091144.

Funding: The study is supported by the project of the Interdisciplinary Scientific and Educational School of Moscow State University "Molecular technologies of living systems and synthetic biology" (# 23-Sh04-45).

References

1. Liebal U.W., Fabry B.A., Ravikrishnan A. et al. Genome-scale model reconstruction of the methylotrophic yeast Ogataea polymorpha. *BMC Biotechnol*. 2021;21:23. doi 10.1186/s12896-021-00675-w
2. Smekalova E.M., Malyavko A.N., Zvereva M.I. et al. Specific features of telomerase RNA from Hansenula polymorpha. *RNA*. 2013;19:1563-1574. doi 10.1261/rna.038612.113
3. Xie L., Yu W., Gao J. et al. Ogataea polymorpha as a next-generation chassis for industrial biotechnology. *Trends Biotechnol*. 2024:S0167-7799(24)00086-6. doi 10.1016/j.tibtech.2024.03.007
4. Ravin N.V., Eldarov M.A., Kadnikov V.V. et al. Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1. *BMC Genomics*. 2013;14:837. doi 10.1186/1471-2164-14-837
5. Kahramanoglou C., Prieto A.I., Khedkar S. et al. Genomics of DNA cytosine methylation in Escherichia coli reveals its role in stationary phase transcription. *Nat Commun*. 2012;3:886. doi 10.1038/ncomms1878