DESIGNING OF CUSTOM BARCODES FOR SEQUENCING ON THE MGI PLATFORM

Shmitko AO✉, Bulusheva IA, Vasiliadis IA, Suchalko ON, Syrko DS, Belova VA, Pavlova AS, Korostin DO

Pirogov Russian National Research Medical University, Russia

The MGI (MGI Tech Co. Ltd., China) next-generation sequencing platform, including the DNBSEQ-G50, -G400, and -T7 sequencers, is being actively adopted in research. Despite its widespread adoption, challenges persist in the form of limitations associated with the manufacturer's provided barcode set for library preparation. These limitations include constraints on the number of samples that can be concurrently sequenced, compatibility issues with barcodes from diverse or incomplete sets, and restrictions on the sample ratio. Purpose: to develop a universal method that allows sequencing of up to 252 samples simultaneously on a single sequencer lane, while eliminating barcode-related limitations. We proposed a “quad method” that provides 4 or 4*n*+2 equilibration of barcodes. This paper also delves into its comprehensive analysis, verification procedures, seamless integration into the sequencing process and validation of the method on the DNBSEQ G-400 platform. The quad method showed efficiency and reliability, allowing sequencing of up to 252 samples simultaneously without compromising data quality. The proposed method optimizes library preparation and improves the flexibility of sequencing on the MGI platform.

Keywords: DNBSEQ, MGI, BGI, NGS, custom barcodes, custom indexes, custom adapters

Funding: this research was funded by the grant №075-15-2019-1789 from the Ministry of Science and Higher Education of the Russian Federation allocated to the Center for Precision Genome Editing and Genetic Technologies for Biomedicine.

Author contribution: Shmitko AO — study planning, data collection, writing Original Draft Preparation; Bulusheva IA — methodology, data analysis, writing Original Draft Preparation; Vasiliadis IA, Suchalko ON, Syrko DS — data analysis, software, visualization; Belova VA — study planning, manuscript review and editing; Pavlova AS — data curation, data analysis, software; Korostin DO — conceptualization, supervision, methodology, manuscript review and editing.

✉ **Correspondence should be addressed:** Anna O. Shmitko
Ostrovityanova, 1/1, Moscow, 117997, Russia; annashmi97@gmail.com

Received: 15.08.2024 **Accepted:** 17.09.2024 **Published online:** 10.10.2024

DOI: 10.24075/brsmu.2024.040

РАЗРАБОТКА КАСТОМНЫХ БАРКОДОВ ДЛЯ СЕКВЕНИРОВАНИЯ НА ПЛАТФОРМЕ MGI

А. О. Шмитко✉, И. А. Булушева, Ю. А. Василиадис, О. Н. Сучалко, Д. С. Сырко, В. А. Белова, А. С. Павлова, Д. О. Коростин

Российский национальный исследовательский медицинский университет имени Н. И. Пирогова, Москва

Платформа секвенирования следующего поколения MGI (MGI Tech Co. Ltd., Китай), включающая секвенаторы DNBSEQ-G50, -G400 и -T7, активно внедряется в научные исследования. Однако сохраняются ограничения, связанные с использованием стандартных баркодов, в частности, на количество одновременно секвенируемых образцов и на соотношение их количества, а также присутствует проблема совместимости баркодов из разных или неполных сетов. Целью работы было разработать универсальный метод, позволяющий секвенировать до 252 образцов одновременно на одной дорожке секвенатора, с устранением ограничений, связанных с баркодами. Мы предложили «метод четверок», обеспечивающий уравновешивание баркодов по принципу 4 или 4*n*+2. Проведена проверка метода на соответствие требованиям к баркодам для секвенирования, а также валидация метода на платформе DNBSEQ G-400. Метод четверок показал эффективность и надежность, позволяет секвенировать до 252 образцов одновременно без снижения качества данных, он оптимизирует процесс подготовки библиотек и повышает гибкость секвенирования на платформе MGI.

Ключевые слова: кастомные баркоды, кастомные индексы, кастомные адаптеры, DNBSEQ, MGI, BGI, NGS

Финансирование: соглашение о предоставлении из федерального бюджета грантов в форме субсидий в соответствии с пунктом 4 статьи 78.1 Бюджетного кодекса Российской Федерации на осуществление государственной поддержки создания и развития центра геномных исследований мирового уровня «Центр высокоточного редактирования и генетических технологий для биомедицины» No 075-15-2019-1789 от 22.11.2019.

Вклад авторов: А. О. Шмитко — планирование исследования, сбор данных, подготовка рукописи; И. А. Булушева — методология, анализ данных, подготовка рукописи; Ю. А. Василиадис, О. Н. Сучалко, Д. С. Сырко — анализ данных, программное обеспечение, визуализация; В. А. Белова — планирование исследования, рецензирование и редактирование рукописи; А. С. Павлова — курация данных, анализ данных, программное обеспечение; Д. О. Коростин — концептуализация, супервизия, методология, рецензирование и редактирование рукописи.

✉ **Для корреспонденции:** Анна Олеговна Шмитко
ул. Островитянова, д. 1, с. 1, г. Москва, 117997, Россия; annashmi97@gmail.com

Статья получена: 15.08.2024 **Статья принята к печати:** 17.09.2024 **Опубликована онлайн:** 10.10.2024

DOI: 10.24075/vrgmu.2024.040

MGI Tech is a relatively new player in the NGS market, founded in 2016 as a subsidiary of BGI Group [1–3]. The company's first sequencing platform, the MGISEQ-2000, was introduced in 2017, followed by the MGISEQ-200RS and MGISEQ-T7 platforms. MGI produces a range of sequencers based on the DNA nanoball technology and cPAS sequencing [4]. It allows for sequencing in single-end or paired-end mode using single or dual barcode conditions. The technology involves barcoding of samples during the ligation of adapters containing barcode sequences. DNA library barcoding is necessary for labeling sequences from different biological samples and read identification during the transformation of temporary sequencing files into the commonly used fastq format. The length of MGI barcodes is 10 bp.

The standard kits for library preparation and sequencing with the mid-throughput sequencer DNBSEQ G-400 are designed for single-indexed sequencing, whereas the dual barcoding mode is optional and requires purchasing additional kits. Currently, MGI provides a kit that includes 96 barcode adapters for the ligation step in DNA library preparation for single-end sequencing. In addition, MGI lists 32 barcode sequences for synthesis.

The G-400 system is sensitive to nucleotide balance at each cycle of barcode sequencing, as the quality drastically

drops if the same position in the barcode sequences from the same lane is occupied by the same nucleotide. This explains why the barcode set from the same lane should meet the criteria for combining their sequences and enable generating compatible sets. The set of 96 adapters provided by MGI allows for forming 11 balanced sets (2 of 4, 8 of 8, and 1 of 24). In practice, however, it is often necessary to combine the samples containing barcodes from different sets, change the number of the samples loaded on a lane, and vary their ratio. In the laboratory routine, it is not uncommon to encounter scenarios where one or several DNA libraries fail to meet the quality control standards in the final stage. To address this issue, a flexible approach to combining samples simplifies the task of pooling libraries for loading to the lane. Additionally, the task of combining samples with different required amounts of output data, such as exomes with different coverage of ×200, ×100, must also be considered. Therefore, the manufacturer imposes limitations on the users of this platform, providing a small number of barcodes and sets, which thus prevents uncovering its true potential for sequencing. This may prove critical when selecting a sequencing platform. Custom solutions for various applications have been developed for the Illumina platform [5–7], whereas for the MGI platform, such solutions have not yet been provided.

We previously developed software that allows choosing the optimal combination of provided barcodes at various ratios and sample numbers for MGI adapter sets [8]. The updated software, including custom barcodes, is available in the GitHub repository (https://github.com/genomecenter/BC-store/tree/custom-adapter-sets). Other software for selecting a balanced ratio of barcodes, depending on the sequencing tasks, has been developed earlier for Illumina NGS Instruments [9–11].

The purpose of this paper is to present the algorithm we have developed, that can generate the required number of barcode sequences for a given study. Using this algorithm, we designed 252 barcodes, forming 63 balanced sets, each comprising 4 barcodes, and allowing any set to be combined with the others.

METHODS

Method formulation and barcode selection

The sequencer has limits in terms of the intensity of the registered signal from the fluorophores corresponding to the nucleotides. If the same position of barcodes contains the same nucleotide, the read quality significantly drops, leading to errors in barcode identification and further assigning reads to the samples [8]. Therefore, we had to design barcodes to generate the most balanced combinations. The algorithm of sequence design is based on the "quad method," which involves adding three barcodes obtained by the consecutive substitutions of bases to each barcode from the MGI set (Fig. 1A, B).

Following this method, each of the 96 barcodes can serve as a root barcode for its quad, resulting in generating 96 * 4 = 384 unique barcodes.

As the percentage of each base at each position is 25%, the resulting combination is perfectly balanced and guarantees the highest quality of sequencing.

Verification of compliance with the criteria

Validation of the compatibility based on the balance

As each quad is perfectly balanced, any number of quads can be combined with each other. The ratio between the quads in a pool can vary; however, the ratios between barcodes in each quad should be equal.

Furthermore, we checked whether it was possible to generate pools containing $4n + 2$ barcodes, where n is the number of quads. We checked the compatibility using the BC-Store software by combining 10 barcodes (as shown in Fig. 2). The nucleotide fraction of each nucleotide at any position in a pool of 10 barcodes has the highest and lowest deviations equal to 0.2 and 0.3, respectively, and meets the criteria for a balanced combination. This is still valid when any of the two barcodes from the same quad are added to *n* quads at a ratio equal to or lower than in quads.

Verifying the compatibility of barcodes based on a mismatch number

At the next step, all quads were checked for compatibility by the number of mismatches. Each sample labeled by a barcode had to be uniquely identified, so the barcode sequences of a certain length should not overlap with others. We selected a threshold of 4 mismatches, as all 96 10 bp barcodes provided by the manufacturer differ by more than 4 bases. The analysis also included the 999 verification MGI barcode (a 10bp technical sequence present in the original software's demultiplexing file). We constructed a graph of incompatible quads (S1 Fig.) and, using an adjacency matrix (S2 Fig.), we selected 63 quads (252 barcodes) compatible with each other based on the number of permitted mismatches (Fig. 3). The sequences of all 252 barcodes are listed in S1 Table.

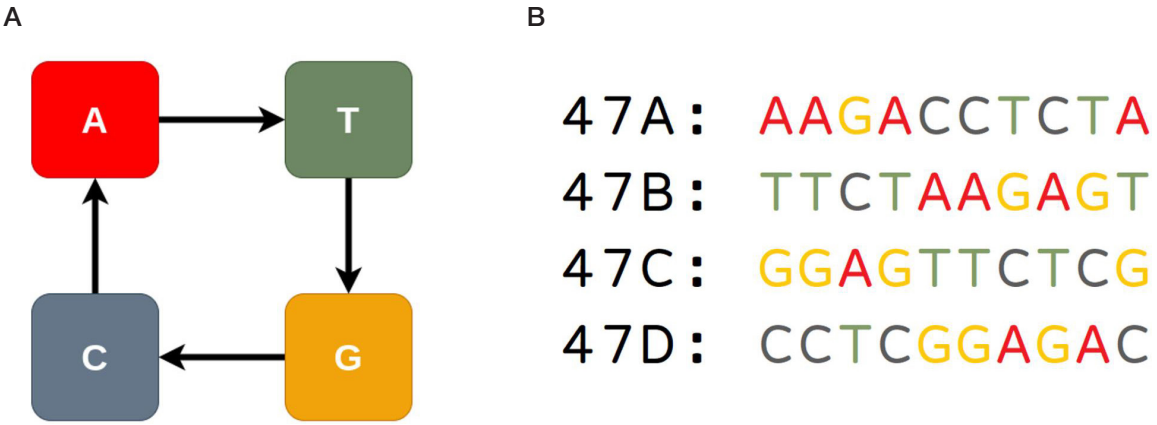

Fig. 1. A. The concept of the quad method. Each MGI barcode serves as a root for the quad, and custom barcodes are generated by sequential changes at each position of the original barcode: A→T, T→G, G→C, C→A. **B**. An example of a barcode quad. 47A is the original MGI barcode, and 47B, 47C, 47D are the custom barcodes generated using the quad method.

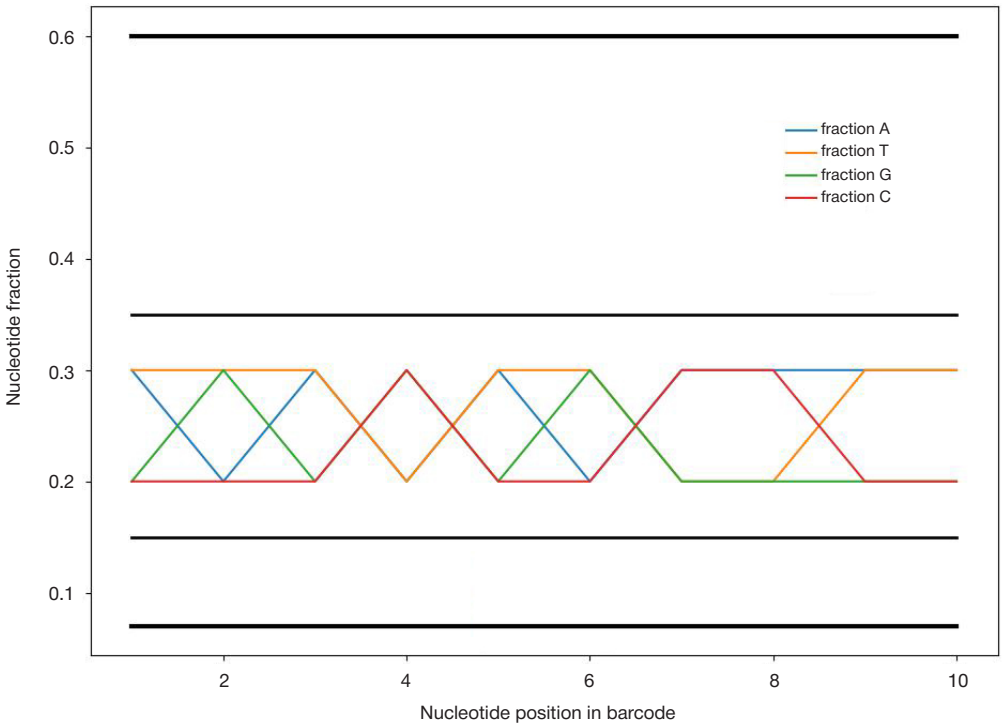

Fig. 2. The nucleotide balance in a pool of $4n + 2$ barcodes. The colored lines represent 295 the nucleotide fractions. The black strong lines represent the boundaries of the weak criterion. 296 The fine lines represent the boundaries of the strong criterion for barcode compatibility [8]

Validating the uniqueness

We checked if the sequences of the designed barcodes are present among the original MGI barcodes. This is necessary for generating the file containing barcodes for automatic demultiplexing. For this purpose, we created a Venn diagram showing the sets of custom and original MGI barcodes. We obtained 63 overlaps, where all 63 barcodes were original MGI barcodes, while the other 189 sequences were unique sequences not coinciding with the MGI barcodes from different kits (Fig. 4).

Preparation for sequencing

Adapter synthesis

According to the manufacturer's instructions, designing an individual adapter requires annealing two oligonucleotides.

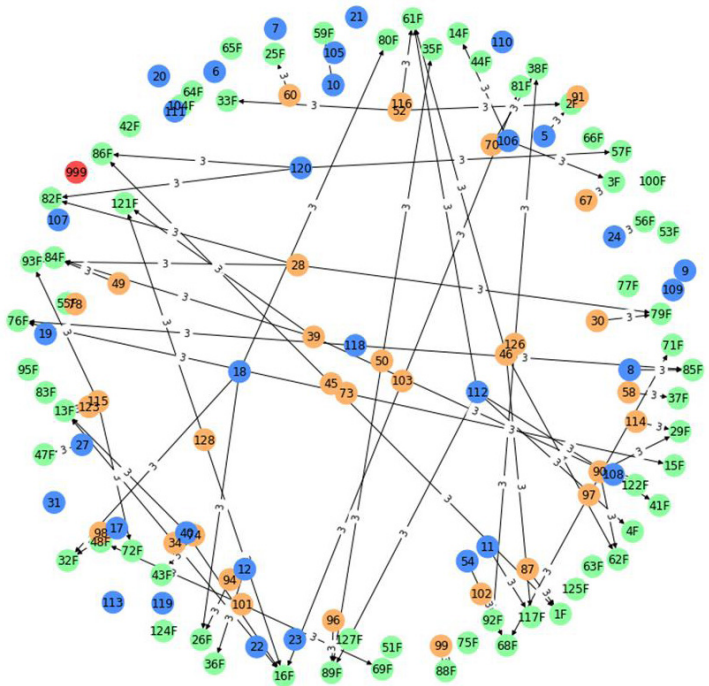

Fig. 3. The incompatibility graph showing 63 quads and MGI barcodes not included in 298 the quads. All quads that passed filtering based on the mismatch number are shown in green, 299 the barcodes from the set of 96 MGI barcodes not included in the quads are shown in orange, 300 the MGI barcodes from the set of 128 MGI barcodes are shown in blue, the 999- manufacturer 301 verification sequence is shown in red. The line connects the incompatible barcodes and quads; 302 the number above the line indicates the lowest number of mismatches between them

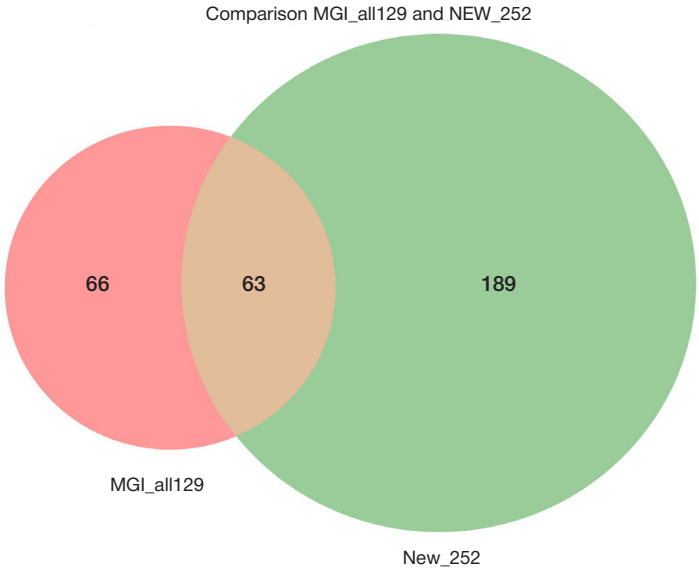

Fig. 4. Venn diagram for comparing the sequences of custom and original MGI barcodes from the 128 barcode set and 999 validation barcode provided by MGI. The custom barcodes are shown in green, the original MGI barcodes which do not overlap with quads are shown in red, and the MGI barcodes overlapping with quads (which were used as roots for the quads) are shown in orange

One of them (top oligonucleotide) contains the barcode sequence and a phosphate at the 5'-end (Ad153_5T_1-index # (1~128) according to the manufacturer), the sequence of bottom oligonucleotide is partially complementary to the top oligonucleotide (Ad153Ω_Bottom_2) (https://en.mgitech.cn/Download/download_file/id/71) [12].

The sequences of oligonucleotides containing the barcodes 1A-1D are shown in Table 1, all sequences containing 252 barcodes are listed in S1 Table.

The sequences of oligonucleotides containing the barcodes 1A-1D are shown in Table 1, all sequences containing 252 barcodes are listed in S1 Table.

To prepare the adapters, a mixture was created by adding 1 µL of 5M NaCl, 10 µL of 200 µM top oligonucleotide, and 10 µL of 200 µM bottom oligonucleotide to 79 µL of LowTE buffer. The mixture was then heated at 95 °C for 2 minutes and gradually cooled to 17°C at a rate of 0.5 °C every 30 seconds.

The algorithm of uploading new barcodes to a sequencer

To automatically demultiplex the sequenced libraries and following the MGI's recommendations, we created a .csv file (S2 Table) containing barcode sequences, including new custom barcodes, the original MGI barcodes, and 999-validation barcode. MGI barcodes that were included in the quads had an nA structure, where n is an adapter number in the original MGI kit, while custom barcodes had *n*B, *n*C, *n*D structure according to the order of quad formation. The format of the original MGI barcodes not included in quads remained unchanged. The barcode numbers were separated from the barcode sequences using commas without spaces.

RESULTS

To validate the designed barcodes, we prepared libraries with the synthesized custom adapters (Evrogen). The libraries, prepared following the standard MGI protocol, were pooled and enriched using the SureSelect Human All Exon v7 kit [13] and then sequenced in the PE100 mode using the DNBSEQ G-400 machine. Fastq demultiplexing was performed by the software built in G-400 provided by MGI basecalllite based on the uploaded file containing the barcode sequencing data. By default, the algorithm considers a read "undecoded" if there are two or more mismatches in a 10 bp barcode sequence. Therefore, the fraction of undecoded reads can be used as a quality metric for the performance of DIY barcode adapters. We compared the fraction of undecoded reads in the complete data from each lane with custom barcodes (44 lanes) and the data from previous runs (44 lanes) that employed MGI barcodes. On average (mean ± SD), the fractions of undecoded reads per lane were 1.08 ± 0.19% and 1.68 ± 0.22% for the MGI adapters and custom adapters, respectively (Fig. 5). Although the proportion of undecoded data increased when utilizing custom barcodes compared to the original barcodes (T = 13.5, df = 83, *p*-value = 1.17E-22), the absolute value relative to the total data output from a single lane is considered to be negligible. The values of undecoded and full data in GB are presented in S3 Table.

Thus, we have developed a viable approach for designing custom barcodes that allows for simultaneous sequencing of more than 96 samples on MGI called the 'quad method.' We obtained 189 custom barcodes that can be combined with the 63 MGI barcodes to generate 63 balanced quads. One barcode from each quad is an original MGI barcode (*n*A, where *n* is a number of an original barcode), and the other three are custom barcodes (*n*B, *n*C, *n*D) that complement it.

Table 1. The list of barcode sequences and full sequences of top and bottom oligonucleotides for adapter preparation. Ad_Bttm -bottom oligo

Name	Barcode/index sequence	Top oligo sequence
1A	TAGGTCCGAT	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAATAGGTCCGATCAACTCCTTGGCTCACA
1B	GTCCGAACTG	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAAGTCCGAACTGCAACTCCTTGGCTCACA
1C	CGAACTTAGC	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAACGAACTTAGCCAACTCCTTGGCTCACA
1D	ACTTAGGTCA	/5Phos/AGTCGGAGGCCAAGCGGTCTTAGGAAGACAAACTTAGGTCACAACTCCTTGGCTCACA
Ad_Bttm		TTGTCTTCCTAAGGAACGACATGGCTACGATCCGACTT

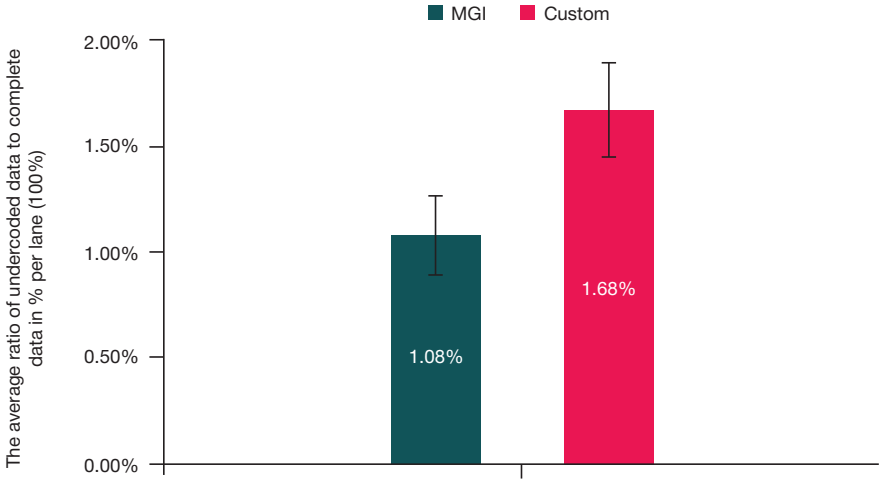

Fig. 5. The average ratio of undecoded reads (%) and complete data (Gb) per lane for 309 the libraries with MGI barcodes (in blue, data from 40 lanes) and with custom barcodes (red, 310 data from 7 lanes)

These quads can be combined with each other at any ratio and number as long as the ratio between the barcodes from the same quad remains equal. It is possible to create library pools with $4n + 2$ barcodes, where n is a number of quads, which can include any two barcodes from the other quad. In this case, the fraction of the last two barcodes should not exceed the fractions of the others.

DISCUSSION

The MGI platform is designed for fast, high-throughput sequencing, offering undeniable benefits yet prone to limitations. We attempted to overcome certain limits resulting from the solutions and kits provided by the manufacturer. Our approach allows for improving the efficiency of sequencing and expanding the possibilities of the MGI platform. However, it is important to bear in mind that the combinations of the quads with some original MGI adapters not included in the quads can fail to meet the compatibility criterion for the mismatch number. That is why we recommend checking whether they are balanced using the BC-Store software. We assume that a higher value of undecoded reads may be related to the insufficient purity of the synthesized oligonucleotides compared to MGI [14]. Previously, we ordered the synthesis of identical barcodes from two different manufacturers and observed that, in the case of one of them, the proportion of undecoded reads was elevated.

CONCLUSIONS

The custom barcodes we devised enable the alteration of the ratio and the number of libraries loaded to a lane depending on the purpose and required data amount. Using the BC-store software we had earlier developed, the libraries can be more easily and quickly pooled for sequencing on the MGI NGS instruments, both in paired-end or single-end modes. Therefore, given all the advantages and disadvantages of this method, it can be used as an additional or alternative solution to the solution provided by MGI.

References

1. Jeon SA, Park JL, Park SJ, Kim JH, Goh SH, Han JY, et al. Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. Genes Genomics. 2021; 43: 713–24. DOI: 10.1007/s13258-021-01096-x.
2. Kim HM, Jeon S, Chung O, Jun JH, Kim HS, Blazyte A, et al. Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing. Gigascience. 2021; 10. DOI: 10.1093/gigascience/giab014.
3. Xu Y, Lin Z, Tang C, Tang Y, Cai Y, Zhong H, et al. A new massively parallel nanoball sequencing platform for whole exome research. BMC Bioinformatics. 2019; 20. DOI: 10.1186/s12859-019-2751-3.
4. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science (1979). 2010; 327: 78–81. DOI: 10.1126/science.1181498.
5. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. 2014. Available from: http://www.microbiomejournal.com/content/2/1/6.
6. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony RK. Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS One. 2015; 10. DOI: 10.1371/journal.pone.0128036.
7. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Appl Environ Microbiol. 2013; 79: 5112–20. DOI: 10.1128/AEM.01043-13.
8. Bulusheva I, Belova V, Nikashin B, Korostin D. BC-store: A program for MGISEQ barcode sets analysis. PLoS One. 2021; 16. DOI: 10.1371/journal.pone.0247532.
9. Faircloth BC, Glenn TC. Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. PLoS One. 2012; 7. DOI: 10.1371/journal.pone.0042543.
10. Trébeau C, De Monvel JB, Tai FWJ, Petit C, Etournay R. DNA Barcode Compatibility: An R-package for optimizing DNA-barcode combinations in multiplex sequencing experiments. Bioinformatics. 2019; 35: 2690–1. DOI: 10.1093/bioinformatics/bty1030.
11. Varet H, Coppée JY. CheckMyIndex: A web-based R/Shiny interface for choosing compatible sequencing indexes. Bioinformatics. 2019; 35: 901–2. DOI: 10.1093/bioinformatics/bty706.
12. Oligos and primers for BGISEQ/DNBSEQ/MGISEQ - MGI Tech. Available from: https://en.mgitech.cn/Download/download_file/id/71.
13. SureSelectXT HS Target Enrichment System For Illumina Multiplexed Sequencing Platforms, Version D1, July 2021. Available from: https://www.agilent.com/cs/library/usermanuals/public/G7530-90000.pdf.

14. Li Q, Zhao X, Zhang W, Wang L, Wang J, Xu D, et al. Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform. BMC Genomics. 2019; 20. DOI: 10.1186/s12864-019-5569-5.

Литература

1. Jeon SA, Park JL, Park SJ, Kim JH, Goh SH, Han JY, et al. Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. Genes Genomics. 2021; 43: 713–24. DOI: 10.1007/s13258-021-01096-x.
2. Kim HM, Jeon S, Chung O, Jun JH, Kim HS, Blazyte A, et al. Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing. Gigascience. 2021; 10. DOI: 10.1093/gigascience/giab014.
3. Xu Y, Lin Z, Tang C, Tang Y, Cai Y, Zhong H, et al. A new massively parallel nanoball sequencing platform for whole exome research. BMC Bioinformatics. 2019; 20. DOI: 10.1186/s12859-019-2751-3.
4. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science (1979). 2010; 327: 78–81. DOI: 10.1126/science.1181498.
5. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. 2014. Available from: http://www.microbiomejournal.com/content/2/1/6.
6. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony RK. Inexpensive multiplexed library preparation for megabase-sized genomes. PLoS One. 2015; 10. DOI: 10.1371/journal.pone.0128036.
7. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Appl Environ Microbiol. 2013; 79: 5112–20. DOI: 10.1128/AEM.01043-13.
8. Bulusheva I, Belova V, Nikashin B, Korostin D. BC-store: A program for MGISEQ barcode sets analysis. PLoS One. 2021; 16. DOI: 10.1371/journal.pone.0247532.
9. Faircloth BC, Glenn TC. Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. PLoS One. 2012; 7. DOI: 10.1371/journal.pone.0042543.
10. Trébeau C, De Monvel JB, Tai FWJ, Petit C, Etournay R. DNA Barcode Compatibility: An R-package for optimizing DNA-barcode combinations in multiplex sequencing experiments. Bioinformatics. 2019; 35: 2690–1. DOI: 10.1093/bioinformatics/bty1030.
11. Varet H, Coppée JY. CheckMyIndex: A web-based R/Shiny interface for choosing compatible sequencing indexes. Bioinformatics. 2019; 35: 901–2. DOI: 10.1093/bioinformatics/bty706.
12. Oligos and primers for BGISEQ/DNBSEQ/MGISEQ - MGI Tech. Available from: https://en.mgitech.cn/Download/download_file/id/71.
13. SureSelectXT HS Target Enrichment System For Illumina Multiplexed Sequencing Platforms, Version D1, July 2021. Available from: https://www.agilent.com/cs/library/usermanuals/public/G7530-90000.pdf.
14. Li Q, Zhao X, Zhang W, Wang L, Wang J, Xu D, et al. Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform. BMC Genomics. 2019; 20. DOI: 10.1186/s12864-019-5569-5.