



HAL
open science

Fully in vitro iterative construction of a 24 kb-long artificial DNA sequence to store digital information

Julien Leblanc, Olivier Boulle, Emeline Roux, Jacques Nicolas, Dominique Lavenier, Yann Audic

► **To cite this version:**

Julien Leblanc, Olivier Boulle, Emeline Roux, Jacques Nicolas, Dominique Lavenier, et al.. Fully in vitro iterative construction of a 24 kb-long artificial DNA sequence to store digital information. *Biotechniques*, 2024, 76 (5), pp.203-215. 10.2144/btn-2023-0109 . hal-04567229

HAL Id: hal-04567229

<https://hal.science/hal-04567229v1>

Submitted on 24 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Fully *in vitro* iterative construction of a 24 kb-long artificial DNA sequence to store digital information

Julien Leblanc, Olivier Boule, Emeline Roux, Jacques Nicolas, Dominique Lavenier & Yann Audic

To cite this article: Julien Leblanc, Olivier Boule, Emeline Roux, Jacques Nicolas, Dominique Lavenier & Yann Audic (2024) Fully *in vitro* iterative construction of a 24 kb-long artificial DNA sequence to store digital information, BioTechniques, 76:5, 207-219, DOI: [10.2144/btn-2023-0109](https://doi.org/10.2144/btn-2023-0109)

To link to this article: <https://doi.org/10.2144/btn-2023-0109>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 04 Jun 2024.



[Submit your article to this journal](#)



Article views: 662



[View related articles](#)



[View Crossmark data](#)

Fully *in vitro* iterative construction of a 24 kb-long artificial DNA sequence to store digital information

Julien Leblanc¹ , Olivier Boule¹ , Emeline Roux² , Jacques Nicolas³, Dominique Lavenier¹  & Yann Audic^{*,4} 

¹University Rennes, Inria, CNRS, IRISA, Campus de Beaulieu, Rennes, France; ²Institut NuMeCan, INRAE, INSERM, University Rennes, France; ³University Rennes, Inria, CNRS, IRISA, Campus de Beaulieu, Rennes, France; ⁴CNRS, University Rennes, Institut de Génétique et Développement de Rennes (IGDR) UMR 6290, Rennes, France;

*Author for correspondence: yann.audic@univ-rennes.fr

BioTechniques 76: 207–219 (May 2024) 10.2144/btn-2023-0109

First draft submitted: 10 November 2023; Accepted for publication: 1 March 2024; Published online: 4 April 2024

ABSTRACT

In the absence of a DNA template, the *ab initio* production of long double-stranded DNA molecules of predefined sequences is particularly challenging. The DNA synthesis step remains a bottleneck for many applications such as functional assessment of ancestral genes, analysis of alternative splicing or DNA-based data storage. In this report we propose a fully *in vitro* protocol to generate very long double-stranded DNA molecules starting from commercially available short DNA blocks in less than 3 days using Golden Gate assembly. This innovative application allowed us to streamline the process to produce a 24 kb-long DNA molecule storing part of the *Declaration of the Rights of Man and of the Citizen* of 1789. The DNA molecule produced can be readily cloned into a suitable host/vector system for amplification and selection.

TWEETABLE ABSTRACT

Minion-controlled *ab initio* production of long DNA molecules of predefined sequences from commercially available short DNA blocks in less than 3 days.

MULTIDISCIPLINARY ABSTRACT

DNA molecules are easily copied from pre-existing DNA molecules. However, in the absence of a pre-existing DNA template, the *ab initio* production of long DNA molecules of chosen sequences is particularly challenging and precludes the usage of such molecule for DNA digital data storage. In this report we propose a protocol that is carried out exclusively in a reaction tube to generate very long DNA molecules starting from synthetic short DNA blocks in less than 3 days using Golden Gate assembly. This innovative application allowed us to streamline the process to produce a 24 kb-long DNA molecule storing part of the *Declaration of the Rights of Man and of the Citizen* of 1789.

METHOD SUMMARY

We developed, using Golden gate assembly, an iterative DNA assembly pipeline to produce long double-stranded DNA molecules of predefined sequences. This pipeline relies on the orderly assembly of multiple 500-nt commercial fragments into medium sized 5-kb BigBlocks that are then assembled in an iterative process. PCR allowed us to select and amplify the full-size molecules. The DNA construction steps were analyzed using single-molecule sequencing, demonstrating that our protocol is highly effective in obtaining correctly assembled molecules. We used this pipeline to store part of the *Declaration of the Rights of Man and of the Citizen* of 1789 on DNA.

KEYWORDS:

DNA assembly • DNA storage • long read sequencing • Oxford nanopore technology • synthetic DNA

For billions of years, double-stranded DNA (dsDNA) molecules have been the molecular support of choice for the storage of biological information and the support of life. In biological systems, DNA replication is a core biological process that occurs at high speed in prokaryotes (~700 nucleotide [nt]/s [1]). In eukaryotes, the process is slower (~15–30 nt/s [2]) but so highly parallelized that it allows to achieve replication of a 1.7-billion nt genome in less than 30 min in early *Xenopus* embryos (~9.4 × 10⁵ nt/s). It is also particularly efficient for the synthesis of collinear DNA molecules of up to hundreds millions of nucleotides, such as lungfish chromosomes [3]. This is made possible by the faithful copy of pre-existing nucleic acid polymers and template-specific DNA polymerases [4]. On the other hand, the *ab initio* production of DNA molecules of only thousands of nucleotides of defined sequence remains a technological and scientific challenge. Thus far, commercial companies generally offer cost-effective chemical synthesis of oligonucleotides up to a few tens of nucleotides and advertise longer DNA molecules of desired sequences for up to 50,000 nt [5,6], albeit at a substantial cost (Table 1). Assemblies of impressively long DNA molecules such as the mouse mitochondrial genome (16.7 kb), T7 bacteriophage

Table 1. Comparison of production costs for the construction of one dsDNA fragment of 24 kb using DNA parts from different suppliers.

Supplier		Integrated DNA Technologies™		Twist		GenArt		GenScript	
Product name		eBlock™	MiniGene	Gene Fragment	Clonal Gene	DNA Fragment	Cloned Genes	Economy	GenBrick
External parameters	Fragment type	ds DNA	ds DNA cloned	ds DNA	ds DNA cloned	ds DNA	ds DNA cloned	ds DNA	ds DNA cloned
	Fragment range size (kb)	0.3–1.5	0.5–5.0	0.3–1.8	0.3–5.0	0.2–3.0	0.1–12.0	<8.0	>8.0–50.0
	Starting price (€/bp)†	0.070	0.650	0.070	0.090	0.134	0.470	0.288	0.450–0.750
	Quantity of each fragment (μg)	0.2	4.0	0.1–1.0	0.1–1.0	0.2	5.0	4.0	4.0
	Shipped business days (n)	10	12	9	15	8	32	10	23
Internal parameters	Minimum number of blocks to assemble	16 blocks of 1.5 kb	5 blocks of 5.0 kb	14 blocks of 1.8 kb	5 blocks of 5.0 kb	8 blocks of 3.0 kb	2 blocks of 12.0 kb	3 blocks of 8.0 kb	1 block of 24 kb
	Assembly wetlab days	3	2	3	2	2	2	2	0
	DNA price (€)	1680	15,600	1680	2160	3220	11,280	6900	10,800
	DNA + reactions + sequencing price (€)	1940	15,840	1940	2400	3455	11,520	6140	10,800

† 2023 prices.
Initial DNA parts are either obtained as linear DNA molecules or as DNA molecules cloned into a vector.

(39.9 kb) or even the *Mycoplasma genitalium* genome (583.0 kb) are documented in the literature, but all rely on hierarchical assembly and molecular cloning steps which are time-consuming and laborious [7–9]. Furthermore, analysis and assessment of the accuracy of the assemblies relies on their cloning into an appropriate vector, which must be transferred into a host and biologically selected. A less laborious procedure has yet to be defined, aimed at producing *in vitro* and cost-effectively long dsDNA molecules of predefined sequences in the absence of a DNA template.

This limitation is unfortunate considering the numerous applications of long DNA molecules in synthetic biology [10,11]. Long synthetic DNA molecules are also a subject of interest in experimental biology, as they can provide insights into several areas of research such as the reconstruction of ancestral DNA genes, regulatory elements and proteins [12] or the investigation of alternative splicing regulation [13]. Moreover, the production of large artificial or semi-artificial DNA molecules, including mutants or variants, is a widespread practice for the remodelling of genetic circuits, the evaluation of gene function, and the identification of functional domains within DNA or encoded proteins [6]. In addition, these DNA molecules are also useful as templates for RNA vaccine production [14]. Their main industrial applications concern *in vitro* screening of mutations for the development of therapeutics and chemical products, including drugs and biofuels [10,15].

Another important use of long DNA molecules takes root in their dense and stable information content. Current data-storage systems, whether magnetic or silicon based, are indeed unable to respond to the rapid increase in archiving needs and have several drawbacks such as a limited lifespan, energy consumption, miniaturization restrictions and environmental impact. DNA molecules are nowadays envisaged as an alternative to store data owing to their tremendous information density and high chemical stability, as evidenced by the recovery and analysis of ancient DNA extracted from fossils [16,17]. However, the main bottleneck for the storage of information on DNA is the DNA synthesis step itself, which is slow, expensive and can only generate short oligonucleotides [18,19]. The reduced size of the oligonucleotides implies fragmenting the digital documents into a large number of small pieces, which must necessarily include indexes allowing the reconstruction of the original document [20]. The size of this index can be important compared with the size of a DNA oligonucleotide and thus significantly affects the effective amount of information stored and the synthesis costs. Incidentally, this fragmentation will also increase the difficulty in recovering the original documents. On the other hand, single-molecule real-time sequencing (Pacific Biosciences, CA, USA) or nanopore sequencing (Oxford Nanopore Technologies, Oxford, UK) are now capable of reading individual long DNA molecules, which is compatible with storing information on longer DNA fragments [21–23]. Such long DNA molecules will only contain a small fraction of indexing information, making them more archival efficient.

Many methods have been developed to construct large DNA fragments from short chemically synthesized DNA oligonucleotides [6,15,24]. The most common are Golden Gate assembly (GGA), Gibson cloning and polymerase cycling assembly. These methods are all based on user-defined overlapping ends, defined as overhangs, and generally allow the seamless joining of five to ten fragments per reaction. The assembly is then cloned into a vector, transferred to a host, selected and amplified.

Herein we propose a fully *in vitro* iterative method based on GGA to faithfully construct a long synthetic DNA molecule starting from commercially available dsDNA. The peculiarity of this method is that it does not rely on the biological properties of the assembled DNA molecule for its selection and therefore the synthetic DNA molecules could be used for any common biological purpose or to store information. The various assembly steps of our procedure were evaluated on an application for storing part of the *Declaration of*

the *Rights of Man and of the Citizen* (26 August 1789) in a single long DNA fragment that was then sequenced using Oxford nanopore technology (ONT) to retrieve the original text. The different steps of the DNA construction were analyzed using ONT long read sequencing technology to quantify the ordering of the building blocks, demonstrating that our protocol is highly effective in obtaining correctly assembled molecules.

Methods

Encoding of a binary file into DNA alphabet

We encoded the first articles of the *Declaration of the Rights of Man and of the Citizen* as a binary string of 4.2 Ko (Supplementary data file 1). The binaries were first randomized by performing a bitwise Exclusive OR operation between the input binary and the output of the hash function (SHA-256) over a constant in a process similar to what is described in [25]. The randomized binary is converted into DNA alphabet using an algorithm that respects synthesis rules of the DNA blocks. Briefly, $60 > \%GC > 40$, no homopolymer >3 nt (Supplementary data file 5). This encoding allows alteration of the DNA sequence at the β -encoded bits to ensure the absence of *BsaI* restriction sites and of inverse repeat regions longer than 10 nt. The 23,400 nt-long DNA sequence (Figure 1C & Supplementary data file 1) is partitioned into 40 blocks of 472 nt (eight internals eBlocks™ for each of the five BigBlocks) and ten blocks of 452 nt (one eBlock for each BigBlock extremity). Each internal block is then framed by a DNA sequence composed of a 15-nt buffer sequence and a specific *BsaI* restriction site. Similarly, each external block is framed by a DNA sequence that includes a 20-nt region following the *BsaI* restriction site, which serves as a target for PCR selection of the BigBlock. Digestion with *BsaI* generates a 5' overhang of 4 nt, which allows ordered assembly. Overhangs are selected with the NEBridge® GetSet™ Tool. Overhangs used are CGCT, ACGA, GCAA, CACC, CCTA, CGGA, TGAA, ACTC and ACAT. Primers used for amplification of assembly products were defined with ITHOS, a submodule of Genofrag [26], with parameters defined in Supplementary data file 6. A 20-nt flanking sequence composed of a *BsaI* site and a 9-nt external buffer sequence was added to each primer to allow for the assembly of BigBlocks to MaxiBlocks (primer sequences are available in Supplementary data file 7). The sequence was controlled to comply to Integrated DNA Technologies™ (IA, USA) eBlock synthesis rules.

DNA fragments

A total of 50- DNA blocks were ordered as 524 bp-long eBlocks from Integrated DNA Technologies. The eBlocks were delivered in a 96 well plate at 10 nmol/ μ l in TE buffer, pH 8.0 (10 mM Tris-HCl/0.1 mM EDTA). Each eBlock was quantified with an AccuGreen™ High Sensitivity dsDNA Quantitation Kit on a Qubit® Fluorometer (Q32851, Invitrogen™, France), aliquoted and stored at -20°C . The eBlock integrities were individually verified by electrophoretic analysis on 3% agarose-TBE gels stained with GelRed® (Biotium, CA, USA).

BigBlock assemblies

Each of the five BigBlocks were comprised of ten oriented eBlocks described above. The assembly reaction was conducted in a 25.0- μ l assembly reaction mix comprised of 1.0- μ l NEBridge Golden Gate Assembly Kit *BsaI*-HF® v2 (NEB E1601), 2.5 μ l T4 DNA ligase buffer 10 \times (NEB B0202), 0.04 pmol of each eBlock. Reactions were conducted for 65 cycles (37°C , 5 min; 16°C , 5 min) and stopped by incubation at 60°C for 5 min. The expected assembly size was 4764 bp.

BigBlocks were PCR-amplified with dedicated primer pairs in a 25.0- μ l PCR reaction mix comprised of 0.25- μ l Q5® Hot Start High-Fidelity DNA Polymerase 2 U/ μ l (NEB® M0493), 5.0 μ l Q5 5 \times buffer (NEB B9027), 0.75 μ l dNTP 10 mM Mix (NEB N0447), 1.25 μ l of each primer (10 μ M) and 1.5 ng of BigBlock as template. Amplification was conducted at 98° for 30 s; 5 \times (98°C , 10 s; *Thyb*, 20 s; 72°C , 2 min 30 s) followed by 10 \times (98°C , 10 s; 72°C , 2 min 30 s) and a terminal extension at 72°C for 2 min. The annealing temperature (*Thyb*) was 62°C , except for BigBlock 1 at 66°C and BigBlock 3 at 66°C . The expected size of each amplified BigBlock was 4796 bp. BigBlock assemblies and PCR products were analyzed by electrophoresis on 1–2% agarose-TBE gels stained with GelRed.

MaxiBlock assembly

The MaxiBlock was assembled from the five BigBlocks (BigBlock 1–5) in a 20.0- μ l reaction mix comprised of 2.0 μ l NEBridge Golden Gate Assembly Kit *BsaI*-HF v2 (NEB E1601), 2.0 μ l T4 DNA ligase buffer 10 \times (NEB B0202) and 200 ng of each BigBlock. The reaction was conducted 65 \times (37°C for 5 min; 16°C for 5 min) and stopped by incubation at 60°C for 5 min. The MaxiBlock size was 23,796 bp.

The MaxiBlock was PCR amplified with MaxiBlock_PCR_Fw and MaxiBlock_PCR_Rv primers in a 25.0- μ l reaction comprised of 1.0 μ l LongAmp® Hot Start Taq DNA polymerase, 2.5 U/ μ l (NEB M0534), 5.0 μ l LongAmp Taq 5 \times buffer (NEB B0323), 0.75 μ l dNTP 10 mM mix (NEB N0447), 0.25- μ l primers pair 5.0 μ M and a 1.0-ng MaxiBlock assembly. Amplification was conducted at 94° for 30 s; 15 \times (48°C for 20 s; 55°C for 20 s; 65°C for 22 min) and a terminal extension at 65°C for 15 min. The expected size of each amplified MaxiBlock was 23,796 bp.

MaxiBlock assembly and PCR products were analyzed on TapeStation 4200 (Agilent Technologies, CA, USA) following the supplier protocol using Genomic DNA reagents (Agilent Technologies, 5067–5366) loaded onto a Genomic DNA ScreenTape (Agilent Technologies, 5067–5365).

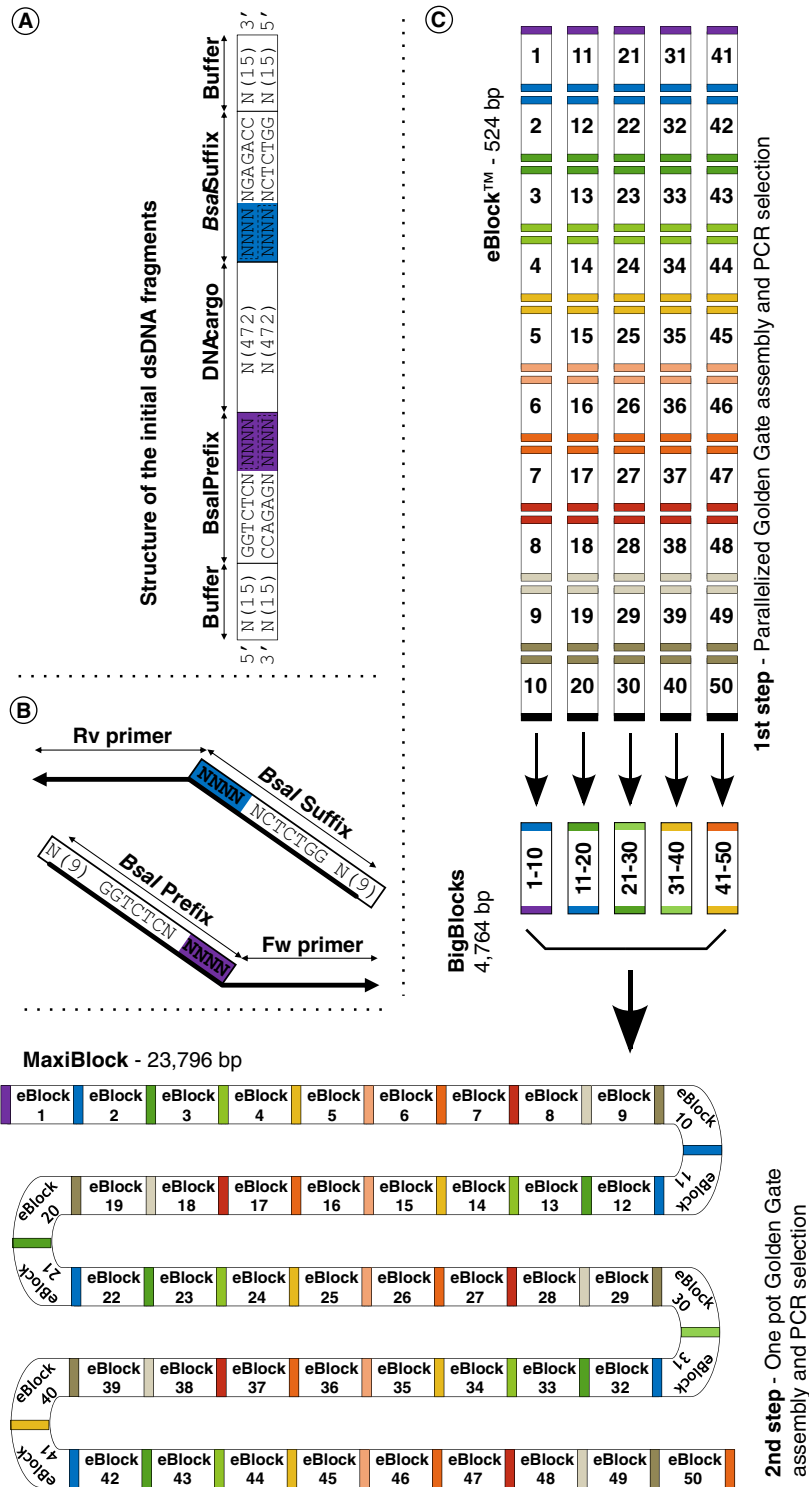


Figure 1. Overview of the strategy for building a 23,796-bp dsDNA from short dsDNA fragments. (A) dsDNA building blocks are comprised of a DNA cargo framed by 2 *BsaI* sites and a buffer sequence at the extremities. The colored tetranucleotides correspond to the overhang produced by the *BsaI* digestion during assembly. (B) Oligonucleotides are comprised from 5' to 3' of a 9-nt buffer sequence, a *BsaI* site and a BigBlock specific sequence. (C) Two-step strategy for building a 23,796-bp dsDNA. Five sets of ten eblocks are assembled into Bigblocks (4764 bp) in parallel using GGA. The expected products are selected, amplified and decorated with *BsaI* sites using specific primer pairs described in B. Bigblocks are mixed in equimolar amounts before proceeding to GGA. The final 23,796-bp product is selected and amplified by PCR.

DNA sequencing

BigBlock and MaxiBlock assemblies and PCR products were sequenced using Oxford Nanopore Technologies MinION Mk1C and GridION Mk1 sequencer hardware, respectively. The sequencing libraries were prepared according to the manufacturer protocol (version: NBE_9065_v109_revAD_14Aug2019) with a Ligation Sequencing Kit (Oxford Nanopore Technologies, SQK-LSK109) and PCR-free Multiplexing Native Barcoding Kit (Oxford Nanopore Technologies, EXP-NBD104). Sequencing libraries were quantified with AccuGreen High Sensitivity dsDNA Quantitation Kit on a Qubit Fluorometer (Invitrogen Q32851) and loading on a Flowcell Flongle (Oxford Nanopore Technologies, FLO-FLG001) for a run time of 24 h.

Data processing

Raw sequencing data were preprocessed using Oxford Nanopore Technologies Guppy software (V6.0.1 and V6.3.9) [27] for super accuracy base calling (see configuration file: Supplementary data file 8) and demultiplexing (front and rear score set >60). The sequencing quality of each demultiplexed sample was controlled with Nanoplot (V1.38.0) [28]. Scripts are available in Supplementary data file 9. Sequencing data have been made available on the European Nucleotide Archive under accession number PRJEB62556.

Bioinformatics results analysis

Each quality-passed long read was analyzed to identify the order and identity of the blocks that comprised it. We used the Smith-Waterman algorithm [29] to test the local alignment of a read to all 50 original block sequences. The highest scoring block was identified and removed from the read. The alignment procedure was then recursively applied to the other parts of the read before and after the aligned subsequence of the read, until the entire read was identified or not identified in blocks. To quantify the association between each block, we determined whether each individual block was followed by an identified block, an unrecognized block or the end of the read. To analyze the global assembly, we classified reads as either *correct* if all the blocks comprising it were correctly ordered, or as *incorrect* if at least one block was misplaced.

Results & discussion

Articles 1 through 9 of the *Declaration of the Rights of Man and of the Citizen* were encoded into a 23.4 kb-long DNA sequence (Supplementary data file 1). To fully build this dsDNA molecule *in vitro*, we started from 50 commercially available 524 nt-long dsDNA molecules (eBlocks) that we assembled into 4764 bp-long dsDNA molecules (BigBlocks) that were then assembled into a 23,796-bp DNA molecule (MaxiBlock; Figure 1). This fully *in vitro* strategy used commercial dsDNA molecules as building blocks, each being a few hundred nucleotides long (eBlocks). The design of these building blocks relied on an architecture comprised of, starting from the end, a 15 nt-long buffer sequence that ensured the integrity of the two *BsaI* prefix and suffix sequences, the four base cleavage sites that allowed directional ligation, and finally the cargo part of the DNA containing the encoded information (Figure 1A). Upon cleavage by *BsaI*, the DNA cargo part remained solely framed by the predefined 4-nt overhangs to allow ordered assembly of ten blocks in a single GGA reaction. Specific primer pairs targeting the extremities of the assembled BigBlocks (Figure 1B) were used to select and amplify the 4764 bp-long assembly products and flank them with sequence-specific type IIS (*BsaI*) restriction sites to allow for the second *in vitro* assembly to generate the MaxiBlocks (Figure 1C). As described in Supplementary data file 2, the different assembly steps were controlled during the course of the experiment. The 23,796-bp MaxiBlocks were PCR amplified before utilization.

First assembly step: eBlocks to BigBlock

To design a 23,796 bp-long DNA molecule, we split the sequence into a total of 50 524-bp parts, which were first assembled in groups of ten to generate five 4764-bp BigBlocks. At the time of our study, purchasing the 524-bp eBlocks at Integrated DNA Technologies was the most cost-effective way to obtain these long dsDNA molecules (Table 1). The five GGA reactions aiming to obtain 4764 bp-long DNA molecules were directly controlled on agarose gels (Figure 2A). Incomplete sequential ligation occurred and a scale of DNA molecules ranging from 480 bp, the initial size of the eBlocks after *BsaI* digestion, to 4764 bp by incremental steps of approximately 500 bp could clearly be observed. We selected the desired 4764-bp assembly using PCR, targeting the extremities of the BigBlocks. As presented in Figure 2B, size fractionation of the PCR product demonstrated that for each of the five BigBlocks, we could specifically amplify the longer 4764-bp DNA molecules. Interestingly, even BB2 that appeared on the gel to have a lower amount of full-length assembled molecules could be very efficiently selected and amplified by PCR.

This demonstrated that approximately 5 kb-long dsDNA molecules can be effectively assembled *in vitro* from commercial 524-bp dsDNA molecules. However, in the absence of functional testing of the DNA molecules, this did not demonstrate that the fragments were correctly ordered in the assembly.

To determine to which extent the assembly process was ordered, we took advantage of ONT sequencing technology to sequence raw assembly products and raw PCR products for BigBlock 1 and 4. We reasoned that single-molecule long-read sequencing technologies such as ONT or PacBio were the most effective at obtaining sequence information about the accuracy of long collinear assemblies. For the assembly of BigBlock 1 (Figure 3A) or BigBlock 4 (Figure 3B), the graphics represent the percentage of transition from one block (n) to another block, to the end of the read (end) or to an unrecognized sequence (unknown). BigBlock 1 graphics were produced from 4935 reads with a median length of 1435 nt and a median quality of 10.9. For BigBlock 4, graphics were generated from 5281 reads with

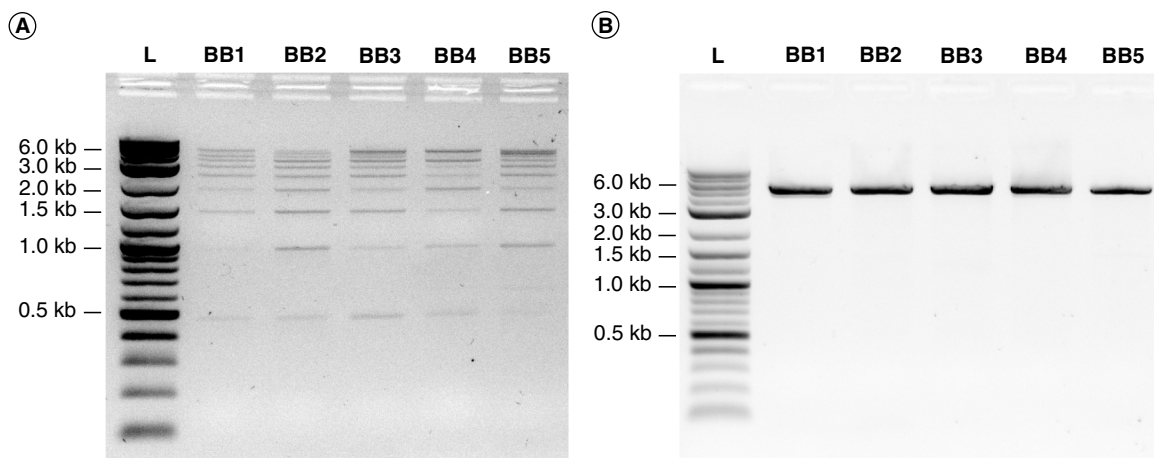


Figure 2. BigBlock™ assemblies before and after PCR analyzed on 2% agarose electrophoresis gels. (A) Five different BigBlocks (BB1–BB5) were assembled from 50 eBlocks™ of 524 bp (Integrated DNA Technologies, IA, USA). Each BigBlock was comprised of ten oriented eBlocks. Expected size = 4764 bp. **(B)** BigBlock products from (A) were PCR-amplified with dedicated flanked primers. Expected size = 4796 bp. L: 1-kb Plus DNA ladder (NEB®).

a median length of 1898 nt and a median quality of 11.1 (Supplementary data file 3). While the median quality is in the expected range for Oxford nanopore sequencing, the median read lengths are about one-third of our expected product size but still in agreement with size fractionation (Figure 2), showing that the reaction is comprised of partially assembled molecules. Directly analyzing the assembly reaction products, we could note that the percentage of correct concatenations of two consecutive eBlocks ranged from 61 to 95% for the assembly of BigBlock 1 and from 74 to 94% for the assembly of BigBlock 4 (left panel in Figure 3A & B, respectively). In both cases and as expected, eBlock 10 was most generally terminal, at 99 and 96%, respectively. In the case of BigBlock 1 assembly, the lower efficiency of assembly of eBlock 5 was associated with either incorrect assembly of eBlock 5 to eBlock 7 (~10%) or eBlock 5 being terminal (~20%). Similarly, in the assembly of BigBlock 4, the lower level of concatenation of eBlock 34 to eBlock 35 was caused by eBlock 34 being terminal.

This analysis indicates a faithful assembly process, as generally the assembly is correct from eBlock (n) to eBlock (n + 1) or otherwise no assembly occurs.

We used PCR to select and amplify the full-length BigBlock assembly (Figure 2B). BigBlock 1 post-PCR graphics were produced from 3000 reads with a median read length of 4757 nt and a median read quality of 10.9. For BigBlock 4 post-PCR, graphics were generated from 1476 reads with a median length of 4752 nt and a median quality of 10.9 (Supplementary data file 3). PCR selection and amplification therefore led to an approximately threefold increase in median read lengths compared with pre-PCR products – 4757 nt compared with 1435 nt for BigBlock 1 and 4752 nt compared with 1898 nt for BigBlock 4. This median size was close to the 4796-nt size of the expected product. This illustrates how effective the PCR amplification step is in selecting full-length assembly products.

After PCR, the BigBlock sequencing results showed a frequency of correctly ordered eBlock pairs ranging from 87–99% for BigBlock 1 assembly and 92–100% for BigBlock 4 assembly, a neat increase in accuracy with respect to previous results (right panel in Figure 3A & B, respectively). The problematic assembly of eBlock 5 to eBlock 7 in BigBlock 1 before PCR remains at a similar level after, while the premature termination of the assembly at eBlock 5 becomes negligible as it cannot be amplified. Similarly, for BigBlock 4, only the premature termination of the assembly at eBlock 9 remains notable (7%), while being reduced by half compared with the assembly without PCR.

This shows that when an eBlock is erroneously terminal this does not affect the final PCR-selected product, while if two eBlocks are erroneously assembled they will remain present in the PCR-selected assembly. It is therefore important to maximize individual ligation efficiency.

While the previous analysis focused on the correct assembly of pairs of eBlocks, it was also important to estimate the overall accuracy of whole BigBlocks assembly. We used the above ONT sequencing data to estimate the distribution of the length of the DNA assembly and to determine the proportion of correct assembly for the different fragment sizes. This also enabled us to quantify the enrichment in correct assemblies following the PCR amplification and selection process (Figure 3C). As observed by gel electrophoresis, the reaction generated all product sizes between one and ten assembled eBlocks (0.5–5 kb). The target molecule comprised of ten eBlocks dominated the reaction products and represented 15 and 22% of the sequenced reads in BigBlock 1 and BigBlock 4 reactions, respectively, with fewer than 3% inaccurate assemblies. The proportion of molecules with more than ten eBlocks was negligible (<1%). Upon PCR amplification and selection, we observed a drastic depletion of BigBlocks comprised of fewer than ten fragments. Approximately 84% of the DNA was included in molecules comprised of ten fragments, almost all being correctly ordered (~99% for both BigBlock 1 and BigBlock 4).

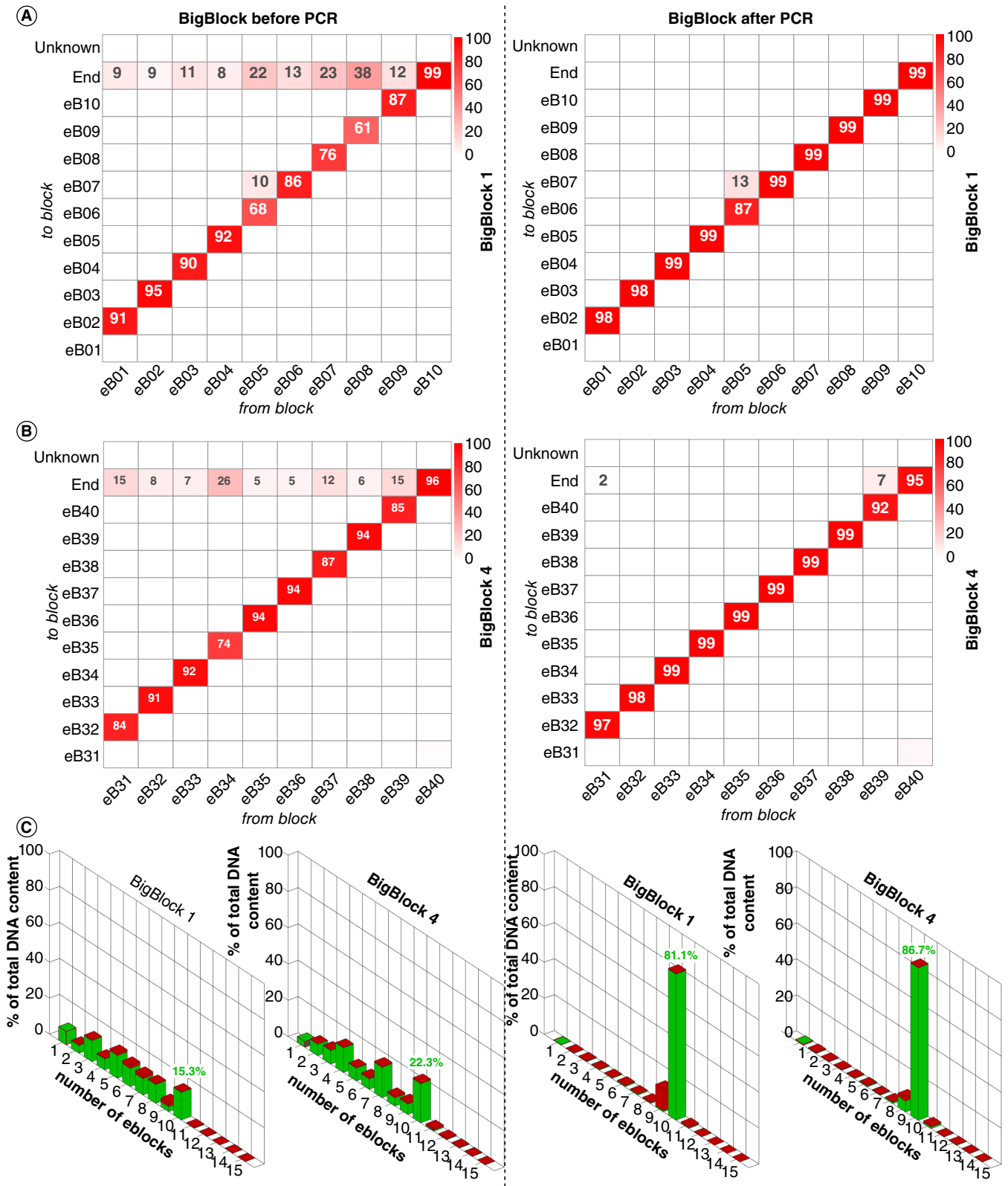


Figure 3. Analysis of the assembly of BigBlock 1 and BigBlock 4 before (left panel) and after (right panel) PCR selection. (A) Heatmap presenting in the assembly of BigBlock 1 the proportion (%) of two ligated consecutive eBlocks determined by Oxford nanopore technology sequencing. (B) Same as (A) for BigBlock 4. Sequencing data were from assemblies shown in Figure 2. Unknown: the eBlock is ligated to an unknown DNA sequence; End: the eBlock is terminal. (C) Distribution of DNA content in the assembly for BigBlock 1 and BigBlock 4 relative to the number of eBlocks comprising the reads. Green: all the eBlocks comprising a read are in the correct order; red: at least one eBlock is misplaced.

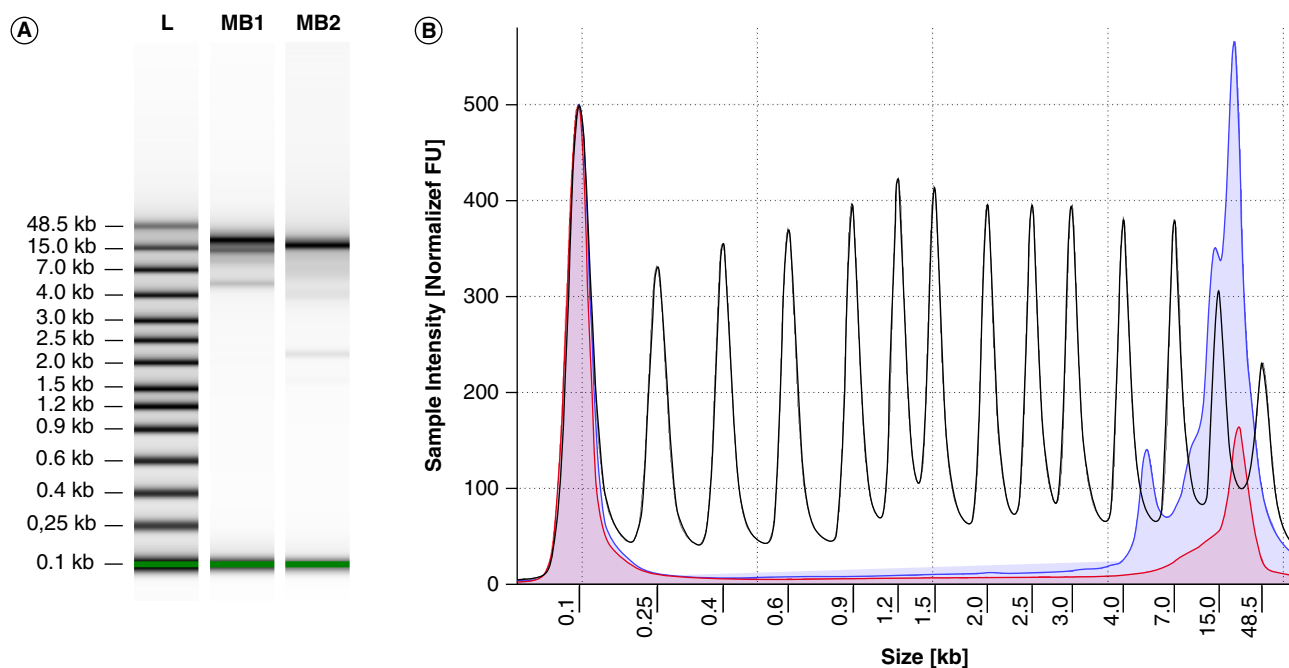


Figure 4. Results of MaxiBlock assembly before and after PCR, analyzed on an Agilent 2200 TapeStation with a Genomic DNA ScreenTape kit. (A) Gel electrophoresis and (B) associated electropherogram. (L) (Black) DNA ladder Agilent; (MB1) (Blue) FinalBlock assembly before PCR; (MB2) (Red) FinalBlock assembly after PCR. The expected size of the MaxiBlock was 23,796 bp.

Taken together, these analyses indicate that the combination of directed assembly without any intermediate purification step and with PCR selection allows the efficient production of long DNA molecules of the expected structure.

Second assembly step: BigBlocks to MaxiBlock

Having established the reliable production of 4796-nt dsDNA molecules (the BigBlocks), we set about assembling five of them into a 23,796-bp MaxiBlock. As shown above, the five BigBlocks were amplified using primers that introduced *Bsa*I restriction sites together with tetramer sequences required to correctly order the fragments (Figure 1B). The assembly reaction of the five BigBlocks shown in Figure 2 was performed using GGA. The reaction products were analyzed either directly after the reaction or after PCR amplification and selection using primers targeting the first (BigBlock 1) and the last (BigBlock 5) BigBlocks of the assembly. Because of the expected size of the DNA assembly, we analyzed the reaction products by capillary electrophoresis on a Tape Station (Agilent Technologies; Figure 4). It is noteworthy that the resolution of the Tape Station does not allow for precise sizing of the DNA molecules in the higher size range. The products of the ligation reactions (Figure 4A; MB1) ranged in size from 4 kb to more than 15 kb in a manner compatible with partial and complete assembled molecules in the reaction mix. After PCR amplification (Figure 4B; MB2) using primers targeting the extremity of the MaxiBlock and LongAmp Hot Start Taq (NEB) DNA polymerase, a single PCR product 15 and 48 kb in size was detected and in agreement with the expected 24-kb DNA product.

To better evaluate the quality of the assembled fragments, we again took advantage of ONT sequencing to compare the reaction products before and after amplification and selection of a 23,796-bp DNA assembly. ONT sequencing of the MaxiBlock assembly produced 25,507 reads with a median length of 3091 nt and a median quality of 11.7 before PCR and 3215 reads with a median length of 1143 nt and a median read quality of 11.9 after PCR (Supplementary data file 3). The limited median size of the sequencing reads of the MaxiBlock assembly takes its root in the presence of 1) numerous short sequencing reads comprised of part of terminal eBlocks 1 or 50; and 2) the presence of sequences that were ill-attributed during demultiplexing (Supplementary data file 4). As shown in Figure 5A (left panel), the proportion of correctly ordered BigBlock pairs ranged from 56 to 85%. As designed, BigBlock 5 is most often (97%) the terminal block. When BigBlock (n) is not ligated with BigBlock (n + 1), then it is usually a terminal block of the assembly. Upon PCR selection, the DNA molecules sequenced were mainly composed of an assembly of BigBlocks in the right order (Figure 5B, right panel). The proportion of correctly ordered BigBlock pairs ranged from 89 to 96% and BigBlock 5 was exclusively terminal. This indicates that upon PCR, full-length products were selected while a large proportion of partial assemblies were not amplified.

We analyzed the number of eBlocks and the correctness of their assembly for all sequenced reads. In the pre-PCR reaction products, a striking pattern of reads comprised of 10, 20, 30, 40 and 50 eBlocks were present and represented the assembly of 1, 2, 3, 4 or 5 BigBlocks together. They respectively represented 12, 4, 11, 4 and 14% of the DNA content of the reaction products. Most notably, the majority of these reads were comprised of correctly ordered fragments (Figure 5B).

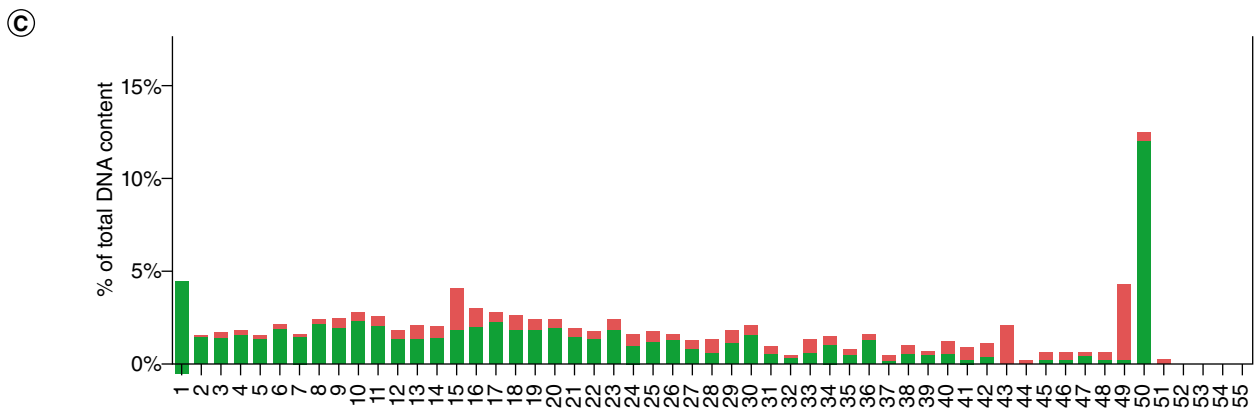
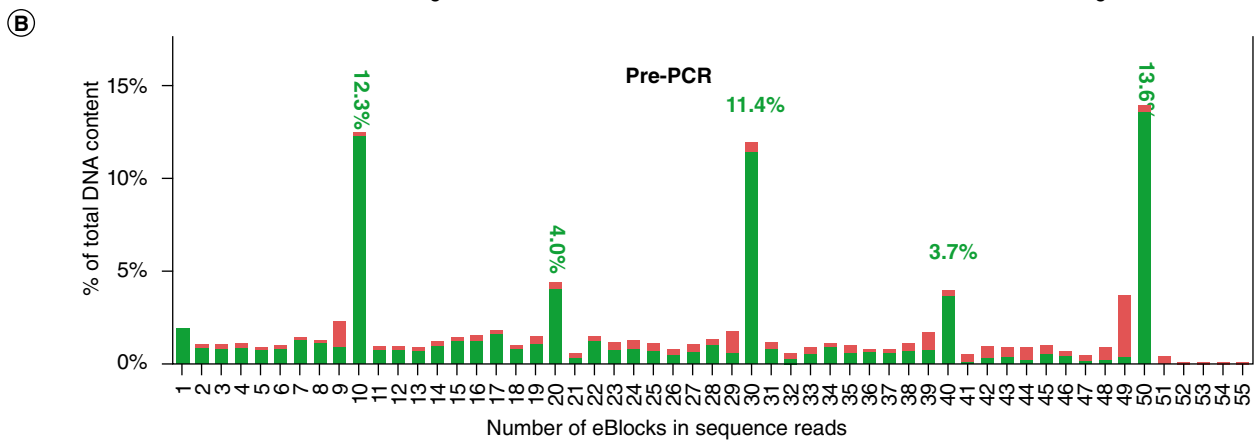
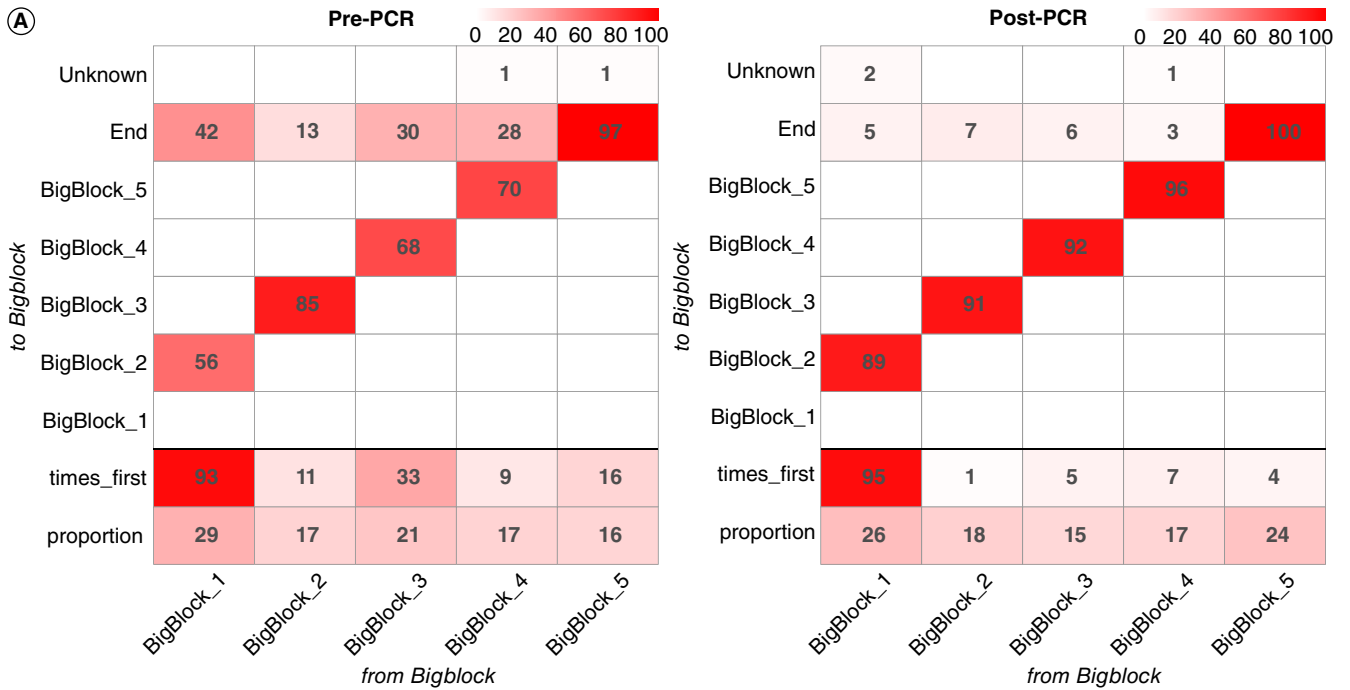


Figure 5. Oxford nanopore technology sequencing analysis of the MaxiBlock assembly. (A) Heatmap showing the proportion of pairs of linked BigBlocks, «from» block on the x-axis «to» block on the y-axis (values >1% are shown) before and after PCR selection. At the bottom of the table the overall proportion of each BigBlock in the set of BigBlocks appearing in the Maxiblocks (row «proportion», sums to 1) can be seen, and the fraction of Maxiblocks containing this BigBlock in the first position (row «times-first», relative to the proportion below). Sequencing data are from assemblies shown in Figure 4. Unknown: BigBlock is ligated to an unknown DNA sequence; end: BigBlock is terminal. **(B)** Distribution of the DNA content of the MaxiBlock before PCR selection assembly relative to the number of eBlocks comprising the reads. **(C)** Same as **(B)** after PCR selection. Green: all the eBlocks comprising a read are in the correct order; red: at least one eBlock is misplaced.

Table 2. Overview of DNA assembly methods for long fragment.

This publication	Pryor et al. [9]; Sikkema et al. [32]	Nozaki [31]	Kirchmaier et al. [33]	Yantsevich et al. [34]	Storch et al. [35]	Casini et al. [36]	Weber et al. [37]	Gibson et al. [38]
Origin of the DNA used for the assembly	Fully synthetic. Low synthesis constraints	DNA is from biological origin, amplified by PCR	DNA is from biological origin amplified by PCR	Fully synthetic	DNA is from biological origin, assembled by Gibson, amplified by PCR	DNA is from biological origin, amplified by PCR	DNA is from biological origin, amplified by PCR	Fully synthetic
Assembly method based on	Golden Gate assembly	iPac (<i>in vitro</i> packaging-assisted DNA assembly)	Golden GATEway Cloning (Golden Gate + Multisite Gateway)	Thermodynamically Balanced Inside-Out (TBI) PCR-based assembly	Biopart Assembly Standard for Idempotent Cloning (BASiC)	MODAL: Modular Overlap-Directed Assembly	MoClo: modular cloning system (based on Golden Gate)	Gibson method
Full <i>in vitro</i> assembly?	No	No	No	Yes	No	No	No	Yes
Assembly size	Up to 40.0 kb	Up to 48.5 kb	Up to 6.9 kb	Up to 0.7 kb	Up to 6.2 kb	Up to 5.0 kb	Up to 33.0 kb	Up to 16.3 kb
Nature of construction	T7 bacteriophage genome	Phage genome and plasmids	Vectors for transgene expression (brainbow 1.0)	<i>EGFP</i>	Expression vectors	Yeast and bacterial plasmids	Vectors	Mouse mitochondrial genome
Selection of assembled DNA	Phenotypic selection of host	Phenotypic selection of host	Phenotypic selection of host	By <i>in vitro</i> final PCR amplification	Phenotypic selection of host or colony PCR	Phenotypic selection of host or colony PCR	Phenotypic selection of host	Transformation in <i>E. coli</i> , DNA sequencing and PCR
Estimated process time	1–2 days	1–2 days	Not reported	1 day	Not reported	Not reported	Not reported	Not reported (many days)
Digital Object Identifier	10.1021/acssynbio.1c00525 - 10.1002/cptz1.882	10.1021/acssynbio.2c00419	10.1371/journal.pone.0076117	10.1177/2472630319850534	10.1021/sb500356d	doi:10.1093/nar/gkt915	doi:10.1371/journal.pone.0016765	doi:10.1016/B978-0-12-385120-8.00015-2

As done for the selection of BigBlocks, PCR on the MaxiBlock aimed to both amplify and select the full-length 23,796-bp assembly product. We used a single DNA polymerase (LongAmp Hot Start Taq, NEB M0534) for amplification of the MaxiBlock. When comparing the PCR-amplified products and the pre-PCR assembly reaction, there is a clear depletion of DNA molecules comprised of 10, 20, 30 and 40 fragments, while the full-length assembly product of 50 eBlocks represents 12% of the total DNA content in the mix (Figure 5C). As could be expected, the efficiency of production of the final 24-kb full-length MaxiBlock was lower than the efficiency of the 5-kb BigBlock assembly (12% compared with ~80%). Additional purification steps based on size fractionation could be envisaged to improve the purity of the final product at the expense of hands-on time. However, our analyses indicate that our fully *in vitro* iterative method could efficiently and faithfully construct long synthetic DNA molecules by assembling commercially available short dsDNA.

Conclusion

The cost of construction of the 23,796-bp molecule, including DNA and enzymatic reactions, is approximately US\$2000 without negotiation on suppliers' list prices (Table 1). One important aspect of our protocol is the relatively short hands-on time: once the DNA sequences have been designed and the materials received, the full-size construction process can be completed within 3 days. On day 1, eBlocks are controlled by agarose gel electrophoresis and quantified. The assembly of the eBlocks into BigBlocks is carried out overnight. On day 2, the assembled BigBlocks are controlled by agarose gel electrophoresis, followed by PCR selection. PCR products are then controlled by agarose gel electrophoresis and quantified by spectrometry. The MaxiBlock assembly is then constructed overnight. On day 3, the MaxiBlock assembly is controlled on a TapeStation and PCR amplified. The final PCR products are ultimately controlled on TapeStation. Higher throughput is easily achievable by robotizing and parallelizing the assembly [30].

Compared with some recent assembly methods (Table 2), our method for producing long DNA molecules offers improvements upon already published GGA-based assembly in three areas: 1) it is fully independent of pre-existing natural DNA molecules as it starts from chemically synthesized DNA; 2) it is conducted solely *in vitro*; and 3) the selection of the assembled DNA molecules does not rely on their biological properties. This differs from previous reports of long DNA assemblies that generally start from PCR products amplified from biological DNA that require some steps to be performed *in vivo* in bacteria or with the help of lambda phages, and usually relies on the biological properties of the reconstructed DNA to all for stringent selection such as plaques forming assays, antibiotic resistance or colorimetric assays [9,31]. However, at the same time these selections impose strong constraints on the DNA being assembled. The method presented here is easily generalizable to any other type of DNA molecule without any requirement for specific biological properties of the molecule, and as such is adapted to the storage of information or assembly of any gene sequence.

The length of the final DNA is limited by the capabilities of the DNA polymerase used for the final PCR selection. In fact, obtaining PCR products larger than 24 kb can be challenging with respect to efficiency and accuracy with standard DNA polymerase such as Taq polymerase. However, some DNA polymerases are known to have higher processivity, extension rate or error correction capabilities, making them more suitable for amplifying longer DNA fragments. LongAmp Hot Start Taq DNA polymerase from NEB has the theoretical capacity to amplify DNA fragments up to 35 kb in length, while Platinum™ SuperFi II DNA Polymerase from Invitrogen can amplify up to 40 kb and TaKaRa LA Taq® DNA Polymerase from Takara Bio (Shiga, Japan) can amplify up to 48 kb. However, it is important to keep in mind that amplifying long DNA fragments can be a difficult task, and successfully obtaining a PCR product larger than 24 kb will depend on various factors such as DNA quality, reaction conditions and the optimization of PCR conditions.

The final PCR product can be directly used for molecular biology processes such as *in vitro* transcription with T7 RNA polymerase if a T7 promoter was included in the PCR primers. It must be highlighted that this product is not clonal and therefore contains a population of different molecules with potential point substitutions or deletions. To obtain a clonal product, the DNA molecule should be purified by performing agarose gel electrophoresis and excising the appropriate sized band, and cloned into a suitable host/vector system for amplification and selection. The size of the final product (~24 kb) remains compatible with cloning into plasmid or cosmid vectors. The *in vitro* process described here allows for the time-efficient construction of long dsDNA molecules. The absence of any subcloning, transformation and selection step allow us to streamline the process of long dsDNA molecule construction. Yet, they can be readily subcloned and selected on the basis of a double PCR screen using two pairs of primers located in the vector backbone and the building blocks at the extremity. Therefore, we feel this protocol may be an approach of choice to obtain artificial building blocks of thousands of nucleotides long that can be used in downstream applications.

Future perspective

The *ab initio* synthesis of DNA molecules of fully artificial composition is a major challenge in DNA digital data storage but also for the reconstruction of ancestral DNA molecules or the production of molecules encoding human or AI-designed proteins. The alternatives for artificial DNA production are scarce and rely either on chemical oligonucleotide synthesis or on enzymatic synthesis of DNA. However, both methods are limited by the maximal product size and must therefore leverage assembly methods to generate longer DNA molecules. The automation of these assembly methods will probably allow for the mass production of several kb long DNA molecules of chosen sequences.

Executive summary

Background

- To define and evaluate a method of assembly of long DNA molecules of artificial sequence.

Experimental

- We devised an iterative method for the *in vitro* two-step construction of a 24 kb-long artificial sequence DNA molecule.
- The assembly process was monitored using single-molecule sequencing on a minion sequencer.

Results & discussion

- This study devised an iterative method that allows for the synthesis of long DNA molecules to be used for digital data storage, with our study producing a DNA molecule encoding the *Declaration of the Rights of Man and of the Citizen* of 1789.
- Evaluation of the assembly steps by minion sequencing allowed us to rapidly evaluate the faithfulness of the assembly process.

Conclusion

- Our fully *in vitro* iterative method can faithfully construct long synthetic DNA molecules by assembling commercially available short dsDNA.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2023-0109

Author contributions

O Boule: Investigation, formal analysis; D Lavenier: Supervision, funding acquisition, project administration, conceptualization; J Nicolas: Conceptualization, review and editing; J Leblanc: Investigation, conceptualization, review and editing; E Roux: Conceptualization, review and editing; Y Audic: original draft preparation, review and editing, supervision, conceptualization

Acknowledgments

The authors gratefully acknowledge Edouard Cadieu and Thomas Derrien for their help with Oxford nanopore technology sequencing and data processing.

Financial disclosure

J Leblanc is supported by Labex CominLabs (dnarXiv project) and O Boule is supported by PEPR MolecularXiv (ANR-22-PEXM-003). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, stock ownership or options and expert testimony.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

References

1. McCarthy D, Minner C, Bernstein H, et al. DNA elongation rates and growing point distributions of wild-type phage T4 and a DNA-delay amber mutant. *J Mol Biol.* 1976;106(4):963–981. doi: 10.1016/0022-2836(76)90346-6
2. Yeeles JTP, Janska A, Early A, et al. How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol Cell.* 2017;65(1):105–116. doi: 10.1016/j.molcel.2016.11.017
3. Meyer A, Schloissnig S, Franchini P, et al. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature.* 2021;590(7845):284–289. doi: 10.1038/s41586-021-03198-8
4. Lehman IR. Discovery of DNA polymerase. *J Biol Chem.* 2003;278(37):34733–34738. doi: 10.1074/jbc.X300002200
5. Hoose A, Vellacott R, Storch M, et al. DNA synthesis technologies to close the gene writing gap. *Nat Rev Chem.* 2023;7(3):144–161. doi: 10.1038/s41570-022-00456-9
6. Hughes RA, Ellington AD. Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harb Perspect Biol.* 2017;9(1):a023812. doi: 10.1101/cshperspect.a023812
7. Gibson DG, Benders GA, Andrews-Pfannkoch C, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science.* 2008;319(5867):1215–1220. doi: 10.1126/science.1151721
8. Gibson DG, Smith HO, Hutchison CA, et al. Chemical synthesis of the mouse mitochondrial genome. *Nat Methods.* 2010;7(11):901–903. doi: 10.1038/nmeth.1515
9. Pryor JM, Potapov V, Bilotti K, et al. Rapid 40 Kb genome construction from 52 parts through data-optimized assembly design. *ACS Synth Biol.* 2022;1(6):2036–2042. doi: 10.1021/acssynbio.1c00525
10. Zhang X-E, Liu C, Dai J, et al. Enabling technology and core theory of synthetic biology. *Sci China Life Sci.* 2023;66:1742–1785. doi: 10.1007/s11427-022-2214-2
11. Kosuri S, Church GM. Large-scale *de novo* DNA synthesis: technologies and applications. *Nat Methods.* 2014;11(5):499–507. doi: 10.1038/nmeth.2918

12. Chang BSW. Ancestral gene reconstruction and synthesis of ancient rhodopsins in the laboratory. *Integr Comp Biol.* 2003;43(4):500–507. doi: 10.1093/icb/43.4.500
13. Tamayo A, Núñez-Moreno G, Ruiz C, et al. Minigene splicing assays and long-read sequencing to unravel pathogenic deep-intronic variants in PAX6 in congenital aniridia. *Int J Mol Sci.* 2023;24(2):1562. doi: 10.3390/ijms24021562
14. de Mey W, De Schrijver P, Autaers D, et al. A synthetic DNA template for fast manufacturing of versatile single epitope mRNA. *Mol Ther – Nucleic Acids.* 2022;29:943–954. doi: 10.1016/j.omtn.2022.08.021
15. Casini A, Storch M, Baldwin GS, et al. Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol.* 2015;16(9):568–576. doi: 10.1038/nrm4014
16. Matange K, Tuck JM, Keung AJ. DNA stability: a central design consideration for DNA data storage systems. *Nat Commun.* 2021;12(1):1358. doi: 10.1038/s41467-021-21587-5
17. Orlando L, Allaby R, Skoglund P, Der Sarkissian C, et al. Ancient DNA analysis. *Nat Rev Methods Primers.* 2021;1(1):14. doi: 10.1038/s43586-020-00011-0
18. Goldman N, Bertone P, Chen S, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature.* 2013;494(7435):77–80. doi: 10.1038/nature11875
19. Ezekannagha C, Becker A, Heider D, et al. Design considerations for advancing data storage with synthetic DNA for long-term archiving. *Mater Today Bio.* 2022;15:100306. doi: 10.1016/j.mtbi.2022.100306
20. Bornholt J, Lopez R, Carmean DM, et al. A DNA-based archival storage system. In: *Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems.* GA, USA: ACM; 2016. p. 637–649. doi: 10.1145/2872362.2872397
21. Song Z, Liang Y, Yang J. Nanopore detection assisted DNA information processing. *Nanomaterials (Basel).* 2022;12(18):3135. doi: 10.3390/nano12183135
22. Rhoads A, Au KF. PacBio sequencing and its applications. *Genom Proteom Bioinform.* 2015;13(5):278–289. doi: 10.1016/j.gpb.2015.08.002
23. Weirather JL, de Cesare M, Wang Y, et al. Comprehensive comparison of pacific biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res.* 2017;6:100. doi: 10.12688/f1000research.10571.2
24. Shevelev GY, Pyshnyi DV. Modern approaches to artificial gene synthesis: aspects of oligonucleotide synthesis, enzymatic assembly, sequence verification and error correction. *Vestn VOGIS.* 2018;22(5):498–506. doi: 10.18699/VJ18.387
25. Park S-J, Lee Y, No J-S. Iterative coding scheme satisfying GC balance and run-length constraints for DNA storage with robustness to error propagation. *J Commun Netw.* 2022;24(3):283–291. doi: 10.23919/JCN.2022.000008
26. Ben Zakour N, Gautier M, Andonov R, et al. GenoFrag: software to design primers optimized for whole genome scanning by long-range PCR amplification. *Nucleic Acids Res.* 2004;32(1):17–24. doi: 10.1093/nar/gkg928
27. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford nanopore sequencing. *Genome Biol.* 2019;20(1):129. doi: 10.1186/s13059-019-1727-y
28. De Coster W, D’Hert S, Schultz DT, et al. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34(15):2666–2669. doi: 10.1093/bioinformatics/bty149
29. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 1981;147(1):195–197. doi: 10.1016/0022-2836(81)90087-5
30. Storch M, Haines MC, Baldwin GS. DNA-BOT: A Low-Cost, Automated DNA Assembly Platform for Synthetic Biology. *Synthet Biol.* 2020;5(1):ysaa010. doi: 10.1093/synbio/ysaa010
31. Nozaki S. Rapid and accurate assembly of large DNA assisted by *in vitro* packaging of bacteriophage. *ACS Synth Biol.* 2022;11,12:4113–4122. doi: 10.1021/acssynbio.2c00419
32. Sikkema AP, Tabatabaei SK, Lee Y-J, et al. High-Complexity One-Pot Golden Gate Assembly. *Curr Protoc.* 2023;3(9):e882. doi: 10.1002/cpz1.8821
33. Kirchmaier S, Lust K, Wittbrodt J. Golden GATEway Cloning – A Combinatorial Approach to Generate Fusion and Recombination Constructs. *PLOS ONE.* 2013;8(10):e76117. doi: 10.1371/journal.pone.0076117
34. Yantsevich AV, Shchur VV, Usanov SA. Oligonucleotide Preparation Approach for Assembly of DNA Synthons. *SLAS Technol.* 2019;24(6):556–568. doi: 10.1177/2472630319850534
35. Storch M, Casini A, Mackrow B, et al. ANew Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology. *ACS Synth. Biol.* 2015;4:781–787. doi: 10.1021/sb500356d
36. Casini A, MacDonald JT, Jonghe JD, et al. One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res.* 2014;42(1):e7. doi: 10.1093/nar/gkt915
37. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE.* 2011;6(2):e16765. doi: 10.1371/journal.pone.0016765
38. Gibson DG. Enzymatic Assembly of Overlapping DNA Fragments. *Methods Enzymol.* 2011;498:349–361. doi: 10.1016/B978-0-12-385120-8.00015-2