Synchronized DNA sources for unconditionally secure cryptography

Sandra Jaudou, Hélène Gasnier, Elias Boudjella, Marc Canève, Victoria Bloquert, Vasily Shenshin, Tilio Pilet, Sacha Gaucher, Soo Hyeon Kim, Philippe Gaborit, et al.

▶ **To cite this version:**

Sandra Jaudou, Hélène Gasnier, Elias Boudjella, Marc Canève, Victoria Bloquert, et al.. Synchronized DNA sources for unconditionally secure cryptography. 2026. ⟨hal-05560338⟩

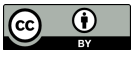
HAL Id: hal-05560338

https://hal.science/hal-05560338v1

Preprint submitted on 20 Mar 2026

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 CC BY 4.0 - Attribution - International License

Synchronized DNA sources for unconditionally secure cryptography

Sandra Jaudou*,[1] Hélène Gasnier*,[2] Elias Boudjella*,[3] Marc Canève,[2] Victoria Bloquert,[1] Vasily Shenshin,[1] Tilio Pilet,[2] Sacha Gaucher,[1] Soo Hyeon Kim,[3,4] Philippe Gaborit,[5] Gouenou Coatrieux,[2] Matthieu Labousse,[1] Anthony Genot,[3] and Yannick Rondelez[1,*]

[1] *Gulliver CNRS, ESPCI Paris, Université PSL, 75005 Paris, France*
[2] *IMT Atlantique, Inserm, LaTIM, Brest, France*
[3] *LIMMS, CNRS-Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan*
[4] *Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan*
[5] *XLIM, University of Limoges, Limoges, France*

Secure communication is the cornerstone of modern infrastructures, from finance and healthcare to defense and elections, yet achieving unconditional security—resistant to any computational attack—remains a fundamental challenge. The One-Time Pad (OTP), proven by Shannon to offer perfect secrecy, requires a shared random key as long as the message, used only once. However, distributing large keys over long distances has been impractical due to the lack of secure and scalable sharing options. Here, we introduce a DNA-based cryptographic primitive that leverages random pools of synthetic DNA to install a synchronized entropy source between distant parties. Our approach uses duplicated DNA molecules—comprising random index-payload pairs—as a shared secret. These molecules are locally sequenced and digitized to generate a common binary mask for OTP encryption, achieving unconditional security without relying on computational assumptions. We experimentally demonstrate this protocol between Tokyo and Paris, using in-house nanopore sequencing, generating a shared secret mask of ~ 400 Mb with a residual error rate of $\sim 5 \times 10^{-5}$, correctable via Bose–Chaudhuri–Hocquenghem (BCH) codes to achieve the usual overall decryption failure rate of 2^{-128}. The min-entropy of the binary mask meets the most recent National Institute of Standards and Technology requirements (SP 800-90B), and is comparable to that of approved cryptographic random number generators. Critically, our system can resist two types of adversarial interference through molecular copy-number statistics, providing an additional layer of security reminiscent of Quantum Key Distribution (QKD), but without distance limitations. This work establishes DNA as a scalable entropy source for long-distance OTP, enabling high-throughput and secure communications in sensitive contexts. By bridging molecular biology and cryptography, DNA-based key distribution opens a promising new route toward unconditional security in global communication networks.

* yannick.rondelez@espci.psl.eu

* Equal contributions

I. MAIN TEXT

Introduction During the encryption of a communication, a sender scrambles a plain message using a shared mask which is known only to the sender (Alice) and receiver (Bob), either with a symmetric or an asymmetric protocol. In practice, a combination of the two approaches is often used in cryptography. To communicate securely, Alice and Bob first use a computationally intensive yet robust asymmetric protocol to exchange a small key (such as, *e.g.*, Diffie-Hellman [1] or quantum-resistant schemes like ML-KEM [2] or HQC-KEM [3]). They then cipher and decipher their large messages with their key via an efficient symmetric algorithm, often AES - *Advanced Encryption Standards* [4]. However, the security of this hybrid approach depends on both the robustness of the asymmetric key exchange protocol and that of the symmetric encryption scheme; ultimately such protocols rely on computational security. With constant increase of computational power available to attackers, security levels must evolve. For instance, in the 80's, keys with 64 bits were considered sufficient, whereas current standards require 128 bits. In practice, it means that data archived 40 years ago with 64 bits hybrid schemes can now be cracked using a modern computer. Basing security guarantees on the computational limits of an attacker introduces a fundamental vulnerability.

An alternative to computational key exchange protocols is Quantum Key Distribution (QKD) [5–7]. Once a quantum channel is established, Alice and Bob can generate a shared secret key using a physical process whose security is grounded in the fundamental laws of quantum mechanics. Accordingly, QKD resists computational attacks and offers provable security guarantees, including the ability to detect any eavesdropping attempts. This property is particularly significant, as it is not achievable with classical computational or physical key exchange methods, where transmitted data may be copied stealthily, leaving no observable trace. However, fiber-based, twin field or device-independent QKD still remains impractical for sending large keys over distances longer than above ~ 1000 km [8–15]. Although some recent techniques are promising candidates for longer distances in the future [16], they face the fundamental limitations of quantum repeaters [17]. Satellite-based QKD [18–20] has achieved long distance ground-to-satellite exchange, but this strategy remains limited by weather conditions. Moreover, the short operational time window of low orbit satellites - about 5 min per day - results in a demonstrated record throughput of about $\approx 10^3$ kbit/day [20]. Finally, the full feasibility of satellite-mediated ground-to-ground QKD has not yet been achieved.

Cryptographic schemes providing unconditional security, *i.e.*, that remain secure against an adversary with unlimited computational power, do exist and have been known for a long time. In 1949, Shannon demonstrated that the One-Time Pad (OTP) cryptosystem, introduced as early as 1882, offers unconditional security [21]. In OTP cryptography (Fig. 1a), Alice and Bob share a secret binary mask. This mask must fulfil the following criteria: be as long as the message to be sent, perfectly random, used only once and destroyed after use. Alice combines the message and the mask with a XOR function. This creates an encrypted message, which can be safely sent over a public channel. At the receiver location, the reverse operation -again a XOR with the shared secret mask- restores the plain message (Fig. 1b). Although simple and efficient, OTP encryption poses two challenges, which in practice, strongly limit its applications: first the generation of large masks with high quality of randomness and, second, their secure sharing at two distant locations.

The holy grail for secure communication would be to create pairs of unclonable and perfectly random sources that remain synchronized independently of the distance between them. Alice and Bob could then each use their local device to generate large random (but identical) numbers, with which they would communicate with unconditional OTP-based security, a method that remains beyond current reach.

Pioneering works [22, 23] suggest that molecular media, in particular synthetic informational polymers such as DNA, display several interesting features with respect to cryptographic applications [24]. First, the synthesis of random DNA is a simple yet massive source of randomness: lab-scale step-by-step chemical synthesis of a few 1 mg of DNA with a balanced mixture of the four A, C, G and T nucleotides can produce exabytes (10^{19} bits) of randomness in a few hours. Second, unlike many physical sources of entropy [25], DNA is natively discrete - canonical DNA is a quaternary alphabet. This discreteness simplifies the mathematical processing of the molecule as an information-carrying medium. Third, large amounts of information stored in DNA pools can be duplicated or amplified by autonomous biomolecular operations. This happens in a massively parallel way (typically 10^{12} parallel operations per mL) and without the need to extract the information stored at the molecular level. Hence, the milligrams of random DNA mentioned above can be replicated by DNA polymerases and distributed so that

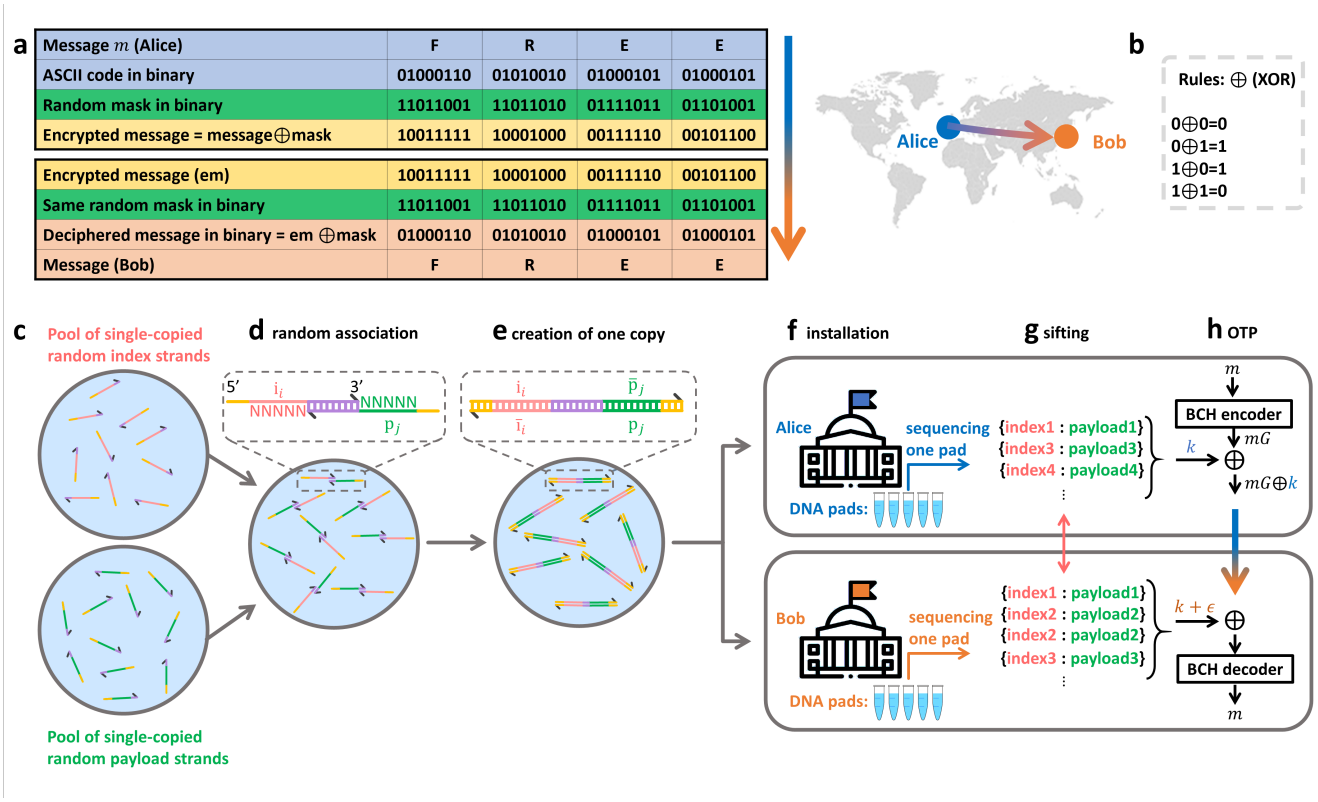


FIG. 1. **DNA-based One-Time-Pad cryptography**. **a** In OTP cryptography, a message, mapped to binary (e.g. using ASCII-code) is encrypted with a random mask using a bit-by-bit XOR function (**b**). The receiver decodes by applying a bit-by-bit XOR with the same random mask. **c**, Two independent pools containing random DNA strands, each of them being unique, playing the role of index and payload, associate at random (**d**), and are extended over each other by a polymerase to form reverse-complemented duplexes (**e**). **f**, The pool is optionally amplified and split into two pads: Alice keeps one and passes the other to Bob. Multiple pads can be duplicated and shared at once, providing synchronized random generation for many future exchanges. **g** Alice and Bob sequence their pad, and publicly share the indices to sift and assemble the corresponding secret payloads into a common binary mask, which they use to communicate via OTP (**h**).

two distant parties share exabytes of common, but still unknown, secret. If the molecular strings are properly stored [26], the secret can be archived for centuries before retrieval [27]. Fourth, the information contained in large molecular DNA archives can be extracted in a digital form by commercial sequencing machines. These machines currently output terabytes of information per run [28], at rates reaching $\simeq 10^8$ basepairs per second [29]. Some sequencing devices are barely larger than a USB stick, and many are push-and-go operations. Lastly, next generation sequencing protocols allow each DNA molecule in a mixture to be uniquely identified and counted [30, 31]. As we demonstrate below, this direct access to the discrete and stochastic nature of molecular processes can be leveraged to provide an additional level of security.

Here, we show that duplicated random DNA pads can be used to synchronize the generation of large random binary numbers at arbitrarily distant locations. We characterize the resulting channel in terms of throughput and the quality of the randomness. We measure the residual mask reconciliation error around 5×10^{-5}, which can be compensated by an error correction code with limited overhead. Using standard lab equipment and a cheap commercial sequencer, we demonstrate that a DNA-based OTP process is currently compatible with intercontinental communication at a rate reaching $10^8 - 10^9$ bits per run and a decryption failure rate below 2^{-128}. Once the sources are installed, key generation happens without transfer of confidential information, limiting interception options. Still, we show that the channel can be further secured against attacks. We experimentally simulate adversarial interference in two different scenarios and provide statistical measures to detect interception.

Installation of synchronized random sources. The creation of duplicated random DNA pads uses three stochastic steps (Methods and supplementary note 1). First, Alice orders the synthesis of two pools of partially random DNA oligonucleotides, called index strands and payload strands, from a commercial manufacturer (Fig. 1c). During the independent syntheses of the random domain of these strands, random bases are selected from among

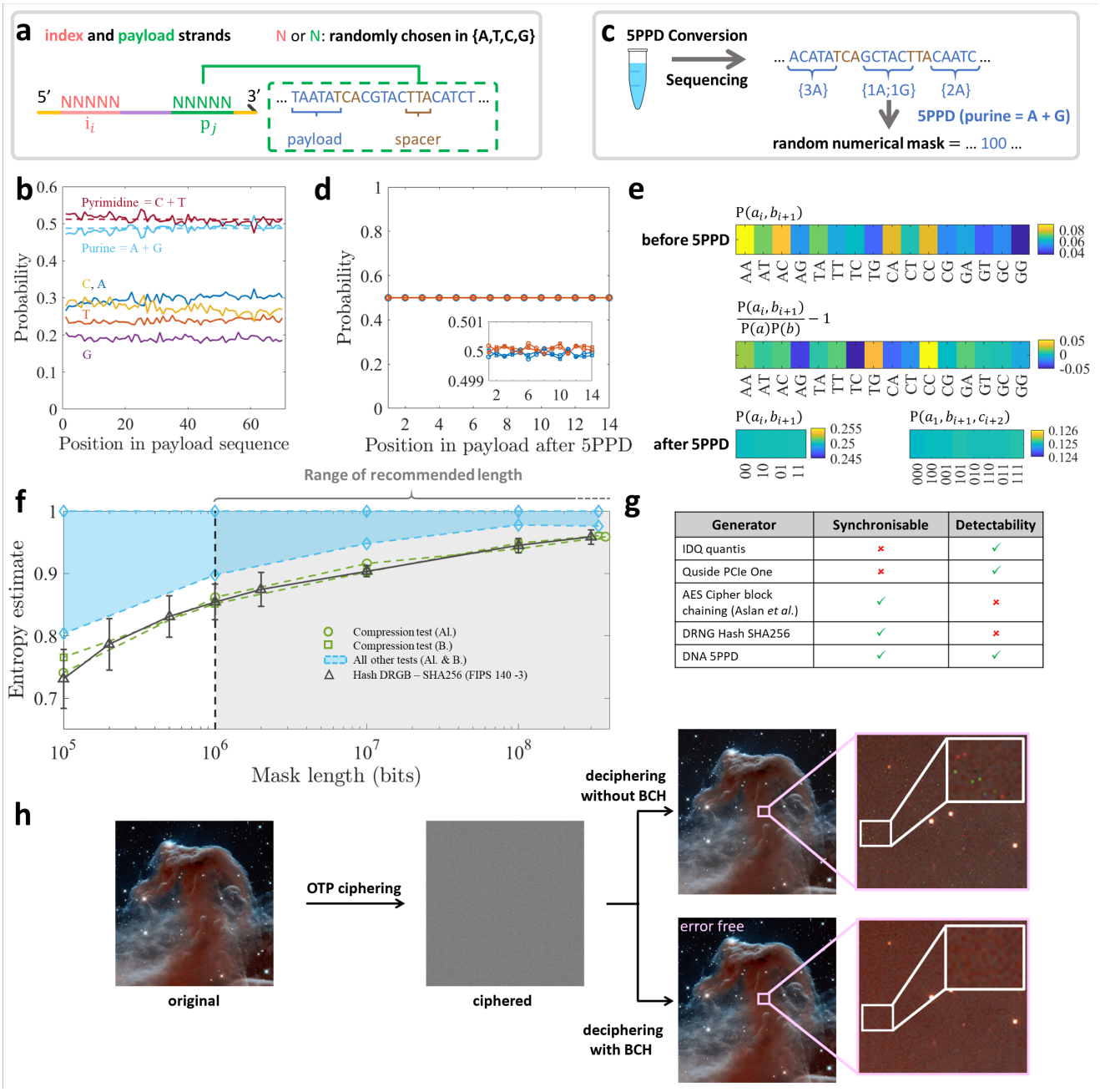
A, C, G, T. A standard order provides about $\sim$ 4 nmol of oligonucleotides, or about 10^{15} unique strands, and much larger syntheses are possible. Second, Alice dilutes the strands to about 100 nM and randomly assemble around 10^{12} index strands with an equivalent number of payload strands (Fig. 1d). A polymerase extends the two oligonucleotides over each other to generate double-stranded index-payload DNA duplexes, which we will call double-stranded DNA keys (Fig. 1e). Third, a defined number of DNA keys, on the order of 10^6-10^9 molecules, is randomly sampled by taking an aliquot from this pool. Each of these step adds a layer of discrete randomness and contribute to the security of the channel. For example, even if the pool of payload strands was not fully random, or the DNA provider is not fully trusted, it would be impossible to guess which payload strand associated with a particular index strand, or which of these combinations were actually sampled in step three. Importantly, in the resulting DNA key pool, sequence information exists in exactly two copies, in the form of two reverse-complemented DNA molecules. Then, the aliquot is physically partitioned, with or without further amplification, which allows to share information between two pads. Alice keeps one and sends the second to Bob (Fig. 1f). The process can be parallelized and repeated to generate multiple duplicated DNA pads, at negligible cost and adaptable capacity (Supplementary note 2). After that stage, the synchronized source is installed and all communications can happen on public channels.

Authentication and creation of a shared secret. When Alice wants to send a message, she selects a pad, informs Bob, and both enter the process of generating a common random mask. (Fig. 1g). Each party independently sequences its pad using standard protocols. The sequencing machines report a list of DNA key sequences, in the form of millions to billions of independent index-payload associations. Alice and Bob's sets overlap, but, because of statistical sampling and biomolecular or sequencing errors, they do not necessarily fully coincide. Bob then publicly sends to Alice his list of index sequences while keeping the associated payloads secret, and Alice compares them to her own list. As the diversity of indices scales exponentially with index length, Alice can now authenticate Bob with a high level of confidence (Supplementary note 3). Then, she computes the intersection between the two sets, decides of a specific (*e.g.*, random) index ordering, and publicly sends back that list to Bob. The corresponding ordered payloads then form the shared secret between Alice and Bob, which they convert to a binary OTP mask (Fig. 1h). This is equivalent to a sifting stage in QKD. Importantly, no information concerning the payload sequences was exchanged in the process.

We localized Alice in Paris and Bob in Tokyo and experimentally tested the full biomolecular protocol. For installation, Alice obtained degenerate oligonucleotides pools from IDT (Integrated DNA Technologies), assembled, amplified by PCR and split the sample in duplicated pads containing approximately 30×10^6 unique DNA keys. The random parts of the DNA keys were composed of 14 domains of length $n = 5$, separated by spacer sequences identical in all strands (Fig. 2a), hence a combinatorial space of more than $2^{140} \approx 10^{42}$ possible DNA keys. This design was selected to facilitate alignment and digitization of the keys (see below). A pad was sent to Tokyo and stored. For communication the two pads were independently sequenced using nanopore technology on local P2 Solo machines, a miniaturized sequencer with a footprint of just 15×11×9 cm. After quality filtering and aligning, the two datasets were clustered, and consensus sequences were extracted for each cluster, along with cluster size and quality metric. These metrics were used for a final filtering stage, after which Alice and Bob retained 26 586 748 and 27 915 041 high-quality DNA key sequences, respectively. When exchanging their list of indices, they found an overlap of 22 603 540 exact correspondences (Supplementary Table S1).

Generating the binary key and assessing the quality of the randomness. The simplest approach to binarizing a DNA sequence operates at the nucleobase level, employing a canonical quaternary code (*e.g.*, A = 00, G = 01, C = 10, T = 11). This encoding scheme maximizes the amount of digital information that can be extracted from DNA—up to 2 bits per base in theory, but closer to 1.83 bits per base once experimental constraints are considered [36]. However, it is not suitable for generating random bits from synthetic DNA due to its sensitivity to biases and correlations commonly associated with degenerate DNA synthesis [22]. For example, Fig. 2b shows the unbalanced distribution of the four nucleobases with a gradual drift along the chemical synthesis direction, observed in Bob's filtered consensus set. In addition, we observe pairwise correlations (Fig. 2e) up to a length of 5 (Supplementary note 4a). Consequently, standard debiasing approaches, such the von Neumann protocol [22] which requires independence of the bits, cannot be directly applied to DNA sequences. Here, we level out position-dependent representation biases, and average over spatial correlations along the polymer chain via blockwise binarization of DNA key sequences (Supplementary note 4c). Among this family of functions which compromise between randomness quality and throughput, we selected the block-5 Purine Parity Digitization (5PPD), which counts modulo 2 the number of purines in each block of 5 degenerate bases (Fig. 2c). The bits are then concatenated column-wise, generating the binary mask (Fig. 2d,e).

The randomness of a binary mask is an essential feature of a secure OTP protocol, and its quality must ad-



Generator	Synchronisable	Detectability
IDQ quantis	×	✓
Quside PCIe One	×	✓
AES Cipher block chaining (Aslan *et al.*)	✓	×
DRNG Hash SHA256	✓	×
DNA 5PPD	✓	✓

FIG. 2. **Generating a shared binary mask using DNA pads**. **a** Index-payload key architecture. **b**, Nucleobase distribution along the payload positions. **c**, Principle of block5 Purine Parity Digitization (5PPD). **d**, Probability of measuring a 1 (red) and 0 (blue) along the binary sequence obtained with a 5PPD of the sequences strands sequenced by Alice (circle) and Bob (square). **e** Pair distribution and correlation in DNA sequences before 5PPD. Pairs and triplets distribution after 5PPD. **f**, Estimated entropy of DNA binary masks according to the NIST standard 800-90B [32] and comparison with a NIST-approved deterministic RNG [33] (see Supplementary Tables S4 and S5). The standard computes ten entropy estimates and retains the minimum value. The min-entropy is dictated by the compression entropy and all the other estimates are grouped in blue for Alice and Bob sequences. **g** Comparison with commercial RNGs [34]. **h** DNA-OTP ciphering of 2704 × 2826 image, 130 Mb, of the Horsehead Nebula in the constellation of Orion. Credits: NASA, ESA, and the Hubble Heritage Team (AURA/STScI) [35]).

here to established cryptographic standards. Here we follow the latest entropy estimation guidelines [32, 37]. Furthermore, we select the most conservative scenario by using the min-entropy metric, which is the minimal values obtained over 10 different entropy tests (Fig. 2f). For our experimental demonstration, applying 5PPD results in a shared binary mask of 316 Mb. We measured the min-entropy for subsequences of various length (Fig. 2f). For the

full-length mask, we obtained min-entropy values of 0.9588 from Alice's side and 0.9604 from Bob. Irrespective of the mask length used, these values are on par with the ones produced by a numerical Random Number Generator (RNG) approved by the Standards FIPS 140-3 [33]), *e.g.,* Hash-DRBG-SHA256, which applies to all sensitive communications among U.S. federal agencies, and with the ones produced by other commercial RNGs (Fig. 2g and comparative table in [38]).

Message sending and correction of residual errors. To avoid any risk of electronic leakage, sequencing, authentication and binary mask generation are performed at the last moment, using a local hardware. The residual error between the masks is treated via a standard layer of error-correcting code. We selected the Bose–Chaudhuri–Hocquenghem (BCH) cyclic code (Supplementary note 5), which is widely accepted for correcting random binary errors [39, 40]. To set the BCH code parameters, Alice needs an estimate of the error rate of the channel -which conceivably may vary with pad storage or experimental conditions. Alice and Bob thus publicly share the 5PPD of an additional random stretch included in the index strand. In the experimental demonstration the two binary differed at 4189/157401800 positions, giving an estimated error rate of 3×10^{-5} (the actual error on the whole shared binary mask was measured at $\approx 5 \times 10^{-5}$). Alice then adjusts the parameters of BCH error correction code such that the probability that Bob cannot reconstruct error-free is lower than the standard cryptographic decryption failure rate of 2^{-128}, and sends the message. To comply to the OTP requirement, immediately after transmission and decoding, all traces of the binary mask are erased on both side. This includes residual DNA in the sequencing chip or experimental waste (such as liquids or contaminated surfaces) which is degraded chemically using standard lab procedures. We tested the full OTP protocol between Paris and Tokyo by OTP-encrypting, sending and deciphering a large color image using the experimentally generated random masks (Fig. 2h and (Supplementary Files 1 for high-resolution pictures).

Securing the DNA-synchronized channel. In DNA-OTP, the random DNA keys are assembled directly by Alice and remain unknown to anybody until sequencing. Thanks to the exceptional information density and stability of DNA, DNA pads are extremely compact and can be easily concealed, transported using physical security measures and securely stored for long times. Accordingly, pad transport can be extremely infrequent, or even happen only at installation. Still, it is conceivable that an attacker (Eve) can get access to a DNA pad during storage. We envision two main scenarios for such an attack (among possibly other options). First, Eve could withdraw a fraction of Bob's pad and sequence it independently, expecting that the partial material loss will go unnoticed. Second, with more time and resources, Eve could steal a full pad, PCR-amplify it, split the amplified solution in two, replace Bob's share and keep the rest for sequencing. We show below that a simple procedure, based on the molecular properties of random DNA pools, is available to resist these two types of attacks.

In this secure version of the protocol, Alice prepares the DNA pads by thermal denaturation and splitting, without PCR preamplification (Fig 3a); the stochastic partitioning process thus applies to molecules present in exactly two copies (one direct and one reverse-complemented). In addition, Alice and Bob add a step in their sequencing protocol, where they initially mark each molecule in the pad with a small Unique Molecular Identifier (UMI, Fig. 3a). Such identifiers are generally made of a short stretch of random DNA and are commonly used in Next Generation Sequencing workflows [41, 42]. Once these identifiers are covalently attached to the individual DNA chains, it becomes possible to use the deep sequencing data to unambiguously access the molecular count of each key in the original sample, regardless of the steps and biases that occur during sample processing. Because Alice's pad was prepared by partitioning a 2-copy sample, she theoretically expects either two UMIs per cluster (when she received both direct and reverse-complemented chains of a given key) or a single UMI (in the case where she received only one of the two chains, that is, among the shared set). By increasing the copy number and introducing an additional partitioning, Eve's copy-and-replace attack will alter the molecular count statistics and become noticeable.

To test this concept, we experimentally prepared 10 denaturated DNA pad pairs, from a sampled diversity of approximately 2 million DNA keys each, and simulated Eve's attacks under the two scenarios above, with various intensities (Fig. 3b-c). Alice and Bob then entered the key sequencing stage as before, except for the addition of the UMI-tagging preliminary step (Supplementary note 6). After sequencing, Alice (or Bob) groups the reads by their index-payload content, and counts the number of different UMI associated with each of these clusters. As expected, the simple partial theft by Eve resulted in a strictly null tripartite shared set in all cases (3d), meaning that the channel remains safe. With a copy-and-replace attack on Bob's pad, Eve could get access to a part of the shared secret, but the statistical analysis of UMI counts within Bob's clusters was clearly affected by the interference (Fig. 3e-f). Due to some imperfections in the biomolecular operations, some clusters with multiplicities greater than 2 were observed even in uncompromised samples. However, internal renormalization between the shared and nonshared set provides a very sensitive "interference index". This index reacted even to the most conservative

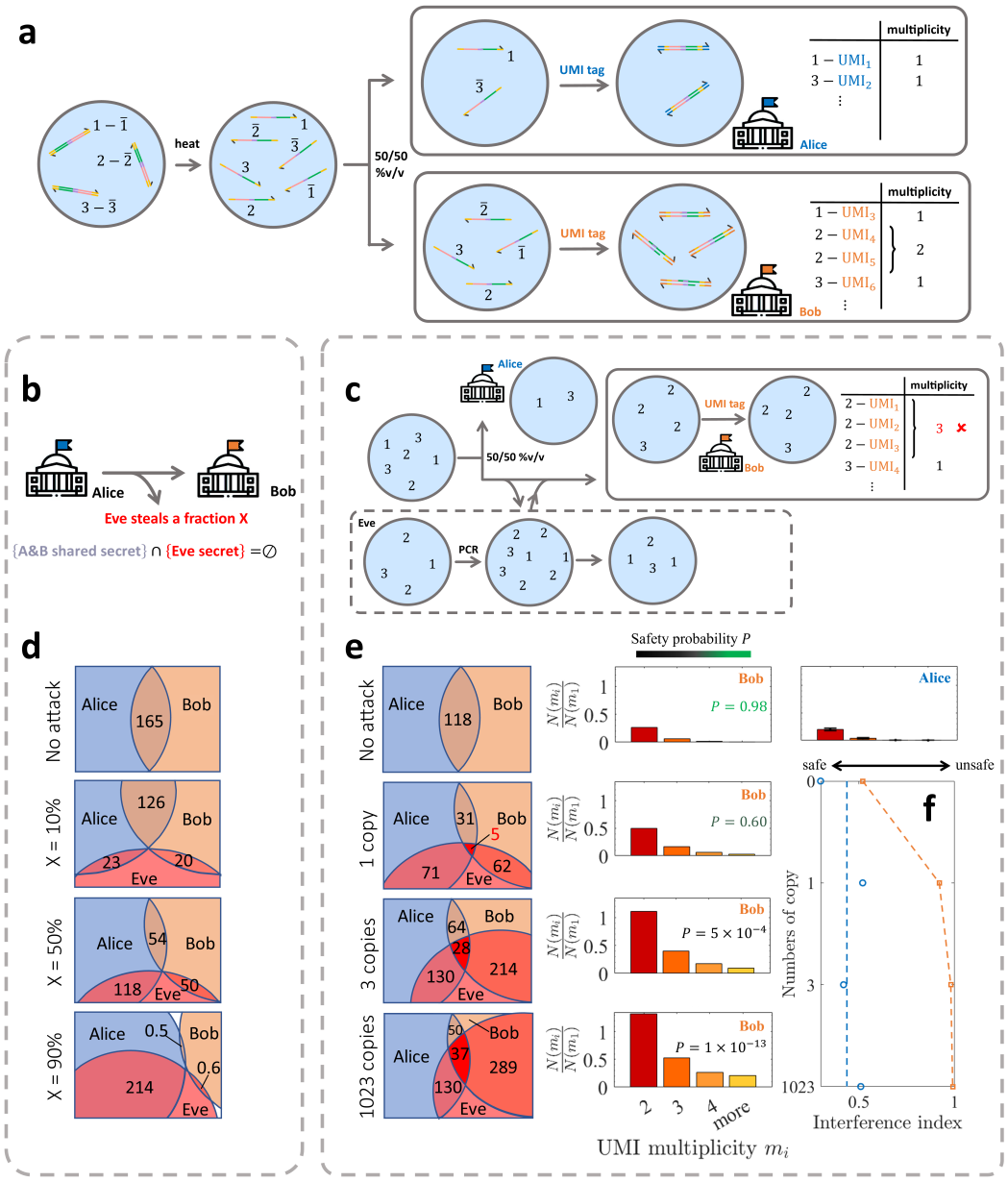


FIG. 3. **Securization and simulation of attacks**. **a** Installation of UMI tags to secure the channel. **b** Scenario 1: Eve steals a fraction of the DNA keys within Bob's pad, without replacement. **c** Scenario 2: Eve steals Bob's pad, amplifies the keys by PCR, splits the solution and replace Bob's pool. **d** Ensemble representation of the shared DNA keys for various fractions of theft in scenario 1, showing $\{\text{Alice} \cap \text{Bob}\} \cap \{\text{Eve}\} = \emptyset$ in all cases. The diagrams indicate the number of shared keys, in thousands. **e** Ensemble representation for various amplification factors by Eve, and normalized distribution of UMI multiplicity mi in clusters in scenario 2. The four replicates for Alice are shown as a single chart with error bars. The safety probability $P = 1 - \alpha$ in inset is calculated from the type-I, critical α of χ^2 test of the difference between native (Alices') and intercepted (Bobs') UMI multiplicity distributions. **f** Interference index defined as $(\sum_{i\geq 2}(N(m_i)/N(m_1))_{\text{unshared keys}})/(\sum_{i\geq 2}(N(m_i)/N(m_1))_{\text{shared keys}})$

one-copy attack, which only provided Eve with 15% of the shared secret, a fraction that would be easily mitigated via standard privacy amplification techniques [43]. The interference is also noticable in the corresponding PCR amplification curves (Supplementary note 7).

Discussion Our work introduces a novel paradigm in secure communication by leveraging synthetic DNA as a medium for generating large shared cryptographic keys, combining the proved security of OTP with the scalability and original properties of random DNA polymers.

At the heart of this approach lie the unique features of DNA, and its associated biotechnological tools. We demonstrate that DNA-based randomness—arising from the statistical incorporation of nucleotides during chemical synthesis; from the inherent stochasticity of biochemical reactions in the assembly of index-payload DNA keys; and from sampling in high-diversity pools—enables the safe generation of high-quality cryptographic masks. Sharability leverage the double helical pairing of DNA molecules, the same property that enable biological heredity. Security rest on limited attack options, and also exploits the discrete nature of molecular pools, where information can exist on molecules with small -possibly single- copy number. When only a single pair of direct and reverse-complement is present, tripartite sharing is naturally forbidden. Attacks then necessarily involve a molecular-level copying process, which may leave scars, such as detectable anomalies in copy-number statistics. Additional security could be provided by converting the DNA to a non-amplifiable—but still sequencable— informational polymer [44], as recently demonstrated [45].

Compared to more explored alternatives for secure key distribution [5–15], a critical advantage of DNA-based system lies in the capacity to generate synchronized random numbers across very distant locations. While physical transport is required for initial installation, mask generation itself happens without material or photon exchange, only public information is transmitted on a classical channel. The high density and long-term stability of DNA allow for exceptionally infrequent installation: a single gram of DNA pads could support petabytes of unconditionally secure transmissions over extended periods. Assuming a one-hour hands-on time to process a pad, we estimate the throughput at $\simeq 10^5$ bit/s with standard equipment and sequencer, a value that compares favorably with KQD [11–15]. Overall, while QKD is grounded in quantum principles but faces challenges in distance and scalability, DNA—and more generally molecular—key generation opens a promising new avenue for cryptographers to explore its operational characteristics [46].

Because of its density and stability, DNA is also actively explored as a medium for massive digital data storage [47]. Beside supporting OTP encryption, DNA-based protocols could be adapted to secure long-term data archiving, where sensitive information stored in DNA databases could only be deciphered by the owner of a matching DNA key.

Despite its promises, several challenges must be addressed to fully realize the potential of DNA in OTP or other cryptographic applications. Latency and cost (Supplementary note 2) remain key considerations, and may be limiting for high-bandwidth communications. Future work could explore faster sequencing technologies, which promise large throughput increase and exponential cost decrease [29]. Combined with automated and packaged workflows for DNA archives manipulation [48], this could make DNA-OTP more accessible. Finally, standardization and interoperability are critical for real-world deployment, requiring the establishment of standardized protocols for DNA pad generation, usage, and destruction.These development may however soon open the widespread application of molecular randomness in securing sensitive communications, financial transactions or archiving [25].

II. METHODS

Oligonucleotide sequence design Index and payload sequences were designed with 14 regions containing 5 random nucleotides (N-blocks) separated by 6-nucleotide defined spacers, summing to 197 and 195 bases, respectively. The 3' end of both oligos are cross-complementary sequences for annealing and extension. The 5' end domains are primer binding site for PCR amplification. The spacers serve two roles: they ease alignment by providing local alignment marks and insulate the variable regions from each other during synthesis, PCR, and sequencing. The dsDNA keys assembly principle is based on the complementarity of 3' end of these two sequences (index strand and payload strand). Hence, the index strand, possess at 3' end a 27-nucleotide long region complementary to the 3' end of the payload strand. Oligonucleotides were ordered at Integrated DNA Technology (IDT) and their sequences are available in Supplementary File 3.

Generating double-stranded Index-payload DNA keys (dsDNA keys). Index strands and payload strands were annealed at 100 nM and extended in a 50 µL reaction mix containing final concentrations of 1 × Q5® reaction buffer (NEB, M0493), 200 µM dNTP (NEB, N0447), 0.2× EvaGreen (BIOTIUM, 31000-T), 1% rAlbumin (NEB, B9200). The following protocol was run on CFX96 Touch Real-Time PCR detection system (BioRad): 25°C for 10 sec and signal acquisition, 98°C for 30 sec, 97°C to 60°C at -0.5°C per min, 60°C for 40 sec and addition of 1% Q5® Hot Start High-Fidelity DNA polymerase (NEB, M0493) at this point, 60°C to 72°C at +0.5°C per min and a final extension at 72°C for 10 min.

Post annealing and extension, dsDNA keys, expected at 365 bases long, were purified using SPRIselect beads

(Beckman Coulter France, B23318) with a beads-to-sample ratio of 1×, following manufacturer's recommendation, except that 85% EtOH was used. dsDNA keys were then quantified using Qubit double-stranded DNA High Sensitivity kit.

To estimate the number of dsDNA keys in the sample post purification, which at this stage correspond to the pool's diversity, an electrophoresis was run on a 4200 TapeStation System (Agilent) using High Sentitivity D1000 reagents and ScreenTape, following manufacturer's instructions. We measured 60.5 pg of DNA per µL for the 365 bp peak, corresponding to 1.5×10^8 dsDNA keys per µL.

Synchronized random number generation using duplicated DNA pads between ESPCI Paris and the University of Tokyo.

Double stranded DNA keys bottlenecking. Purified dsDNA keys (quantified at 1.5×10^8 molecules per µL) were diluted and sampled to obtain approximately 30 million of molecules in 2 µL.

Amplifying dsDNA keys. To amplify dsDNA key sample, a PCR was run using a mix containing 1× Q5 reaction buffer (NEB, M0493), 200 µM dNTP (NEB, N0447), 500 nM forward-index and reverse-payload primers, 0.2× EvaGreen (BIOTIUM, 31000-T), 1% rAlbumin (NEB, B9200), 1% Q5® HotStart High-Fidelity DNA polymerase (NEB, M0493) in a total volume of 20 µL. Amplification was realized on CFX96 Touch Real-Time PCR Detection System using the following protocol: first denaturation step at 95°C for 30 sec, 39 cycles of 95°C 30 sec, 70°C 30 sec and 72°C 1 min and a final extension step at 72°C for 5 min. To avoid heteroduplexes formation, PCR was followed in real time by fluorescent tracking and stopped at the end of the exponential phase by skipping step after 30 sec of extension at 72°C. PCR product was purified using SPRIselect beads as previously mentioned using a 0.95× beads-to-sample ratio.

Estimation of the diversity. Prior to sequencing the whole sample, a part was sequenced using Oxford Nanopore Technology (ONT) on a Flongle flow cell to estimate key diversity. The remaining sample was kept at +4°C until ready to use. Library preparation step was realized on amplified and purified dsDNA keys using the LSK-SQK114 ligation kit, with some modifications. The full protocol is available in Supplementary File 4a. Libraries were sequenced on FLO-FLG114 flow cell. Due to a sequencing crash, left-over libraries were loaded on a new flow cell. The reads were filtered, clustered and the diversity estimated by fitting the cluster size distribution to a Poisson law.

Splitting amplified dsDNA keys. The remaining sample was end-prepped following the supplier's protocol provided in Supplementary File 4b. The sample was purified and was split in two parts of 31 µL. The first part was kept in Paris for library preparation, while the other half was sent to Tokyo (LIMMS laboratory, Komaba Campus Tokyo University, Japan) at room temperature where it was kept at 4°C until sequencing, roughly 1 month later. The adapter ligation step was prepared just before sequencing.

PromethION sequencing in Paris and Tokyo. A total of 130 fmol of libraries was loaded on PRO-MIN114 flow cell and sequenced on PromethION Solo 2 platform, generating 195.55M reads. The same protocol was applied in Tokyo, except that the left-over libraries ($\sim$ 50 fmol) were loaded on a second flow cell. First Tokyo run generated a total of 178.21M reads, while the second run generated 143.42M reads. After aligning and filtering, Alice and Bob's runs retained 146 033 874 and 201 264 655 reads, respectively.

Secure data sharing protocol: DNA key sample splitting at single copy stage and UMI tagging.

Denaturation of dsDNA keys. We first measured the melting temperature of dsDNA keys and found an experimental Tm of 74°C. dsDNA keys were prepared as mentioned previously but using the following protocol: 25°C for 10 sec and signal acquisition, 85°C for 1 min, 84°C to 60°C at -0.5°C per min, 60°C for 40 sec and addition of 1% Q5® Hot Start High-Fidelity DNA polymerase (NEB, M0493) at this point, 60°C to 72°C at +0.5°C per min and a final extension at 72°C for 10 min. The prepared dsDNA keys were bottlenecked via dilution and sampling to the targeted diversity and diluted in Milli-Q containing the double-stranded circular plasmid pUC19 (NEB, N3041), used as a carrier, at a final concentration of 4.7 nM. This sample was then denaturated by heating to 85°C for 30 sec and cooling down to 24°C with a rate of -3°C per min. The denaturated sample was separated in two 2.9 µL aliquots, for Alice and Bob.

Tagging the denaturated DNA keys with Unique Molecular Identifiers. The forward and reverse UMI-primers were designed with 3 domains. From 5' to 3': a tail domain corresponding to the sequence of the external amplification primers (forward and reverse reamplification primers), used for library amplification, a short N_5 UMI domain;

a 3' head identical to standard amplification primers (forward-index and reverse-payload primers, Supplementary File 4). The DNA keys were submitted to two-cycles PCR using UMI-primers in a mix containing: 1× Q5® reaction buffer (NEB, M0493), 200 µM dNTP (NEB, N0447), 200 nM forward and reverse UMI primers, 0.2× EvaGreen (BIOTIUM, 31000-T), 1% rAlbumin (NEB, B9200), 1% Q5® Hot Start High-Fidelity DNA polymerase (NEB, M0493). The following protocol was used: first denaturation at 98°C for 15 sec, 2 cycles: 98°C 10 sec, 70°C for 30 sec, 72°C for 1 min and a final extension step at 72°C for 30 sec. Although Q5 enzyme requires a final concentration of primers of 500 nM, optimization had shown that the efficiency of this PCR does not deteriorate down to 200 nM of each primers.

Excess UMI primers were then enzymatically digested. ExonucleaseI thermolabile enzyme (NEB, M0568) was diluted 6 times as follow: 6.3 µL Milli-Q, 2 µL Q5 reaction buffer 5X and 1.68 µL exonucleaseI thermolabile. One µL of this solution was added to the PCR tube and incubated at 20°C for 10 min. After the reaction, the enzyme was deactivated for 1 min at 80°C. To minimize pipetting and avoid losing molecules on surfaces, these reactions were conducted in the same PCR tube.

Amplifying UMI-tagged DNA keys for sequencing. A PCR using external reamplification primers were realized with 1× Q5 reaction buffer (NEB, M0493), 200 µM dNTP (NEB, N0447), 500 nM forward and reverse reamplification primers (Supplementary File 5), 0.2× EvaGreen (BIOTIUM, 31000-T), 1% rAlbumin (NEB, B9200), 1% Q5® Hot Start High-Fidelity DNA polymerase (NEB, M0493). The PCR was run with a first denaturation step at 98°C for 15 sec, 39 cycles as follow: 98°C for 10 sec, 65°C for 30 sec and 72°C for 1 min; and a last extension step at 72°C for 5 min. The PCR was followed in real time by fluorescent tracking and stopped at the end of the exponential phase by skipping step after 30 sec of extension at 72°C. Final extension was performed for 5 min at 72°C. Lastly, PCR products were purified using SPRIselect magnetic beads using a beads-to-sample ratio of 0.95x and 85% EtOH.

Attack scenario 1: Eve steals part of the message without replacement. First, four single copy DNA pad pairs with diversity 2 million were created as mentioned above (See Denaturation of dsDNA keys). Four stealing fractions were tested: 0% (no-stealing control), 10%, 50% and 90%. Eve sampled the corresponding volume (0.29 µL, 1.95 µL and 2.61 µL) from Bob's pad and replaced it with the same amount of Milli-Q water.

Eve's amplified the stolen sample via a standard DNA key amplification (See Amplifying dsDNA keys) but using the following protocol: first denaturation step at 98°C for 10 sec, 39 cycles of 98°C 15 sec, 70°C 30 sec and 72°C 1 min and a final extension step at 72°C for 5 min. This protocol is later referred as "Eve Stealing Protocol" (ESP). She then prepared the sample for sequencing without UMI tagging.

Alice and Bob tagged their denatured and split DNA keys with UMI and amplified their UMI keys as mentioned previously (See UMI-tagging and amplifying UMI-tagged sections). ONT libraries were prepared on stolen amplified keys as well as Alice and Bob keys using the SQK-NBD114-24 library preparation kit with native barcoding as mentioned in Supplementary File 4c. Libraries (60 fmol) were loaded on a promethION flow cell and run on a PromethION Solo 2 platform.

Attack scenario 2: Eve steals Bob's pads, amplifies it by PCR, splits and replaces. First, four single copy DNA pad pairs with diversity 2 million were created as above. Eve took the entirety of Bob's pads and performed PCR using the ESP protocol, butlimiting the cyle number to 1, 2 or 10 PCR cycles in order to adjust the copy number. One of Bob's pads was left untouched as a control.

To avoid being uncovered by left-over primers, Eve then degraded her primers using exonucleaseI thermolabile/6 following the protocol mentioned in *Tagging denaturated DNA keys with Unique Molecular Identifiers.* Then, she sampled her PCR product to collect 2 million DNA strands for Bob, amounting to half of her sample in the 1 cycle case,1/4 of for 2 cycles and 1/1000 of for the 10 cycles case. Additionally, she adjusted the volume restituted to Bob to 2.9 µL. Prior to sequencing, Eve performed as described above (See ESP of Attack 1 section). On their side, Alice and Bob performed UMI-tagging and sequencing as previously described (See *UMI-tagging* and *amplifying UMI-tDNA* sections). A total of 130 fmol of libraries was loaded onto a PromethION flow cell.

Basecalling and sequence processing for mask generation.

Basecalling. Raw data were acquired using live basecalling except for the run of Tokyo. Basecalling for all experiments was performed using dorado v .1.1.1+3c7eef9 with the following model: dna_r10.4.1_e8.2_400bps_sup@v5.0.0 or v5.2.0. In the case of pooled sequencing, the different experiments were tagged with the native barcoding kit SQK-NBD114-24, and live demultiplexed using dorado demux and the *–barcode-both-ends* parameter.

All sequence processing followed the same pipeline, using custom code written in Mathematica or Python. The raw basecalls were filtered by median Qscore and length. (-) reads were converted to (+) reads and all were then aligned on a reference where the expected random regions were represented by N, allowing the aligning algorithm to match any bases at these positions without penalities. After aligning, the non-constant regions (including sequencing barcodes and random blocks -such as UMI, index blocks or payload blocks) were extracted from the sequence, along with their associated Qscores. Insertions were replaced with the first base at the corresponding position in alignment (and given the minimal value of the associated Qscores), to allow a dense tabular format. Qscores for deletions were computed as the min of the two nearest attributed Qscores.

In the case of indexed sequencing (*i.e.*, using the native barcoding kit on ONT to pool multiple samples in the same sequencing run), we then extracted, for each barcode, only the reads with proper matching barcodes on both sides (allowing an edit distance of 3 for barcode attribution).

Clustering and consensus calling. We then performed clustering of all reads according to their index and payload blocks (and ignoring UMIs in case UMIs were present) via an iterative process. The median Qscore M_q of each block, averaged over all reads, was computed and ordered. We then selected the 6 blocks with the highest M_q and concatenated the corresponding sequences to generate an i_1 read identifier (of length 30 nt). We then grouped the reads by perfect i_1 match. Within each group, we extracted the list of sequences for the second-best group of 6 blocks, and computed a consensus via simple majority voting, to generate i_1. If a group contained only one read, the raw sequence was used as i_2. We then iterated the grouping, combining all groups that had the same i_2. The process was repeated until all blocks had been used.

For example, the number of clusters and the number of clusters containing more than one read along the iterative process for Bob's sequencing shown in Fig.2 in the main text, is given in Supplementary Table S6.

We then retained only the clusters with more than one read and computed a complete consensus, while also estimating the error probability (as a consensus Qscore) at each position. For each position, the consensus base was selected by weighted majority voting, where each weight was the Qscore provided by the basecaller for that base. Indeed, as the Qscore are defined on a logarithmic scale, the most likely base is the one with the highest total Qscore. Its consensus Qscore can then be approximated as the sum of all Qscores associated with the winning base, minus the sum of all Qscores associated with non-majority bases, minus a penalty of 4.8 per non majority base (Supplementary note 8 for a derivation).

Consensus filtering. Once all consensuses were computed, the data was organized in an index/payload format by fusing blocks 8-14 as an index and 15-28 as a payload (that is, 6 blocks originating from the index strand as index, and all blocks from the payload strand as payload). This design offered an indexing capacity of $4^{30} \sim 10^{18}$ sequences. The blocks 1-7 were reserved for error estimation. The consensuses were then filtered, retaining only those with a minimum Qscore in the payload above 30. To filter out PCR errors (which are expected to generate a point-wise defect in the consensus per-base quality scores), we also computed the min value of the Qscores normalized by the block median and filtered out consensuses containing at least one value below 0.5. Finally we checked the unicity of each index and discarded the small fraction ($\sim$1/10000) of consensuses with non-unique indices, which may originate from PCR artifacts.

Synchronization and mask generation. Bob sent the full list of his indices to Alice. Alice computed the intersection between this list and her own set of indices. For the large sequencing presented in Fig.2, Alice found the intersection of indices to represent $\approx 82\%$ of her total and sent that intersection in a random ordering O_r back to Bob. Both Alice and Bob organized their set of payloads according to the ordering O_r (thus discarding the keys whose index are not in the intersection) and created a file with the corresponding ordered payloads. They then grouped the payload sequences in blocks of 5, applied 5PPD, and concatenated the result column-wise to obtain a single large binary string. The true error rate can be computed by comparing the mask computed independently by Alice and Bob. We found that the error was slightly higher ($\sim$4 folds) in protocol without PCR amplification before sample partitioning. Since the sequencing depth we used was typically higher for the later runs, this higher error rate does not originate from the sequencing. We speculated that this mild desynchronization originates from

the PCR errors occurring during the two parallel amplifications from single molecules.

[1] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[2] R. Avanzi, J. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J. M. Schanck, P. Schwabe, G. Seiler, D. Stehlé, *et al.*, "Crystals-kyber algorithm specifications and supporting documentation," *NIST PQC Round*, vol. 2, no. 4, pp. 1–43, 2019.

[3] P. Gaborit, C. Aguilar-Melchor, N. Aragon, S. Bettaieb, L. Bidoux, O. Blazy, J.-C. Deneuville, E. Persichetti, G. Zémor, J. Bos, A. Dion, J. Lacan, J.-M. Robert, P. Véron, P. Barreto, S. Ghosh, S. Gueron, T. Güneysu, R. Misoczki, R. Richter-Brokmann, N. Sendrier, J.-P. Tillich, and V. Vasseur, "Hamming quasi-cyclic," *NIST Post-Quantum Standardization, 4th Round; National Institute of Standards and Technology*, 2025.

[4] D. Joan and R. Vincent, "The design of rijndael, aes - the advanced encryption standard," 2002.

[5] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Theoretical Computer Science*, vol. 560, pp. 7–11, 2014. Theoretical Aspects of Quantum Cryptography – celebrating 30 years of BB84.

[6] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," *Rev. Mod. Phys.*, vol. 81, pp. 1301–1350, Sep 2009.

[7] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, "Secure quantum key distribution with realistic devices," *Rev. Mod. Phys.*, vol. 92, p. 025002, May 2020.

[8] B. Korzh, C. C. W. Lim, R. Houlmann, N. Gisin, M. J. Li, D. Nolan, B. Sanguinetti, R. Thew, and H. Zbinden, "Provably secure and practical quantum key distribution over 307 km of optical fibre," *Nature Photonics*, vol. 9, pp. 163–168, Mar 2015.

[9] M. Lucamarini, Z. L. Yuan, J. F. Dynes, and A. J. Shields, "Overcoming the rate–distance limit of quantum key distribution without quantum repeaters," *Nature*, vol. 557, pp. 400–403, May 2018.

[10] M. Pittaluga, M. Minder, M. Lucamarini, M. Sanzaro, R. I. Woodward, M.-J. Li, Z. Yuan, and A. J. Shields, "600-km repeater-like quantum communications with dual-band stabilization," *Nature Photonics*, vol. 15, pp. 530–535, Jul 2021.

[11] S. Wang, Z.-Q. Yin, D.-Y. He, W. Chen, R.-Q. Wang, P. Ye, Y. Zhou, G.-J. Fan-Yuan, F.-X. Wang, Y.-G. Zhu, P. V. Morozov, A. V. Divochiy, Z. Zhou, G.-C. Guo, and Z.-F. Han, "Twin-field quantum key distribution over 830-km fibre," *Nature Photonics*, vol. 16, pp. 154–161, Feb 2022.

[12] L. Zhou, J. Lin, Y. Jing, and Z. Yuan, "Twin-field quantum key distribution without optical frequency dissemination," *Nature Communications*, vol. 14, p. 928, Feb 2023.

[13] Y. Liu, W.-J. Zhang, C. Jiang, J.-P. Chen, C. Zhang, W.-X. Pan, D. Ma, H. Dong, J.-M. Xiong, C.-J. Zhang, H. Li, R.-C. Wang, J. Wu, T.-Y. Chen, L. You, X.-B. Wang, Q. Zhang, and J.-W. Pan, "Experimental twin-field quantum key distribution over 1000 km fiber distance," *Phys. Rev. Lett.*, vol. 130, p. 210801, May 2023.

[14] Y. Zheng, H. Wang, X. Jia, J. Huang, H. Yuan, C. Zhai, J. Dai, J. Shi, L. Zhang, X. Zhang, M. Zhuang, J. Liu, J. Mao, T. Dai, Z. Fu, Y. Jiao, Y. Shi, D. Dai, X. Wang, Y. Li, Q. Gong, Z. Yuan, L. Chang, and J. Wang, "Large-scale quantum communication networks with integrated photonics," *Nature*, Feb 2026.

[15] B.-W. Lu, C.-W. Yang, R.-Q. Wang, B.-F. Gao, Y.-Z. Zhen, Z.-G. Wang, J.-K. Shi, Z.-Q. Ren, T. A. Hahn, E. Y.-Z. Tan, X.-P. Xie, M.-Y. Zheng, X. Jiang, J. Zhang, F. Xu, Q. Zhang, X.-H. Bao, and J.-W. Pan, "Device-independent quantum key distribution over 100 km with single atoms," *Science*, vol. 391, no. 6785, pp. 592–597, 2026.

[16] S. Gupta, Y. Huang, S. Liu, Y. Pei, Q. Gao, S. Yang, N. Tomm, R. J. Warburton, and T. Zhong, "Dual epitaxial telecom spin-photon interfaces with long-lived coherence," *Nature Communications*, vol. 16, p. 9814, Nov 2025.

[17] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, "Fundamental limits of repeaterless quantum communications," *Nature Communications*, vol. 8, p. 15043, Apr 2017.

[18] S.-K. Liao, W.-Q. Cai, J. Handsteiner, B. Liu, J. Yin, L. Zhang, D. Rauch, M. Fink, J.-G. Ren, W.-Y. Liu, Y. Li, Q. Shen, Y. Cao, F.-Z. Li, J.-F. Wang, Y.-M. Huang, L. Deng, T. Xi, L. Ma, T. Hu, L. Li, N.-L. Liu, F. Koidl, P. Wang, Y.-A. Chen, X.-B. Wang, M. Steindorfer, G. Kirchner, C.-Y. Lu, R. Shu, R. Ursin, T. Scheidl, C.-Z. Peng, J.-Y. Wang, A. Zeilinger, and J.-W. Pan, "Satellite-relayed intercontinental quantum network," *Phys. Rev. Lett.*, vol. 120, p. 030501, Jan 2018.

[19] S.-K. Liao, W.-Q. Cai, W.-Y. Liu, L. Zhang, Y. Li, J.-G. Ren, J. Yin, Q. Shen, Y. Cao, Z.-P. Li, F.-Z. Li, X.-W. Chen, L.-H. Sun, J.-J. Jia, J.-C. Wu, X.-J. Jiang, J.-F. Wang, Y.-M. Huang, Q. Wang, Y.-L. Zhou, L. Deng, T. Xi, L. Ma, T. Hu, Q. Zhang, Y.-A. Chen, N.-L. Liu, X.-B. Wang, Z.-C. Zhu, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, "Satellite-to-ground quantum key distribution," *Nature*, vol. 549, pp. 43–47, Sep 2017.

[20] Y. Li, W.-Q. Cai, J.-G. Ren, C.-Z. Wang, M. Yang, L. Zhang, H.-Y. Wu, L. Chang, J.-C. Wu, B. Jin, H.-J. Xue, X.-J. Li, H. Liu, G.-W. Yu, X.-Y. Tao, T. Chen, C.-F. Liu, W.-B. Luo, J. Zhou, H.-L. Yong, Y.-H. Li, F.-Z. Li, C. Jiang, H.-Z. Chen, C. Wu, X.-H. Tong, S.-J. Xie, F. Zhou, W.-Y. Liu, Y. Ismail, F. Petruccione, N.-L. Liu, L. Li, F. Xu, Y. Cao, J. Yin, R. Shu, X.-B. Wang, Q. Zhang, J.-Y. Wang, S.-K. Liao, C.-Z. Peng, and J.-W. Pan, "Microsatellite-based real-time quantum key distribution," *Nature*, vol. 640, pp. 47–54, Apr 2025.

[21] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[22] L. C. Meiser, J. Koch, P. L. Antkowiak, W. J. Stark, R. Heckel, and R. N. Grass, "Dna synthesis for true random number generation," *Nature Communications*, vol. 11, p. 5869, Nov 2020.

[23] A. M. Luescher, A. L. Gimpel, W. J. Stark, R. Heckel, and R. N. Grass, "Chemical unclonable functions based on operable random dna pools," *Nature Communications*, vol. 15, p. 2955, Apr 2024.

[24] G. B. M. Wisna, D. Sukhareva, J. Zhao, P. Chopade, D. Satyabola, M. Matthies, S. Roy, C. Wang, P. Šulc, H. Yan, and R. F. Hariadi, "High-speed 3d dna paint and unsupervised clustering for unlocking 3d dna origami cryptography," *Nature Communications*, vol. 16, p. 11514, Dec 2025.

[25] O. Amer, S. Chakrabarti, K. Chakraborty, S. Eloul, N. Kumar, C. Lim, M. Liu, P. Niroula, Y. Satsangi, R. Shaydulin, and M. Pistoia, "Applications of certified randomness," *Nature Reviews Physics*, vol. 7, pp. 514–524, Sep 2025.

[26] S. D. Lemaire, D. Turek, D. Landsman, M. Colotte, and T. F. A. de Greef, "Challenges and opportunities in dna computing and data storage," *Nature Nanotechnology*, vol. 20, pp. 710–714, Jun 2025.

[27] K. H. Kjær, M. Winther Pedersen, B. De Sanctis, B. De Cahsan, T. S. Korneliussen, C. S. Michelsen, K. K. Sand, S. Jelavić, A. H. Ruter, A. M. A. Schmidt, K. K. Kjeldsen, A. S. Tesakov, I. Snowball, J. C. Gosse, I. G. Alsos, Y. Wang, C. Dockter, M. Rasmussen, M. E. Jørgensen, B. Skadhauge, A. Prohaska, J. Å. Kristensen, M. Bjerager, M. E. Allentoft, E. Coissac, I. G. Alsos, A. Rouillard, A. Simakova, A. Fernandez-Guerra, C. Bowler, M. Macias-Fauria, L. Vinner, J. J. Welch, A. J. Hidy, M. Sikora, M. J. Collins, R. Durbin, N. K. Larsen, E. Willerslev, and P. Consortium, "A 2-million-year-old ecosystem in greenland uncovered by environmental dna," *Nature*, vol. 612, pp. 283–291, Dec 2022.

[28] W. KA, "Dna sequencing costs: Data from the nhgri genome sequencing program (gsp)," 2022. www.genome.gov/sequencingcostsdata Accessed [22 Jan. 2026].

[29] B. Hajieghrari and S. Nejati-Jahromi, "Next generation sequencing and beyond: a review of genomic sequencing methods," *Functional & Integrative Genomics*, vol. 25, p. 242, Nov 2025.

[30] J. König, K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe, and J. Ule, "iclip reveals the function of hnrnp particles in splicing at individual nucleotide resolution," *Nature Structural & Molecular Biology*, vol. 17, pp. 909–915, Jul 2010.

[31] T. Kivioja, A. Vähärautio, K. Karlsson, M. Bonke, M. Enge, S. Linnarsson, and J. Taipale, "Counting absolute numbers of molecules using unique molecular identifiers," *Nature Methods*, vol. 9, pp. 72–74, Jan 2012.

[32] M. Sönmez Turan, E. Barker, J. Kelsey, K. McKay, M. Baish, and M. Boyle, "Recommendation for the entropy sources used for random bit generation, nist special publication 800-90b," 2018.

[33] "Security requirements for cryptographic modules," 2019.

[34] M. Aslan, A. Doğanaksoy, Z. Saygı, M. Sönmez Turan, and F. Sulak, "Observations on nist sp 800-90b entropy estimators," *Cryptography and Communications*, Jan 2025.

[35] NASA, ESA, and the Hubble Heritage Team, "New infrared view of the horsehead nebula — hubble's 23rd anniversary image. heic1307a," 2013. https://esahubble.org/images/heic1307a/ Accessed [13 Feb.2026].

[36] Y. Erlich and D. Zielinski, "Dna fountain enables a robust and efficient storage architecture," *Science*, vol. 355, no. 6328, pp. 950–954, 2017.

[37] M.-J. O. Saarinen, "Sp 800–22 and gm/t 0005–2012 tests: Clearly obsolete, possibly harmful," in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS & PW)*, pp. 31–37, 2022.

[38] G. Vrana, D. Lou, and R. Kuang, "Raw qpp-rng randomness via system jitter across platforms: a nist sp 800-90b evaluation," *Scientific Reports*, vol. 15, p. 27718, Jul 2025.

[39] A. Hocquenghem, *Codes correcteurs d'erreurs*. Chiffres (Paris), Association française de calcul, 1956.

[40] R. Bose and D. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, no. 1, pp. 68–79, 1960.

[41] O. Stegle, S. A. Teichmann, and J. C. Marioni, "Computational and analytical challenges in single-cell transcriptomics," *Nature Reviews Genetics*, vol. 16, pp. 133–145, Mar 2015.

[42] H. Yeom, N. Kim, A. C. Lee, J. Kim, H. Kim, H. Choi, S. W. Song, S. Kwon, and Y. Choi, "Highly accurate sequence- and position-independent error profiling of dna synthesis and sequencing," *ACS Synthetic Biology*, vol. 12, pp. 3567–3577, Dec 2023.

[43] A. Bittau, U. Erlingsson, P. Maniatis, I. Mironov, A. Raghunathan, D. Lie, M. Rudominer, U. Kode, J. Tinnes, and B. Seefeld, "Prochlo: Strong privacy for analytics in the crowd," in *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, (New York, NY, USA), p. 441–459, Association for Computing Machinery, 2017.

[44] J.-F. Lutz, M. Ouchi, D. R. Liu, and M. Sawamoto, "Sequence-controlled polymers," *Science*, vol. 341, no. 6146, p. 1238149, 2013.

[45] M. Kokoris, R. McRuer, M. Nabavi, A. Jacobs, M. Prindle, C. Cech, K. Berg, T. Lehmann, C. Machacek, J. Tabone, J. Chandrasekar, L. McGee, M. Lopez, T. Reid, C. Williams, S. Barrett, A. Lehmann, M. Kovarik, R. Busam, S. Miller, B. Banasik, B. Kesic, A. Arryman, M. Rogers-Peckham, A. Kimura, M. LeProwse, M. Wolfin, S. Kritzer, J. Leadbetter, M. Babazadeh, J. Chase, G. Thiessen, W. Lint, D. Goodman, D. O'Connell, N. Lumanpauw, J. Hoffman, S. Vellucci, K. Collins, J. Vellucci, A. Taylor, M. Murphy, M. Lee, and M. Corning, "Sequencing by expansion (sbx) – a novel, high-throughput single-molecule sequencing technology," *bioRxiv*, 2025.

[46] F. Walter, H. Narayanan, J. Bariffi, A. Lüscher, R. Bitar, R. Grass, A. Wachter-Zeh, and Z. Yakhini, "A security framework for chemical functions," 2026.

[47] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in dna," *Science*, vol. 337, no. 6102, pp. 1628–1628, 2012.

[48] C. N. Takahashi, B. H. Nguyen, K. Strauss, and L. Ceze, "Demonstration of end-to-end automation of dna data storage," *Scientific Reports*, vol. 9, p. 4998, Mar 2019.

III. ACKNOWLEDGEMENTS

Dr. A.G. passed away on April 2025 Before his passing, the late A.G. was deeply involved in conceptualizing and implementing the present studies. This manuscript is based on a preliminary outline that he produced. We are incredibly grateful for his contributions and would like to dedicate this work to his memory. The authors also thank Yannick Tauran, Kévin Ricard for their help in sequencing, Guillaume Gines for critical comments, Nicolas Clément, Masahiro Nomura, Bruno Le Pioufle, the Service pour la Science et la Technologie and the press service of the French Embassy in Japan for his their kind assistance and support. We would like to thank Yuri Klebanov and Naoto Takayama from the IIS Tokyo Design Lab for the design of a DNA capsule. This work is funded by the ANR DNASec (ANR-24-CE39-3908-04), and the PEPR MoleculArXiv (ANR-22-PEXM-0002).

IV. AUTHORS CONTRIBUTIONS

SJ performed the experiments and analysed the results with YR and VS. HG and MC performed the entropy analysis. EB performed preliminary experiments. VB and VS contributed to the experimental workflow. TP, SG performed additional statistical analysis. All authors participated in data acquisition, writing and critical revision of the manuscript. YR, PG, ML, SHK, GC and AG were responsible for, conceptualization, supervision and funding acquisition. All authors (excepted AG) read and approved the final version of the manuscript.

V. DECLARATION OF INTEREST

A patent reporting some of the methods was filed by the CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, INSTITUT MINES TELECOM, UNIVERSITE DE LIMOGES and ÉCOLE SUPÉRIEURE DE PHYSIQUE ET DE CHIMIE INDUSTRIELLES DE LA VILLE DE PARIS.