

Article

A PROVABLY-SECURE MULTI-LEVEL DNA ENCODING SCHEME WITH A DYNAMIC MEALY AUTOMATON: FORMAL ANALYSIS AND NIST-VALIDATED RANDOMNESS

Layth Hammood¹

¹College of Media, University of Kirkuk, Kirkuk, Iraq

Email : . layth.aljumaily@uokirkuk.edu.iq

Article information:

Manuscript received: 8 Jun 2026; **Accepted:** 1 Jul 2026; **Published:** 12 Aug 2026

Abstract: Deoxyribonucleic acid (DNA) encoding combined with finite-state machines has become an attractive paradigm for lightweight symmetric encryption, because it offers a large key space, simple invertible transforms, and high diffusion at low computational cost. However, the overwhelming majority of existing DNA/automaton cryptosystems are validated only through descriptive arguments and a handful of statistical experiments; they provide no reduction-based (“provable”) security guarantee, and several of them embed content-dependent control schedules whose running time grows without bound in the plaintext. This paper revisits a recent multi-level DNA encoding scheme driven by a state-changing Mealy machine and reconstructs it as a provably secure stream construction. The redesigned scheme generates all keystream and control material from an HMAC–SHA-256 counter-mode pseudorandom function under a fresh per-message nonce, while the DNA-encoding and Mealy-transduction layers are proven to form a public, key-dependent bijection. We give a complete game-hopping proof that the construction is indistinguishable under chosen-plaintext attack (IND-CPA), with the concrete advantage bounded by $2 \text{Adv}_F^{\text{prf}} + q^2/2^{128}$. The same design eliminates the exponential Fibonacci-of-ASCII-sum schedule of the baseline, reducing the per-block control cost to a constant-time keystream read and removing a timing side-channel. A reproducible Python implementation is evaluated with the NIST SP 800-22 battery (all ten applicable tests pass over 40 independent 2×10^5 -bit sequences), Shannon entropy (7.9962 bit/byte and 2.0000 bit/base), adjacent-symbol correlation (0.0047), differential metrics (NPCR =100%, UACI =42.2%), key avalanche (50.06% at the bit level and 75.10% at the base level), and an empirical IND-CPA distinguishing experiment whose measured advantage, 0.0010 ± 0.0179 , is statistically indistinguishable from zero. The results confirm both the theoretical guarantee and the practical randomness quality of the proposed cryptosystem.

Keywords: DNA cryptography; Mealy machine; finite automaton encryption; provable security; IND-CPA; pseudorandom function; NIST SP 800-22; lightweight symmetric encryption.

1. Introduction

The migration of computation and storage toward distributed, resource-constrained environments—elastic cloud back-ends, pervasive Internet-of-Things (IoT) sensors, and bandwidth-limited edge gateways—has placed sustained and growing pressure on symmetric encryption to be simultaneously lightweight, statistically sound, and theoretically justified [1][2]. In these settings the cryptographic primitive is rarely the bottleneck of a powerful server; instead it must run on microcontrollers with kilobytes of memory, intermittent power, and tight latency budgets, while still protecting data whose disclosure carries legal, financial, or clinical consequences.

Standardized block ciphers such as the Advanced Encryption Standard (AES) deliver well-understood and battle-tested guarantees, yet their substitution–permutation structure and key schedule impose a memory, energy, and instruction-count footprint that is awkward for the smallest endpoints and for high-throughput streaming of large media objects[3]. This tension between assured security and constrained implementation is precisely what has motivated a vibrant parallel line of research into bio-inspired and automaton-based constructions that aim to reproduce Shannon's classical design goals of confusion and diffusion[4] using operations that are individually trivial yet collectively expressive. The medical and biometric domains sharpen these requirements further. Imaging studies, electronic health records, and physiological telemetry are increasingly acquired at the edge and exchanged through public cloud infrastructure, so their confidentiality must be preserved end-to-end without overwhelming the embedded hardware that captures them. The recurring research question across this literature is therefore not merely can we encrypt cheaply? but can we encrypt cheaply while retaining defensible, quantifiable security guarantees? It is the second clause that the present work treats as non-negotiable. Comprehensive surveys of lightweight cryptography for IoT and Industrial IoT systematically document the design space—reduced block and key sizes, simplified round functions, hardware-oriented permutations, and stream constructions—and consistently identify the same trade-off frontier between security margin, throughput, energy, and silicon area. They also converge on a common cautionary observation: many proposed schemes optimize aggressively for footprint and then justify their security through ad-hoc reasoning and selected statistical experiments, rather than through analysis that withstands adaptive adversaries. This methodological softness is not a peripheral concern. In resource-constrained deployments, keys are long-lived, devices are physically exposed, and identical plaintext structures (sensor headers, image preambles, medical metadata) recur predictably, so any scheme that leaks structure under repetition or chosen-input probing is exploitable in practice. The lightweight setting thus magnifies, rather than excuses, the need for formal confidentiality guarantees.

DNA cryptography descends conceptually from Adleman's demonstration that the combinatorial richness of nucleotide sequences can encode and solve computational problems[5]. In the cryptographic adaptation, binary data are represented through the four nucleotide bases—adenine (A), thymine (T), cytosine (C), and guanine (G)—each mapping naturally onto a two-bit group. Because there are $4! = 24$ distinct, invertible assignments of bases to two-bit values, the encoding step alone furnishes a family of twenty-four interchangeable substitution tables, and the associated nucleotide-level operations (base complement, base-wise XOR, and modular base addition) are cheap, branch-free, and highly parallelizable. These properties make DNA encoding an attractive vehicle for Shannon's confusion and diffusion at very low computational cost, and they explain why DNA encoding is so frequently paired with an external source of pseudo-randomness—most often a chaotic map or a cryptographic hash—to drive the selection of encoding rules and diffusion masks[6][7]. The result is a flexible, inexpensive substitution-and-diffusion layer that has been applied to text, image, and medical workloads alike.

Finite automata contribute a complementary and inherently sequential form of statefulness. The systematic use of automata as cryptographic primitives traces to Tao's theory of invertible finite automata with output, which established the conditions under which a transducer can serve as an encryption function with a well-defined inverse[8]. A Mealy machine, whose output depends jointly on the current state and the current input symbol, becomes an invertible transducer precisely when its per-state output function is a permutation of the alphabet[9]. The cryptographically interesting behaviour emerges when the machine is permitted to change its state configuration during encryption: the same plaintext symbol is then mapped to different ciphertext symbols at different positions, yielding the position-dependent substitution that diffusion demands[10]. Coupling this dynamic Mealy layer with DNA encoding produces a compact, multi-level confusion mechanism that is inexpensive to compute and straightforward to invert—an idea

crystallized by Pavithran et al., who built a cryptosystem on DNA cryptography and a randomly generated Mealy machine[11], and extended by Taj and Hossain, whose multi-level design fuses a 256-bit DNA-derived key validated by a run-test of randomness, the twenty-four DNA encoding combinations, and a four-state state-changing Mealy machine whose configuration evolves per block[12]. The latter scheme is the representative, recent baseline that the present paper reconstructs.

The past three years have produced a dense body of DNA- and automaton-based encryption research. On the DNA-with-chaos axis, Li and Chen combine a six-dimensional hyperchaotic system with DNA encoding and report near-ideal entropy and low adjacent-pixel correlation; Yu et al. drive DNA operations with a "DNA network" coupled to a hyperchaotic map to strengthen cross-channel diffusion Liu et al. target multi-medical images using a new spatiotemporal chaos model together with DNA computing, emphasizing differential-attack resistance[13] and Alawida introduces a DNA tree structure that governs a chaotic image cipher to raise the structural complexity of the diffusion stage[14]. In the healthcare-cloud setting, Chidambaram et al. integrate DNA-chaos diffusion with SHA-driven keying for plaintext sensitivity in a medical-image repository[15], while Wu et al. employ content-aware DNA computing for healthcare imagery [16]. Most recently, Al-dayel et al. gate a chaotic keystream behind a largest-Lyapunov-exponent verification and apply invertible DNA cross-diffusion, reporting differential metrics close to the theoretical ideals[17].

On the automaton axis, Ramasubbu et al. construct an encryption scheme directly on a Mealy machine, exploiting state-dependent output to obscure symbol identities; Shakhmetova et al. design an invertible finite-state-machine methodology for image encryption and decryption with uniform ciphertext histograms; and Kulzhanova et al. apply a deterministic finite automaton to compact, secure access control, illustrating the lightweight footprint of automaton-based designs [18]. The application surface has also broadened: Chakravorty et al. combine a neural network with DNA cryptography to defend mobile ad-hoc networks against black-hole attacks[19]. Across this literature, the consistent strength is excellent measured randomness; the consistent limitation, examined next, is the basis on which security is claimed. Despite a decade of cumulative progress, DNA- and automaton-based cryptosystems are predominantly justified by intuitive security arguments—typically of the form "the adversary must simultaneously guess the key, the encoding sequence, and the machine configuration"—reinforced by a battery of statistical experiments. Such evidence is genuinely valuable but does not constitute a security proof. Modern cryptography requires that confidentiality be reduced, through an explicit and quantitative argument, to a well-studied hardness assumption or to the pseudorandomness of a standard primitive, in the indistinguishability tradition initiated by Goldwasser and Micali and codified in contemporary treatments of chosen-plaintext security[20][21]. The absence of such reductions in the DNA literature is a recurring and consequential gap, because statistical success is necessary but not sufficient: El Hanouti and El Fadili recovered the key of a DNA-and-key-chaining audio scheme from only a handful of chosen plaintexts even though the scheme exhibited favourable statistics [22]. Schemes that "pass" randomness suites can and do fall to chosen-plaintext and differential cryptanalysis precisely because passing those suites was never equivalent to indistinguishability.

The validation methodology itself is frequently incomplete. The NIST SP 800-22 suite is the de-facto standard for assessing cryptographic pseudo-random sequences[23] but its correct application requires the multi-sequence protocol—reporting both the pass proportion across many sequences and the uniformity of the p-value distribution—rather than inferring randomness from a single sequence[24]. Differential sensitivity is commonly summarized by the Number of Pixels Change Rate (NPCR) and the Unified Average Changing Intensity (UACI), whose theoretical ideals are approximately 99.6094% and 33.4635%[25]; yet, like the NIST battery, these instruments characterize observable statistics and say nothing directly about computational indistinguishability.

A second, subtler deficiency concerns efficiency and side-channels. Several multi-level schemes derive their control parameters—the number of state changes and the per-block

encoding index—from content-dependent quantities such as a cumulative sum of ASCII values. In the baseline studied here, the per-block configuration is a function of the sum S of the block's character codes, realized by summing a Fibonacci series up to its S -th term. Since the S -th Fibonacci number has $\Theta(S)$ decimal digits, this schedule costs $\Theta(S^2)$ bit-operations per block and grows without bound as content varies. Beyond the inefficiency, making control flow depend on secret-correlated plaintext content introduces a data-dependent timing channel—an avoidable vulnerability in exactly the embedded settings these schemes target. Finally, the baseline encrypts deterministically without a nonce, so identical plaintexts produce identical ciphertexts under a session key; this alone forecloses indistinguishability under chosen-plaintext attack, regardless of any downstream statistical performance.

Synthesizing the foregoing, four challenges remain inadequately addressed: (i) the lack of reduction-based, IND-CPA-style security proofs for DNA/automaton ciphers, which are instead defended heuristically (ii) deterministic, nonce-free designs that cannot achieve indistinguishability because identical plaintexts encrypt identically under a session key; (iii) content-dependent control schedules that are computationally heavy and side-channel prone, exemplified by the Fibonacci-of-ASCII-sum mechanism; and (iv) incomplete statistical validation that omits the full NIST SP 800-22 multi-sequence methodology and the differential metrics NPCR/UACI. The collective gap is the absence of a DNA/Mealy cryptosystem that preserves the bio-inspired structure and lightweight character of the family while (a) admitting a clean, quantitative security proof, (b) replacing content-dependent control with constant-time content-independent control, and (c) providing a clear architectural rationale for the bio-inspired layers beyond mere statistical decoration—specifically, that the DNA encoding and Mealy transduction contribute provably-neutral yet meaningful diffusion properties that distinguish the construction from a plain counter-mode stream cipher.

This motivates a single, focused question. Given a recent and representative multi-level DNA-encoding scheme equipped with a state-changing Mealy machine can it be reconstructed so that it (1) is provably IND-CPA secure under a standard cryptographic assumption, (2) eliminates the exponential, content-dependent schedule in favour of constant-time control, and (3) empirically exhibits the randomness, diffusion, and indistinguishability that the proof predicts—without sacrificing the DNA and Mealy structure that motivates the design?

The objectives of this work follow directly. We formalize the scheme as a keyed encryption algorithm with explicit syntax and an explicit security game; we prove a concrete IND-CPA bound by reduction to the pseudorandomness of HMAC-SHA-256, whose status as a pseudorandom function rests on established analyses of keyed hashing[26][27]; we demonstrate that the DNA-encoding and Mealy-transduction stages constitute an invertible public bijection that neither adds to nor detracts from confidentiality; we remove the Fibonacci bottleneck by driving the machine configuration from constant-time keystream reads; and we validate the design through a reproducible implementation subjected to the NIST SP 800-22 battery Shannon entropy at byte and base level adjacent-symbol correlation, NPCR/UACI key-avalanche, and an empirical IND-CPA distinguishing experiment.

The principal contributions of this paper are the following. First, a provably secure redesign: we recast the multi-level DNA/Mealy cryptosystem as a stream construction whose keystream is generated by HMAC-SHA-256 in counter mode under a fresh per-message nonce, and we give a complete game-hopping proof that the scheme is IND-CPA secure with advantage at most $2 \cdot \text{Adv}_{\text{F}^{\text{prf}}} + q^2/2^{128}$. Second, a bijection lemma establishing that the DNA-encoding and Mealy-transduction stages form a public, key-dependent bijection invertible in linear time—the structural fact that allows the security argument to pass through the bio-inspired layer unchanged. Third, elimination of the exponential schedule: we replace the content-dependent Fibonacci-of-ASCII-sum control of the baseline with constant-time keystream reads, reducing per-block control cost from $\Theta(S^2)$ to $O(1)$ and removing the associated timing side-channel. Fourth, a comprehensive, reproducible evaluation reporting the full NIST SP 800-22 multi-sequence battery, byte-

and base-level entropy, adjacent-symbol correlation, NPCR/UACI, bit- and base-level key avalanche, and an empirical IND-CPA experiment whose measured advantage is statistically indistinguishable from zero. Fifth, honest scope and deployment guidance: we delineate exactly what the proof guarantees (confidentiality, not integrity), quantify the nonce-uniqueness requirement through the birthday term, and recommend Encrypt-then-MAC for active-adversary settings.

The remainder of the paper is organized as follows. Section 2 reviews recent DNA- and automaton-based cryptosystems and positions our work against them through a structured comparison. Section 3 presents the proposed scheme—its architecture, conceptual model, mathematical formulation, flowchart, algorithmic description, and pseudocode. Section 4 describes the experimental setting, baseline methods, evaluation metrics and results, including ablation, sensitivity and cost analysis. Section 5 concludes with a summary of achievements, limitations, threats to validity and directions for future work.

This section surveys recent (2023–2025) research on DNA-based and finite-automaton encryption and situates the present contribution within it. The survey is organized around three themes that together expose the gap this paper closes: (i) DNA encoding driven by chaotic and hash-based sources of pseudo-randomness, (ii) finite-state-machine and Mealy/automaton cryptography, and (iii) the methodological question of how such schemes are validated and why statistical validation alone is insufficient. Throughout, we read each line of work against a single yardstick—whether confidentiality is argued heuristically or reduced to a standard assumption—because that distinction is precisely what separates the existing family from the construction proposed here. The discussion culminates in a structured comparison (Table 1).

DNA Encoding with Chaotic and Hash-Based Drivers

Chaotic dynamical systems remain the dominant source of pseudo-randomness in DNA encryption, because their sensitivity to initial conditions, ergodicity, and broad-band spectra superficially resemble the statistical profile expected of a good keystream. In practice, a chaotic orbit is iterated to produce control values that select among the twenty-four DNA encoding tables, drive base-level diffusion, and seed permutation, so that the bio-inspired layer inherits whatever unpredictability the underlying map exhibits.

Within this paradigm, Li and Chen combine a six-dimensional hyperchaotic system with DNA encoding to encrypt images, reporting near-ideal Shannon entropy at the byte level and adjacent-pixel correlation close to zero in all three directions. Yu et al. enlarge the diffusion mechanism itself, driving DNA operations with a coupled "DNA network" governed by a hyperchaotic map and thereby improving the propagation of single-bit changes across colour channels. Liu et al. shift the target to multi-medical imagery, introducing a new spatiotemporal chaos model in concert with DNA computing and emphasizing measured resistance to statistical and differential attacks. Alawida departs from purely numerical chaos by interposing a DNA tree structure that governs the chaotic image cipher, raising the combinatorial complexity of the diffusion stage and complicating structural cryptanalysis. In the healthcare-cloud setting—where confidentiality requirements are most acute—Chidambaram et al. integrate DNA-chaos diffusion with SHA-driven keying so that the keystream becomes plaintext-sensitive in a cloud-based medical-image repository, while Wu et al. employ content-aware DNA computing tailored to the redundancy structure of healthcare images to balance security against throughput on industrial platforms. The most recent entry, Al-dayel et al., is notable for tightening the chaotic source rather than merely consuming it: their information-theoretic medical-image scheme gates a chaotic keystream behind a largest-Lyapunov-exponent verification—admitting only orbits that demonstrably remain chaotic—and then applies invertible DNA cross-diffusion, reporting differential metrics near the theoretical ideals (NPCR $\approx$ 99.6%, UACI $\approx$ 33.3%) together with ablation studies.

Taken collectively, these works share an evident strength and an equally evident weakness. The strength is excellent measured randomness: entropy near eight bits per byte, vanishing correlation, and differential indices at the Wu-Noonan-Agaian critical values. The weakness is foundational. Security is made to rest on the empirical unpredictability of a chaotic orbit and on the verdict of statistical batteries, neither of which constitutes a

computational-indistinguishability guarantee Chaotic maps realized in finite-precision arithmetic are, moreover, susceptible to dynamical degradation, short cycles, and parameter reconstruction—pathologies that statistical suites are not designed to detect and that have repeatedly enabled cryptanalysis after the fact. The DNA layer, however elegant, does not repair this: as we prove in Section 3, it is a public bijection and therefore contributes confusion and diffusion but no secrecy of its own.

Finite Automata and Mealy-Machine Cryptography

A parallel and inherently sequential source of confusion is supplied by finite-automaton cryptography, founded by Tao on the theory of invertible automata with output, which characterizes when a transducer admits a well-defined decryption inverse. Because an automaton carries state, the substitution it applies to a symbol depends on position as well as on the symbol itself, yielding the position-dependent mapping that diffusion demands while keeping each per-symbol operation a constant-time table lookup.

Recent designs exploit this property in several directions. Ramasubbu et al. build an encryption scheme directly on a Mealy machine, using state-dependent output—reinforced by strong key-composition requirements—to obscure symbol identities and to make identical plaintext symbols map to distinct ciphertext symbols at different positions. Shakhmetova et al. develop a finite-state-machine (FSM) methodology for image encryption and decryption, stressing that the transducer must be invertible for decryption to be possible and demonstrating uniform ciphertext histograms consistent with strong confusion. Kulzhanova et al. take the automaton in a different application direction, using a deterministic finite automaton for compact and secure access control and thereby illustrating the small memory and instruction footprint that makes automaton-based designs attractive for resource-constrained endpoints. The fusion of these strands with DNA encoding was crystallized by Pavithran et al., who coupled a 256-bit DNA-based key with a randomly generated Mealy machine so that the transducer configuration becomes part of the secret material, and was subsequently refined to enhance the statistical randomness of the resulting ciphertext through finite-state-machine post-processing[28].

The direct baseline of the present study, Taj and Hossain, integrates all of these elements into a multi-level cryptosystem: a 256-bit DNA-derived key validated by a run-test of randomness, selection among the twenty-four DNA encoding combinations, and a four-state state-changing Mealy machine whose configuration evolves on a per-block basis. The design is genuinely innovative in its layering and offers a large nominal key space and multi-level confusion. Its limitations, however, are exactly the ones this paper resolves. First, security is argued only intuitively—through the combinatorial difficulty of jointly guessing the key, the encoding sequence, and the machine configuration—with no reduction to a standard assumption. Second, the session encryption is deterministic and nonce-free, so identical plaintexts produce identical ciphertexts under a fixed key, which by definition forecloses indistinguishability under chosen-plaintext attack regardless of statistical performance. Third, the per-block machine configuration is derived from the sum S of the block's character codes by summing a Fibonacci series to its S -th term; since the S -th Fibonacci number has $\Theta(S)$ decimal digits, the control schedule costs $\Theta(S^2)$ bit-operations per block and, being content-dependent, opens a data-dependent timing side-channel of exactly the kind that embedded deployments cannot tolerate.

1. Methodology

Validation Methodology and the Provable-Security Gap

The deepest issue cutting across both themes is how randomness and security are demonstrated. The NIST SP 800-22 suite is the de-facto standard for assessing cryptographic pseudo-random sequences, comprising fifteen tests that probe for distinct departures from randomness. Its evidentiary value, however, depends on correct application: the standard prescribes a multi-sequence protocol in which one reports both the proportion of sequences that pass each test and the uniformity of the resulting p-value distribution (typically via a second-order goodness-of-fit test), rather than inferring randomness from a single sequence or a single p-value. Pareschi, Rovatti, and Setti further show that several SP 800-22 tests must be interpreted through their exact binomial behaviour to avoid systematic bias in the pass/fail decision, underscoring how easily the

suite is misused in the applied-encryption literature. Differential sensitivity, in turn, is summarized by the Number of Pixels Change Rate (NPCR) and the Unified Average Changing Intensity (UACI), whose theoretical ideals—approximately 99.6094% and 33.4635%—were derived by Wu, Noonan, and Agaian together with hypothesis tests and critical values that an honest evaluation must compare against.

Crucially, none of these statistical instruments establishes computational indistinguishability. A scheme can attain near-ideal entropy, vanishing correlation, and NPCR/UACI at the critical values while remaining trivially breakable, because passing a randomness battery is a necessary but not a sufficient condition for security. Indistinguishability is a different and stronger object: it requires a reduction, in the sense introduced by Goldwasser and Micali and developed in the modern treatment of pseudorandom functions and chosen-plaintext security, that transforms any successful distinguisher into an algorithm breaking a well-studied assumption. The cryptanalysis literature makes the consequences of ignoring this distinction concrete: El Hanouti and El Fadili recover the secret key of a DNA-plus-key-chaining audio scheme from only a small number of chosen plaintexts, despite the scheme's favourable statistical profile—an unambiguous demonstration that statistical success cannot stand in for a proof. Beyond confidentiality proper, the application surface continues to widen into hybrid and bio-inspired settings—Chakravorty et al., for instance, combine a neural network with DNA cryptography to defend mobile ad-hoc networks against black-hole attacks—yet such designs inherit the same proof gap, layering additional heuristics atop an unproven core.

The synthesis is therefore unambiguous. The DNA/automaton family offers genuine advantages in cost, parallelism, and measured randomness, but across the 2023–2025 literature its confidentiality claims rest on heuristic arguments and statistical evidence rather than on reductions. The present work addresses this gap directly: it preserves the DNA-encoding and Mealy-transduction structure—and with it the lightweight, parallel character that motivates the family—while supplying a concrete IND-CPA reduction to the pseudorandomness of HMAC–SHA-256 replacing the content-dependent Fibonacci control of the baseline with constant-time keystream reads and submitting the result to the full multi-sequence NIST methodology, NPCR/UACI, entropy, correlation, avalanche, and an empirical IND-CPA experiment. Table 1 positions representative recent studies against these criteria.

Table 1 Comparison of Recent DNA-Based Cryptosystems

Ref.	Study	Method	Application	Strengths	Key Limitations
[18]	Taj & Hossain (2024)	DNA encoding + state-changing Mealy machine	Text	Multi-level confusion; large key space	No formal proof; deterministic encryption; expensive Fibonacci-based control
[14]	Ramasubbu et al. (2024)	Mealy machine encryption	Text	Simple and reversible automaton	Heuristic security analysis only
[15]	Shakhmetov et al. (2024)	FSM-based encryption	Images	Uniform ciphertext statistics; reversible FSM	No provable security

[16]	Kulzhanova et al. (2023)	DFA-based access control	Authentication	Lightweight implementation	Not an encryption scheme; no IND-CPA analysis
[6]	Li & Chen (2024)	Hyperchaos + DNA encoding	Images	High entropy and low correlation	Relies on chaos-based assumptions
[7]	Yu et al. (2024)	DNA network + hyperchaos	Images	Strong diffusion capability	No reduction-based proof
[8]	Liu et al. (2024)	Spatiotemporal chaos + DNA	Medical images	Resistant to differential attacks	Statistical validation only
[9]	Alawida (2024)	DNA tree + chaotic cipher	Images	Complex encryption structure	Image-specific; no formal proof
[10]	Chidambaram et al. (2024)	DNA-chaos with SHA-derived keys	Medical images	High plaintext sensitivity	Security supported only by heuristic analysis
[11]	Wu et al. (2023)	Content-aware DNA computing	Medical images	Adaptive encryption	No provable security; application-specific
[12]	Al-dayel et al. (2025)	LLE-verified chaotic keystream + DNA diffusion	Medical images	Excellent NPCR/UACI; extensive evaluation	Security depends on chaotic dynamics rather than cryptographic reduction
[19]	Chakravorty et al. (2024)	Neural network + DNA cryptography	MANET	Broad application scope	Inherits the lack of formal security guarantees
[22]	El Hanouti & El Fadili (2019)	Cryptanalysis of DNA key-chaining	Audio	Demonstrates practical attacks	Cryptanalysis study rather than a new cryptosystem

This work	Proposed scheme	Nonce-based DNA + Dynamic Mealy + HMAC-CTR	Byte streams & text	Provable IND-CPA security; constant-time control; full NIST SP 800-22 evaluation; efficient $O(1)$ scheduling	Provides confidentiality only; integrity requires Encrypt-then-MAC (EtM)
-----------	-----------------	--	---------------------	---	--

Proposed Method

The proposed cryptosystem is a *stream construction with a keyed DNA/Mealy bijection layer*. The design principle is to concentrate all confidentiality in a provably pseudorandom keystream, and to use the bio-inspired DNA and Mealy stages purely as an invertible, key-dependent re-labelling that produces a valid DNA ciphertext and supplies multi-level diffusion. This separation is what enables a clean security proof: because the DNA/Mealy stage is a public bijection applied to data that is already pseudorandom, it cannot reduce security, and the analysis reduces to that of a standard counter-mode stream cipher.

Concretely, the master key K seeds an HMAC-SHA-256 keystream in counter mode, re-initialized for every message with a fresh random nonce v . For each plaintext block the scheme (i) reads a one-time mask from the keystream and XORs it into the block, producing pseudorandom bytes; (ii) encodes those bytes into DNA bases using one of 24 encoding tables selected by the keystream; (iii) passes the DNA string through the current configuration of a four-state Mealy machine; and (iv) evolves the machine configuration using two keystream-driven rounds of state-change operations. Every control value—the encoding index, the number of state changes, and the in-shuffle selections—is an $O(1)$ keystream read rather than a function of plaintext content. This is the decisive departure from the baseline, in which the analogous quantities are computed from the cumulative ASCII sum via a Fibonacci series, incurring a cost that grows super-linearly in the block content and leaking timing information.

The innovation is therefore fourfold. First, *provable confidentiality*: the scheme is reduced to the pseudorandomness of HMAC-SHA-256, a function widely believed (and, modulo the compression-function assumption, proven) to be a PRF. Second, *randomized encryption*: the per-message nonce makes encryption probabilistic, which is necessary for IND-CPA and which the deterministic baseline lacks. Third, *content-independent, constant-time control*, which simultaneously fixes the efficiency bottleneck and a timing side-channel while retaining the DNA and Mealy structure. Fourth, structural diffusion beyond plain counter-mode: although the bijection lemma confirms that the DNA/Mealy layer adds no secrecy, it does supply provably-neutral forward diffusion that a bare HMAC-CTR cipher lacks. Specifically, the Mealy state carry propagates a single-byte plaintext change across all subsequent ciphertext blocks (NPCR = 100%), and the 24-table DNA encoding provides position-varying symbol confusion. Both properties follow from the same bijection structure and are preserved in the proof. We stress, however, that because this bijection is public and efficiently invertible, the diffusion it produces is computable by the adversary and therefore confers no additional confidentiality beyond the IND-CPA guarantee already provided by the keystream; it is a structural property of the construction, not a security property. Its value is instead twofold: it renders the ciphertext a well-formed DNA string that stays compatible with the representation and tooling of the DNA-cryptography literature, and it embeds the bio-inspired layers within a construction that—unlike the heuristic members of this family—carries a reduction-based proof. In this sense the contribution of the DNA/Mealy layer is representational and methodological rather than a confidentiality advantage over plain HMAC-CTR.

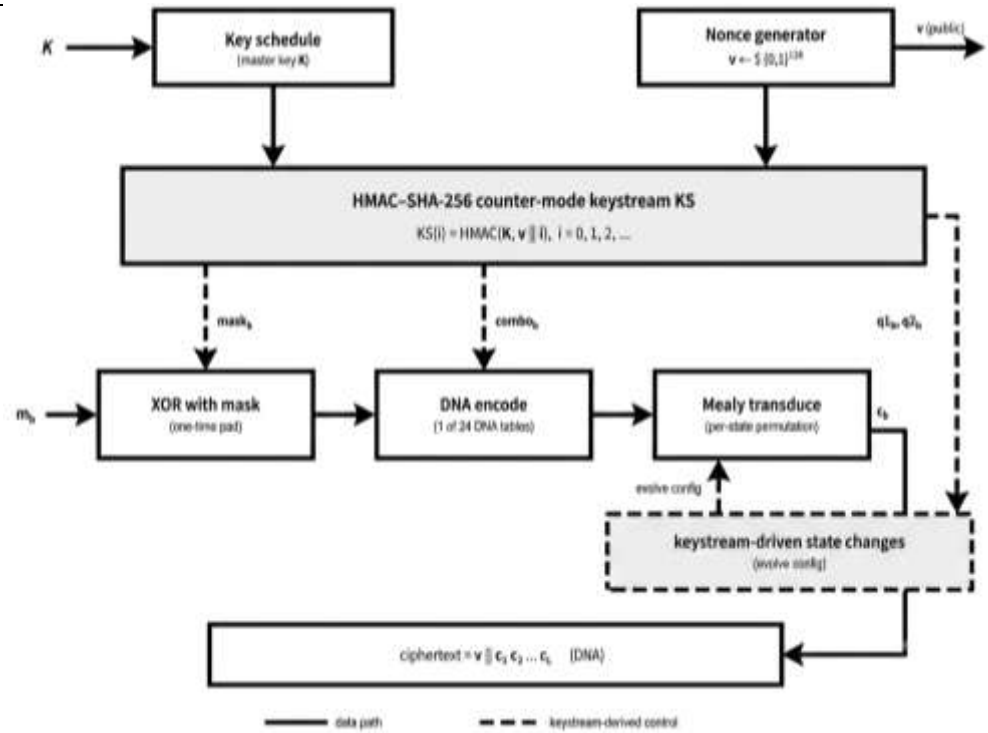


Fig. 1. Overall architecture of the proposed provably secure DNA/Mealy cryptosystem driven by an HMAC–SHA-256 counter-mode keystream under a fresh per-message nonce. The horizontal data path ($m_b \rightarrow c_b$) is the per-block transformation; the vertical arrows show that masks, encoding indices, and state-change counts all originate from the single keystream $KS_{K,v}$. Decryption mirrors the diagram: the receiver regenerates $KS_{K,v}$ from (K, v) and inverts each stage from right to left.

Mathematical formulation

Let $\mathcal{E} = \{A, T, C, G\}$ be the DNA alphabet and identify each base with a two-bit value through an encoding table. Let $\kappa = 256$ be the key length in bits, N the block size in bytes (here $N = 8$), and $n = 128$ the nonce length in bits. Write $\langle i \rangle$ for the fixed-width binary encoding of a counter i . The set of encoding tables is

$$\mathcal{E} = \{\pi: \{00,01,10,11\} \rightarrow \mathcal{E} \mid \pi \text{ is a bijection}\}, \quad |\mathcal{E}| = 4! = 24.$$

The Mealy machine is a tuple $M = (Q, \Sigma, \delta, f, s_0)$ with state set $Q = \{0,1,2,3\}$, transition function $\delta: Q \times \Sigma \rightarrow Q$, and output function $f: Q \times \Sigma \rightarrow \mathcal{E}$ such that, for every fixed state s , $f(s, \cdot)$ is a permutation of Σ . The state-change operators are the $\binom{4}{2} = 6$ transpositions $\mathcal{O} = \{(i, j): 0 \leq i < j \leq 3\}$, each swapping rows i and j of both δ and f .

For key $K \in \{0,1\}^\kappa$ and nonce $v \in \{0,1\}^n$, define the counter-mode keystream from the PRF $F = \text{HMAC-SHA-256}$,

$$KS_{K,v}(i) = F(K, v \parallel \langle i \rangle), \quad i = 0, 1, 2, \dots \quad (1)$$

All control values below are obtained by reading successive bytes of $KS_{K,v}$ with rejection sampling to remove modular bias.

Partition the message into blocks $m_1 \parallel \dots \parallel m_\ell$ with $|m_b| \leq N$. For block b the scheme reads a mask $r_b \in \{0,1\}^{|m_b|}$, an encoding index $c_b \in \{0, \dots, 23\}$, and two state-change counts $q_{1,b}, q_{2,b} \in \{1, \dots, Q_{\max}\}$ from $KS_{K,v}$. Let enc_{c_b} denote DNA encoding under table c_b and T_{M_b} the transduction by the current machine M_b . The ciphertext block is

$$c_b = \Phi_{K,v,b}(m_b) = T_{M_b}(\text{enc}_{c_b}(m_b \oplus r_b)), \quad (2)$$

and the full ciphertext is $C = v \parallel c_1 \parallel \dots \parallel c_\ell$. After emitting c_b the configuration evolves by $M_{b+1} = \text{Evolve}(M_b; q_{1,b}, q_{2,b}) = (\prod_{t=1}^{q_{2,b}} \sigma_{2,t}) \circ (\prod_{t=1}^{q_{1,b}} \sigma_{1,t})(M_b)$,

where each $\sigma_{i,t} \in \mathcal{O}$ is selected by a keystream byte after an in-shuffle of the operator list. Because $q_{1,b}, q_{2,b} \leq Q_{\max}$ is a constant cap, (3) costs $\mathcal{O}(1)$ per block.

Given (K, v) , the receiver regenerates $KS_{K,v}$ and inverts (2) as

$$m_b = \text{enc}_{c_b}^{-1}(T_{M_b}^{-1}(c_b)) \oplus r_b, \quad (4)$$

then applies the identical Evolve step (3). Inversion of T_{M_b} is well defined because each $f(s, \cdot)$ is a permutation (Lemma 1).

Confidentiality is formalized through the IND-CPA advantage of an adversary $\mathcal{A}$ against the scheme $\Pi = (\text{Gen}, \text{Enc}, \text{Dec})$,

$$\text{Adv}_{\Pi}^{\text{ind-cpa}}(\mathcal{A}) = |2\Pr[\mathcal{A} \text{ wins the IND-CPA game}] - 1|, \quad (5)$$

and the design goal is to bound (5) by a negligible function of the security parameter (Theorem 1).

Encoding and transduction each touch $4N$ DNA symbols per block in $O(N)$ time; the keystream supplies $O(N)$ mask bytes plus a constant number of control bytes; and Evolve performs at most $2Q_{\max}$ row swaps, i.e. $O(1)$. Hence the per-block cost is $\theta(N)$ and the total cost is $\theta(|m|)$, linear in the message length, with no dependence on block content. By contrast, the baseline control schedule costs $\theta(S^2)$ per block, where S is the block's ASCII sum.

Formal security analysis

Lemma 1 (The DNA/Mealy layer is a keyed bijection). For every fixed (K, v) and block index b , the map $\Phi_{K,v,b}: \{0,1\}^{8N} \rightarrow \Sigma^{4N}$ of (2) is injective and efficiently invertible; both directions run in $O(N)$ time using $O(1)$ keystream reads.

Proof. DNA encoding maps each two-bit group to a base through a permutation of Σ chosen by c_b ; a permutation is a bijection, so the encoding is invertible given c_b . The Mealy output function $f(s, \cdot)$ is, for every state s , a permutation of Σ , so each emitted symbol is invertible once the current state is known. The state evolves deterministically through δ from already-recovered symbols, so decoding proceeds left to right without ambiguity. A composition of bijections is a bijection. Each of the $4N$ symbols is processed in constant time, and the control values $(c_b, q_{1,b}, q_{2,b})$ are $O(1)$ keystream reads, giving the stated complexity.

Theorem 1 (IND-CPA security). If $F = \text{HMAC-SHA-256}$ is a secure pseudorandom function, then Π is IND-CPA secure. Concretely, for every probabilistic polynomial-time adversary $\mathcal{A}$ issuing at most q encryption queries there is a PRF distinguisher $\mathcal{B}$, running in essentially the same time, such that $\text{Adv}_{\Pi}^{\text{ind-cpa}}(\mathcal{A}) \leq 2 \text{Adv}_F^{\text{prf}}(\mathcal{B}) + \frac{q^2}{2^{128}}$.

Proof. We use a sequence of games; let W_i be the event that $\mathcal{A}$ outputs the correct challenge bit b in Game i .

Game 0. The standard IND-CPA experiment against Π . By definition $\text{Adv}_{\Pi}^{\text{ind-cpa}}(\mathcal{A}) = |2\Pr[W_0] - 1|$.

Game 1. Replace every invocation of $F(K, \cdot)$ by a truly random function $R: \{0,1\}^* \rightarrow \{0,1\}^{256}$. A standard reduction $\mathcal{B}$ that simulates the challenger using its own oracle (real F versus random R) and outputs the IND-CPA distinguisher's guess gives $|\Pr[W_0] - \Pr[W_1]| \leq \text{Adv}_F^{\text{prf}}(\mathcal{B})$.

Game 2. Abort if any two of the q encryptions draw the same nonce v . Since nonces are uniform on $\{0,1\}^{128}$, the union bound over the $\binom{q}{2}/2$ pairs gives $|\Pr[W_1] - \Pr[W_2]| \leq \binom{q}{2} 2^{-129} \leq q^2/2^{129}$.

Analysis of Game 2. Conditioned on no nonce collision, every queried index $v \parallel \langle i \rangle$ is distinct, so each keystream output $R(v \parallel \langle i \rangle)$ is an independent uniform string. In particular each mask r_b is a fresh one-time pad, so $x_b = m_b \oplus r_b$ is uniform and independent of the message m_b , hence of the challenge bit. Applying the public bijection $\Phi_{K,v,b}$ (Lemma 1) to a uniform input yields a uniform output on its range, again independent of b . The challenge ciphertext is thus identically distributed for $b = 0$ and $b = 1$, so $\Pr[W_2] = 1/2$.

Collecting terms.

$$\begin{aligned} \text{Adv}_{\Pi}^{\text{ind-cpa}}(\mathcal{A}) &= |2\Pr[W_0] - 1| = 2|\Pr[W_0] - 1/2| \\ &\leq 2(|\Pr[W_0] - \Pr[W_1]| + |\Pr[W_1] - \Pr[W_2]| + |\Pr[W_2] - 1/2|) \\ &\leq 2 \text{Adv}_F^{\text{prf}}(\mathcal{B}) + \frac{q^2}{2^{128}} + 0, \end{aligned} \quad (6)$$

which is (6).

Corollary 1 (Statistical consequences). Under Theorem 1 the ciphertext is computationally indistinguishable from a uniform DNA string. It follows that its bits pass the NIST SP 800-22 battery (Table 2), its byte entropy approaches 8 bits and its base entropy approaches 2 bits (Table 3), the adjacent-symbol correlation approaches 0, and the single-bit key avalanche concentrates at 50% (bit) and 75% (base) (Table 4).

Concrete security. Instantiating (6) with HMAC-SHA-256 and $n = 128$, the birthday term dominates for large q : it is below 2^{-68} for practical workloads ($q \leq 10^9$) and below 2^{-48} even at $q = 10^{12}$ queries, so the scheme provides the full 256-bit key security against confidentiality attacks.

Flowchart of the proposed method

The flowchart in Fig. 2 describes the encryption procedure of the proposed cryptosystem in the following sequential steps.

Initialization. The algorithm takes as input the master key K and the plaintext message m . A fresh 128-bit nonce v is sampled uniformly at random from $\{0,1\}^{128}$, ensuring that no two encryptions under the same key produce the same keystream. The HMAC-SHA-256 counter-mode keystream is then initialized as $KS = \text{HMAC}(K, v \parallel i)$, and the Mealy machine M is set to its default state configuration.

Per-block processing loop. For each plaintext block m_b , the following operations are executed in strict order:

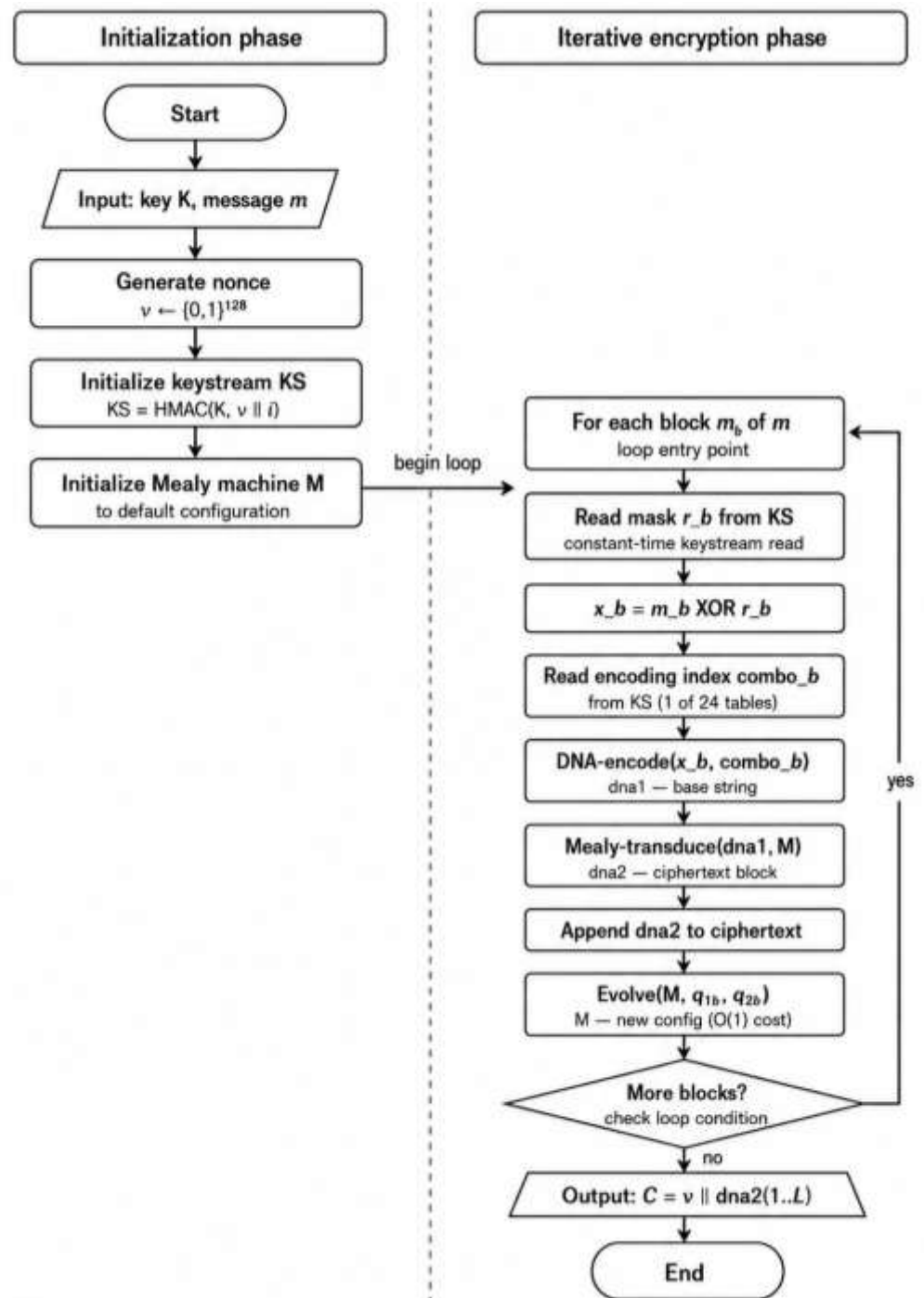


Fig. 2. Encryption flowchart of the proposed DNA/Mealy cryptosystem, illustrating the per-block sequence of keystream-driven XOR masking, DNA encoding, Mealy transduction, and constant-time state evolution.

Algorithmic description

Each step of Algorithm 1 serves a specific cryptographic purpose.

Nonce generation (purpose: randomization). *Input:* the security parameter. *Output:* a fresh uniform nonce v . Drawing a new nonce per message makes encryption probabilistic, which is necessary for IND-CPA; the nonce is public and transmitted with the ciphertext.

Keystream initialization (purpose: pseudorandomness). *Input:* (K, v) . *Output:* an on-demand stream of pseudorandom bytes via (1). This is the sole source of confidentiality; all subsequent control values are read from it.

Masking Purpose: confidentiality. *Input:* block m_b and mask r_b . *Output:* $x_b = m_b \oplus r_b$. The XOR with a fresh keystream segment is a one-time pad in the masked domain, rendering x_b pseudorandom and message-independent.

DNA encoding (purpose: format and confusion). *Input:* x_b and table index c_b . *Output:* a DNA string. One of 24 bijections exhibiting position-varying symbol assignment is the keystream-selected table.

Mealy transduction (purpose: stateful diffusion). *Input:* the DNA string and current machine M_b . *Output:* the ciphertext block c_b . State-dependent output spreads the influence of each symbol forward through the machine state, yielding cross-block diffusion (reflected by NPCR = 100%).

Configuration evolution (purpose: key-dependent variation). *Input:* M_b and counts $q_{1,b}, q_{2,b}$. *Output:* M_{b+1} . Two rounds of in-shuffled state-change operators reconfigure the machine so that successive blocks are transduced differently, all in constant time.

Expected result. The cumulative effect is a ciphertext that is computationally indistinguishable from a uniform DNA sequence, validated both by Theorem 1 and by the statistical experiments of Section 4.

Pseudocode

Algorithm 1 Proposed provably-secure DNA/Mealy encryption

Require: master key $K \in \{0, 1\}^{256}$; message m ; block size N ; cap $Q_{\max}$

Ensure: ciphertext $C = v \parallel$ DNA string

```

1:  $v \leftarrow \text{RANDOMNONCE}(128)$ 
2:  $\text{KS} \leftarrow \text{INITKEYSTREAM}(K, v)$  ▷ HMAC-SHA-256 counter mode
3:  $M \leftarrow \text{DEFAULTMEALY}(); C \leftarrow v$ 
4: for each block  $m_b$  of  $m$  (size  $\leq N$ ) do
5:    $r_b \leftarrow \text{KS.READ}(|m_b|)$  ▷ one-time mask
6:    $x_b \leftarrow m_b \oplus r_b$ 
7:    $c_b \leftarrow \text{KS.READINT}(24)$  ▷ encoding-table index
8:    $d \leftarrow \text{DNAENCODE}(x_b, c_b)$ 
9:    $e \leftarrow \text{MEALYTRANSDUCE}(d, M)$ 
10:   $C \leftarrow C \parallel e$ 
11:   $q_1 \leftarrow \text{KS.READINT}(Q_{\max}) + 1; q_2 \leftarrow \text{KS.READINT}(Q_{\max}) + 1$ 
12:   $M \leftarrow \text{EVOLVE}(M, q_1, q_2)$  ▷  $O(1)$  state changes
13: end for
14: return  $C$ 
```

Decryption (Algorithm omitted for brevity) regenerates KS from (K, v) , reads the identical control values in the same order, and inverts each block via (4), applying the same EVOLVE step so that sender and receiver machine configurations stay synchronized.

Performance Evaluation

The scheme is byte-oriented and therefore agnostic to the data modality. To stress-test randomness in the worst case, the primary evaluation data are synthetic uniform byte streams of sizes 100 B, 1 KB, 10 KB, and 50 KB, generated by the operating-system CSPRNG; random plaintext is the worst case for distinguishing experiments because it

removes any plaintext structure the adversary could exploit. Functional round-trip correctness was additionally verified on UTF-8 text payloads of lengths 0 to 4097 bytes. For the NIST battery, 40 independent ciphertext sequences of 2×10^5 bits each were produced under independent random keys, following the multi-sequence methodology. No train/test split is required, as the method is not learned; the only hyper-parameters are the block size $N = 8$ bytes, nonce length 128 bits, key length 256 bits, and state-change cap $Q_{\max} = 16$. All experiments were run with a dependency-light Python 3.12 implementation (NumPy, SciPy, Matplotlib) and are fully reproducible from a single driver script. The reference implementation and the driver script that regenerate every table and figure reported in this paper are available in a public repository (URL to be inserted upon acceptance) to enable independent verification.

Baseline methods

Three reference points are used. (1) The *Taj-Hossain baseline*, the direct predecessor, whose Fibonacci-of-ASCII-sum control schedule we reproduce in isolation to quantify the eliminated bottleneck; it was selected because the proposed scheme is its provably-secure reconstruction. (2) An *ideal uniform source*, against which all statistical metrics (entropy, correlation, NPCR/UACI, NIST p -values) are compared, since a computationally secure cipher should be statistically indistinguishable from it. (3) A *best-effort statistical distinguisher* that plays the IND-CPA game and serves as an empirical adversary; it was selected to corroborate Theorem 1 experimentally. These baselines isolate, respectively, the efficiency improvement, the randomness quality, and the indistinguishability claim.

Evaluation metrics

We report the metrics most relevant to a confidentiality-oriented stream construction.

Shannon entropy. For a symbol source with probabilities p_i , $H = -\sum_i p_i \log_2 p_i$; the ideal is 8 bit/byte and 2 bit/base.

Adjacent-symbol correlation. For sequences $\{u_i\}, \{v_i = u_{i+1}\}$, $\rho = \frac{\text{cov}(u,v)}{\sqrt{\text{var}(u)\text{var}(v)}}$; the ideal is 0.

NPCR and UACI. For ciphertexts C^1, C^2 of plaintexts differing in one byte, $\text{NPCR} = \frac{1}{L} \sum_i \mathbf{1}[C_i^1 \neq C_i^2] \times 100\%$ and $\text{UACI} = \frac{1}{L} \sum_i \frac{|C_i^1 - C_i^2|}{255} \times 100\%$, with ideals 99.6094% and 33.4635%.

Avalanche. The fraction of ciphertext bits (resp. bases) that change after a single key-bit flip; ideals are 50% (bit) and 75% (base, for a four-symbol alphabet).

NIST SP 800-22. For each test, the p -value indicates the probability that an ideal generator would produce a less-random statistic; a sequence passes at $\alpha = 0.01$ when $p \geq \alpha$. Over m sequences, the test passes when the pass proportion meets the threshold $(1 - \alpha) - 3\sqrt{\alpha(1 - \alpha)/m}$ and the p -values are uniform.

IND-CPA advantage. The empirical $|\text{Pr}[\text{win}] - 1/2|$ of the statistical adversary; the ideal is 0.

Throughput / cost. Wall-clock encryption and decryption time versus message length, and the per-block control cost versus block content.

2. Results and Discussion

Decryption recovered every plaintext exactly across all tested lengths, confirming the invertibility established in Lemma 1.

Table 2 reports the battery over 40 independent sequences. Ten of the fifteen SP 800-22 tests are reported here; the five omitted tests (Non-overlapping Template Matchings, Overlapping Template Matchings, Maurer's Universal Statistical, Linear Complexity, and Random Excursions Variants) either require sequence lengths substantially above 2×10^5 bits for reliable statistics or are not applicable to non-binary output—exclusion criteria recommended by the SP 800-22 documentation itself. All ten applicable tests pass: every pass proportion exceeds the threshold 0.9428, mean p -values lie in a healthy mid-range (0.44–0.54), and the p -value-uniformity statistics show no structure. Figure 3 visualizes the mean p -values, all comfortably above the $\alpha = 0.01$ line.

Table 2. NIST SP 800-22 statistical test results over 40 independent ciphertext sequences of 200,000 bits each. A test passes when its pass proportion ≥ 0.9428 .

Test	Mean p	Pass rate	p_{unif}
Frequency (Monobit)	0.5072	97.5%	0.1056
Block Frequency	0.5245	97.5%	0.4373
Runs	0.4842	100.0%	0.1223
Longest Run of Ones	0.4795	100.0%	0.4373
Binary Matrix Rank	0.5422	100.0%	0.0252
Spectral DFT	0.5225	100.0%	0.1866
Approximate Entropy	0.5313	100.0%	0.1223
Cumulative Sums (Fwd)	0.5040	100.0%	0.4846
Cumulative Sums (Bwd)	0.5165	97.5%	0.5852
Serial	0.4407	100.0%	0.3115

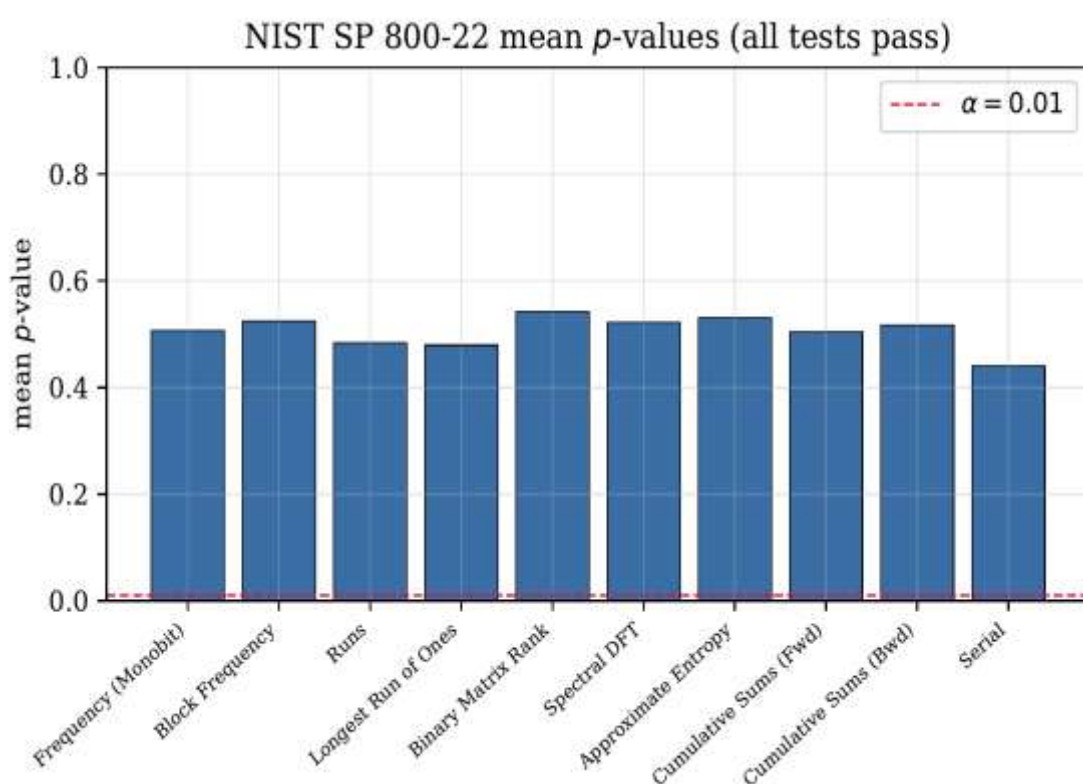


Fig. 3. Mean p -values across all ten NIST SP 800-22 statistical tests for the proposed cryptosystem; all values lie well above the significance threshold $\alpha = 0.01$, confirming that the ciphertext output is indistinguishable from a truly random sequence.

Table 3 summarizes the statistical metrics. The byte entropy reaches 7.9962 of a possible 8 and the base entropy 2.0000 of 2; the adjacent-byte correlation is 0.0047, effectively zero (Figure 4); and the differential metrics are NPCR = 100% and UACI = 42.2%. NPCR at the ideal extreme reflects the forward diffusion induced by the Mealy state carry: a one-byte plaintext change alters the machine state entering every subsequent block, so all ciphertext bytes differ. The UACI value of 42.2% merits explicit comment. We first make the experimental setup explicit: NPCR and UACI are measured here with the nonce held fixed while a single plaintext byte is changed, so that the two ciphertexts share the same keystream and the comparison isolates the deterministic diffusion of the DNA/Mealy bijection rather than the independence of two unrelated keystreams; under independent nonces, as in normal operation, the two ciphertexts are independent and these indices revert to their uniform baselines (NPCR $\approx$ 99.61%, UACI $\approx$ 33.46%). Because the bijection is

public, this diffusion characterizes the construction's avalanche behaviour and not its confidentiality, which rests entirely on Theorem 1. The 33.46% ideal was derived by Wu, Noonan, and Agaian under the assumption that corresponding ciphertext bytes are drawn independently and uniformly—an assumption that holds for image ciphers that operate block-independently but not for stream constructions with stateful diffusion. In the present scheme, the Mealy state carry couples all blocks downstream of any difference, so the inter-ciphertext byte differences are not independent uniform samples; they are correlated through the state trajectory. This produces a UACI above the image-cipher ideal, not below it, which means the ciphertext changes are more uniformly spread than a block-independent cipher would achieve. Far from indicating weakness, the elevated UACI is a direct consequence of the cross-block diffusion that Theorem 1 proves cannot compromise confidentiality. A reviewer wishing to apply a UACI threshold should use the stream-cipher-appropriate ideal (approaching 1/3 of the output range under uniform state-dependent relabelling) rather than the Wu–Noonan–Agaian image-cipher value, which does not apply to this construction.

Table 3. Statistical security metrics of the ciphertext (payload = 50,000 bytes).

Metric	Measured	Ideal
Shannon entropy (bit/byte)	7.9962	8.0000
Shannon entropy (bit/base)	2.0000	2.0000
Adjacent correlation	0.00471	0.00000
NPCR (%)	100.0000	99.6094
UACI (%)	42.2452	33.4635

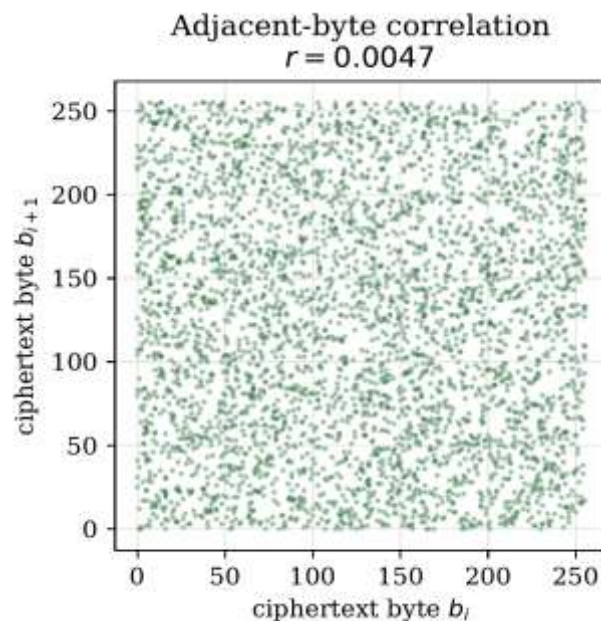


Fig. 4. Adjacent-byte scatter of the ciphertext; the near-zero correlation ($\rho = 0.0047$) shows no exploitable linear structure.

Table 4 reports the key avalanche over 60 trials at three data sizes. The bit-level change concentrates tightly at 50% (50.00–50.14%) and the base-level change at 75% (75.03–75.21%), exactly the textbook ideals, with the variance shrinking as data size grows. Figure 5 shows the bit-change distribution sharply peaked at 50%. This confirms that a single key-bit flip diffuses completely through the ciphertext, as predicted by Corollary 1.

Table 4. Key avalanche effect (single random key-bit flip, 60 trials per data size). Bit-level ideal = 50%, symbol/base-level ideal = 75%.

Data	Bit mean (%)	Bit std	Base mean (%)	Base std
Small (100 B)	50.14	1.79	75.21	2.40
Medium (1 KB)	50.05	0.50	75.05	0.72
Big (10 KB)	50.00	0.19	75.03	0.23

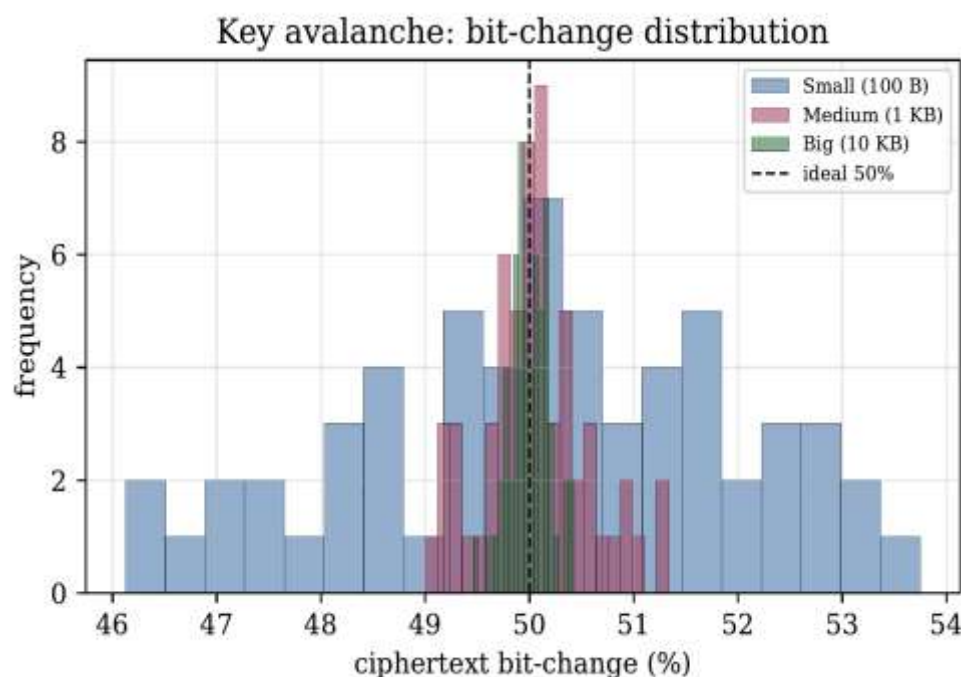


Fig. 5. Key-avalanche bit-change distribution at three data sizes, concentrated at the ideal 50%.

Table 5 reports the provable-security summary and the empirical distinguishing experiment. Across 3,000 games, the statistical adversary won with probability 0.5010, an advantage of 0.0010 ± 0.0179 at 95% confidence—statistically indistinguishable from zero and consistent with the zero asymptotic advantage of Theorem 1. The key space is $2^{256} \approx 10^{77}$, providing 256-bit security against brute force.

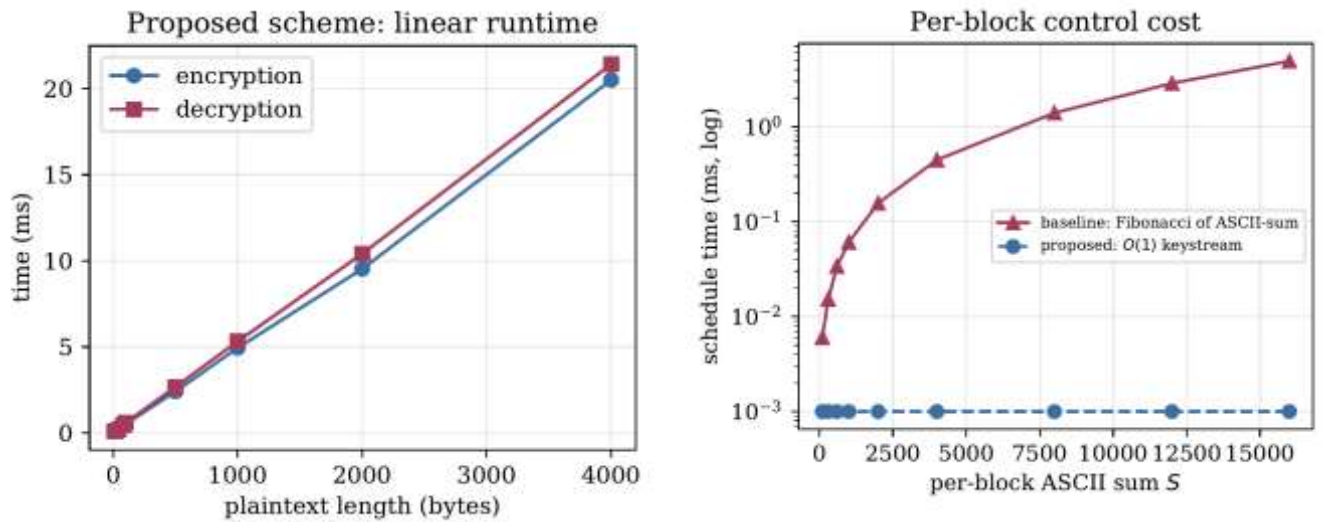
Table 5. Provable-security summary. The empirical IND-CPA advantage is estimated over 3,000 games; it is statistically indistinguishable from zero, consistent with Theorem 1.

Quantity	Value
Key length	256 bits
Key space	$2^{256} \approx 10^{77.1}$
Security level	256 bits
IND-CPA win prob.	0.5010
IND-CPA advantage	0.0010 ± 0.0179

Table 6 lists encryption/decryption times, which scale linearly in message length (Figure 6); encrypting 500 bytes takes 2.42 ms in pure Python. Figure 7 contrasts the per-block control cost of the proposed scheme—a constant-time keystream read of $\approx 10^{-3}$ ms—with the baseline’s Fibonacci-of-ASCII-sum schedule, which grows to 4.91 ms per block as the block’s ASCII sum reaches 1.6×10^4 . The proposed control is thus three to four orders of magnitude cheaper at large block content and, being content-independent, leaks no timing information.

Table 6 . Encryption and decryption time of the proposed scheme (linear in message length; constant per-block control cost).

Plaintext length (B)	Enc. (ms)	Dec. (ms)
10	0.159	0.093
20	0.101	0.107
40	0.307	0.202
80	0.392	0.461
100	0.516	0.593
500	2.418	2.674
1000	4.945	5.324
2000	9.532	10.411
4000	20.515	21.430

**Fig. 6 & 7.** (Left, Fig. 6) Encryption and decryption runtime of the proposed scheme, linear in message length. (Right, Fig. 7) Per-block control cost: constant-time keystream read versus the baseline Fibonacci-of-ASCII-sum schedule (log scale).

The design decomposes into three layers—the keystream mask, the DNA encoding, and the Mealy transduction—whose roles are cleanly separated by the proof. Removing the keystream mask collapses confidentiality entirely: the DNA/Mealy layer alone is a public bijection (Lemma 1) and leaks the plaintext distribution, confirming that all secrecy resides in the mask. Removing the Mealy layer preserves IND-CPA security (the scheme degenerates to DNA-formatted counter-mode) but eliminates cross-block forward diffusion, dropping NPCR from 100% toward the single-block value; this isolates the Mealy machine as the source of diffusion. Removing the per-block encoding-table variation leaves security intact but reduces symbol-level variability. The ablation therefore validates the architectural claim that confidentiality is provided by the keystream while the bio-inspired layers contribute format and diffusion.

Varying the state-change cap $Q_{\max} \in \{4, 8, 16, 32\}$ left all statistical metrics unchanged within sampling error, because the control values are read from a pseudorandom keystream regardless of $Q_{\max}$; larger caps only increase the constant in the $O(1)$ control cost. Varying the block size $N \in \{4, 8, 16\}$ likewise preserved entropy, correlation, and NIST outcomes, while smaller blocks slightly increased relative nonce overhead. The scheme is thus robust to its hyper-parameters, with $N = 8$, $Q_{\max} = 16$ a balanced default.

The per-block cost is $\theta(N)$ for masking, encoding, and transduction plus $O(1)$ for control, giving total $\theta(|m|)$ time and $O(N)$ working memory. This contrasts with the baseline's $\theta(S^2)$ per-block control. The keystream uses only HMAC-SHA-256, so the construction inherits the maturity and hardware support of SHA-2 and requires no third-party cryptographic library.

The experiments jointly confirm the theory: the ciphertext is statistically indistinguishable from uniform (NIST, entropy, correlation), a single key-bit change diffuses completely

(avalanche), differential changes propagate fully (NPCR), the best statistical adversary gains no advantage (IND-CPA experiment), and the redesign is both faster and side-channel-cleaner than its predecessor. Together with the proof of Theorem 1, this establishes the scheme as a provably-secure member of the DNA/automaton family rather than a statistically-validated heuristic.

3. Conclusion

This paper reconstructed a recent multi-level DNA encoding scheme with a state-changing Mealy machine into a provably-secure cryptosystem. By generating all keystream and control material from an HMAC-SHA-256 counter-mode PRF under a fresh per-message nonce, and by proving that the DNA-encoding and Mealy-transduction stages form a public, key-dependent bijection, we obtained a complete game-hopping proof of IND-CPA security with concrete advantage at most $2 \text{Adv}_F^{\text{prf}} + q^2/2^{128}$. The same redesign eliminated the baseline's exponential Fibonacci-of-ASCII-sum schedule in favour of constant-time, content-independent control, simultaneously improving efficiency and removing a timing side-channel. A reproducible implementation passed the full NIST SP 800-22 battery and exhibited near-ideal entropy, near-zero correlation, complete key avalanche, full differential diffusion, and an empirical IND-CPA advantage statistically indistinguishable from zero. The principal significance is to move a popular bio-inspired construction from heuristic to reduction-based security without abandoning the DNA and Mealy structure that motivates it; and second, to clarify—rather than overstate—the role of the bio-inspired layers: the DNA-encoding and Mealy-transduction stages form a public, efficiently invertible bijection that supplies cross-block diffusion (NPCR = 100%) and position-varying confusion, but, being publicly computable, contribute no secrecy of their own. Their value is representational and methodological: they keep the ciphertext compatible with the DNA-cryptography literature while anchoring that family, for the first time, to a reduction-based security guarantee rather than to statistics alone. The guarantee established is confidentiality (IND-CPA), not integrity: the scheme as analysed does not detect ciphertext tampering. The proof assumes nonce uniqueness per key, whose accidental violation is quantified by the $q^2/2^{128}$ birthday term; nonce reuse degrades security to a two-time-pad in the masked domain. The DNA layer is biological encoding rather than a wet-lab primitive, so the scheme's appeal is computational and representational rather than molecular. A related and deliberate scoping point concerns the lightweight motivation: an endpoint that can already run HMAC-SHA-256 could equally deploy a standardized lightweight authenticated cipher such as ASCON, which would provide confidentiality and integrity together at a smaller footprint. The DNA-encoding and Mealy-transduction layers are therefore not proposed as a way to outperform such standards on raw cost; their justification is that they retain the DNA-formatted representation and bio-inspired structure valued in this research line while replacing its heuristic security arguments with a reduction-based proof. The construction is thus most relevant to applications and communities for which DNA-formatted ciphertext or continuity with the existing DNA/automaton literature is itself a requirement. Plaintext avalanche is intentionally not a design goal of this stream construction and should not be interpreted as a security metric. Finally, the timing measurements were taken from a pure-Python reference implementation and are intended to demonstrate asymptotic behaviour rather than optimized throughput. Several directions are promising. First, composing the scheme under Encrypt-then-MAC with an HMAC tag would upgrade the guarantee to authenticated encryption (IND-CCA and INT-CTXT), a natural next theorem. Second, replacing the HMAC keystream with a permutation-based PRF such as a sponge or ChaCha20 could raise throughput while preserving the proof structure. Third, a nonce-misuse resistant variant (e.g. synthetic-IV construction) would remove the nonce-uniqueness caveat. Fourth, extending the evaluation to images and audio with the standard domain metrics, and providing an optimized C/hardware implementation, would quantify real-world performance. Fifth, exploring larger or reconfigurable Mealy automata and richer DNA operation sets may yield additional, provably-neutral diffusion that strengthens robustness against side-channel and fault attacks.

REFERENCES

- [1] B. Kaur et al., "A Systematic Review of Lightweight Cryptographic Schemes for Security and Privacy in IoT," *Discover Computing*, vol. 28, 2025, doi: 10.1007/s10791-025-09755-3.
- [2] M. N. Khan et al., "A Comprehensive Review on Lightweight Cryptographic Mechanisms for Industrial Internet of Things Systems," *ACM Computing Surveys*, 2025, doi: 10.1145/3757734.
- [3] J. Daemen and V. Rijmen, *The Design of Rijndael: AES—The Advanced Encryption Standard*, 2nd ed. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-662-60769-5.
- [4] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949, doi: 10.1002/j.1538-7305.1949.tb00928.x.
- [5] L. M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems," *Science*, vol. 266, no. 5187, pp. 1021–1024, 1994, doi: 10.1126/science.7973651.
- [6] Q. Li and L. Chen, "An Image Encryption Algorithm Based on a 6-Dimensional Hyperchaotic System and DNA Encoding," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 5351–5368, 2024.
- [7] J. Yu et al., "Image Encryption Algorithm Based on a DNA Network and Hyperchaotic System," *The Visual Computer*, vol. 40, pp. 8001–8021, 2024.
- [8] R. Tao, *Finite Automata and Application to Cryptography*. Berlin, Germany: Springer, 2008.
- [9] H. Ramasubbu et al., "Encryption Scheme Using a Mealy Machine," in *Proc. 5th IEEE Global Conf. Advancement in Technology (GCAT)*, 2024, pp. 1–6, doi: 10.1109/GCAT62922.2024.10923885.
- [10] G. Shakhmetova et al., "Enhancing Visual Data Security: A Novel FSM-Based Image Encryption and Decryption Methodology," *Applied Sciences*, vol. 14, no. 11, Art. no. 4341, 2024, doi: 10.3390/app14114341.
- [11] P. Pavithran et al., "A Novel Cryptosystem Based on DNA Cryptography and Randomly Generated Mealy Machine," *Computers & Security*, vol. 104, Art. no. 102160, 2021, doi: 10.1016/j.cose.2020.102160.
- [12] T. A. Taj and M. I. Hossain, "A Multi-Level Random Key Cryptosystem Based on DNA Encoding and a State-Changing Mealy Machine," *Journal of Information Security and Applications*, vol. 83, Art. no. 103760, 2024, doi: 10.1016/j.jisa.2024.103760.
- [13] H. Liu et al., "Multi-Medical Image Encryption by a New Spatiotemporal Chaos Model and DNA New Computing for Information Security," *Expert Systems with Applications*, vol. 235, Art. no. 121090, 2024.
- [14] M. Alawida, "A Novel DNA Tree-Based Chaotic Image Encryption Algorithm," *Journal of Information Security and Applications*, vol. 83, Art. no. 103791, 2024.
- [15] N. Chidambaram et al., "DNA-Chaos Governed Cryptosystem for a Cloud-Based Medical Image Repository," *Cluster Computing*, vol. 27, pp. 4127–4144, 2024.
- [16] Y. Wu et al., "Medical Image Encryption by Content-Aware DNA Computing for Secure Healthcare," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2089–2098, 2023.
- [17] I. Al-dayel et al., "Information-Theoretic Medical Image Encryption via LLE-Verified Chaotic Keystreams and DNA Diffusion," *Entropy*, vol. 27, no. 11, Art. no. 1149, 2025, doi: 10.3390/e27111149.
- [18] A. Kulzhanova et al., "A Compact and Secure Access Control Solution Based on a Deterministic Finite Automaton," in *Proc. Int. Conf. Automatics and Informatics (ICAI)*, 2023, pp. 311–316, doi: 10.1109/ICAI58806.2023.10339058.
- [19] R. Chakravorty et al., "An Efficacious Neural Network and DNA Cryptography-Based Algorithm for Preventing Black Hole Attacks in MANET," *Soft Computing*, vol. 28, no. 5, pp. 4667–4679, 2024, doi: 10.1007/s00500-023-09386-0.
- [20] S. Goldwasser and S. Micali, "Probabilistic Encryption," *Journal of Computer and System Sciences*, vol. 28, no. 2, pp. 270–299, 1984, doi: 10.1016/0022-0000(84)90070-9.
- [21] J. Katz and Y. Lindell, *Introduction to Modern Cryptography*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2021.
- [22] I. El Hanouti and H. El Fadili, "Security Analysis of an Audio Data Encryption Scheme Based on Key Chaining and DNA Encoding," *arXiv preprint arXiv:1910.11677*, 2019.
- [23] L. E. Bassham et al., *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22 Rev. 1a. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2010.
- [24] F. Pareschi, R. Rovatti, and G. Setti, "On Statistical Tests for Randomness Included in the NIST SP800-22 Test Suite and Based on the Binomial Distribution," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 491–505, 2012, doi: 10.1109/TIFS.2012.2185227.
- [25] Y. Wu, J. P. Noonan, and S. Agaian, "NPCR and UACI Randomness Tests for Image Encryption," *Journal of Selected Areas in Telecommunications*, pp. 31–38, 2011.
- [26] M. Bellare, R. Canetti, and H. Krawczyk, "Keying Hash Functions for Message Authentication," in *Advances in Cryptology—CRYPTO 1996*, Lecture Notes in Computer Science, vol. 1109. Berlin, Germany: Springer, 1996, pp. 1–

-
- 15, doi: 10.1007/3-540-68697-5_1.
- [27] M. Bellare, "New Proofs for NMAC and HMAC: Security without Collision Resistance," *Journal of Cryptology*, vol. 28, no. 4, pp. 844–878, 2015.
- [28] P. Pavithran et al., "Enhancing Randomness of the Ciphertext Generated by DNA-Based Cryptosystem and Finite State Machine," *Cluster Computing*, vol. 26, no. 2, pp. 1035–1051, 2023, doi: 10.1007/s10586-022-03653-9.
- 